

Aggregation of Homologous Protein Families (HPFs)

FOR MAPPING THEM ONTO AN EVOLUTIONARY TREE

Boris Mirkin¹, Renata Camargo¹, Trevor Fenner¹, George Louizou¹ and Paul Kellam²

¹School of Computer Science and Information Systems, Birkbeck University of London

²Department of Infection, Winer Institute of Medical Sciences, University College London

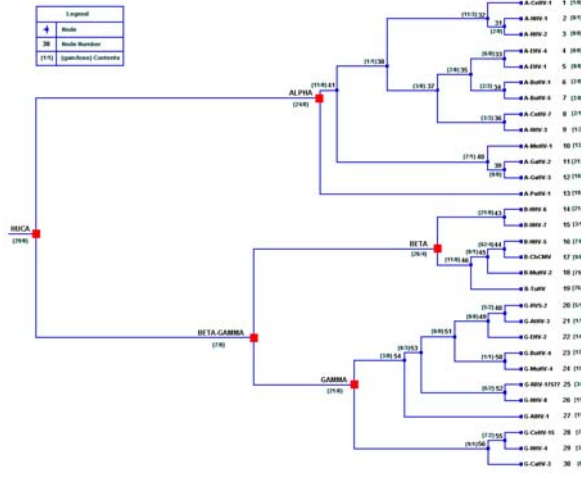
Input

DATA: 30 complete herpesvirus genomes extracted from the virus database VIDA^[1], and 740 homologous protein families (HPFs) representing the orthologous genes from herpesvirus genomes.

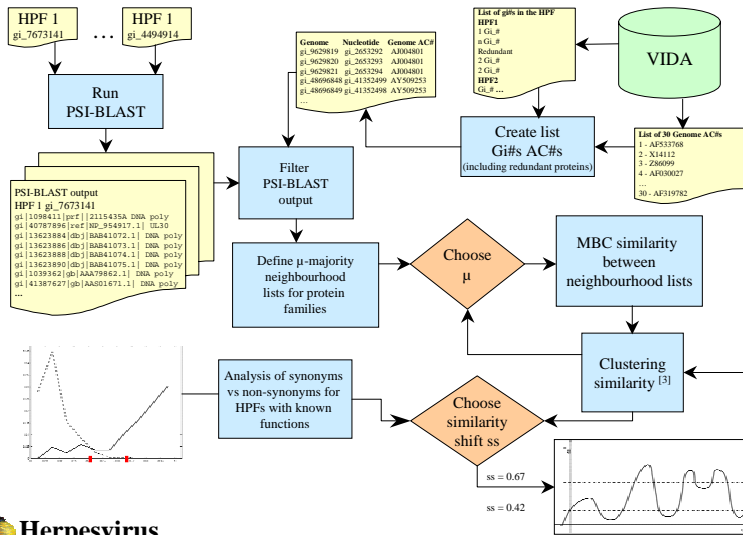
AIM: Mapping the histories of HPFs onto the evolutionary tree of the herpesviruses, and analysing the evolution of function.

ISSUE: HPFs are based on solitary conserved fragments and should be aggregated to reflect multi-functional nature of some proteins.

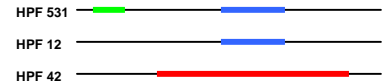
Annotated Evolutionary Tree



Aggregation Method



The Structure of an Aggregate Protein Family (APF)



Results

- Annotation of the tree with parsimonious gain and loss events for 740 HPFs.^[2]
- No difference in contents of HUCA, α , β , γ reconstructions for $ss=0.42$ and $ss=0.67$, except for the glycoproteins C, I and L that are absent from α at $ss=0.67$ and present at $ss=0.42$.
- Composition of HUCA broadly supports the reconstruction by Davison^[4].
- The function of 70% of the contents of the β , γ , β and γ ancestors are unknown.

Future Work

- Interpretation of functional differences between α , β and γ subfamilies.
- Fine tuning the solutions using the principle of maximum likelihood.

Herpesvirus Genomes

α subfamily

Mapping	HPF	Gene	Davison's List ^[4]
Control and modulation			
HUCA	29	UL13	Serine-threonine protein kinase
HUCA	APF 67	UL54	Post-transcriptional regulator
DNA replication machinery			
HUCA	16	UL5	Component of DNA helicase-primase
HUCA	15	UL29	Single-stranded DNA-binding
HUCA	1	UL30	Catalytic subunit of DNA polymerase
HUCA	32	UL52	Component of DNA helicase-primase

β subfamily

ALPHA	72	UL8	Component of DNA helicase-primase
ALPHA	APF 11	UL9	Binds to origins of DNA synthesis
NODE 41	APF 43	UL42	Processivity subunit of DNA polymerase
Peripheral enzymes			
HUCA	8	UL2	Uracil-DNA glycosylase
HUCA	24	UL39	Ribonucleotide reductase (large)
HUCA	33	UL40	Ribonucleotide reductase (small)
HUCA	APF 10	UL23	Thymidine Kinase
HUCA	43	UL50	dUTPase

γ subfamily

Processing and packaging of DNA			
HUCA	28	UL6	Minor capsid protein
HUCA	APF 12	UL12	Doxorubicinase
HUCA	APF 13	UL15	ATPase subunit of terminase
HUCA	APF 14	UL25	-
HUCA	7	UL28	Subunit of terminase
HUCA	34	UL32	-
HUCA	39	UL33	-
ALPHA	108	UL17	Tegument protein

Herpesvirus Subfamily Ancestors

Mapping	HPF	Gene	Davison's List ^[4]
Capsid assembly and structure			
HUCA	25	UL18	Component of interspersed triplex
HUCA	19	UL19	Major capsid protein
HUCA	APF 1	UL26	Protease&minor capsid scaffold protein
HUCA	APF 1	UL26.5	Major capsid scaffold protein
ALPHA	77	UL38	Component of interspersed triplex
NODE 41	129	UL35	Located on tips of hexons
Egress of capsids from nucleus			
HUCA	17	UL31	Nuclear matrix protein
ALPHA	132	UL34	Inner nuclear membrane protein
Tegument			
HUCA	23	UL7	-
HUCA	APF 32	UL16	-
HUCA	APF 76	UL36	-
ALPHA	79	UL14	-
ALPHA	96	UL37	-
NODE 32	204	UL11	Myristylated protein
ALPHA	APF 39	UL51	-
Surface and Membrane			
HUCA	20	UL10	Glycoprotein M
HUCA	3	UL27	Glycoprotein B
HUCA	APF 3	UL22	Glycoprotein H
HUCA	267	UL9A	Glycoprotein N
NODE 41	47	UL1	Glycoprotein L
Unknown			
HUCA	37	UL24	-

Acknowledgements

Wellcome Trust (Funding 2004–2006)

Dr Mar Alba (Universitat Pompeu Fabra, Barcelona, Spain)

Dr Eugene Koonin (NCBI, NLM, NIH, Bethesda, USA)

References

- M. Alba, D. Lee, F. Pearl, A. Shepherd, N. Martin, C. Orengo and P. Kellam (2001) VIDA: a virus database system for the organization of animal virus genome open reading frames, *Nucleic Acids Research*, 29(1), 133-136.
- B. Mirkin, T. Fenner, M. Galperin and E. Koonin (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes, *BMC Evolutionary Biology*, 3:2.
- B. Mirkin (2005) *Clustering for Data Mining: A Data Recovery Approach*, Chapman & Hall/CRC.
- A. Davison (2002) *Evolution of the herpesviruses*, *Veterinary Microbiology* 86, 69-88.