Intervirology, 2008; 51,101-111, 2008, Epub 2008 May 21.

GENERAL PREADAPTATION OF VIRAL INFECTORS TO THEIR HOSTS

(accompanying file)

by
I. Barrai, G. Salvatorelli, E. Mamolini, S. De Lorenzo,
A. Carrieri, A. Rodriguez-Larralde[1], and C. Scapoli
Department of Biology, University of Ferrara, 44100 Ferrara, Italy
[1]Laboratorio de Genetica Humana, Departamento de Medicina Experimental, IVIC, Caracas

ABSTRACT

We tested the hypothesis of optimal adaptation of viral infectors to their eukaryotic hosts, using 1) linear correlation in codon and aminoacid usage between pairs of organisms, and 2) canonical correlation between groups of hosts and infectors. The frequencies of the 61 codons and of the 20 aminoacids were counted for 69,481,022 codons in 179,770 coding sequences of six hosts, man, pig, chicken, catfish, potato, rice, and seven parasites, HIV1, African swine fever virus, Avian infectious bronchitis virus, Ictalurid herpesvirus, Potato virus X, Rice stripe virus, and Rice blast fungus. This last was included as a control of viral parasites, since it is not an endocellular invader. The codon correlations between parasites and their hosts vary considerably, being very low between Swine and African Swine Fever virus (r=0.18), and high and significant between Potato and Potato virus X (r=0.60). These correlations might indicate different stages of evolution toward optimal adaptation of the parasite codon distribution to the host tRNA pools. The aminoacid correlations are consistently high, varying from r=.71 between pig and ASF, to .88 between catfish and its herpesvirus. It was observed that both in virus and hosts, there is a high negative correlation between frequency of an aminoacid and its molecular weight. Correcting for the dominant effect of molecular weight decreases only slightly most of the correlations involving aminoacids. Therefore it was advanced that viral infectors might be preadapted to their hosts because of the general similarities of the tRNA pools of hosts, and that further evolution toward optimization would be dependent on the size of the divergence between the codon distributions of infector and host. It was also advanced that pre-adaptation does not necessarily imply origin of the virus by lateral transfer from the present host, since the correlation of the molecular weight of aminoacids with their abundance in proteins is a general phenomenon.

Table 1. The database used in the present study.

Organism	Sequences	Codons	Codons/Sequences
1) Hosts:			
Homo sapiens	84949	36349745	428
Sus scrofa	2028	798957	394
Gallus gallus	5579	2541006	455
Ictalurus	255	67311	263
Solanum tuberosum	839	347079	414
Oryza sativa	70055	24804673	354
All host	163705	64908771	397
2) parasites			
Human immunodeficiency 1	14408	3928725	273
African swine fever virus	572	165790	290
Chicken infectious bronchitis virus	426	204853	481
Ictalurid herpes virus	93	38973	419
Potato virus X	91	33487	368
Rice stripe virus	300	86072	287
Rice blast fungus	161	89356	555
All parasites	16051	4547256	283
All 13 organisms	179756	69456027	386

Table 2- Correlations between molecular weight (MW) and degeneracy (DY, coded as 1,2,3,4 and 6) and the frequencies of aminoacid usage in 13 organisms. Six hosts, six viral parasites and an eukaryotic parasite. First row MW, 2nd row degeneracy, 3rd, 4th and 5th row codon effective number of the species.

	MAN	HIV	SUS	ASF	CHK	CHB	IKT	IKV	POT	POX	ORY	RSV	RBF
MW	-.57	-.47	-.57	-.33	-.57	-.61	-.56	-.60	-.58	-.58	-.62	-.48	-.65
DY	.80	.71	.79	.48	.74	.65	.77	.82	.70	.62	.82	.55	.81
N_E	54.1	47.3	51.2	56.8	54.8	46.4	53.6	48.6	52.3	54.5	56.9	49.0	51.4
N_V	49.6	44.6	46.8	45.6	49.2	48.7	48.5	44.6	47.4	48.7	52.3	44.0	47.2
N_Q	54.7	50.4	53.0	52.8	54.8	48.7	54.1	51.1	52.8	53.7	56.4	49.8	53.0

Table 3- Correlations between the frequencies in codon and aminoacid usage in 13 organisms. Six hosts, six viral parasites and an eukaryotic parasite. First row, codons, second row, aminoacids

	MAN	HIV	SUS	ASF	CHK	CHB	IKT	IKV	POT	POX	ORY	RSV	RBF
MAN	1.00	.31	.97	.30	.97	.09	.91	.54	.37	.66	.74	.58	.70
		.87	.88	.69	.99	.81	.95	.88	.94	.85	.93	.79	.94
HIV		1.00	.17	.54	.39	.47	.28	-.15	.57	.53	.08	.46	.02
			.88	.71	.88	.72	.86	.83	.86	.78	.83	.69	.84
SUS			1.00	.18	.93	-.06	.90	.67	.21	.59	.76	.49	.78
				.71	.98	.85	.95	.90	.95	.85	.92	.80	.94
ASF				1.00	.38	.64	.31	-.07	.68	.39	.08	.46	.02
					.73	.75	.78	.61	.86	.71	.52	.84	.84
CHK					1.00	.19	.94	.52	.48	.72	.74	.66	.68
						.83	.97	.88	.97	.88	.88	.83	.94
CHB						1.00	.18	-.36	.82	.32	-.04	.57	-.17
							.90	.86	.90	.78	.77	.84	.83
IKT							1.00	.56	.44	.71	.70	.70	.71
								.88	.97	.87	.88	.90	.91
IKV								1.00	-.13	.33	.71	.18	.72
									.88	.87	.92	.76	.93
POT									1.00	.60	.27	.76	.09
										.87	.88	.90	.92
POX										1.00	.58	.72	.54
											.83	.68	.88
ORY											1.00	.43	.86
												.68	.97
RSV												1.00	.32
													.76
RBF													1.00

MAN=Homo sapiens HIV=Himmunodeficiency virus type 1
SUS=Sus scrofa ASF=African swine fever virus
CHK=Gallus gallus CHB=Chicken infectious bronchitis virus
IKT=Ictalurus sp. IKV=Ictalurid herpes virus
POT=Solanum tuberosum POX=Potato virus X
ORY=Oryza sativa RSV=Rice stripe virus
RBF=Rice blast fungus

Table 4. Percent of variance due to the first three eigenvalues of the dispersion matrices of frequencies in 1) hosts alone 2) parasite alone 3) all the organisms studied. Rice blast fungus excluded.

A) CODONS

Component	Hosts	Parasites	Both
1	76.9	50.6	54.0
2	14.7	24.1	24.8
3	5.1	9.9	5.6
First 3 components	97.7	84.6	84.4

B) AMINOACIDS

Component	Hosts	Parasites	Both
1	95.5	80.7	85.8
2	2.5	7.8	5.9
3	1.1	5.0	2.6
First 3 components	99.1	93.5	94.3

Table 5. Correlations in frequencies between Hosts and Infectors.

a) Linear correlation between individual hosts and their infector

Codons	Aminoacids		Host	Infector
	+MW	-MW		
0.60*	0.87*	0.80*	Potato	Potato virus X
0.56*	0.88*	0.82*	Catfish	Ictalurus virus
0.43*	0.68*	0.56*	Rice	Rice stripe virus
0.31*	0.87*	0.84*	Man	HIV1
0.19	0.83*	0.75*	Chicken	Chicken infect. bronch. virus
0.18	0.71*	0.68*	Pig	African swine fever
0.86*	0.97 *	0.96 *	Rice	Rice blast fungus

-MW: Frequencies corrected for the effect of molecular weight.

c) Canonical correlations between hosts and intracellular infectors

	A. Codons	B. Aminoacids	
		+MW	-MW
First axis	.939	.992	.989
Second axis	.888	.919	.909
Third axis	.516	.662	.661

Legends to figures

Figure 1.

The regressions of the aminoacid frequencies on molecular weight for the organisms studied here.

Figure 2.

Correlations of codon and aminoacid frequencies for the pairs Pig-ASF (2a,2b), Man-HIV1 (2c,2d), and Potato-PVX (2e,2f). Under the Subak-Sharpe hypothesis, the system Pig-ASF is the most recent and the system Potato-PVX the most ancient.

Figure 3.

Dendrograms of infectors and hosts from the correlation matrices for codons (3a) and for aminoacids (3b). Note the cluster Man-Chicken for codons and the cluster Man-Pig for aminoacids.

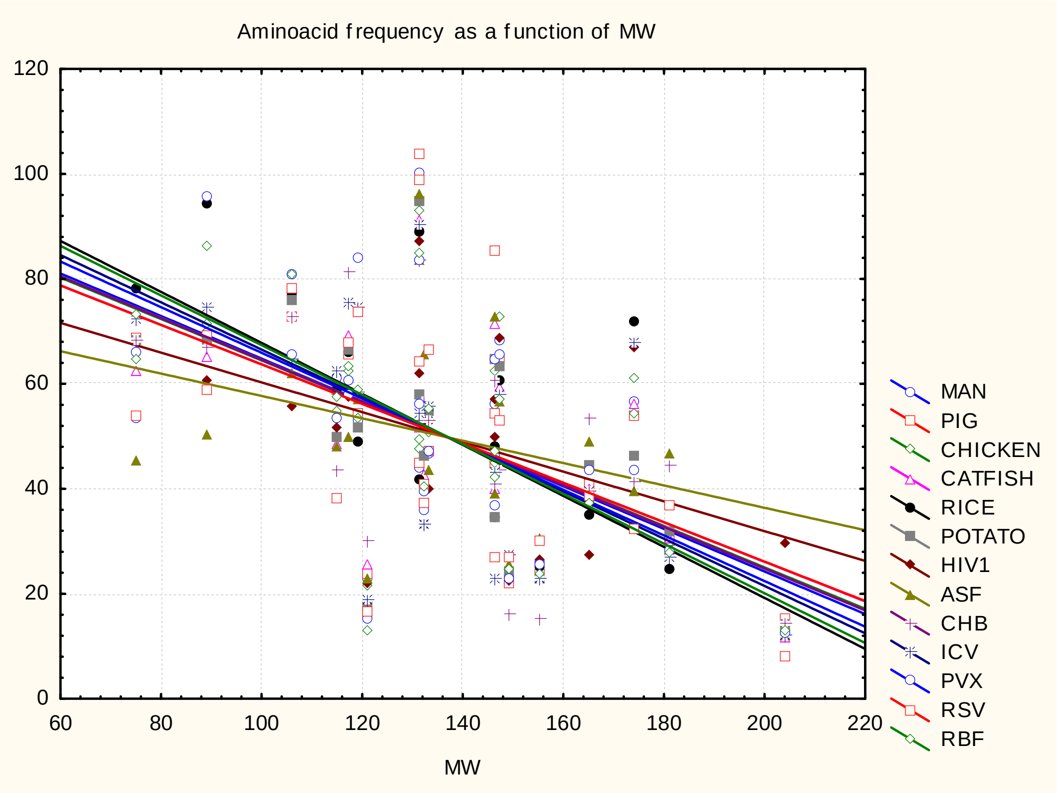
Figure 4.

Correlations between the projections on the first axes of the codon swarms for hosts and infectors.

Figure 5.

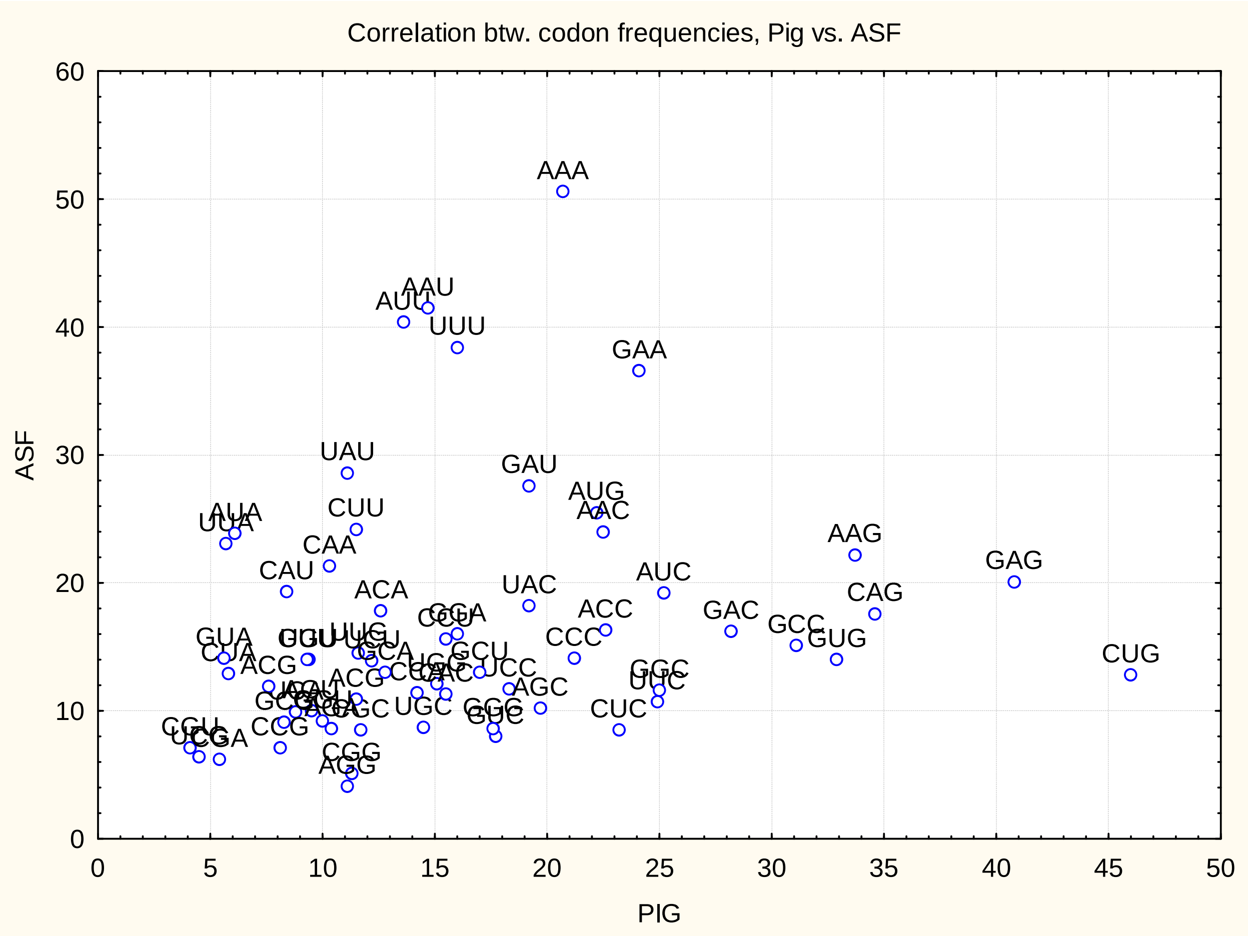
Correlations between the projections on the first axes of the aminoacid swarms for hosts and infectors, before (5a) and after (5b) correction for the effect of molecular weight. Note the considerable effect of MW on the position of triptophan, glycine, and alanine.

Figure 1



The regressions of the aminoacid frequencies on molecular weight for the organisms studied here.

Figure 2a


Correlation btw. codon frequencies, Pig vs. ASF
ASF
PIG
0
10
20
30
40
50
60
0
5
10
15
20
25
30
35
40
45
50
AAA
AAU
AUU
UUU
GAA
UAU
GAU
AUG
AAC
AUA
UUA
CUU
CAA
AAG
GAG
CAU
UAC
AUC
ACA
CGA
CCU
ACC
GAC
CAG
GCC
GUG
CUG
GUA
CUA
GGU
UUG
UCU
GCA
CCC
GCU
ACG
UGG
CCA
UAC
UCC
GGC
UUC
AGC
ACG
GCG
AGU
CGC
UGC
GGG
GUC
CUC
CCG
CGU
UCG
CGA
CGG
AGG

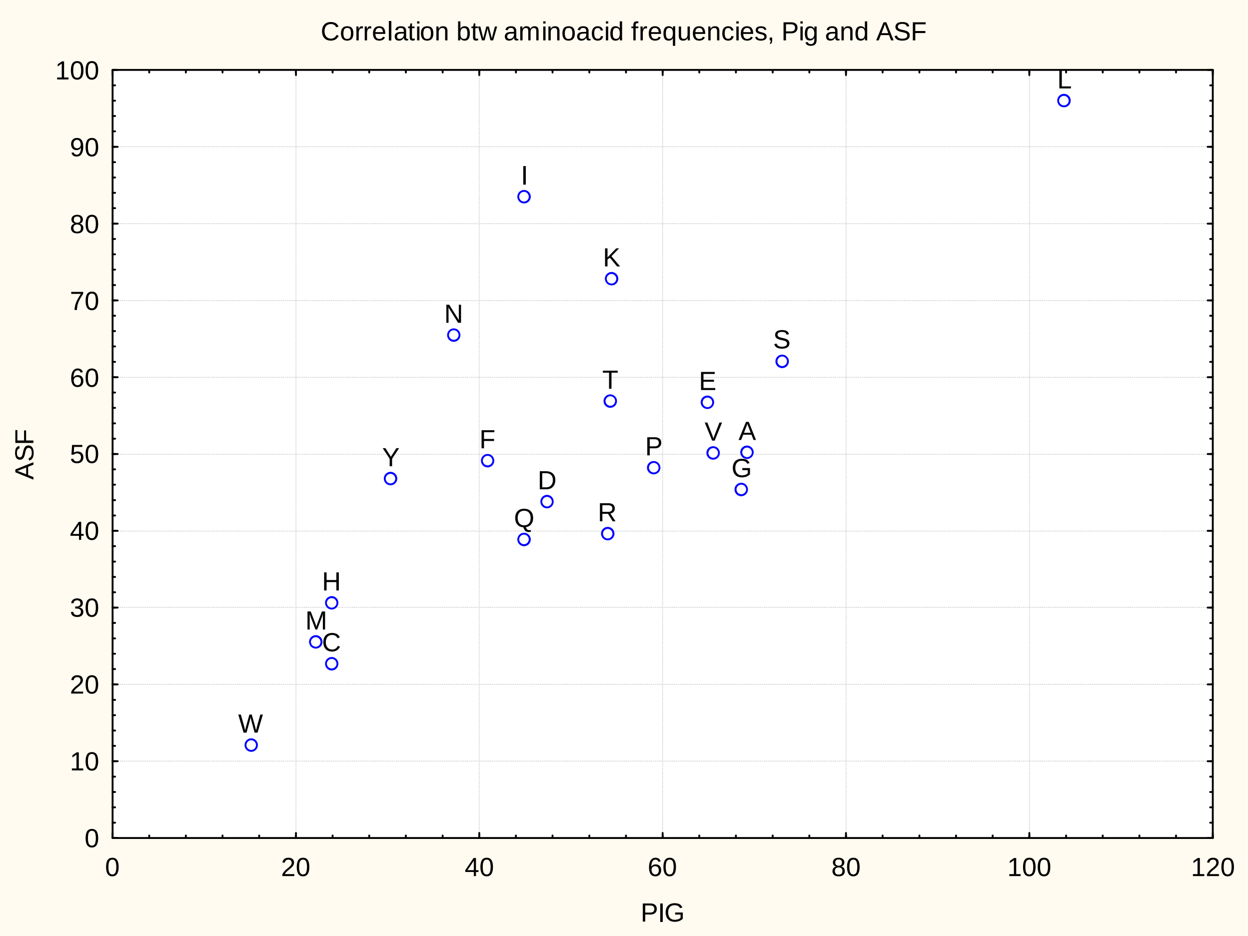

Correlation btw aminoacid frequencies, Pig and ASF
ASF
PIG
0
10
20
30
40
50
60
70
80
90
100
0
20
40
60
80
100
120
L
I
K
N
S
T
E
F
P
V
A
Y
G
D
R
Q
H
M
C
W

Figure 2b

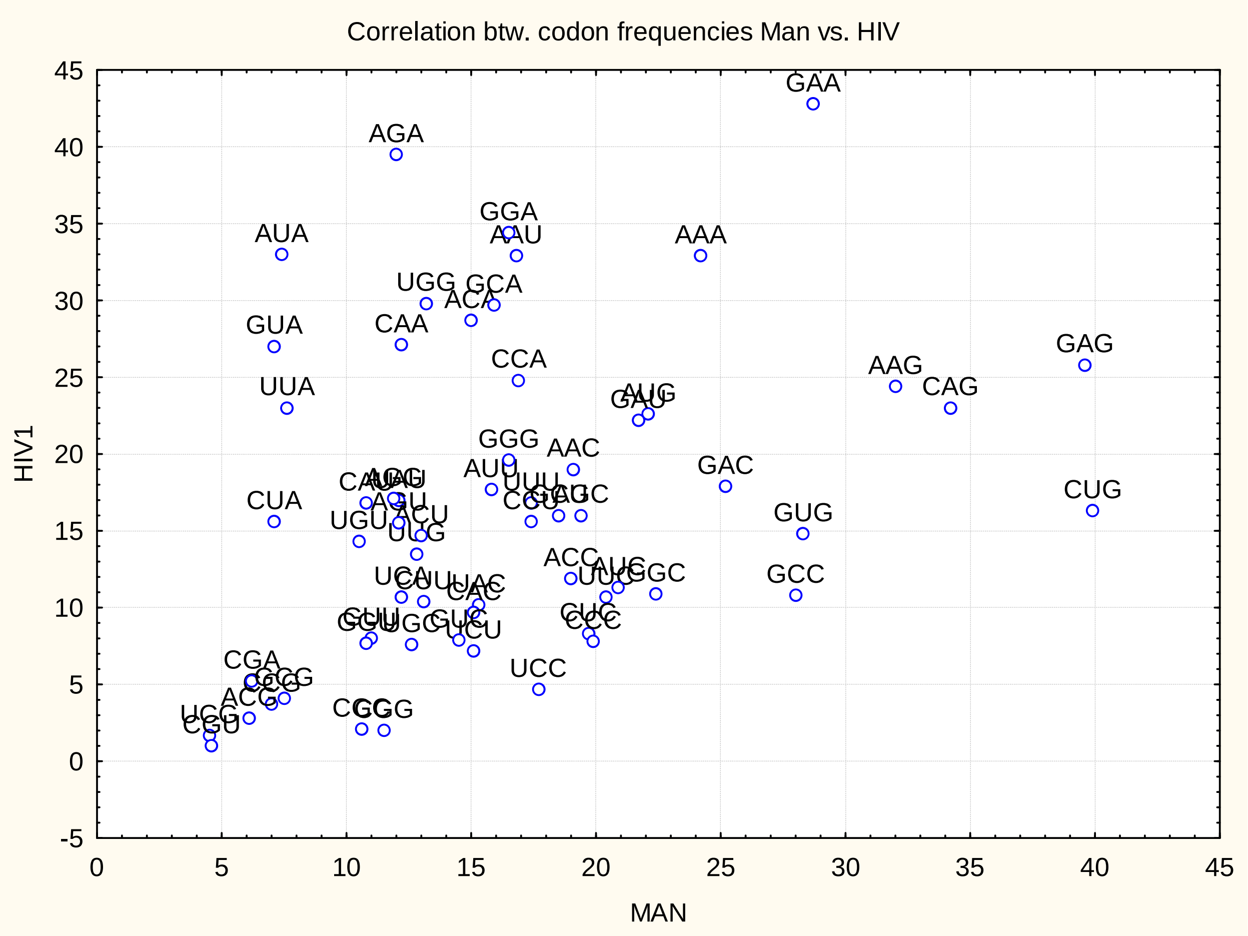

Correlation btw. codon frequencies Man vs. HIV
HIV1
MAN
45
40
35
30
25
20
15
10
5
0
-5
0
5
10
15
20
25
30
35
40
45
GAA
AGA
GGA
AAU
AUA
AAA
UGG
GCA
ACA
GUA
CAA
GAG
CCA
AAG
CAG
UUA
AUG
GAU
GGG
AAC
GAC
CUG
CAU
AGG
UAU
AUU
UUU
GCU
AGC
CCU
AGU
CUA
ACU
UGU
UUG
GUG
ACC
AUC
UUC
GGC
GCC
UCA
CUU
UAC
CAC
GUU
GGU
UGC
GUC
UCU
CUC
CCC
CGA
CGC
CCG
GCG
UCC
ACG
UCG
CGU
CGG
GCG

Figure 2c

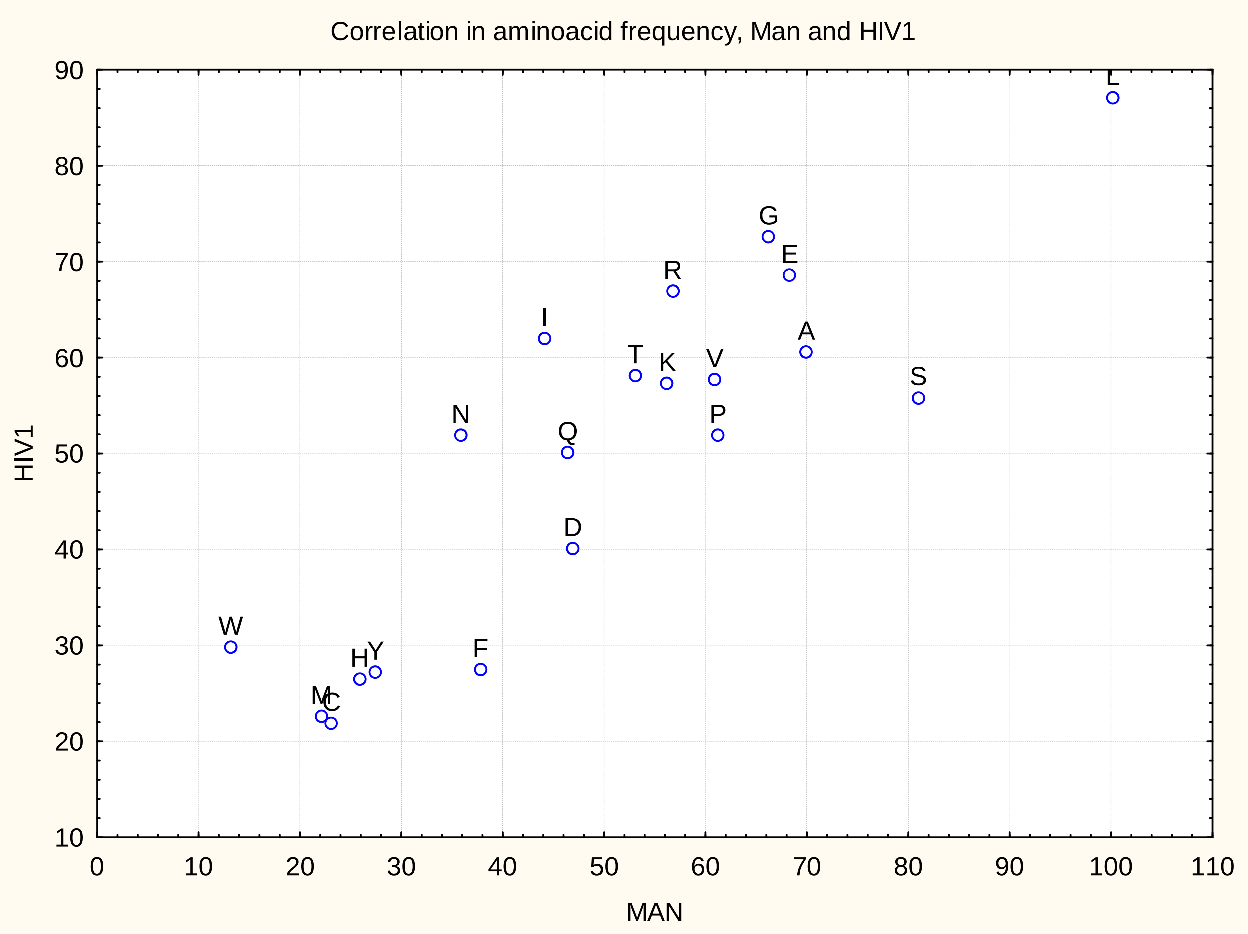

Correlation in aminoacid frequency, Man and HIV1
HIV1
MAN
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
110
L
G
E
R
I
A
T
K
V
S
N
P
Q
D
W
H
Y
F
M
C

Figure 2d

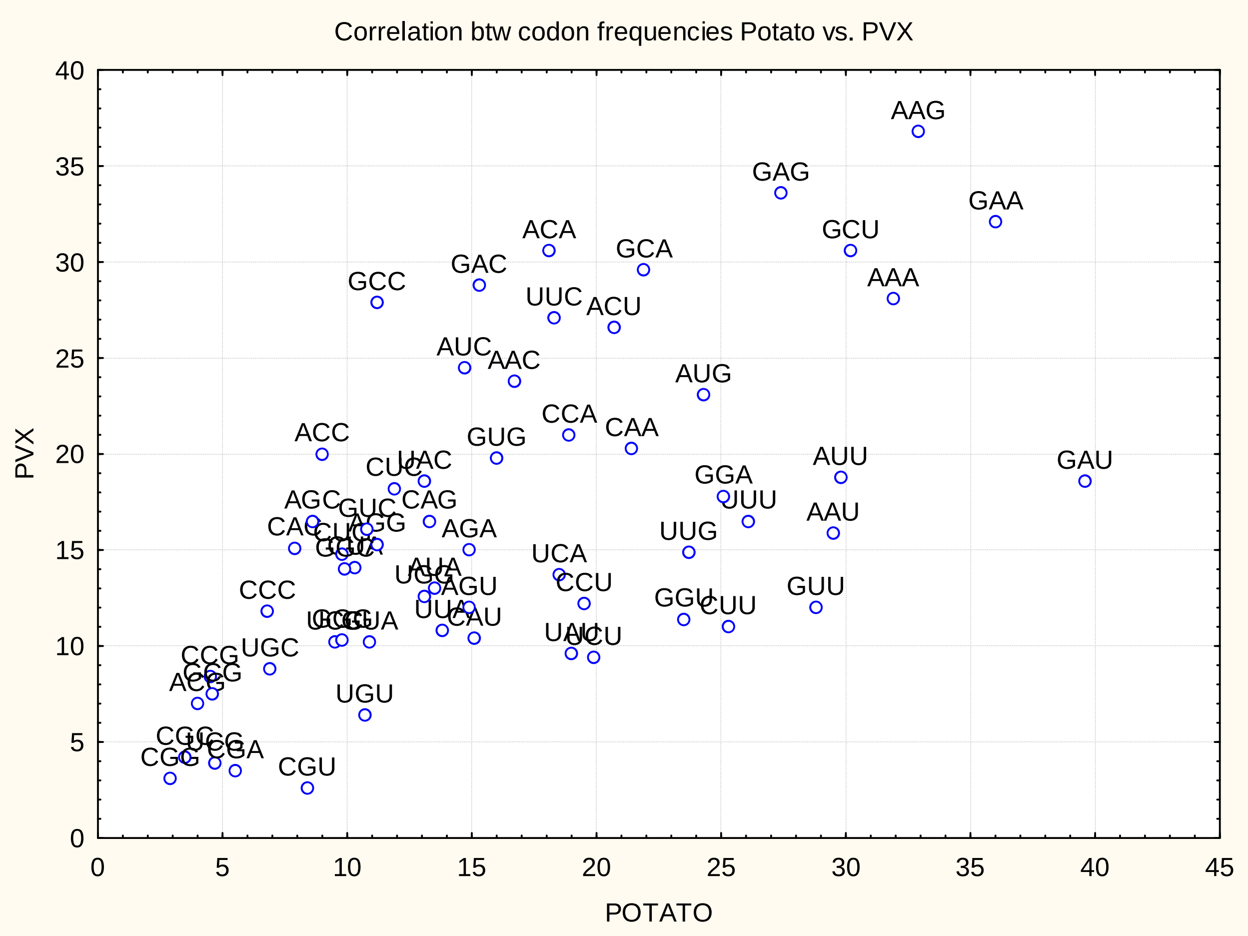

Correlation btw codon frequencies Potato vs. PVX
PVX
POTATO
0
5
10
15
20
25
30
35
40
45
AAG
GAG
GAA
ACA
GCU
GCA
GAC
AAA
GCC
UUC
ACU
AUC
AAC
AUG
CCA
CAA
ACC
GUG
CUC
UAC
AUU
GAU
GGA
AGC
UUU
GUC
CAG
AGG
AAU
CAC
AGA
UUG
UCA
AUA
UGG
AGU
CCU
CCC
GGU
GUU
CUU
UUA
CAU
GUA
UAU
UCU
UGC
CCG
CGG
ACG
UGU
CGC
UCG
CGA
CGG
CGU

Figure 2e

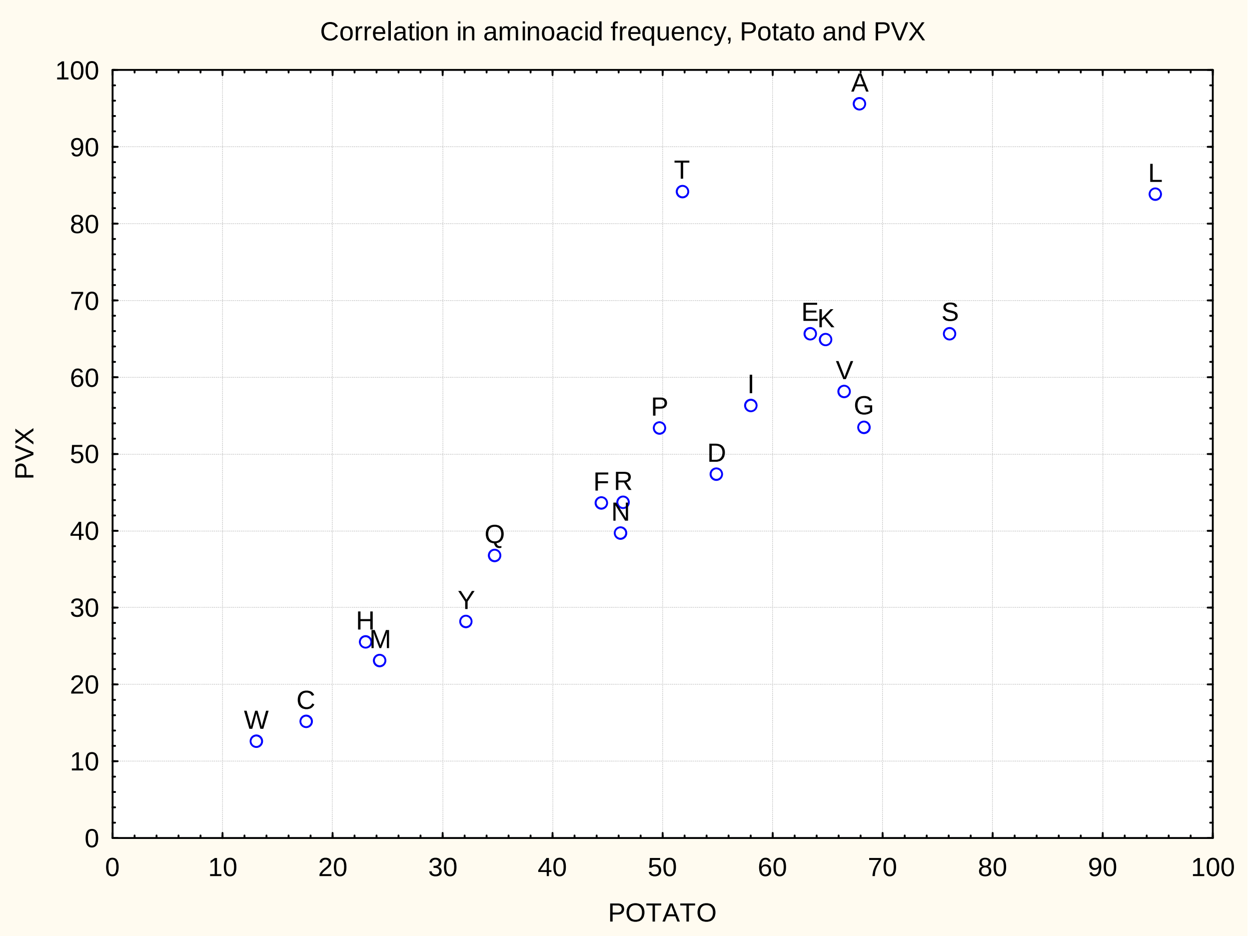


Figure 2f

Correlations of codon and aminoacid frequencies for the pairs Pig-ASF (2a,2b), Man-HIV1 (2c,2d), and Potato-PVX (2e,2f). Under the Subak-Sharpe hypothesis, the system Pig-ASF is the most recent and the system Potato-PVX the most ancient.

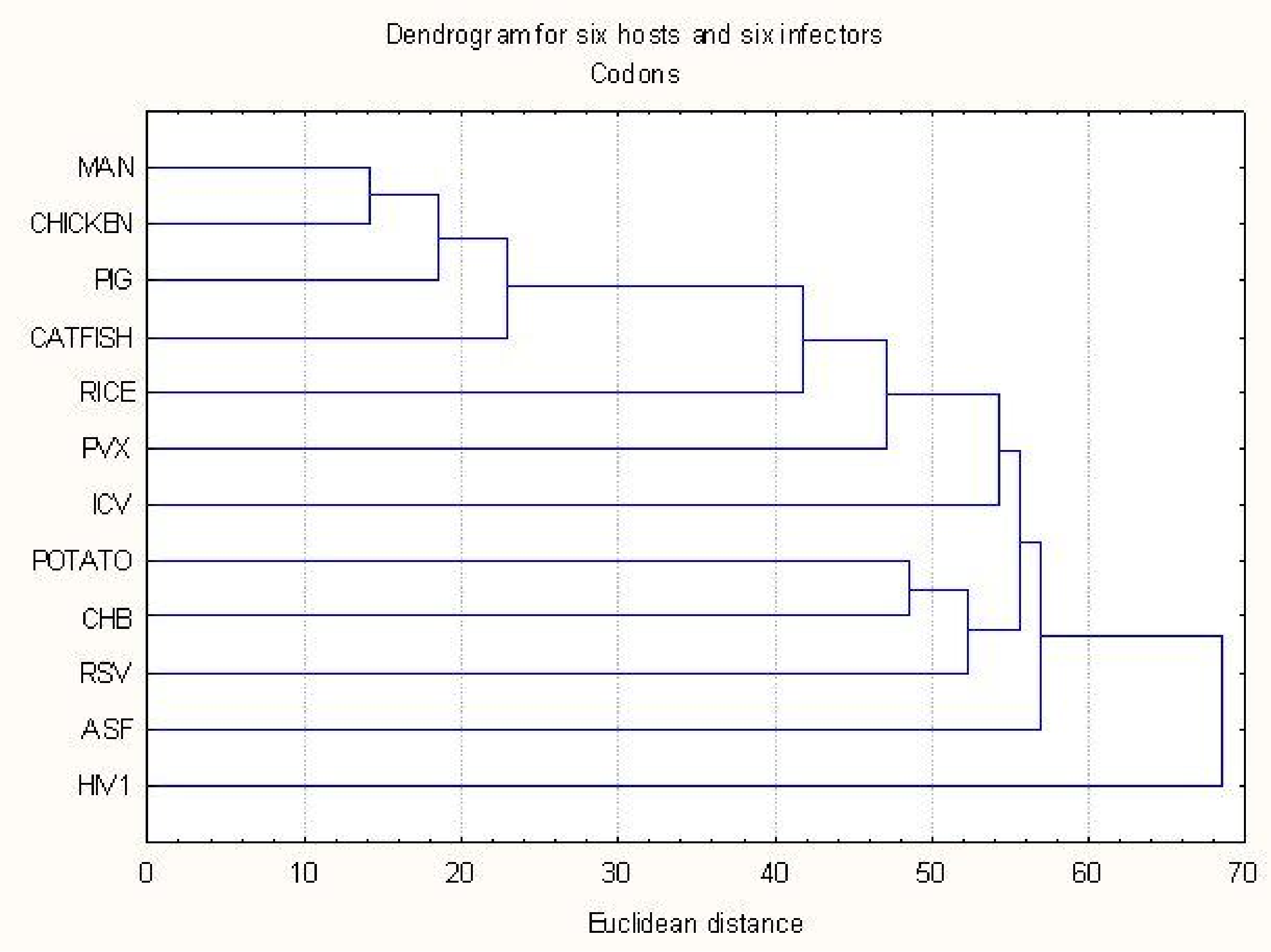

Dendrogram for six hosts and six infectors
Codons
MAN
CHICKEN
PIG
CATFISH
RICE
PVX
ICV
POTATO
CHB
RSV
ASF
HIV1
0
10
20
30
40
50
60
70
Euclidean distance

Figure 3a

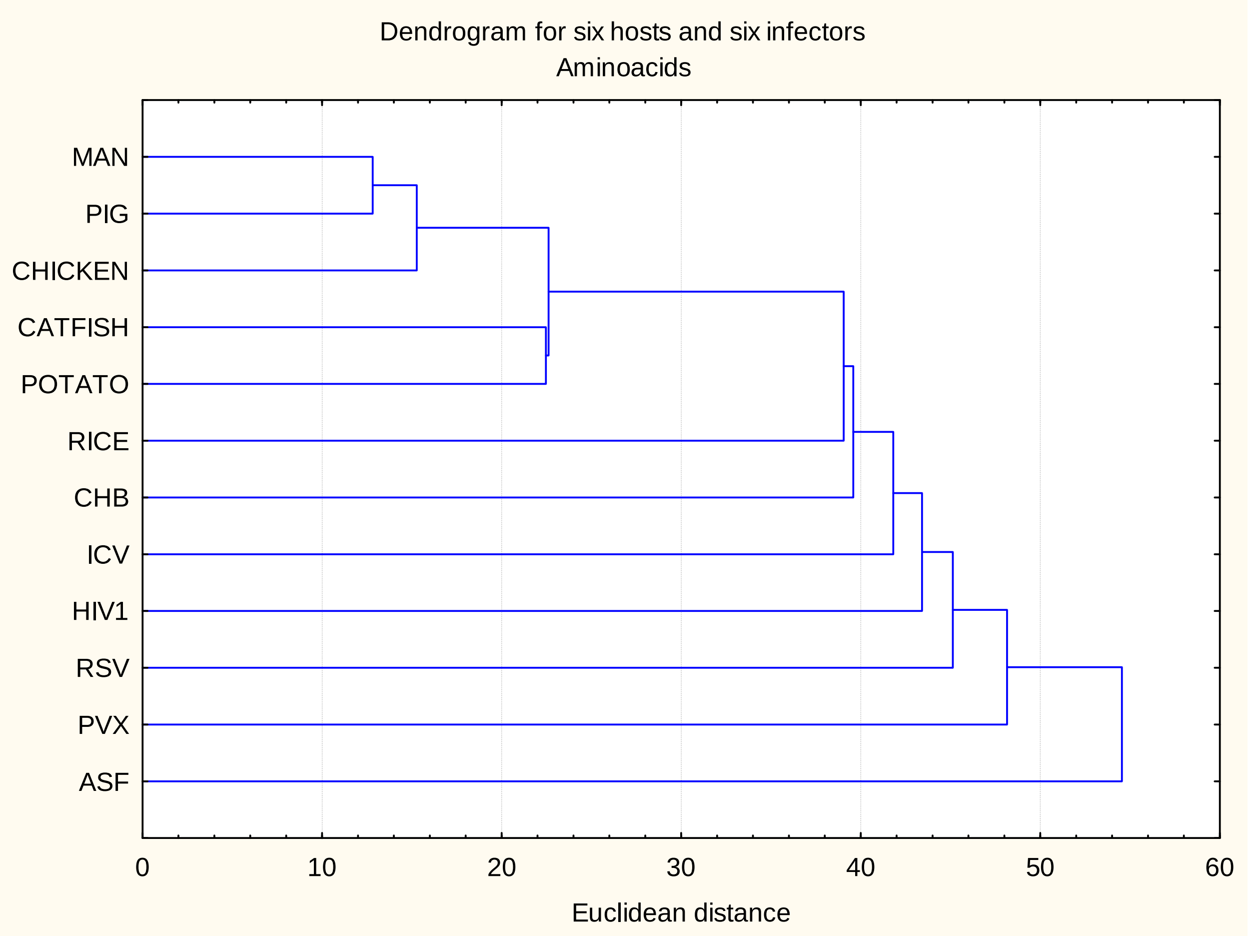


Figure 3b

Dendrograms of infectors and hosts from the correlation matrices for codons (3a) and for aminoacids (3b). Note the cluster Man-Chicken for codons and the cluster Man-Pig for aminoacids.

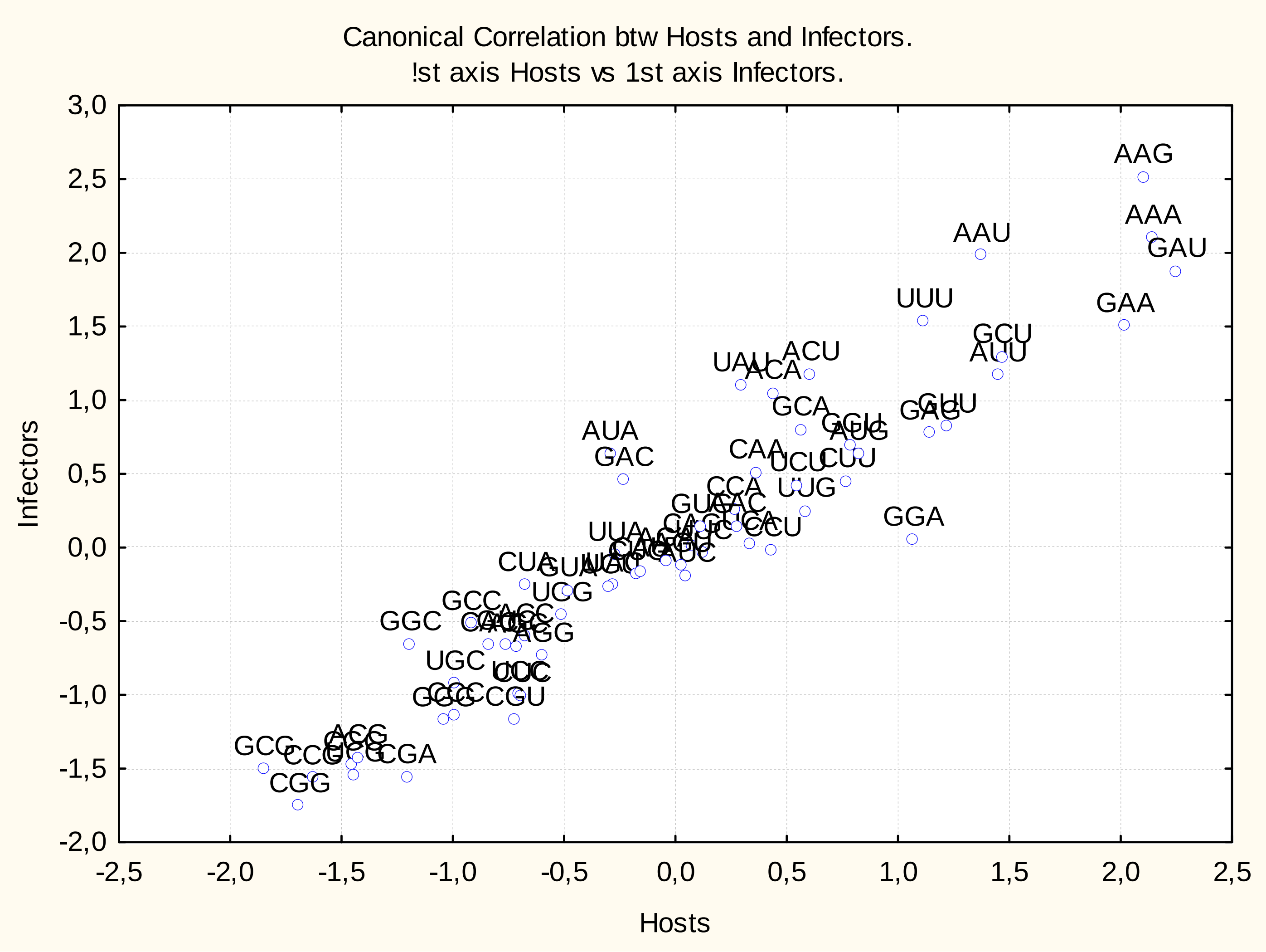


Figure 4

Correlations between the projections on the first axes of the codon swarms for hosts and infectors.

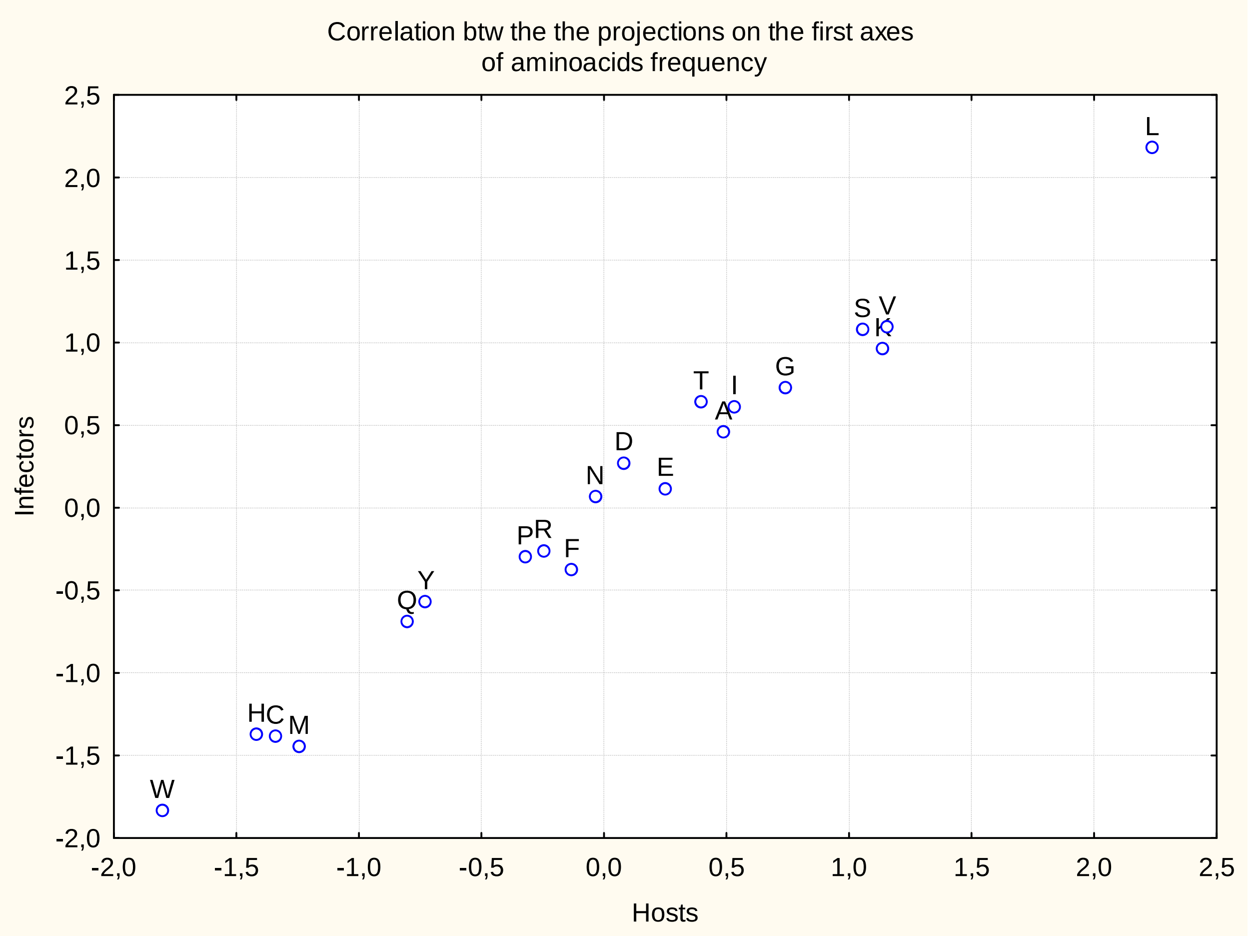

Correlation btw the the projections on the first axes
of aminoacids frequency
Infectors
Hosts
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0
2,5
L
S
V
K
G
T
I
A
D
N
E
P
R
F
Y
Q
H
C
M
W

Figure 5a

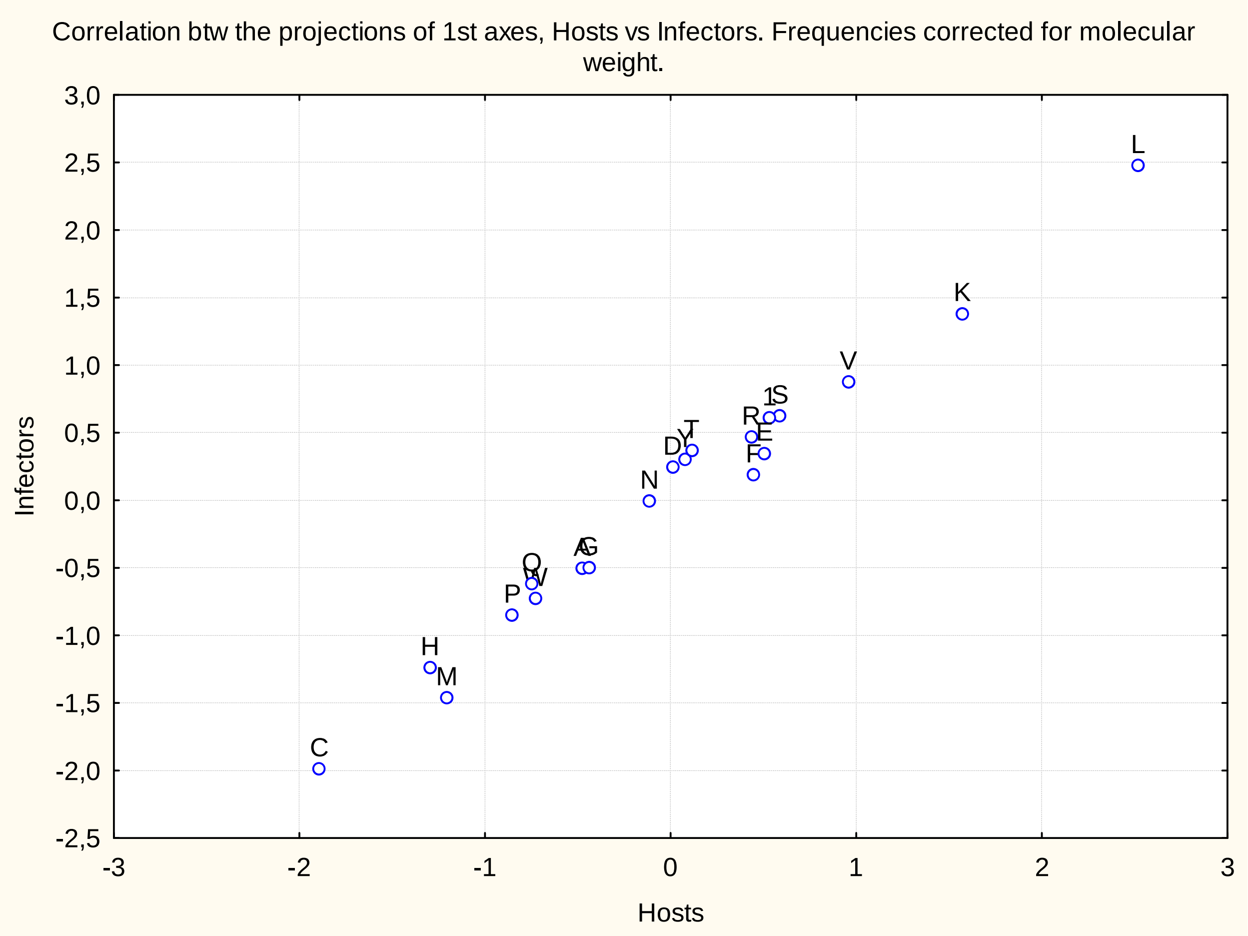


Figure 5b

Correlations between the projections on the first axes of the aminoacid swarms for hosts and infectors, before (5a) and after (5b) correction for the effect of molecular weight. Note the considerable effect of MW on the position of triptophan, glycine, and alanine.