THE MOLECULAR AND EVOLUTIONARY DETERMINANTS OF HOST-SWITCHING VIRUSES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Karin Hoelzer

May 2009

THE MOLECULAR AND EVOLUTIONARY DETERMINANTS OF HOST-SWITCHING VIRUSES

Karin Hoelzer, Ph. D.

Cornell University 2009

Emerging infectious diseases represent imminent threats to human and animal health worldwide and can impose immense economic burdens, particularly on the world's most impoverished regions. Cross-species transmission of pathogens represents a common path towards disease emergence, but the mechanisms that lead to successful host jumping are incompletely understood.

Here I have addressed some of the pathogen-associated factors involved in successful host jumping and have analyzed the evolutionary dynamics of the novel pathogen in the new host after cross-species transfer. These studies were performed using two emerging diseases of dogs: Canine Parvovirus (CPV) and Canine Influenza virus (CIV). While CPV emerged in the mid-1970s from an endemic virus of cats and has since become endemic in the global dog population, CIV was first described as a pathogen of dogs in 2004 - thus representing a new, likely not completely host-adapted virus. Contrasting the evolutionary dynamics of these two pathogens which share the same host species allowed a powerful analysis of the determinants and risk factors for disease emergence. In particular, I focused these studies on understanding the viral population dynamics in infected animals - contrasting where applicable the dynamics in the novel and the ancesteral host species, and dissecting host-associated effects and the impact of cross-immunity. For the case of CPV, I further analyzed the

evolutionary dynamics of the novel virus on a population level since its emergence - contrasting those to the dynamics observed in the ancesteral virus. Finally, I analyzed the effect of potentially host-specific codon usage and codon usage bias for the evolution of CPV. Because CpG methylation might represent an important driver of codon bias in the CPV genome, some work was dedicated to obtain a more comprehensive understanding of this major driver in DNA virus evolution.

The results presented here show similarities but also marked differences between the evolutionary dynamics of CPV and CIV, and between the novel and ancesteral viruses. Further research is needed but it appears that key drivers differ between pathogens and between time intervals after emergence. However, some common mechanisms appear to be shared between viruses and others appear conserved between hosts.

BIOGRAPHICAL SKETCH

Karin Hoelzer received her Veterinary Medical degree (D.V.M. equivalent) from the University of Veterinary Medicine in Hannover, Germany in 2005. Karin joined Colin Parrish's group at Cornell University that same year, returning to the laboratory group that had hosted her during her participation in the "Leadership Program for Veterinary Students" in 2001. Her research focuses on the molecular and evolutionary mechanisms that allow viruses to change hosts, aiming to better predict and thus prevent emerging infectious diseases in the future.

To Stefan.

῾ότι δ῾ ῾οίόμενοι δεῖν ζητεῖν ῾ά μή τις ῾οῖδεν βελτίους ῾άν ῾εῖμεν καί ῾ανδρικώτεροι καί ῾ήττον ῾αργοί ῾ή ῾ει ῾οιοίμεθα ῾α μή ῾επιστάμεθα μηδέ δυνατόν ῾εῖναι ῾ευρεῖν μηδέ δείν ζητεῖν περί τούτου πάνυ ῾άν διαμαχοίμην, ᾽έι ῾οῖός τε᾽ ῾είην, καί λόγω καί᾽ έργω

Plato, Menon 86b,c

ACKNOWLEDGMENTS

This dissertation would not have been possible without the untiring support and encouragement of many faculty members, staff, friends and family. First and foremost I thank my graduate advisor, Dr. Colin R. Parrish, for his continuous support, guidance and encouragement throughout the years. His inspiring enthusiasm for scientific research and his ceaseless scientific curiosity have inspired me and he has taught me countless valuable lessons that I will doubtlessly cherish throughout the years to come.

Furthermore, I thank Dr. Edward C. Holmes at the Pennsylvania State University for being a mentor and an incredible source of knowledge and support. Without his contribution, none of the work presented here would have been possible and I am very grateful for his tireless support, motivation and insight.

Moreover, I am very grateful to Dr. Hollis N. Erb, Dr. Carlos D. Bustamante and Dr. Martin Wiedmann for serving on my academic committee and for their continuous support, advice and insights over the years.

Likewise I am very grateful to Dr. Klaus Osterrieder, Dr. Cynthia Leifert and Dr. Edward Dubovi as well as Dr. James Wood at Cambridge University for their collaborative support throughout the years. I have tremendously benefitted from their help and advice and they have had major impacts on the studies presented here.

Over the years I have also been fortunate to be able to collaborate with several extraordinary colleagues at Cornell, throughout the United States and abroad and I am indebted to each and every one of them. In particular, I thank Dr. Laura A. Shackelton for her inspiring friendship and her incredible support, especially as I took my first steps in the scientific field. Likewise, I am extraordinarily grateful to Dr. Stephan

Metzger, Dr. Pablo Murcia and Dr. Greg Baillie for the part they played in enabling extremely fruitful collaborations. Moreover, I have been particularly fortunate to have the support of extraordinary group-mates, colleagues and friends at the James A. Baker Institute for Animal Health: Laura Palermo, Christian Nelson, Karla Stucker, Leela Noronah, Yeun Hee Kim, Meagan Wisnewski, Vimal Selvara, Atsushi Asano and Caroline Coffey to name just a few. Their support and friendship helped me excel academically and also enriched my graduate studies in so many other ways.

I also thank all the wonderful support staff and undergraduate students that I have had the privilege to work with at the Baker Institute. First and foremost I thank Virginia Scarpino and Wendy Weichert for their tireless and unceasing support, which was truly instrumental to the research presented here. I also thank the students that have tremendously contributed to my work and from whom I have learned how to teach; in particular Jason Kaelber whose enthusiasm and fascination for research inspires me. Finally, I thank Dr. Douglas McGregor and Dr. Susan Bliss for their continuous friendship, advice and support. I am certain their insights will continue to serve as valuable assets throughout my career.

Last but not least I thank my family and close friends for their caring affection and unconditional support. And I am extremely grateful to Stefan for his ceaseless love, understanding and endless support which guides my path through life.
Finally, I thank the Cornell University College of Veterinary Medicine and the Morris Animal Foundation for providing salary support during my PhD studies. The research presented here was supported by MAF grant D08FE-403 to K.H. and NIH grants GM080533 to E.C. H. and AI028385 to C.R.P.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

HOST JUMPING AND DISEASE EMERGENCE – A UNIFYING VIEW

1.1. The global importance of emerging infectious diseases

Emerging infectious diseases (EIDs) represent important economic and public health burdens worldwide, which challenge particularly the world's most impoverished regions (5, 33). Some evidence suggests that the frequency of EIDs has significantly increased over past decades, and anthropogenic ecosystem perturbations - manifested for example as climate change or deforestation - might be partially responsible, even though this topic is still subject to debate (22, 23, 30, 32, 48). As our general understanding in the area of disease emergence is growing, several common risk factors begin to be indisputably recognized (see for example (1, 6-8, 11, 20)). Numerous socio-economic and ecological factors clearly influence emergence probabilities, but their nature and relative impact appear strongly correlated with the EID origin (22).

1.2. The different paths to disease emergence

The World Health Organization (WHO) defines EIDs as diseases which recently increased in incidence, or for which the incidence is likely to increase in the near future, thus encompassing three distinct origins of emerging diseases: (i) novel diseases recognized for the first time, (ii) diseases that were previously controlled through chemotherapeutic drugs or (iii) diseases with changed distribution pattern (Figure 1.1). With or without the acquisition of genetic mutations, diseases might suddenly appear in new geographic locations, might increase in incidence due to behavioral changes in the host population, or might become more frequent due to an increase in immune-suppressed hosts. The probably best-recognized examples thereof include the recent West Nile and Lyme disease epidemics in the Western Hemisphere and the global increase of opportunistic infections associated with the HIV epidemic (4, 13, 16, 33, 35). These diseases, frequently caused by parasites or vector-borne

viruses, appear strongly impacted by anthropogenic ecosystem perturbations or naturally occurring climate changes such as the El Nino/ Southern Oscillation phenomenon (22, 34).

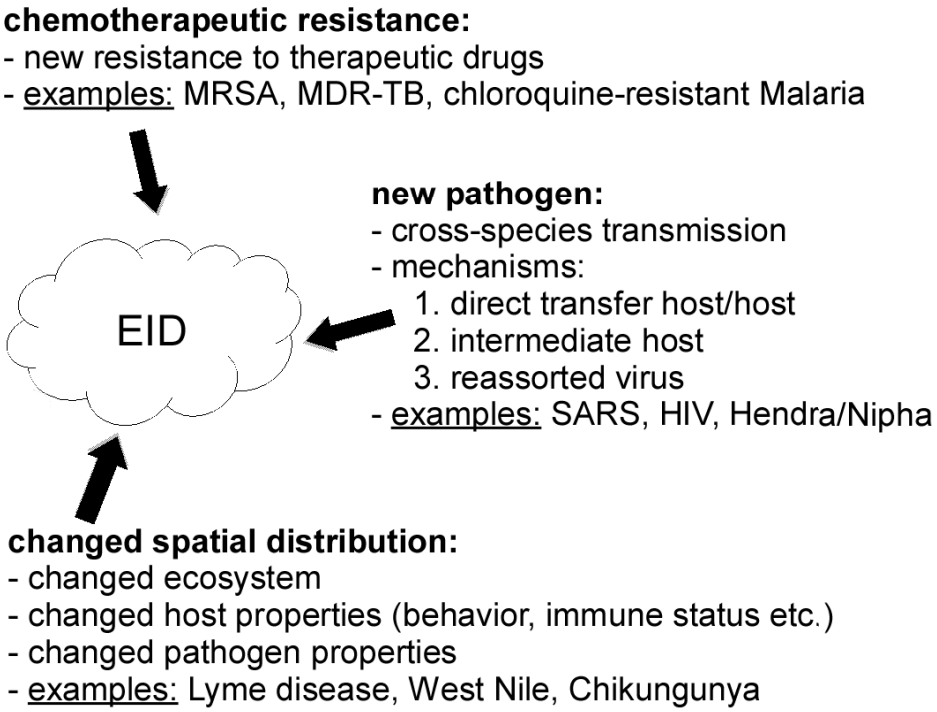


Figure 1.1. The different mechanisms of disease emergence and some of the important drivers associated with the route of emergence.

Alternatively, disease incidence can increase as a consequence of resistance to antimicrobial drugs, for example multi-resistant tuberculosis, methicillin - resistant Staphylococcus aureus or chloroquine-resistant malaria (see for example (14, 21, 29, 36, 40)). The emergence of resistance to antimicrobial drugs can in many cases be linked to their extensive - and often inadequate - use. High population density, agricultural practices which favor the use of chemotherapeutic drugs, and suboptimal antibiotic treatment regimes therefore represent important risk factors for this category of emerging diseases (22). Many of the pathogens in this category are bacteria, even though some important exceptions do exist.

1.3. Disease emergence after cross-species transmission

In some other cases, novel diseases emerge after cross-species transmission from an alternative host. Intriguingly, nearly all EIDs that result from such host jumps are caused by viruses - a phenomenon frequently attributed to their comparably small genome sizes, relatively high mutation rates, large population sizes and, in most cases, their inability for error-correction during genome replication (38, 51). EIDs of viral origin have constituted between 25 and 44% of all documented emerging diseases in the last century, several of which originated from species jumps (8, 22, 44, 51). Cross-species transmissions from animal reservoirs apparently gave rise to many devastating human viruses in the past, such as the SARS coronavirus, Niphavirus, Human immunodeficiency virus (HIV) 1 and 2 and Measles virus- and even the 'Spanish flu' is believed to have jumped from birds into humans (recently reviewed for example in (38, 45, 49)). The importance of animal and particularly wildlife reservoirs for human disease emergence is exemplified by the fact that approximately 75% of human EIDs are of a zoonotic origin and an estimated 43% of all EIDs during the past century arose from wildlife (22, 44).

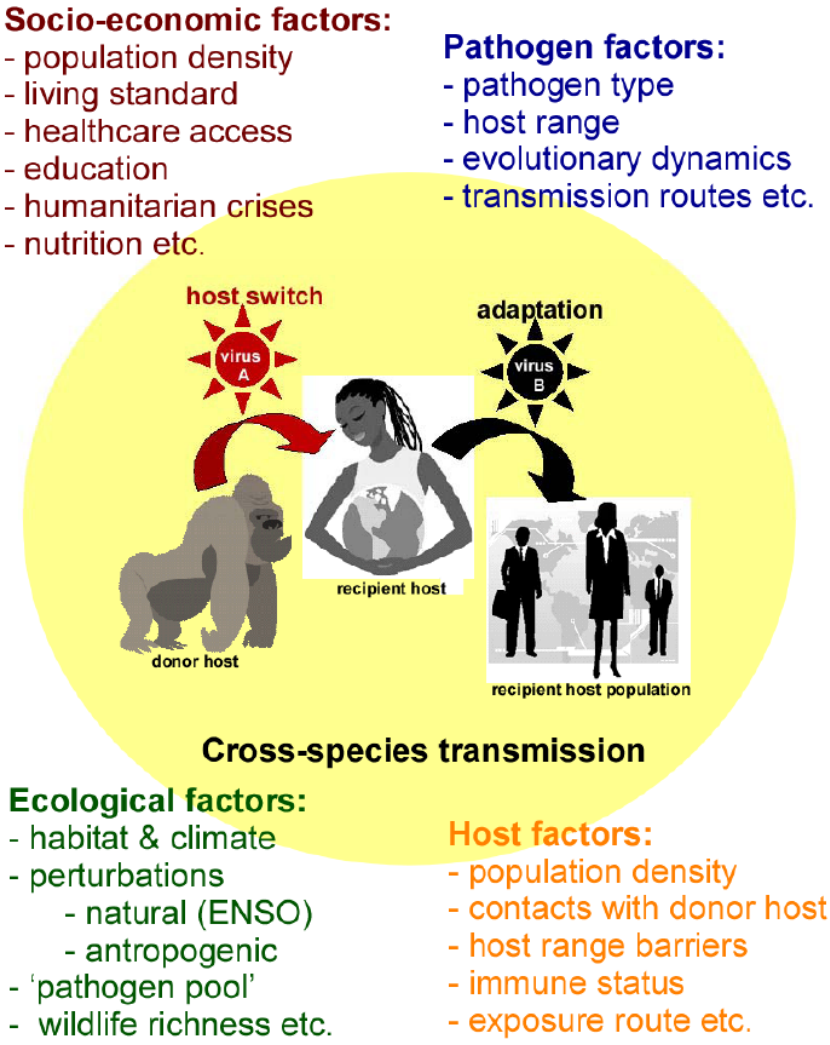


Figure 1.2: Overview of some common risk factors associated with disease emergence through cross-species transmission.

1.4 Risk factors for successful host jumping

In general, EIDs in the tropics seem to arise more frequently from wildlife than domestic animals while that relationship appears reversed in temperate zones, but the relative abundance of domesticated animals and particularly the presence of non-

human primates in different geographic regions might confound this relationship (49). The probability of disease emergence from wildlife reservoirs seems to be determined by the frequency of human-wildlife contacts, the “zoonotic pool” of microorganisms present in a region and the degree of habitat perturbations as well as their effects on indigenous wildlife (48) (Figure 1.2). Similar factors probably impact the likelihood of emergence from domestic animals, and viruses with broad host range, particularly those that infect both domestic animals and wildlife, are more likely to give rise to human EIDs than those viruses with narrow host range, though the underlying factors are still incompletely understood (8). The route of disease transmission might also impact emergence probabilities and direct or indirect transmission might facilitate the emergence of pathogens, but this topic is still subject to discussion (50).

Furthermore, the type and frequency of contact between host species probably impacts the likelihood with which host jumps occur, but again many questions remain unanswered. Some types of contact such as fighting, hunting or consumption might be more likely to lead to successful host jumps than other, more casual contacts, but the underlying molecular determinants are still unknown (20, 48). Historically the vast majority of human EIDs seem to have emerged from ungulate, carnivore or rodent species, which have been living in close proximity to humans since the beginning of human civilization (8, 50). Indeed, several human diseases such as measles, mumps, pertussis and smallpox have been linked to the beginning of agriculture, which marked the foundation of densely populated human settlements and for the first time gathered large human and animal populations in close proximity (49). However, a substantial number of human EIDs have been associated with other vertebrate sources such as primates, bats and birds (8, 50). Additional factors (for example, the phylogenetic distance between the host species) might also influence the likelihood of successful species jumps- a notion that mainly stems from anecdotal evidence that a

disproportionate number of recently emerged human viruses probably arose from non-human primates (49).

1.5. What we do and do not know about host-range barriers

The molecular and physiological determinants of host-range barriers have rarely been defined in detail. However, it appears likely that a combination of different factors restrict the replication and spread of viruses in alternative hosts, which likely include physical barriers, immunological barriers and biochemical blockades (recently reviewed in (24, 38)). The binding of species-specific cellular receptors or co-receptors represents one of the best-studied host-range factors, which appears to be important for the emergence of numerous EIDs such as HIV, the SARS coronavirus, H5N1 avian influenza virus, Venezuelan equine encephalitis virus, canine distemper virus and canine parvovirus (2, 18, 19, 25-28, 31, 37, 41, 42, 52). Indeed, multi-host pathogens, which seem to be more likely sources of EIDs than pathogens with narrow host ranges, appear to preferentially use highly conserved receptors such as the Coxsackie virus and adenovirus receptor but further studies are needed to obtain conclusive results (50). Other host-range restrictions probably occur at the intracellular level, where viruses can for instance induce species-specific interferon responses, might be subjected to APOBEC 3-G related deamination or might be blocked through other species-specific host proteins (recently reviewed in (38)). Moreover, viral nucleotide composition and codon usage might further restrict host ranges. In several virus families (for instance influenza A, the asteroviridae and potentially the irioviridae) nucleotide composition and codon usage appear to correspond surprisingly well with those of the host species, and have been proposed as one mechanism of viral adaptation after cross-species transmission (15, 46, 47)

1.6. The evolutionary dynamics of cross-species transmission

In most cases examined, a pathogen had to acquire several mutations to cross these host-range barriers (see for example (12, 24, 39)). It is still debated whether the phylogenetic relationship between the hosts is correlated with the required number of adaptive mutations, and whether these mutations have to be pre-formed in the donor host or can arise after accidental spill-over into the recipient host (9, 10, 38). Pre-adapted viruses might form in the donor host (a process potentially facilitated by relaxation of selective constraints) then be transmitted to the recipient host and subsequently colonize the new host. Alternatively, adaptive mutations may arise in the recipient host after accidental spill-over. In this latter case, the original virus accidentally infects the recipient host during spill-over infections, after which the virus population is exposed to altered selection pressures. Selection will favor beneficial mutations in the new environment and the virus population will therefore rapidly accumulate adaptive mutations. Unfortunately, the currently available data are generally insufficient to determine conclusively which host-range mutations were generated de-novo in the recipient host and which had been pre-existing in the donor host. Moreover, the strong transmission bottleneck during cross-species transmission and the large number of deleterious or slightly deleterious mutations that arise simultaneously with a relatively small number of beneficial mutations complicate the analysis of these evolutionary dynamics, and complex epistatic interactions seem to frequently dominate the adaptation process ((43); reviewed in (24, 38)). The success of the new pathogen appears to depend on the fitness cost associated with the 'intermediate' viruses and the degree of genetic diversity present among viruses in the infected host, with more variable viral populations being particularly prone to generate successful emerging infections (3, 8, 50). This intra-host diversity is a result of error-prone viral replication and selection against mutation-bearing viruses. The structure of

viral populations in infected hosts and their major determinants, however, are rarely understood. It is further unclear how the emerging viruses acquire the necessary sets of host-range mutations. Given the potentially large size of viral sequence space, the specific sets of host-range mutations are very unlikely to arise simultaneously. Single mutations, on the contrary, are most likely deleterious as they will probably incur a fitness cost in the donor host and do not promote an expanded host spectrum, and are therefore likely selected against (10). Therefore, sequential accumulation of adaptive mutations appears also an unlikely event, raising the question of how viruses with altered host- range overcome this “fitness trade-off” (reviewed for example in (38)). Several evolutionary mechanisms such as recombination or somatic hypermutation might allow an expedited accumulation of mutations, but the key evolutionary conundrum of successful disease emergence remains unsolved.

1.7. Onward transmission in the recipient host species

The complete adaptation of pathogens to new host species is thought to be a sequential process involving several intermediate steps (3). After crossing the species barrier, successful onward transmission in the recipient host species appears to be the greatest hurdle for many EIDs. It is believed that most emerging viruses are initially unfit and frequently only cause stuttering, self-limiting transmission chains which are greatly affected by stochastic events. The population dynamics of the recipient host - and in particular its population density, contact rates and frequency of replenishment of naïve hosts - therefore strongly impact EID dynamics at this stage. Indeed, it has been shown through simulation approaches that novel pathogens with initial basic reproductive ratio (i.e. the number of secondary cases caused by the introduction of an index case into a naive, panmictic population) far lower than one can be successfully sustained in the recipient host population if the mutation rate is sufficiently large,

exemplifying the potentially strong impact of stochastic events and the central role of the speed with which adaptive mutations can be harbored (3). Thus, only if further host adaptation occurs can the virus continue to circulate in the recipient host species and many novel viruses seem to fall short at this step. The basic reproductive ratio most likely increases during this adaptation period and many successful novel pathogens seems to have been circulating at low frequency in the recipient host species for years before causing widespread epidemics. For example, HIV strains dating back to the 1960s and 1970s have been described, often predating the HIV epidemic of the mid-1980s by more than a decade (reviewed for example in (17)). Wolfe et al. (2007) describe five distinct phases that human EIDs traverse as they emerge from zoonotic origins: (i) transmission restricted to animal hosts; (ii) spill-over into humans without onward transmission; (iii) spill-over with limited onward transmission; (iv) sustained transmission in human and animal species and (v) transmission restricted to humans. The authors hypothesize that diseases can stagnate at any of these stages and that this has happened for several well-recognized human diseases in the past. The authors further subdivide the fourth stage into three sub-stages, depending on the relative importance of sylvatic cycles and direct human-to-human transmission. Conversely, from the viewpoint of cross-species transmission, evolutionary stages two and three appear to be of greatest interest. During these periods with short, stuttering transmission chains, the detection and control of novel human EIDs appears most likely. Hendra and Nipha virus represent classical examples of stage 2 viruses while several hemorrhagic fevers (such as Ebola, Marburg and Lassa) represent viruses at stage 3 (Table 1.1). Avian influenza appears to be at the interface between steps two and three where only very limited human-to-human transmission appears possible.

Table 1.1: Overview of several human EIDs currently in stages 2 or 3of emergence as classified by Wolfe et al. (2007). The list of pathogens has been modified from a previously published list of human EIDs (Jones *et al.*, 2008) and additional information regarding the host species and virus properties was added.

Virus	Host		1st description		Virus family	Evolutionary stage (Wolfe et al.2007)
	donor	recipient	Year	Location		
Alkhurma virus	camel/sheep	human	1995	Saudi Arabia	Flaviviridae (+ssRNA)	2
Andes virus	rodents	human	1995	Argentinia	Bunyaviridae (-ssRNA)	3
Australian bat lyssa	bat	human	1996	Australia	Rhabdoviridae (-ssRNA)	2
Banna virus	unclear	human	1987	China	Reoviridae (dsRNA)	2
Barmah forest virus	birds or marsupials	human	1974	Australia	Togaviridae (+ssRNA)	2
California encephalitis virus	unclear	human	1945	US (California)	Bunyaviridae (-ssRNA)	2
Chikungunya virus	primates (?)	human	1952	Tanzania	Togaviridae (+ssRNA)	3 to 4
Crimean-Congo Hemorrhagic fever	rodents or hedgehogs	human	1944	Ukraine	Bunyaviridae (-ssRNA)	3
Ebola virus	unclear	human	1976	Sudan	Filoviridae (-ssRNA)	3
Tick-borne encephalitis virus	ungulates or rodents	human	1993	Scandinavia; Japan	Flaviviridae (+ssRNA)	2
Guama virus	unclear	humans	1954	Brazil	Bunyaviridae (-ssRNA)	2
Guanarito virus	rodents	humans	1989	Venezuela	Arenaviridae (-ssRNA)	2
Hantaanvirus	rodents	humans	1941	China	Bunyaviridae (-ssRNA)	2
Hendra virus	bats	humans; horses	1994	Australia	Paramyxoviridae(-ssRNA)	2
H5N1 avian influenza virus	birds	humans	1997	Hong Kong	Orthomyxoviridae (-ssRNA)	3
Jamestown Canyon virus	deer	humans	1960	US (Wisconsin)	Bunyaviridae (-ssRNA)	2
Japanese encephalitis virus	pigs; birds	humans	1989	Papua New Guinea	Flaviviridae (+ssRNA)	2
Junin virus	rodents	humans	1958	Argentinia	Arenaviridae (-ssRNA)	3
Kyasanur forest diseases virus	rodents	humans; primates	1957	India	Flaviviridae (+ssRNA)	2
LaCrosse virus	chipmunks; squirrels	humans	1960	US (Wisconsin)	Bunyaviridae (-ssRNA)	2
Laguna Negra virus	rodents	humans	1995	Paraguay	Bunyaviridae (-ssRNA)	2
Lassa virus	rodents	humans	1969	Nigeria	Arenaviridae (-ssRNA)	3
Machupo virus	rodents	humans	1959	Bolivia	Arenaviridae (-ssRNA)	3
Marburg virus	unclear	humans	1967	Germany	Filoviridae (-ssRNA)	3
Mayaro virus	unclear	humans	1954	Trinidad	Togaviridae (+ssRNA)	2
Menangle virus	bats	humans	1997	Australia	Paramyxoviridae (-ssRNA)	2
Monkeypox virus	primates	humans	1970	Congo	Poxviridae (dsDNA)	3
Murray valley encephalitis virus	birds	humans	1950	Australia	Flaviviridae (+ssRNA)	2
Nipha virus	bats	humans; pigs	1998	Malaysia	Paramyxoviridae(-ssRNA)	2
Picobirnavirus	birds (?)	humans	1984	Brazil	Birnaviridae (dsRNA)	2
Puumala virus	rodents	humans	1995	Croatia	Bunyaviridae (-ssRNA)	2
Rabies virus	carnivores; ungulates	humans	3000BC	Eurasia	Rhabdoviridae (-ssRNA)	2
Rift valley fever virus	ungulates	humans	1977	Egypt	Bunyaviridae (-ssRNA)	2
Sabia virus	unclear	humans	1990	Brazil	Arenaviridae (-ssRNA)	3
SARS coronavirus	unclear (civets?)	humans	2002	China	Coronaviridae (+ssRNA)	3
Seoul virus	rodents	humans	1989	US (Maryland)	Bunyaviridae (-ssRNA)	3
Sin Nombre virus	rodents	humans	1978	US (Idaho)	Bunyaviridae (-ssRNA)	1
Sindbis virus	birds	humans	1952	Egypt	Togaviridae (+ssRNA)	2
Venezuelan equine encephalitis virus	equines	humans	1943	Trinidad	Togaviridae (+ssRNA)	2
Wesselsbron virus	ungulates	humans	1989	Madagascar	Flaviviridae (+ssRNA)	2
West Nile virus	birds	humans; horses	1937	Uganda	Flaviviridae (+ssRNA)	3
Whitewater Arroyo virus	rodents	humans	1999	US (California)	Arenaviridae (-ssRNA)	2

1.8. Some unifying conclusions

In conclusion, despite a significant increase in our understanding of EIDs over the past decades, several molecular, evolutionary, epidemiological and ecological drivers remain to be elucidated. A more complete understanding of host-range barriers, and of the evolutionary and population dynamics that allow EIDs to emerge and be sustained, might lead to better-targeted surveillance and control programs. Moreover, a better appreciation of the impact anthropogenic perturbations have on disease emergence will prove crucial as we face the global ecological challenges ahead.

REFERENCES

1. Aguirre, A. A., and G. M. Tabor. 2008. Global factors driving emerging infectious diseases. Ann N Y Acad Sci **1149:**1-3.

2. Anishchenko, M., R. A. Bowen, S. Paessler, L. Austgen, I. P. Greene, and S. C. Weaver. 2006. Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. Proc Natl Acad Sci USA **103:**4994-9.

3. Antia, R., R. R. Regoes, J. C. Koella, and C. T. Bergstrom. 2003. The role of evolution in the emergence of infectious diseases. Nature **426:**658-61.

4. Asnis, D. S., R. Conetta, A. A. Teixeira, G. Waldman, and B. A. Sampson. 2000. The West Nile Virus outbreak of 1999 in New York: the Flushing Hospital experience. Clin Infect Dis **30:**413-8.

5. Binder, S., A. M. Levitt, J. J. Sacks, and J. M. Hughes. 1999. Emerging infectious diseases: public health issues for the 21st century. Science **284:**1311-3.

6. Blancou, J., B. B. Chomel, A. Belotto, and F. X. Meslin. 2005. Emerging or re-emerging bacterial zoonoses: factors of emergence, surveillance and control. Vet Res **36:**507-22.

7. Bridges, V. E., J. Akkina, J. Grannis, C. Johnson, R. Johnson, and C. Tuszynski. 2007. A qualitative assessment tool for the potential of infectious disease emergence and spread. Prev Vet Med **81:**80-91.

8. Cleaveland, S., M. K. Laurenson, and L. H. Taylor. 2001. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. Philos Trans R Soc Lond B Biol Sci **356:**991-9.

9. Domingo, E., and J. J. Holland. 1997. RNA virus mutations and fitness for survival. Annu Rev Microbiol **51:**151-78.

10. Domingo, E., L. Menendez-Arias, and J. J. Holland. 1997. RNA virus fitness. Rev Med Virol **7:**87-96.

11. Epstein, P. R. 1995. Emerging diseases and ecosystem instability: new threats to public health. Am J Public Health **85:**168-72.

12. Fauci, A. S. 2006. Emerging and re-emerging infectious diseases: influenza as a prototype of the host-pathogen balancing act. Cell **124:**665-70.

13. Feldmann, H., M. Czub, S. Jones, D. Dick, M. Garbutt, A. Grolla, and H. Artsob. 2002. Emerging and re-emerging infectious diseases. Med Microbiol Immunol **191:**63-74.

14. Fluit, A. C., F. J. Schmitz, and J. Verhoef. 2001. Multi-resistance to antimicrobial agents for the ten most frequently isolated bacterial pathogens. Int J Antimicrob Agents **18:**147-60.

15. Greenbaum, B. D., A. J. Levine, G. Bhanot, and R. Rabadan. 2008. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS Pathog **4:**e1000079.

16. Hayes, E. B., N. Komar, R. S. Nasci, S. P. Montgomery, D. R. O'Leary, and G. L. Campbell. 2005. Epidemiology and transmission dynamics of West Nile virus disease. Emerg Infect Dis **11:**1167-73.

17. Holmes, E. C. 2001. On the origin and evolution of the human immunodeficiency virus (HIV). Biol Rev Camb Philos Soc **76:**239-54.

18. Hueffer, K., L. Govindasamy, M. Agbandje-McKenna, and C. R. Parrish. 2003. Combinations of two capsid regions controlling canine host range determine canine transferrin receptor binding by canine and feline parvoviruses. J Virol **77:**10099-105.

19. Hueffer, K., J. S. Parker, W. S. Weichert, R. E. Geisel, J. Y. Sgro, and C. R. Parrish. 2003. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. J Virol **77:**1718-26.

20. Hui, E. K. 2006. Reasons for the increase in emerging and re-emerging viral infectious diseases. Microbes Infect **8:**905-16.

21. Jansen, W. T., J. T. van der Bruggen, J. Verhoef, and A. C. Fluit. 2006. Bacterial resistance: a sensitive issue complexity of the challenge and containment strategy in Europe. Drug Resist Updat **9:**123-33.

22. Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. Global trends in emerging infectious diseases. Nature **451:**990-3.

23. Karesh, W. B., R. A. Cook, E. L. Bennett, and J. Newcomb. 2005. Wildlife trade and global disease emergence. Emerg Infect Dis **11:**1000-2.

24. Kuiken, T., E. C. Holmes, J. McCauley, G. F. Rimmelzwaan, C. S. Williams, and B. T. Grenfell. 2006. Host species barriers to influenza virus infections. Science **312:**394-7.

25. Li, F. 2008. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. J Virol **82:**6984-91.

26. Li, W., T. C. Greenough, M. J. Moore, N. Vasilieva, M. Somasundaran, J. L. Sullivan, M. Farzan, and H. Choe. 2004. Efficient replication of severe acute respiratory syndrome coronavirus in mouse cells is limited by murine angiotensin-converting enzyme 2. J Virol **78:**11429-33.

27. Li, W., M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature **426:**450-4.

28. Li, W., C. Zhang, J. Sui, J. H. Kuhn, M. J. Moore, S. Luo, S. K. Wong, I. C. Huang, K. Xu, N. Vasilieva, A. Murakami, Y. He, W. A. Marasco, Y. Guan, H. Choe, and M. Farzan. 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. Embo J **24:**1634-43.

29. Macias, A. E., and S. Ponce-de-Leon. 2005. Infection control: old problems and new challenges. Arch Med Res **36:**637-45.

30. Marano, N., P. M. Arguin, and M. Pappaioanou. 2007. Impact of globalization and animal trade on infectious disease ecology. Emerg Infect Dis **13:**1807-9.

31. McCarthy, A. J., M. A. Shaw, and S. J. Goodman. 2007. Pathogen evolution and disease emergence in carnivores. Proc Biol Sci **274:**3165-74.

32. McMichael, A. J., R. E. Woodruff, and S. Hales. 2006. Climate change and human health: present and future risks. Lancet **367:**859-69.

33. Morens, D. M., G. K. Folkers, and A. S. Fauci. 2004. The challenge of emerging and re-emerging infectious diseases. Nature **430:**242-9.

34. Morens, D. M., G. K. Folkers, and A. S. Fauci. 2008. Emerging infections: a perpetual challenge. Lancet Infect Dis **8:**710-9.

35. Nash, D., F. Mostashari, A. Fine, J. Miller, D. O'Leary, K. Murray, A. Huang, A. Rosenberg, A. Greenberg, M. Sherman, S. Wong, and M. Layton. 2001. The outbreak of West Nile virus infection in the New York City area in 1999. N Engl J Med **344:**1807-14.

36. Neu, H. C. 1992. The crisis in antibiotic resistance. Science **257:**1064-73.

37. Palermo, L. M., S. L. Hafenstein, and C. R. Parrish. 2006. Purified feline and canine transferrin receptors reveal complex interactions with the capsids of canine and feline parvoviruses that correspond to their host ranges. J Virol **80:**8482-92.

38. Parrish, C. R., E. C. Holmes, D. M. Morens, E. C. Park, D. S. Burke, C. H. Calisher, C. A. Laughlin, L. J. Saif, and P. Daszak. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev **72:**457-70.

39. Parrish, C. R., and Y. Kawaoka. 2005. The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. Annu Rev Microbiol **59:**553-86.

40. Pfaller, M. A., R. N. Jones, G. V. Doern, A. C. Fluit, J. Verhoef, H. S. Sader, S. A. Messer, A. Houston, S. Coffman, and R. J. Hollis. 1999. International surveillance of blood stream infections due to Candida species in the European SENTRY Program: species distribution and antifungal susceptibility including the investigational triazole and echinocandin agents. SENTRY Participant Group (Europe). Diagn Microbiol Infect Dis **35:**19-25.

41. Philpott, S. M. 2003. HIV-1 coreceptor usage, transmission, and disease progression. Curr HIV Res **1:**217-27.

42. Regoes, R. R., and S. Bonhoeffer. 2005. The HIV coreceptor switch: a population dynamical perspective. Trends Microbiol **13:**269-77.

43. Sanjuan, R., A. Moya, and S. F. Elena**.** 2004. The contribution of epistasis to the architecture of fitness in an RNA virus. Proc Natl Acad Sci USA **101:**15376-9.

44. Taylor, L. H., S. M. Latham, and M. E. Woolhouse. 2001. Risk factors for human disease emergence. Philos Trans R Soc Lond B Biol Sci **356:**983-9.

45. Torres-Velez, F., and C. Brown. 2004. Emerging infections in animals--potential new zoonoses? Clin Lab Med **24:**825-38, viii.

46. Tsai, C. T., C. H. Lin, and C. Y. Chang. 2007. Analysis of codon usage bias and base compositional constraints in iridovirus genomes. Virus Res **126:**196-206.

47. van Hemert, F. J., B. Berkhout, and V. V. Lukashov. 2007. Host-related nucleotide composition and codon usage as driving forces in the recent evolution of the Astroviridae. Virology **361:**447-54.

48. Wolfe, N. D., P. Daszak, A. M. Kilpatrick, and D. S. Burke. 2005. Bushmeat hunting, deforestation, and prediction of zoonoses emergence. Emerg Infect Dis **11:**1822-7.

49. Wolfe, N. D., C. P. Dunavan, and J. Diamond. 2007. Origins of major human infectious diseases. Nature **447:**279-83.

50. Woolhouse, M. E. 2002. Population biology of emerging and re-emerging pathogens. Trends Microbiol **10:**S3-7.

51. Woolhouse, M. E., and S. Gowtage-Sequeria. 2005. Host range and emerging and reemerging pathogens. Emerg Infect Dis **11:**1842-7.

52. Yamada, S., Y. Suzuki, T. Suzuki, M. Q. Le, C. A. Nidom, Y. Sakai-Tagawa, Y. Muramoto, M. Ito, M. Kiso, T. Horimoto, K. Shinya, T. Sawada, M. Kiso, T. Usui, T. Murata, Y. Lin, A. Hay, L. F. Haire, D. J. Stevens, R. J. Russell, S. J. Gamblin, J. J. Skehel, and Y. Kawaoka. 2006. Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. Nature **444:**378-82.

CHAPTER 2

EVOLUTION AND VARIATION OF THE PARVOVIRUSES

From: Hoelzer, K. and Parrish, C.R. Evolution and variation of the parvoviruses. In: Origin and Evolution of Viruses, 2nd Ed. Domingo, E., Holland, J. and Parrish, C.R. (Eds). Elsevier, London 2008, with permission.

2.1. Summary

The parvoviruses are small non-enveloped viruses that contain ~5,000 bases of linear ssDNA, and they are widespread in nature, infecting many different animals, from crustaceans to primates. The viruses contain two large genes that encode proteins associated with the control of DNA replication and the capsid proteins, as well as variants of those proteins and some small accessory proteins. These viruses replicate using the host cell polymerases, or for the AAVs with helper virus polymerase, but the amount of variation seen in nature or after tissue culture passage appears high, and the rates of sequence change of the viruses can be similar to that of some RNA viruses. Recombination clearly occurs among the parvoviruses, and may be an underappreciated evolutionary mechanism. In specific examples the parvoviruses are shown to have different and often complex evolutionary processes. The viruses appear to be remarkably adaptable in some cases; canine parvovirus for example was able to jump between hosts in a process that involved complex evolutionary mechanisms.

2.2. Parvoviruses and their properties

Parvoviruses are a family of small viruses with a non-enveloped capsid that contains a linear single stranded (ss) DNA genome of between 4,500 and 5,250 nts. The viruses are very widely distributed in nature, and likely infect most vertebrate and invertebrate hosts (40, 115, 149). The two subfamilies within the family Parvoviridae are the *Parvovirinae* which infect vertebrate hosts and the *Densovirinae* which infect invertebrates.

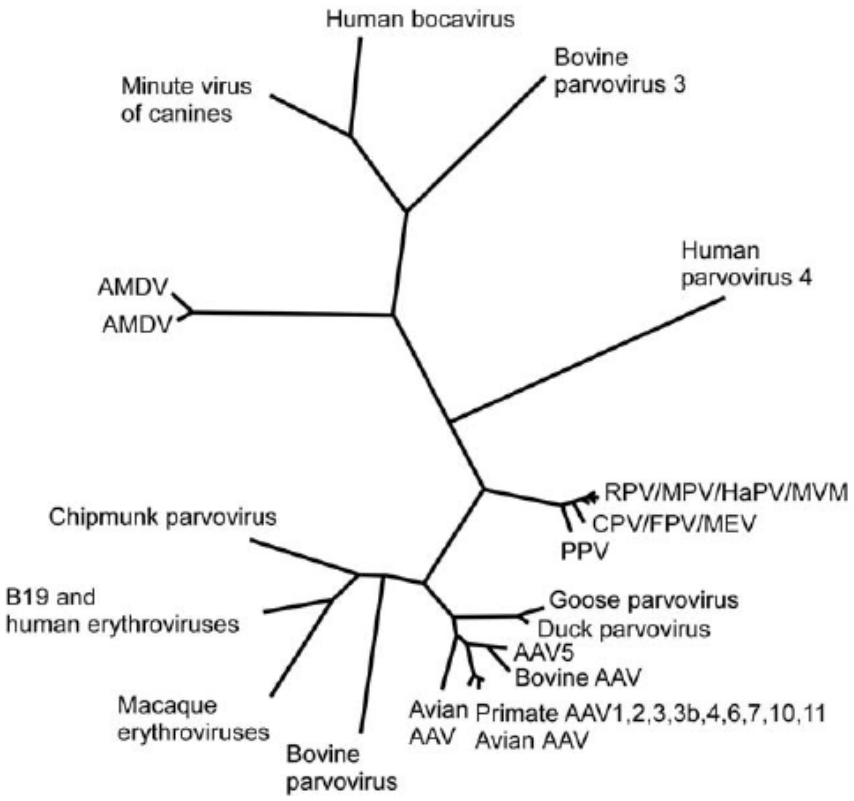


Figure 2.1: The evolutionary relationship among the parvoviruses. A conserved region of the NS1 amino acid sequence of representative parvoviruses of vertebrates (i.e. residues 277 to 387 of MVM sequence NP_041242) was aligned by eye and using the Clustal W algorithm available in MegAlign. Phylogenetic relationships were determined using PAUP version 4.0b10 for Unix to construct neighbor - joining trees.

Within the Parvovirinae the phylogenetic relationships can be shown using a conserved region of the NS1 gene sequences. The parvovirinae fall into several distinct, well separated clades (Figure 2.1). The five recognized genera are the *parvoviruses* , which include various rodent parvoviruses such as Minute Virus of Mice (MVM) and H1 as well as the parvoviruses of carnivores such as canine parvovirus (CPV), feline panleukopenia virus (FPV) and mink enteritis virus (MEV); the *erythroviruses* (human B19 virus and related viruses of primates); *bocaviruses*

(bovine parvovirus type 1, minute virus of canines (MVC) and human bocavirus); *Amdoviruses* (Aleutian mink disease virus (AMDV)) and the *dependoviruses* (adeno-associated viruses (AAV)) of humans and other hosts, which primarily replicate in cells which are coinfected with an adenovirus or herpesvirus. In the Densovirinae the four genera are the Densoviruses, Iteraviruses, Pefudensoviruses and Brevidensoviruses. Densovirinae infect many different invertebrate hosts from the Classes Insecta and Crustacea. Viruses have been isolated from various insect hosts, including members of the Orders Lepidoptera, Diptera, and Orthoptera. Several viruses infect members of the crustacean order Decapoda, but those viruses are still less well characterized. The Densovirinae are considerably more diverged than the Parvovirinae.

2.2.1. Gene Structure and Replication.

The parvovirus virion contains a single ssDNA molecule, and most members of the parvovirus genus, such as MVM, H1 or CPV preferentially package the negative strand, while others such as the dependoviruses and erythroviruses and the parvovirus LuIII package ssDNA of both polarities with similar frequencies. The bovine parvovirus type 1 preferentially packages negative sense DNA, but approximately 10% of viral capsids contain positive strand DNA (31). The detailed processes involved in ssDNA packaging and the molecular determinants responsible for the packaging bias involve the relative efficiency of the left- and right-hand origins of replication in initiating replication (42). Parvoviruses use parts of the host cell polymerase complex for viral DNA replication including the α and/or δ polymerases, perhaps among others, under various circumstances (12, 15, 36, 89) as well as a variety of cellular proteins including Cyclin A and chaperone-associated proteins. Viral DNA replication depends on the host cell passing through S-phase (41). Some

parvovirus proteins appear to affect the host cell cycle (113, 114, 126), generally by leading to cell cycle arrest in G1 or G2. The ssDNA genome is filled in through reactions that may involve the cellular DNA repair responses. In the case of the AAVs the viral genome is converted into circular and concatameric forms in the absence of active DNA replication. However, parvoviruses do not directly induce mitosis, but must wait until the host cell enters S-phase (125). Dependoviruses require additional factors for productive infection, although genotoxic stimuli such as UV light, cyclohexamine treatment and heat exposure can lead to permissive replication (reviewed in (15)), but dependoviruses replicate mostly in the presence of early gene products of their helper viruses.

Replication initiates from the incoming genome which is converted into a double-stranded (ds) DNA intermediate within the nucleus, and then proceeds through a modified rolling-circle mechanism (rolling-hairpin), where the palindromes located in each end of the genome are used for replication of the virus templates (16, 38, 41, 74). The dependoviruses and erythroviruses have identical palindromes at the 3' and 5' end (inverted terminal repeats (ITRs)), while the parvoviruses, bocaviruses and ambdoviruses have different palindromes, leading to more complex replication mechanisms. The 3'-end of the genome forms a base paired structure that primes second-strand synthesis, leading to the generation of the first replicative form (RF) DNA which extends and unfolds the 5' palindrome. Subsequent replication is NS1 dependent and creates origins by nicking specific sites in the sequence to create a new 5' end, and the NS1 helicase activity unfolds the 5'-end of the DNA, and then that is replicated and refolds, a process termed hairpin transfer, giving inverted palindromes. Subsequently replication proceeds through the formation of dimeric or tetrameric RF DNA, where nicking by the viral NS1 protein and strand exchange is involved in the resolution of the DNA replication intermediates (43). There are a number of aspects

of this replication scheme that would favor higher substitution rates compared to the normal cellular DNA replication. The parvovirus replication may not require the complete polymerase complex used for the synthesis of the host DNA, the genome is likely not methylated so that the template strand is not identified during replication, the single stranded form of the genome would be vulnerable to base conversion and the initial fill-in is not error corrected.

2.2.2. Viral gene functions.

The parvoviruses are genetically simple, and their genomes contain two large open reading frames (ORFs), where the left-hand ORF encodes the non-structural genes and the right hand ORF encodes the capsid proteins (Figure 2.2). The genomes contain between 1 and 3 promoters, depending on the virus. Alternative splicing, alternative transcription start sites, or alternative initiation of translation can give rise to additional gene products, and the two major ORFs give rise to messages for between one and 4 non-structural proteins, and between 2 and 4 capsid proteins. The bocaviruses contain a third ORF, located between the other two ORFs which encodes a phosphorylated NP1 non-structural protein of unknown function (5, 32, 92, 141).

2.2.3. Non-structural proteins.

The NS1 and Rep proteins in the parvoviruses and dependoviruses, respectively, are multifunctional proteins required for DNA replication, for regulation of viral gene expression, and for integration of some AAV genomes into the host DNA, and they have site-specific nickase activity, ATPase, ligase and helicase activities. During replication they are covalently attached to the 5' end of the viral DNAs. NS1 interacts with various cellular proteins including transcription regulators and members of the replication machinery, and it induces cytopathic effects. The phosphorylation status of

the protein is intricately involved in regulating the various NS1 functions (91, 121, 122), and cytopathic potential (46).

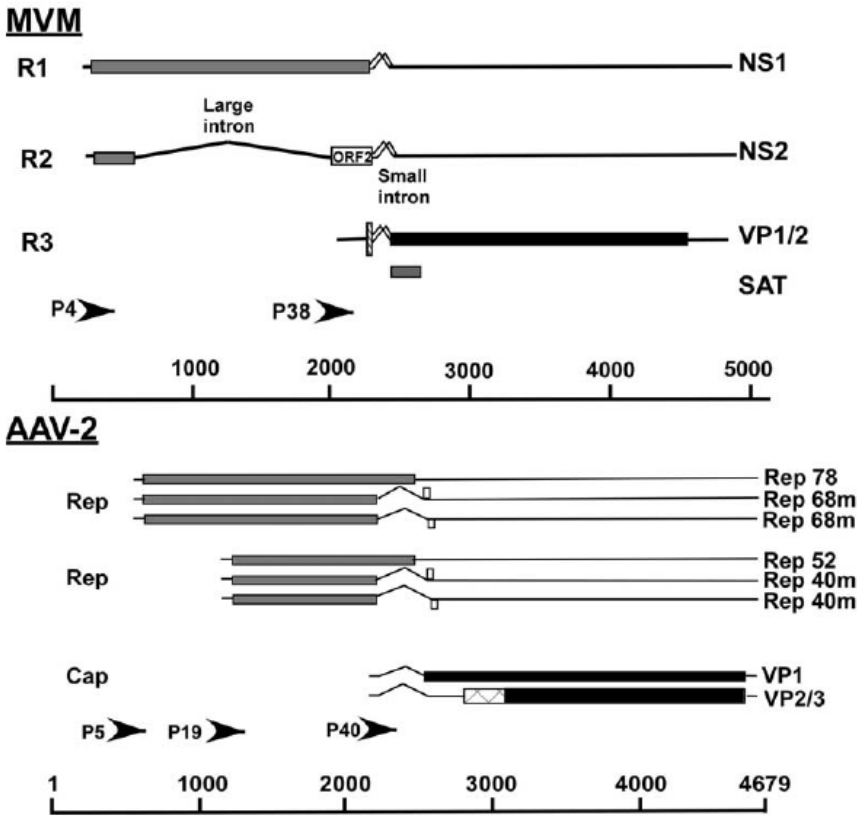


Figure 2.2: The genome structure of autonomous parvoviruses (MVM) and Dependoviruses (AAV-2). The members of the Parvoviridae vary in having from 1 to 3 transcriptional promoters, one or more sites of poly-A addition, in the presence of smaller non-structural protein genes and in having identical or different terminal palindromes. R= transcriptional products; NS or Rep= non-structural proteins involved in replication; V= viral capsid protein; SAT= small open reading frame of largely unknown function.

The smaller Rep proteins lack the DNA-binding and replication functions of the larger Rep proteins, but do retain the helicase and ATPase function, and they stimulate the production of ssDNA and are likely to be involved in packaging of the viral DNA (34, 51, 145). The NS2 proteins of MVM or the rat virus LuIII appear dispensable for viral replication in host cells other than mouse or rat cells respectively,

but are required for efficient translation or assembly of the virus capsid in the natural host cells (35, 39, 45, 52, 94, 117). The functions of NS2 are not completely defined, but the protein interacts with 14-3-3 family member proteins and with the Crm1 proteins involved in nuclear export (22, 24, 107). NS2 is itself regulated by phosphorylation and may be involved in phosphorylation of the capsid proteins and nuclear export of capsids (45). In CPV NS2 does not appear required for virus replication in dog and cat cells in tissue culture or in dogs (161).

2.2.4. Capsid proteins and genes.

VP1 and VP2 are produced by a variety of strategies, including alternative splicing of the viral mRNA producing ~10% VP1 and 90% VP2 (40). The proteins overlap in sequence so that the entire 62-70 kDa VP2 is contained within the VP1 sequence, with VP1 having a unique N-terminal extension (~ 120 and 150 amino acids depending on the virus). Erythroviruses and Ambdoviruses have capsids containing only VP1 and VP2 while the capsids of Dependoviruses contain VP1, 2 and 3 which are obtained by alternative splicing of the mRNA. The full (DNA containing) capsids of members of the parvovirus genus contain a smaller VP3, which is generated from the VP2 protein by proteolytic cleavages of the N-terminus.

Located in the VP1 unique region of most parvoviruses is a phospholipase 2 (PLA2) enzyme domain required for cell infection (56, 63, 169) most likely because it modifies the endosomal membrane and allows the viral particle to penetrate into the cytoplasm (56, 147).

2.2.5. Structure of the viral capsid.

The parvovirus capsids are small and appear to contain only a limited number of antigenic sites. However, those have mostly been defined using analysis of naturally

or selected antibody escape mutants or by analysis of peptide binding, and it is not clear how accurately those represent the true binding sites of the normal host antibodies. Most antigenic sites appear to be conformation dependent, but linear epitopes defined include the N-terminus of VP2 which is exposed to the exterior of CPV full (DNA containing) capsids, and the N-terminus of VP1 for the B19 parvoviruses (95, 139, 146, 167) and AAV2 capsids (162, 163). The role of antigenic variation and immune selection in the evolution of most parvoviruses is not well understood, and in some cases the analysis is complicated as many of the changes which result in antigenic variation also alter host range or other properties of the virus by altering binding to the cellular receptor (30, 128). Genetically distinct viruses can be distinguished by serology with polyclonal sera, but within many virus groups little antigenic variation has been described even among viruses that have been separated by decades. Variation of capsid epitopes of natural CPV or MEV isolates has been identified using monoclonal antibody analysis, and some variants have become widely distributed around the world (103, 118, 130, 146). For MVM antigenic variants were selected in tissue culture after selection with neutralizing monoclonal antibodies, but it is not known whether such variation is common among the viruses in nature (95). For many parvoviruses animals that have recovered from infection resist re-infection by antigenically related viruses, even those variant at one or more epitopes. However, in natural transmission cycles low levels of waning maternal antibody may allow infection and provide an environment allowing selection of antigenic variation.

2.3. Epidemiology and antiviral immunity

The epidemiology and pathogenesis of the virus infections varies widely, and likely influences the variation and evolution of the viruses. Many autonomous parvoviruses, including canine parvovirus (CPV), feline panleukopenia virus (FPV), porcine

parvovirus (PPV), and B19 human parvovirus, normally cause acute infections of their hosts lasting <10 days. Virus is usually cleared by the host immune response within this time span, after which infectious virus is not present, and the recovered hosts are no longer infectious for other animals (116, 129, 156). However, prolonged replication and persistence may occur in some of those hosts, particularly during fetal infection or when immune suppressed. Persistent infections may affect the development of intra-host variation, permit mixed infections, and could result in prolonged shedding with the re-introduction of older viruses back into circulation. Persistent infection and shedding occurs for some viruses, including rodent parvoviruses where persistence in the kidneys results in shedding in the urine. AMDV-infected adult mink are persistently infected and the virus continues to replicate in a number of tissues for long periods (2, 81-83, 134, 137). In neonatal mink AMDV causes an acute disease without persistence developing. Viral replication in neonates occurs in type II pneumocytes, while in older animals replication is limited to macrophages in the lymph nodes with reduced transcription and replication. In many genotypes of mink a chronic immune complex mediated disease develops (1, 3, 84). A small proportion of chronic and persistent B19 infections occur in humans who are immune suppressed, or who do not develop effective immunity (27, 57, 168). Even in normal infections there may be low-level persistence of B19 DNA in the presence of antibodies, although that DNA may not be infectious (29, 93, 102). Secondary infections by variant erythroviruses have been reported (67, 87).

The epidemiology of the newly discovered human bocavirus (101) is not well understood, but clinical disease appears to occur in the presence of other respiratory virus infections (58). Whether human bocavirus is persistent is not yet known. Some bovine parvoviruses can be recovered from fetal bovine sera, suggesting that those may also cause longer term or fetal infections (4). The recently discovered human

parvoviruses PARV4 and PARV5 appear generally apathogenic, although an association with HIV infection is possible (102).

2.4. Mechanisms of transmission

Viruses similar to CPV use a fecal-oral transmission cycle, but replicate systemically in the host tissues before invading the intestine (99, 100, 129). Some rodent parvoviruses also replicate in the intestine, but may be transmitted through urine after replication in the kidney (11, 81). Although the human B19 virus replicates primarily in the bone marrow, it is thought to be transmitted by respiratory routes (26). Parenteral transmission is possible as B19 virus is inefficiently inactivated in blood and blood products (8, 20, 67, 133). Veneral transmission might be important for porcine parvovirus (PPV) as PPV or viral DNA can be isolated from pig semen (66). Vertical transmission has been shown for several members of the parvovirus genus including several rodent parvoviruses, FPV and PPV (85, 106). Transplacental transmission can occur for many parvoviruses, although the fetus might be protected by maternal antibodies, depending on the host species and type of placentation (25, 47, 53, 68). Since the non-enveloped parvovirus capsid is very resistant in the environment, viral spread through vectors or formites occurs and might explain the global spread observed for some parvoviruses such as CPV and B19.

2.5. Immune response and protection

There is clearly an important role for cell mediated immunity in recovery from infection and there may be limited numbers of T cell epitopes in the viral proteins (86). However, for many vertebrate viruses humoral immunity, including maternal antibody, protects against infection by antigenically related viruses. Antibody treatments can arrest CPV replication in dogs, and can terminate chronic human

infections by the B19 parvovirus (26, 27). In the case of CPV, FPV, PPV and B19 the virus is functionally inactivated within a few days of the antibody response developing, while in contrast (as described previously) AMDV forms persistent infections despite the strong immune response (2, 137).

2.6. Genetic relationships and variation among the members of the family parvovirus

Comparing the sequences of conserved regions of the genome shows that all parvoviruses have sequences in common, and are related through a distant common ancestor (Figure 2.1). The viruses are divide into distinct clades which may show some correlations to the hosts of origin (98). As more viruses are collected and analyzed, a variety of more-or-less related viruses has been seen to infect many species. For example, three distinct bovine parvoviruses (types 1, 2, and 3) infect cattle (4), and there are two distantly related parvoviruses infecting dogs (141). The several erythroviruses from primates (B19 and related human and primate parvoviruses) are most closely related to each other and to the chipmunk parvovirus (98, 166). Most viruses from rats, mice and hamsters were found to be within the same clade as CPV and the related viruses of carnivores and PPV, while AMDV was found to be quite distantly related to all the other vertebrate viruses (Figure 2.1). The LuIII virus proved to be a recombinant between two different rodent viruses, most likely between MPV and Hamster PV (98), although whether that occurred in nature or during passage of the virus in tissue culture is not clear (Figure 2.3). True times of divergence of the various viruses are not known as no long term molecular clock can be estimated. Whether particular vertebrate parvoviruses have co-evolved along with their hosts or have crossed between different hosts in more recent times is also not known, although both origins are likely for different viruses.

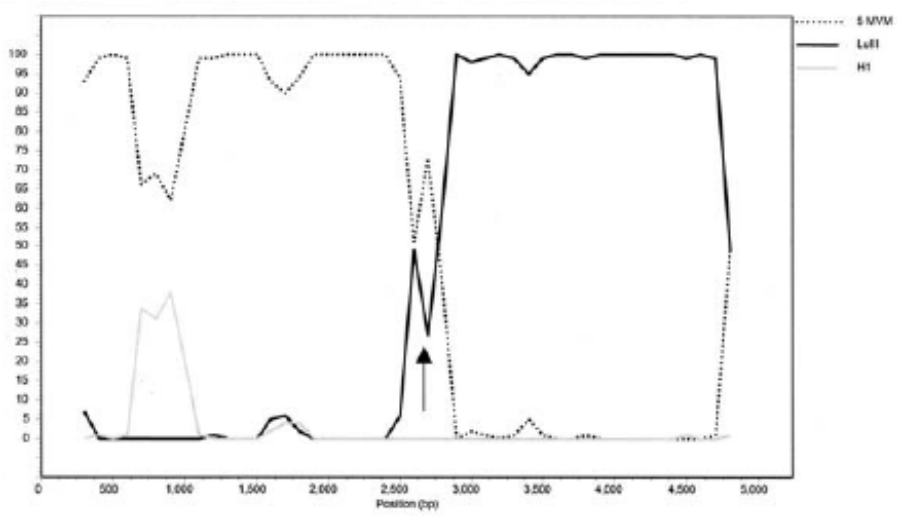


Figure 2.3: Analysis of the genome of the parvovirus LuIII, compared to the genome of mouse parvoviruses and the Hamster parvovirus, or to the genomes of MVM-related viruses, showing the position of recombination (arrows) around the middle of the genome. From (Lukashov *et al.*, 2001), with permission.

The LuIII virus proved to be a recombinant between two different rodent viruses, most likely between MPV and Hamster PV (98), although whether that occurred in nature or during passage of the virus in tissue culture is not clear (Figure 2.3). True times of divergence of the various viruses are not known as no long term molecular clock can be estimated. Whether particular vertebrate parvoviruses have co-evolved along with their hosts or have crossed between different hosts in more recent times is also not known, although both origins are likely for different viruses.

2.6.1. Human B19 and related erythroviruses. There are several human erythrovirus strains that are related to B19. Strain variation of the B19 related viruses so far shows no association with specific clinical signs or tissue tropisms (50, 55, 60,

70, 140). Three genotypes have been identified (142), with the prototype B19 viruses classified as genotype 1 (Figure 2.4). Genotype 2 was first identified as the LaLi strain (originally isolated from a Finnish skin sample), and also comprises the A6 strain, both of which are ~11% divergent from the B19 reference strains at the nucleotide level (73, 142, 164). Genotype 3 was originally isolated from a patient with aplastic crisis in Paris in 1995 (as the V9 virus) (120), and is ~12% divergent from B19 at the nucleotide level. The different human erythroviruses are between 5 and 20% different at the DNA sequence level (29, 60, 73, 142), but most mutations are synonymous and the viruses are >96% homologous in amino acid sequence. Antibodies against genotypes 1 or 2 are cross-reactive (54, 69). When examining the B19-related viruses, between 1 and 4% sequence variation was found between isolates within each clade. The simian erythroviruses were 40-50% divergent at the nucleotide level from human erythroviruses.

2.6.2. CPV and related viruses.

DNA sequences of viruses from dogs, cats, raccoons, mink, and arctic foxes have been examined in a number of studies (76, 80, 144, 155). CPV was recognized in 1978 as the cause of new diseases of dogs, and all canine isolates derived from a common ancestor which arose during the late 1960s or early 1970s (144, 155) (Figure 2. 5). The sequences of viruses from cats, mink, raccoons or foxes all fell within a single clade, suggesting that interspecies transmission of those viruses occurs.

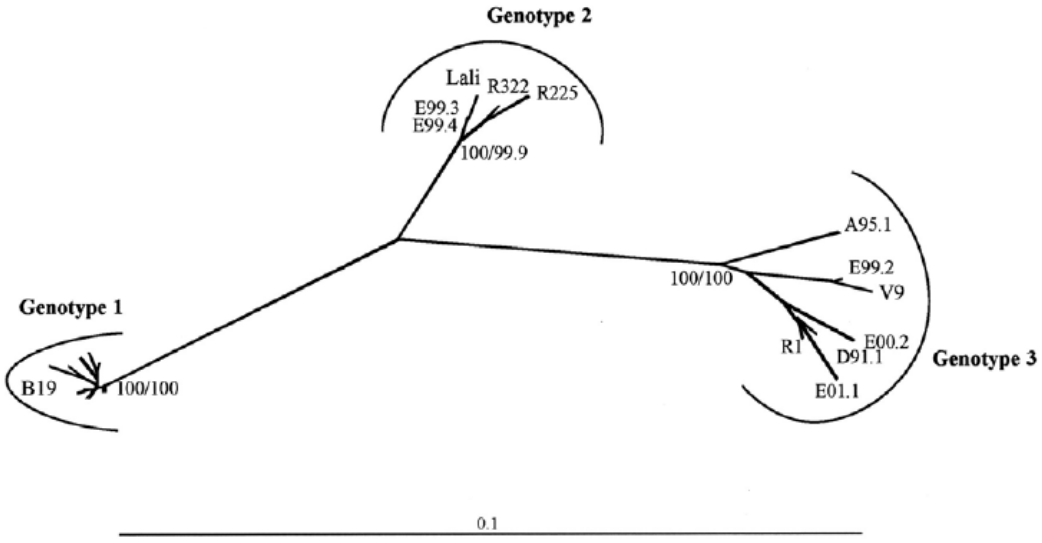


Figure 2.4: The diversity in the sequences of human erythroviruses related to the prototype B19 virus, showing the three distinct genotypes that have been characterized. From (Servant *et al.,* 2002), with permission.

The specific origin of CPV has not been completely defined but the most closely related virus was an isolate from an Arctic fox in Finland (BFPV isolate, Figure 2.5). It was suggested early on that CPV might be derived from vaccine strains of FPV (152), but DNA sequence analysis showed the CPV isolates were unrelated to vaccine strains tested (154). Serology and analysis of viral DNA from paraffin-embedded tissues (158) show that CPV was circulating in Europe between 1974 and 1976. Many of the substitutions in CPV were located in the capsid structure (Figure 2.6). Significant levels of antigenic variation have been demonstrated by monoclonal antibody (MAb) analysis, which showed that CPV isolates differed from viruses from cats, mink, or raccoons in two specific neutralizing epitopes- one present only on CPV, and the other present on FPV isolates (109, 130, 131, 146).

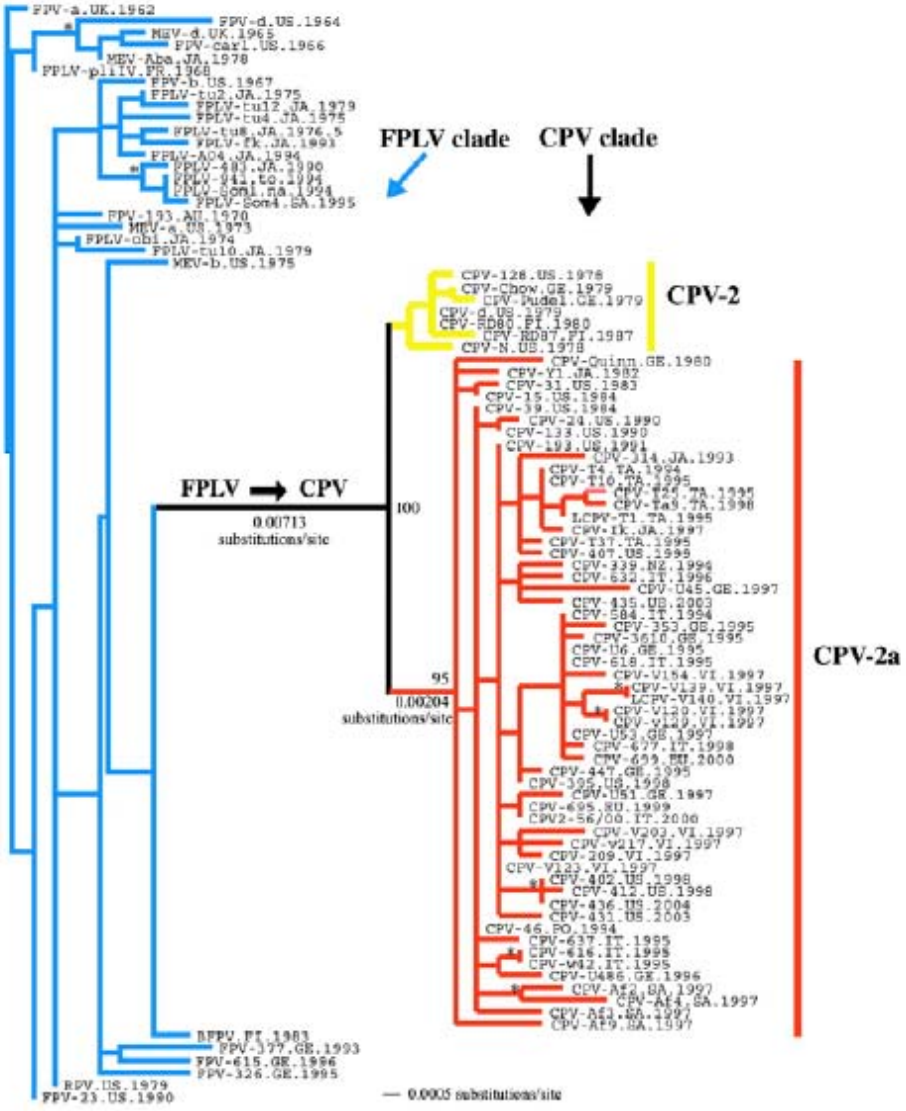


Figure 2.5: Phylogenetic tree of 91 VP2 gene sequences from carnivore parvoviruses, rooted with the oldest sampled sequence. Bootstrap values are shown for relevant nodes, and nodes with > 70% support are marked with an asterisk. Horizontal branch length are drawn to scale. The name of each isolates is followed by the location and year of isolation. Locations are coded as follows: UK, United Kingdom; US, United States; JA, Japan; FR, France; AU, Australia, FI, Finland; GE, Germany; TA, Taiwan, NZ, New Zealand; VI, Vietnam; EU, Europe (no further information available); PO; Poland; IT, Italy; SA, South Africa. The FPV clade is shown in blue, the CPV-2 subclade is shown in yellow, and the CPV-2a subclade is shown in red. BFPV, blue fox parvovirus. From (Shackelton *et al.*, 2005); with permission.

During the later evolution of CPV a number of mutations altered the antigenic structure of the capsid. However, many of the changes which altered the antigenic structure of the capsid also changed the TfR binding sites, making it difficult to determine which functions are under selection.

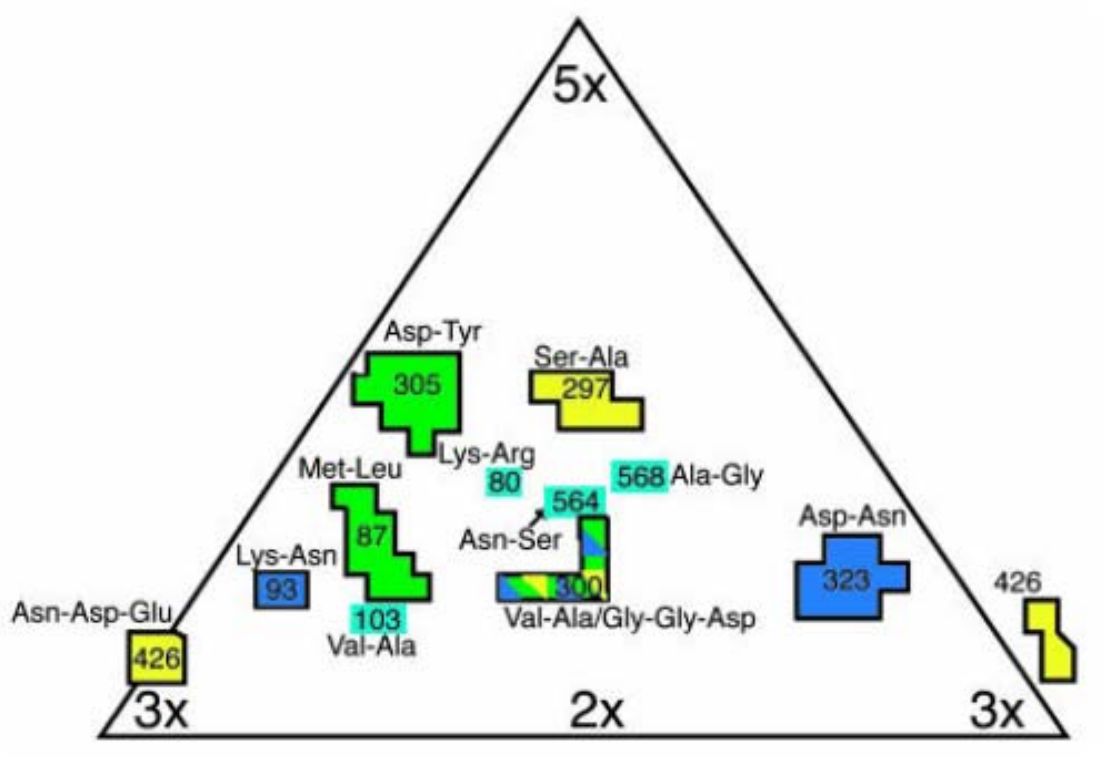


Figure 2.6: A model of one asymmetric unit of the CPV capsid, showing the locations of substituted amino acids. Sites that changed along the FPV-CPV branch are in blue, with the darker shade indicating surface-exposed residues and the lighter shade showing the relative position of the surface sites. Sites that underwent substitutions along the CPV-2 - CPV-2a branch are in green, and sites responsible for variants within the CPV-2a subclade are in yellow; all surface-exposed. From (Shackelton *et al.*, 2005), with permission.

The various FPV and CPV isolates differ in their abilities to infect cells in culture and animals, and the subsequent evolution in dogs further changed their natural host ranges. The CPV type-2 isolates did not replicate in cats, but later viruses (CPV type-2a and its derivatives) replicated efficiently in cats (153). This was a natural host range for the viruses as CPV type-2a and related viruses were isolated from 10 to 20% of cats which had natural parvovirus disease in Asia, Germany, Italy and the USA during the 1990s (13, 79, 108, 157).

2.6.3. Aleutian Mink Disease Virus (AMDV).

There is significant variation in the genomes of AMDV isolates, and the genetic variability of the viruses from a single farm or region was found to be up to 16% at the nucleotide sequence level (124). AMDV isolates normally grow only in animals, but certain strains have been adapted to grow in feline cells in tissue culture. A 2.5% sequence difference was seen between a wild type pathogenic strain of AMDV and a tissue culture adapted virus. Intriguingly, a hypervariable sequence was detected within the capsid protein gene, most likely located on the surface of the VP2 protein structure (21, 105, 123). AMDV has also been recovered from ferrets, skunks, and raccoons. The raccoon and skunk viruses appear similar to viruses isolated from farmed mink, although it is not known whether there is natural transmission between those hosts (123).

2.6.4. Rodent parvoviruses.

The known and suspected rodent parvoviruses are all >70% identical in DNA sequence, while viruses that are considered to be of the same type and given the same name are generally >95% identical (Figure 2.7). The rodent parvoviruses MVM, LuIII

and H1 have been intensively studied for their genetic and biochemical properties, but there is less information about the variation and evolution of their strains in nature.

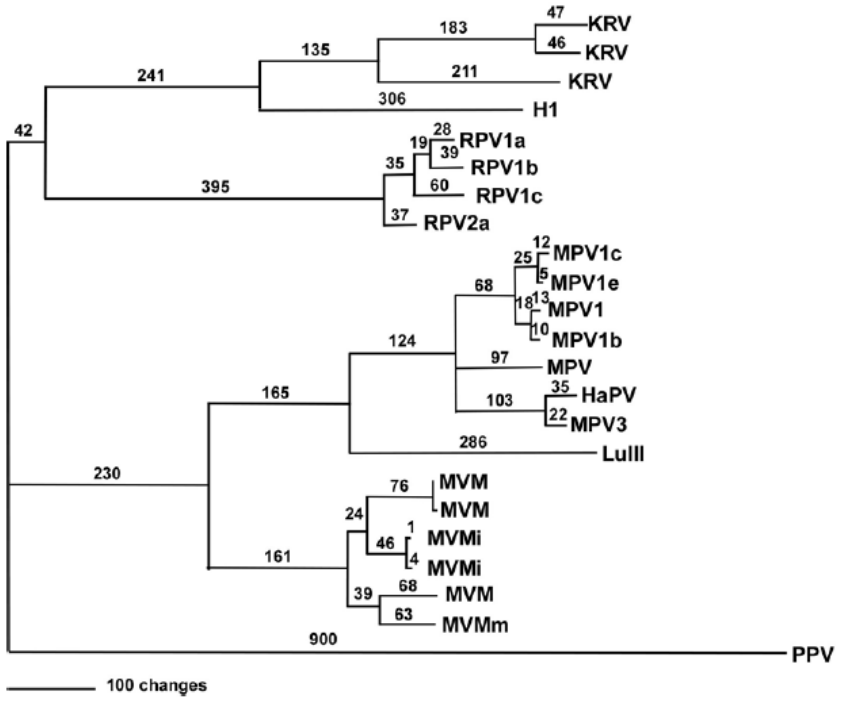


Figure 2.7: Relationship between the rodent parvoviruses, determined from the almost-complete genome sequences. The phylogeny was prepared by maximum likelihood using PAUP version 4.0 for Unix. Branch lengths are proportional to the distances between the different taxa, and the numbers on the branches indicate the number of substitutions. Viruses are: KRV, Kilham rat virus; H1, parvovirus H1; RPV, rat parvovirus; MPV, mouse parvovirus; MVM, minute virus of mice, which are divided into distinct clades. The porcine parvovirus (PPV) was used as an outgroup in the analysis.

Many early strains were isolated from tissue cultures or transplantable tumors, and their true origin and degree of tissue culture or host adaptation are not fully understood. Those viruses include the MVM prototype strain (MVMp) isolated from a murine adenovirus stock (44), H1 virus from the HEP-1 human tumor transplanted into rats and likely a rat virus (151) and LuIII from human cells. Tissue culture

isolates of MVM isolated from lymphocytes in 1976 and immune suppressive in vitro were named MVMi (23) while MVM-Cutter (MVM-(c)) was a contaminant of BHK21 cells (17). A number of viruses have been detected directly in mice, including mouse parvovirus (MPV)) strains 1a, 1b, 1c, 2 and 3, a new MVM strain MVMm which is pathogenic in non-obese diabetic (NOD) mice, as well as viruses of hamsters (hamster parvovirus (HaPV)), and rats (rat parvovirus (RPV)) (11, 18, 19, 159, 160). These viruses appear common in wild rodents (14), and until these were identified and testing introduced they were also widespread in rodent colonies (10, 11, 18, 81, 160). Many of the rodent parvoviruses are difficult to grow in tissue culture and the adaptations allowing tissue culture growth are poorly understood.

Apart from the relationships between the viruses at the sequence level, little is known about the details of their evolutionary histories or any host range differences between these viruses, or about their temporal variation. Recombination has been demonstrated among the rodent parvoviruses (Figures 2.3 and 2.8) and porcine parvovirus and may be a common event.

2.6.5. Adeno-associated viruses (AAV).

AAVs have been isolated from primates and from many other mammalian and avian hosts. Most depend upon a helper virus for replication, while some related viruses replicate independent of helper viruses. AAV was discovered in the 1960s as contaminant of Adenovirus preparations (7, 72).

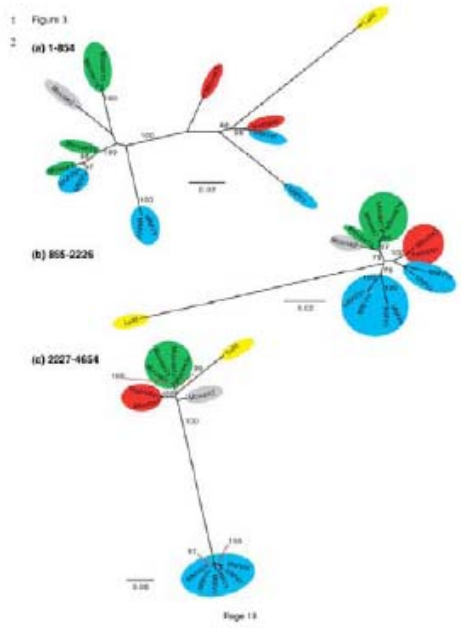


Figure 2. 8: Phylogenetic analyses showing recombination among the rodent parvoviruses. Colors indicate different rodent virus types. Nodes with > 70% bootstrap support are shown and branch lengths are scaled according to nucleotide substitutions per site. From (Shackelton *et al.*, 2007), with permission.

The AAVs of humans and non-human primates are currently divided into 5 species (AAV 1-5) and two tentative species (AAV 7 and 8), with the AAV1 species containing 2 strains (AAV strain 1 and 6) and all other species containing one AAV strain (149) (Figure 2.9). Screening of tissue samples from humans and other primates by PCR detected large numbers of DNA sequences, and in human and non-human primate tissue the AAV prevalence was ~19% (61, 62). AAV 1 and 6 were closely related, and sequence identity between the other strains ranged between 75% and 82%. The primate AAVs fell into a number of distinct clades (62).

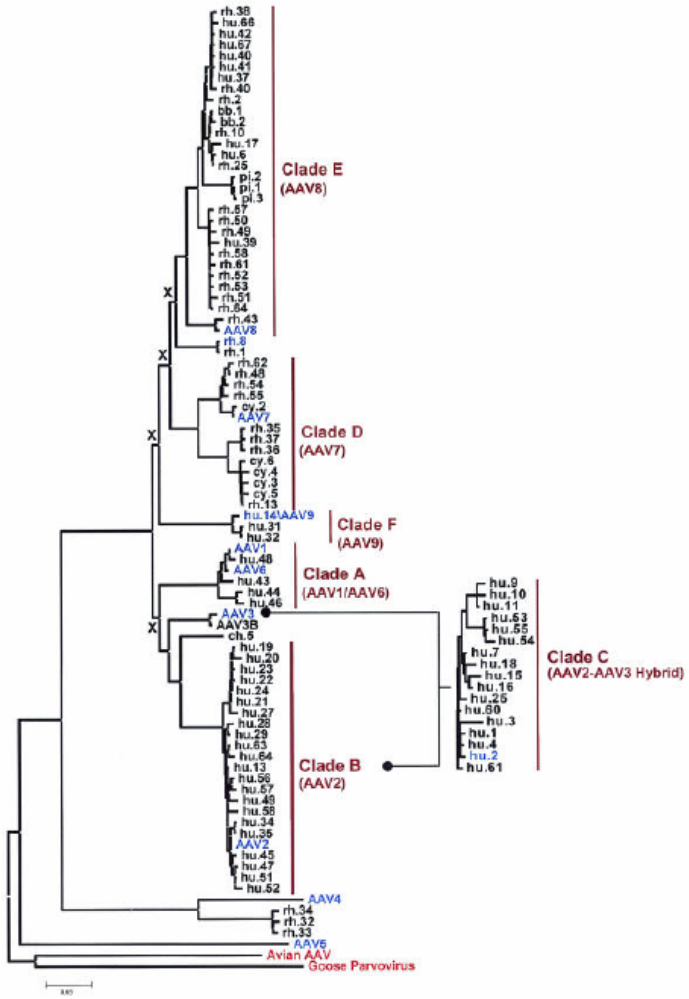


Figure 2.9: Phylogenetic relationships between the VP1 protein sequences of primate AAVs. X = major nodes with bootstrap values of <75. Goose parvovirus and an avian AAV were used as the outgroup. Clades are indicated by name and by vertical lines. Viruses are identified by the serotype name or a reference to the species source (hu, human; rh, rhesus macaque; cy, cynomolgus macaque; bb, baboon; pi, pigtailed macaque; ch, chimpanzee), number = order of sequencing. Clade C originated through the recombination of known clades. The AAV2-AAV3 hybrid clade originated after one recombination event, and its phylogeny is shown. From (Gao *et al.*, 2004), with permission.

Clades B and C seemed most prevalent in humans, Clades A and F were only found in humans and clade D in macaques, while clade E was isolated from both humans and non-human primates. Evolution of AAVs might be at least partly driven

by their helper viruses, and some AAVs may also have co-evolved with their hosts but experimental or phylogenetic evidence is missing. Recombination between AAV genomes seems to occur frequently and can give rise to replication-competent vectors following homologous recombination (6, 61).

2.7. Genetic variation and replication error rate

The level of variation of the parvoviruses can be high when measured over defined time periods. This has been seen in studies of variation for CPV and B19 parvovirus in nature (143, 144), in the growth of CPV through serial passage in tissue culture (9), and in the analysis of MVM during replication under selective conditions such as replication in alterative hosts (97) or monoclonal antibody selection (95). However, it is not know how much variation occurs during the replication of the parvovirus genome. Although the viral DNA is replicated using host cell DNA polymerases, the fidelity of replication is likely lower than that seen for the host chromosomal DNA replication. If DNA repair mechanisms are activated during parvovirus replication those might also lead to higher mutation rates, perhaps due to the use of alternative DNA polymerases, or to the filling in of the replicated ssDNA forms (90).

2.8. Intra-host diversity during natural infections

The degree of intra-host diversity has been assessed for B19, MVM and AMDV infections, and most recently also for one CPV infected cat. While there are probably differences in selective pressures between the chronic infections of ADV and MVM and the acute infection of CPV, all of these studies showed intra-host variation or diversity.

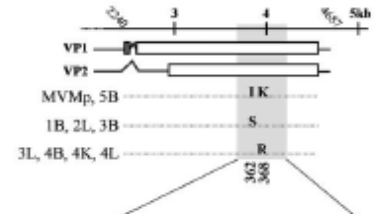
Analyzing individual virus genomes isolated from one CPV infected cat revealed diversity in a 1745 bp fragment of the VP2 gene (13). Sequences were 99.5

to 99.9% identical in nucleotide sequence and 99.3 to 99.8 % identical in amino acid sequence, but 10 distinct sequences were observed among 14 analyzed viral clones. Two antigenically distinct CPV variants (CPV type-2a and a variant with the VP2 D426E substitution, named CPV type-2c) were isolated from this animal, indicating that this was likely a co-infection by co-circulating strains. The ratio of non-synonymous to synonymous changes (dN/dS ratio), when compared to extant CPV strains circulating in Italy, ranged between 0.08 and 0.4 for individual clones, suggesting purifying selection on the capsid gene.

A spectrum of AMDV sequences was seen in experimentally infected Danish mink, with several different sequences in mink inoculated with a single inoculum. The sequences differed by up to 5% and also differed in the highly variable region (64). Since AMDV can establish persistent infections, it is likely that mixed infections occur, and viral inocula used in these studies potentially contained multiple virus strains. Comparing viral non-structural gene sequences (nts. 123-2208) from different AMDV isolates showed extensive variation in the different samples (65).

Experimental MVM infections in immune deficient mice using molecularly cloned virus stocks showed emergence of variation over a period of days or weeks (95, 96, 138). The original viruses used had likely been tissue culture adapted, and MVMp variants contained non-synonymous changes in the VP1/2 gene at sites which gave lower affinity binding of sialic acid (96, 119) (Figure 2.10). When treating severe and combined immunodeficient (SCID) mice with a neutralizing monoclonal antibody, escape mutants rapidly arose and quickly dominated the viral population, leading to a delayed onset of clinical disease. Substitutions were located close to a raised region at the 3fold axis of symmetry (95). However, polyclonal anti-capsid antibodies did not allow escape mutations of the MVMi strain to be selected, and non-synonymous changes were restricted to the NS2 protein, along with two changes in both NS1 and

NS2. NS2 changes clustered into two regions and altered the biochemical properties of the protein (97).



Mouse/ Organ	n	325 W V S	362 D I T Q G V D	368 K E
1/B	2	-	S	-
	1	-	-	R
2/L	6	-	S	-
	3	M	-	-
3/B	2	-	S	-
	1	M	S	-
	1	M	-	-
	1	-	-	R
3/L	3	-	-	R
4/B	4	-	-	R
4/K	3	-	-	R
4/L	3	-	-	R
	1	M	-	-
5/B	6	-	-	-
6/B	3	-	-	R
6/K	4	-	-	R
6/L	2	-	-	R
7/K	2	-	-	-

Figure 2.10: Sequence variation found in a part of the genome of the minute virus of mice prototype strain (MVMp) after growth in immunodeficient mice. (Top) Amino acid substitutions in the entire capsid protein (VP) gene of isolated MVMp clones from various organs (B=brain, K=kidney, L= liver). (Bottom) Distribution of amino acid changes in the collection of MVMp clones where a region of the genome (nt. 3710 to 4200) from 48 viral clones from 7 mice (numbered 1-7) was amplified. The amino acid changes at residues 325, 362 and 368 of the VP2 sequence are outlined. N=number of clones with identical genotypes in this region. From (Lopez-Bueno *et al.,* 2004), with permission.

Adeno-associated viruses show sequence diversity both within hosts and within the population. Mutations are primarily located in variable regions of the capsid proteins which are exposed to the outside of the capsid and some represent antigenic sites or receptor binding sites (127, 165). Immune responses against AAVs in vivo suggest a potential role of immune escape (reviewed by (110)).

2.9. Virus variation at a population level (spatial heterogeneity)

Sequence variation of the B19 viruses in humans has been analyzed comparing viruses from different regions of the world or from chronic and acute infections. There appears to be a global distribution of the viruses - there was close similarity between some isolates collected from various regions of the world, and also between viruses collected at various times over the past two decades. However, viruses collected from one geographic area are generally more similar to each other. For B19 viruses, viruses from patients with persistent infections appear to have a higher level of variation compared to viruses from patients with acute infections. Comparing 7 isolates collected in Italy between 1989 and 1994 from one geographic area showed maximum variation of 0.61% in a 1000 nt. sequence within the amino terminal end of the VP genes (59). Those viruses were 0.7% and 0.77% different on average from the prototype Wi and Au sequences of viruses collected in the UK and USA, respectively, and there were 9 non-synonymous changes, 7 between residues 4 and 114 in VP1 sequence and 2 within the VP1/VP2 common region (59). Viruses from Vietnam showed the presence of two genotypes of the type 1 B19 in that country (150)

The sequences of the complete VP1 and VP2 gene region (2343 nts) of 29 isolates from 25 infected patients in various regions of the world were compared to each other, and to the 2 published Wi and Au sequences. Those viruses included 10 from an outbreak in Ohio, USA, one a mother-child pair, and other isolates from the USA, UK, Brazil, Ireland, Venezuela, Korea, Japan and China (55). The sequences differed by between 2 and 99 nucleotides (4.2%), and by 0 and 13 (1.7%) in amino acid sequence. Nucleotide variation was found throughout the VP1 and VP2 genes, but coding changes clustered in three regions, the VP1-unique region, around the junction of the VP1 and VP2 coding regions, and within the VP1 and VP2 overlapping

region. Isolates from the outbreak in Ohio were divided into two classes, represented by 7 and 2 samples each (55). Within each group the sequences differed by only a few nucleotides, and where multiple isolates were collected from 4 individuals only a single difference was found in the sequences from each person. There was some geographic clustering of the strains- viruses from China formed a distinct clade, while those from the USA were generally clustered, as were the isolates from Korea.

Viruses recovered from chronic and acute infections were examined in several studies, and in most cases no specific correlation was seen between any particular disease syndrome and any virus type (59, 88, 111, 112, 148). Higher genomic variability has been seen in the B19 sequences recovered from cases of persistent infection compared to those from acute infections (78).

2.9.1. Spatial heterogeneity of CPV and related viruses.

CPV and the closely related viruses have also been examined closely. CPV-like viruses were first circulating in Europe in the mid to late 1970s, and early CPV-2 isolates were essentially identical in nucleotide sequence, and were all replaced by the variant CPV-2a strain which spread worldwide in 1979 and 1980. Other mutations have been seen worldwide in various countries since 1980, including a mutation changing VP2 residue 426 (referred to as CPV-2b) from Asn to Asp first seen around 1984 which spread worldwide. More recently viruses with residue 426 as a Glu (referred to as CPV type 2c) have been identified in Europe (48, 103), Vietnam (118), and in North and South America (136). The reasons for the global spread of some mutations and the time of their emergence are hard to explain. CPV types 2a and 2b co-circulated throughout the world (although in different proportions) over the 20 years without either becoming fixed. CPV2a and 2b co-circulated in Brazil as early as 1986 (135) while CPV-2b was the predominant variant circulating there between 1995

and 2001 (37). The dominant CPV type circulating in Italy at that time was CPV-2a (104), while both CPV2a and 2b circulated at similar frequencies in 1999/2000. There appear to be Taiwanese-Japanese and Indian CPV populations which are phylogenetically distinct from US American and Vietnamese populations (33, 49, 71). Additional mutations that have been observed include Ser297Ala and Glu300Asp, which may affect receptor binding or host range properties of the viruses. However, an analysis of CPV isolates from Brazil and other regions failed to detect clear geographic pattern in extant parvovirus isolates (135).

2.10. Temporal variation

Good temporal data analyzing the evolution of parvoviruses are still relatively rare. The best defined are those that examine the evolution of CPV after the host jump into dogs, where all viruses have evolved from a single common ancestor with an accumulation of mutations. Early time points after the colonization of a new host are thought to coincide with accelerated mutation rates due to host adaptation, while other times in the CPV pandemic appear to be characterized by selective constraints. During the emergence of the CPV lineage the viruses showed a steady rate of change of around 10^{-5} /site/year (144). Many of the capsid protein gene differences between CPV and the FPV-like viruses are associated with changes in host range, antigenicity, and sialic acid binding properties of the viruses, suggesting they are under strong selection (30, 75, 77). The CPV lineage split into two major variants during the mid-1970s, with the CPV type-2 strains emerging worldwide in 1978, and the CPV type-2a variant replacing the CPV type-2 viruses during 1979 and 1980 (Figure 2. 5). The rate of variation of FPV was lower than that seen for CPV (76, 144). Analysis of Brazilian CPV2 sequences indicated neutral evolution in the early phase of the pandemic (analyzing isolates collected around 1980), while later time points (isolates collected

around and after 1990) were associated with strong purifying selection, with the exception of a change of VP2 residue 297, which was under strong positive selection. A Bayesian Skyline plot analysis showed an initial very rapid spread with a subsequent decreasing spread (135) (Figure 2.11).

CPV isolates collected during 1978 and early 1979 were all antigenically identical worldwide, but the CPV type-2a strain differed in the VP1/VP2 protein by 4-5 amino acids compared to CPV type-2 isolates. CPV type-2a also differed from CPV type-2 in having lost two antigenic epitopes and gained two epitopes on the capsid (146). Other antigenic variants emerging around 1984, and during the later 1990s and early 2000s each contained a single amino acid sequence difference within neutralizing epitopes in the capsid (28, 79, 80, 118). Despite the small numbers of differences those viruses each became globally distributed within a year or two of being detected. Natural antigenic variation of FPV and MEV isolates has also been described (109, 131, 132). It is unclear what the consequences are of the antigenic variation that is seen in some of the viruses, but in general there appears to be good cross-protection between antigenically variant strains.

2.11. Conclusions

Our understanding of the evolution of the parvoviruses is still being developed, but there are some interesting conclusions that can be drawn. All of the parvoviruses appear to derive from a common ancestral virus, but the different genera are mostly well separated genetically, suggesting long divergence times. Surveys for non-host DNA sequences often show sequences of previously unknown parvoviruses, suggesting that there are many different viruses still to be discovered.

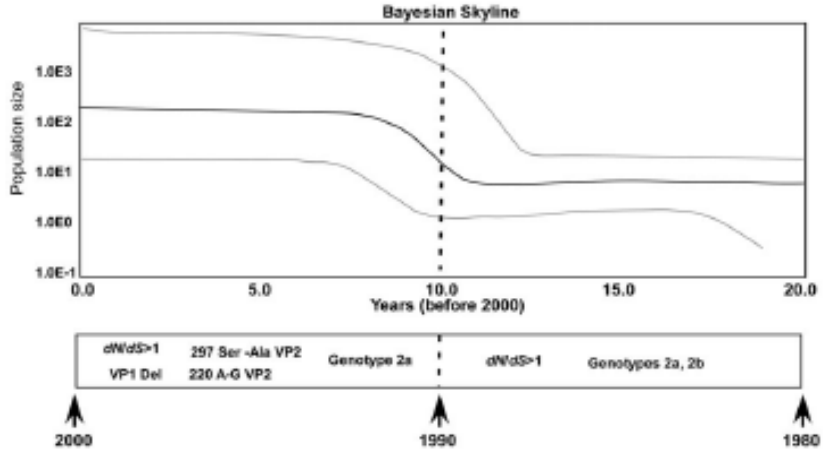


Figure 2.11: Evolutionary dynamics of canine parvovirus in Brazil. (A) Bayesian skyline plot obtained for the VP2 gene sequences of CPV strains from Brazil. The X axis is in years before 2000, the black line represents the mean and the grey line the 95% highest probability density limits. (B) Major genetic events occurring during the period of CPV epidemics. The vertical dashed line delineates the distinct patters of CPV evolution characterized by two intervals (1) the period between 1980 and 1990 with stochastic allele fixation and the occurrence of antigenic types CPV-2a and -2b, and (2) the period after 1990 distinguished by deterministic allele fixation, presence of only the antigenic type CPV-2a strain with positive mutation such as VP1 intron deletion, and a synonymous mutation in codon 220 of theVP2 gene region. From (Pereira, *et al.,* 2007), with permission.

Levels of sequence variation within a virus genus appears to differ significantly between different parvoviruses, and this may be due both to different life styles (such as whether they cause acute or persistent infections), and to differences in the levels of variation that are tolerated, particularly in the capsid protein gene. Selection is likely to be on both the gene products and on the DNA sequences themselves. Mixed infections by closely related strains of the same virus or by genetically or serologically different viruses may be common for some viruses, but these are overlooked unless subpopulations in DNA mixtures are specifically tested for. High rates of variation may be seen among many parvoviruses when they are placed under selection. The

replication fidelity and rates of mutation are not known. Although replicated by host or helper virus DNA polymerases, fidelity is likely considerably lower than those seen during replication of the host genomes. Global spread of genomic variants of some viruses occurs, and has been seen for the distribution of CPV during the early stages of the pandemic, or B19 virus in the human population. Transmission between related host species might be frequent for some parvoviruses (rodent parvoviruses, parvoviruses of the feline parvovirus group, or AAVs of the E clade), while host restrictions appear to be strict for others.

REFERENCES

1. Alexandersen, S. 1986. Acute interstitial pneumonia in mink kits: experimental reproduction of the disease. Vet Pathol **23:** 579-588.

2. Alexandersen, S., M. E. Bloom, and J. Wolfinbarger. 1988. Evidence of restricted viral replication in adult mink infected with Aleutian disease of mink parvovirus. J Virol **62:**1495-1507.

3. Alexandersen, S., S. Larsen, B. Aasted, A. Uttenthal, M. E. Bloom, and M. Hansen. 1994. Acute interstitial pneumonia in mink kits inoculated with defined isolates of Aleutian mink disease parvovirus. Vet Pathol **31:**216-28.

4. Allander, T., S. U. Emerson, R. E. Engle, R. H. Purcell, and J. Bukh. 2001. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. Proc Natl Acad Sci USA **98:**11609-14.

5. Allander, T., M. T. Tammi, M. Eriksson, A. Bjerkner, A. Tiveljung-Lindell, and B. Andersson. 2005. Cloning of a human parvovirus by molecular screening of respiratory tract samples. Proc Natl Acad Sci USA **102:**12891-6.

6. Allen, J. M., D. J. Debelak, T. C. Reynolds, and A. D. Miller. 1997. Identification and elimination of replication-competent adeno-associated virus (AAV) that can arise by nonhomologous recombination during AAV vector production. J Virol **71:**6816-6822.

7. Atchison, R. W., B. C. Casto, and W. M. Hammon. 1965. Adenovirus-Associated Defective Virus Particles. Science **149:**754-6.

8. Aubin, J. T., C. Defer, M. Vidaud, M. Maniez Montreuil, and B. Flan. 2000. Large-scale screening for human parvovirus B19 DNA by PCR: application to the quality control of plasma for fractionation. Vox Sang **78:**7-12.

9. Badgett, M. R., A. Auer, L. E. Carmichael, C. R. Parrish, and J. J. Bull. 2002. Evolutionary dynamics of viral attenuation. J Virol **76:**10524-10529.

10. Ball-Goodrich, L. J., and E. Johnson. 1994. Molecular characterization of a newly recognized mouse parvovirus. J Virol **68:**6476-86.

11. Ball-Goodrich, L. J., S. E. Leland, E. A. Johnson, F. X. Paturzo, and R. O. Jacoby. 1998. Rat parvovirus type 1: the prototype for a new rodent parvovirus serogroup. J Virol **72:**3289-99.

12. Bashir, T., R. Horlein, J. Rommelaere, and K. Willwand. 2000. Cyclin A activates the DNA polymerase delta -dependent elongation machinery in vitro: A parvovirus DNA replication model. Proc Natl Acad Sci USA **97:**5522-7.

13. Battilani, M., A. Scagliarini, S. Ciulli, L. Morganti, and S. Prosperi. 2006. High genetic diversity of the VP2 gene of a canine parvovirus strain detected in a domestic cat. Virology **352:**22-6.

14. Becker, S. D., M. Bennett, J. P. Stewart, and J. L. Hurst. 2007. Serological survey of virus infection among wild house mice (Mus domesticus) in the UK. Lab Anim **41:**229-38.

15. Berns, K. I. 1990. Parvovirus replication. Microbiol Rev **54:**316-329.

16. Berns, K. I., and R. M. Linden. 1995. The cryptic life style of adeno-associated virus. Bioessays **17:**237-45.

17. Besselsen, D. G., C. L. Besch-Williford, D. J. Pintel, C. L. Franklin, R. R. Hook, Jr., and L. K. Riley. 1995. Detection of newly recognized rodent parvoviruses by PCR. J Clin Microbiol **33:**2859-63.

18. Besselsen, D. G., D. J. Pintel, G. A. Purdy, C. L. Besch-Williford, C. L. Franklin, R. R. Hook, Jr., and L. K. Riley. 1996. Molecular characterization of newly recognized rodent parvoviruses. J Gen Virol **77 (Pt 5):**899-911.

19. Besselsen, D. G., M. J. Romero, A. M. Wagner, K. S. Henderson, and R. S. Livingston. 2006. Identification of novel murine parvovirus strains by epidemiological analysis of naturally infected mice. The Journal of general virology **87:**1543-56.

20. Bianchi, M., I. Rago, G. Zini, G. D'Onofrio, and G. Leone. 2007. Parvovirus b19 infection after plasma exchange for myasthenia gravis. Lab Hematol **13:**34-8.

21. Bloom, M. E., O.-R. Kaaden, E. Huggans, A. Cohn, and J. B. Wolfinbarger. 1988. Molecular comparisons of in vivo and in vitro derived strains of Aleutian disease of mink parvovirus. J Virol **62:**132-138.

22. Bodendorf, U., C. Cziepluch, J. C. Jauniaux, J. Rommelaere, and N. Salome. 1999. Nuclear export factor CRM1 interacts with nonstructural proteins NS2 from parvovirus minute virus of mice. J Virol **73:**7769-79.

23. Bonnard, G. D., E. K. Manders, D. A. Campbell, Jr., R. B. Herberman, and M. J. Collins, Jr. 1976. Immunosuppressive activity of a subline of the mouse EL-

4 lymphoma. Evidence for minute virus of mice causing the inhibition. J Exp Med **143:**187-205.

24. Brockhaus, K., S. Plaza, D. J. Pintel, J. Rommelaere, and N. Salome. 1996. Nonstructural proteins NS2 of minute virus of mice associate in vivo with 14-3-3 protein family members. J Virol **70:**7527-7534.

25. Broll, S., and S. Alexandersen. 1996. Investigation of the pathogenesis of transplacental transmission of Aleutian mink disease parvovirus in experimentally infected mink. J Virol **70:**1455-1466.

26. Brown, K. E., and N. S. Young. 1995. Parvovirus B19 infection and hematopoiesis. Blood Rev **9:**176-82.

27. Brown, K. E., N. S. Young, and J. M. Liu. 1994. Molecular, cellular and clinical aspects of parvovirus B19 infection. Crit Rev Oncol Hematol **16:**1-31.

28. Buonavoglia, C., V. Martella, A. Pratelli, M. Tempesta, A. Cavalli, D. Buonavoglia, G. Bozzo, G. Elia, N. Decaro, and L. Carmichael. 2001. Evidence for evolution of canine parvovirus type 2 in Italy. J Gen Virol **82:**3021-5.

29. Candotti, D., N. Etiz, A. Parsyan, and J. P. Allain. 2004. Identification and characterization of persistent human erythrovirus infection in blood donor samples. J Virol **78:**12169-78.

30. Chang, S. F., J. Y. Sgro, and C. R. Parrish. 1992. Multiple amino acids in the capsid structure of canine parvovirus coordinately determine the canine host range and specific antigenic and hemagglutination properties. J Virol **66:**6858-6567.

31. Chen, K. C., B. C. Shull, M. Lederman, E. R. Stout, and R. C. Bates. 1988. Analysis of the termini of the DNA of bovine parvovirus: demonstration of sequence inversion at the left terminus and its implication for the replication model. J Virol **62:**3807-3813.

32. Chen, K. C., B. C. Shull, E. A. Moses, M. Lederman, E. R. Stout, and R. C. Bates. 1986. Complete nucleotide sequence and genome organisation of bovine parvovirus. J.Virol. **60:**1085-1097.

33. Chinchkar, S. R., B. Mohana Subramanian, N. Hanumantha Rao, P. N. Rangarajan, D. Thiagarajan, and V. A. Srinivasan. 2006. Analysis of VP2 gene sequences of canine parvovirus isolates in India. Arch Virol. **151:** 1881-7.

34. Chiorini, J. A., L. Yang, B. Safer, and R. M. Kotin. 1995. Determination of adeno-associated virus Rep68 and Rep78 binding sites by random sequence oligonucleotide selection. J Virol **69:**7334-8.

35. Choi, E. Y., A. E. Newman, L. Burger, and D. Pintel. 2005. Replication of minute virus of mice DNA is critically dependent on accumulated levels of NS2. J Virol **79:**12375-81.

36. Cossons, N., E. A. Faust, and M. Zannis-Hadjopoulos. 1996. DNA polymerase delta-dependent formation of a hairpin structure at the 5' terminal palindrome of the minute virus of mice genome. Virology **216:**258-64.

37. Costa, A. P., J. P. Leite, N. V. Labarthe, and R. C. Garcia. 2005. Genomic typing of canine parvovirus circulating in the state of rio de janeiro, Brazil from 1995 to 2001 using polymerase chain reaction assay. Vet Res Commun **29:**735-43.

38. Cotmore, S. F. 1996. Parvovirus DNA replication. In: DePamphilis ML, ed. DNA Replicaton in Eukaryotic Cells. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1996:799.

39. Cotmore, S. F., A. M. D'Abramo, Jr., L. F. Carbonell, J. Bratton, and P. Tattersall. 1997. The NS2 polypeptide of parvovirus MVM is required for capsid assembly in murine cells. Virology **231:**267-280.

40. Cotmore, S. F., and P. Tattersall. 1987. The autonomously replicating parvoviruses of vertebrates. Adv.Virus Res. **33:**91-174.

41. Cotmore, S. F., and P. Tattersall. 1995. DNA replication in the autonomous parvoviruses. Seminars in Virology **6:**271-281.

42. Cotmore, S. F., and P. Tattersall. 2005. Genome packaging sense is controlled by the efficiency of the nick site in the right-end replication origin of parvoviruses minute virus of mice and LuIII. J Virol **79:**2287-2300.

43. Cotmore, S. F., and P. Tattersall. 2003. Resolution of parvovirus dimer junctions proceeds through a novel heterocruciform intermediate. J Virol **77:**6245-6254.

44. Crawford, L. V. 1966. A minute virus of mice. Virology **29:**605-612.

45. D'Abramo, A. M., Jr., A. A. Ali, F. Wang, S. F. Cotmore, and P. Tattersall. 2005. Host range mutants of minute virus of mice with a single VP2 amino acid change require additional silent mutations that regulate NS2 accumulation. Virology **340:**143-54.

46. Daeffler, L., R. Horlein, J. Rommelaere, and J. P. Nuesch. 2003. Modulation of minute virus of mice cytotoxic activities through site-directed mutagenesis within the NS coding region. J Virol **77:**12466-12478.

47. de Haan, T. R., M. F. Beersma, E. C. Claas, D. Oepkes, A. C. Kroes, and F. J. Walther. 2007. Parvovirus B19 infection in pregnancy studied by maternal viral load and immune responses. Fetal Diagn Ther **22:**55-62.

48. Decaro, N., G. Elia, V. Martella, C. Desario, M. Campolo, L. D. Trani, E. Tarsitano, M. Tempesta, and C. Buonavoglia. 2005. A real-time PCR assay for rapid detection and quantitation of canine parvovirus type 2 in the feces of dogs. Vet Microbiol **105:**19-28.

49. Doki, M., K. Fujita, R. Miura, M. Yoneda, Y. Ishikawa, A. Taneno, and C. Kai. 2006. Sequence analysis of VP2 gene of canine parvovirus isolated from domestic dogs in Japan in 1999 and 2000. Comparative immunology, microbiology and infectious diseases **29:**199-206.

50. Dorsch, S., B. Kaufmann, U. Schaible, E. Prohaska, H. Wolf, and S. Modrow. 2001. The VP1-unique region of parvovirus B19: amino acid variability and antigenic stability. J Gen Virol **82:**191-199.

51. Dubielzig, R., J. A. King, S. Weger, A. Kern, and J. A. Kleinschmidt. 1999. Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes. J Virol **73:**8989-98.

52. Eichwald, V., L. Daeffler, M. Klein, J. Rommelaere, and N. Salome. 2002. The NS2 proteins of parvovirus minute virus of mice are required for efficient nuclear egress of progeny virions in mouse cells. J Virol **76:**10307-19.

53. Eis-Hubinger, A. M., D. Dieck, R. Schild, M. Hansmann, and K. E. Schneweis. 1998. Parvovirus B19 infection in pregnancy. Intervirology **41:**178-84.

54. Ekman, A., K. Hokynar, L. Kakkola, K. Kantola, L. Hedman, H. Bonden, M. Gessner, C. Aberham, P. Norja, S. Miettinen, K. Hedman, and M. Soderlund-Venermo. 2007. Biological and immunological relation of human parvovirus B19 genotypes 1-3. J Virol.**81**: 6927-35.

55. Erdman, D. D., E. L. Durigon, Q.-Y. Wang, and L. J. Anderson. 1996. Genetic diversity of human parvovirus B19: sequence analysis of the VP1/VP2 gene from multiple isolates. J Gen Virol **77:**2767-2774.

56. Farr, G. A., L. G. Zhang, and P. Tattersall. 2005. Parvoviral virions deploy a capsid-tethered lipolytic enzyme to breach the endosomal membrane during cell entry. Proc Natl Acad Sci USA **102:**17148-53.

57. Frickhofen, N., and N. S. Young. 1989. Persistent parvovirus B19 infections in humans. Microbial Pathogenesis **7:**319-327.

58. Fry, A. M., X. Lu, M. Chittaganpitch, T. Peret, J. Fischer, S. F. Dowell, L. J. Anderson, D. Erdman, and S. J. Olsen. 2007. Human bocavirus: a novel parvovirus epidemiologically associated with pneumonia requiring hospitalization in Thailand. J Infect Dis **195:**1038-45.

59. Gallinella, G., S. Venturoli, G. Gentilomi, M. Musiani, and M. Zerbini. 1995. Extent of sequence variability in a genomic region coding for capsid proteins of B19 parvovirus. Arch Virol **140:**1119-25.

60. Gallinella, G., S. Venturoli, E. Manaresi, M. Musiani, and M. Zerbini. 2003. B19 virus genome diversity: epidemiological and clinical correlations. J Clin Virol **28:**1-13.

61. Gao, G., M. R. Alvira, S. Somanathan, Y. Lu, L. H. Vandenberghe, J. J. Rux, R. Calcedo, J. Sanmiguel, Z. Abbas, and J. M. Wilson. 2003. Adeno-associated viruses undergo substantial evolution in primates during natural infections. Proc Natl Acad Sci USA **100:**6081-6.

62. Gao, G., L. H. Vandenberghe, M. R. Alvira, Y. Lu, R. Calcedo, X. Zhou, and J. M. Wilson. 2004. Clades of Adeno-associated viruses are widely disseminated in human tissues. J Virol **78:**6381-8.

63. Girod, A., C. E. Wobus, Z. Zadori, M. Ried, K. Leike, P. Tijssen, J. A. Kleinschmidt, and M. Hallek. 2002. The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. J Gen Virol **83:**973-8.

64. Gottschalck, E., S. Alexandersen, A. Cohn, L. A. Poulsen, M. E. Bloom, and B. Aasted. 1991. Nucleotide sequence analysis of Aleutian mink disease parvovirus shows that multiple virus types are present in infected mink. J Virol **65:**4378-4386.

65. Gottschalck, E., S. Alexandersen, T. Storgaard, M. E. Bloom, and B. Aasted. 1994. Sequence comparison of the non-structural genes of four different types of Aleutian mink disease parvovirus indicates an unusual degree of variability. Arch Virol **138:**213-31.

66. Guerin, B., and N. Pozzi. 2005. Viruses in boar semen: detection and clinical as well as epidemiological consequences regarding disease transmission by artificial insemination. Theriogenology **63:**556-72.

67. Hattori, S., M. Yunoki, M. Tsujikawa, T. Urayama, Y. Tachibana, I. Yamamoto, S. Yamamoto, and K. Ikuta. 2007. Variability of parvovirus B19 to inactivation by liquid heating in plasma products. Vox Sang **92:**121-4.

68. Hayder, H. A., J. Storz, and S. Young. 1983. Antigenicity of bovine parvovirus in fetal infections. Am J Vet Res **44:**558-63.

69. Heegaard, E. D., and E. B. Brown. 2002. Human Parvovirus B19. Clinical Microbiology Reviews **15:**485-505.

70. Hemauer, A., A. von Pobotzki, A. Gigler, P. Cassinotti, G. Siegl, H. Wolf, and S. Modrow. 1996. Sequence variability among different parvovirus B19 isolates`. J Gen Virol **77:**1781-1785.

71. Hirayama, K., R. Kano, T. Hosokawa-Kanai, K. Tuchiya, S. Tsuyama, Y. Nakamura, Y. Sasaki, and A. Hasegawa. 2005. VP2 gene of a canine parvovirus isolate from stool of a puppy. J Vet Med Sci **67:**139-43.

72. Hoggan, M. D., N. R. Blacklow, and W. P. Rowe. 1966. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. Proc Natl Acad Sci USA **55:**1467-74.

73. Hokynar, K., M. Soderlund-Venermo, M. Pesonen, A. Ranki, O. Kiviluoto, E. K. Partio, and K. Hedman. 2002. A new parvovirus genotype persistent in human skin. Virology **302:**224-8.

74. Hong, G., P. Ward, and K. I. Berns. 1994. Intermediates of adeno-associated virus DNA replication in vitro. J Virol **68:**2011-5.

75. Horiuchi, M., H. Goto, N. Ishiguro, and M. Shinagawa. 1994. Mapping of determinants of the host range for canine cells in the genome of canine parvovirus using canine parvovirus/mink enteritis virus chimeric viruses. J Gen Virol **75:**1319-1328.

76. Horiuchi, M., Y. Yamaguchi, T. Gojobori, M. Mochizuki, H. Nagasawa, Y. Toyoda, N. Ishiguro, and M. Shinagawa. 1998. Differences in the evolutionary pattern of feline panleukopenia virus and canine parvovirus. Virology **249:**440-52.

77. Hueffer, K., L. Govindasamy, M. Agbandje-McKenna, and C. R. Parrish. 2003. Combinations of two capsid regions controlling canine host range

determine canine transferrin receptor binding by canine and feline parvoviruses. J Virol **77:**10099-10105.

78. Hung, C. C., W. H. Sheng, K. L. Lee, S. J. Yang, and M. Y. Chen. 2006. Genetic drift of parvovirus B19 is found in AIDS patients with persistent B19 infection. J Med Virol **78:**1374-84.

79. Ikeda, Y., M. Mochizuki, R. Naito, K. Nakamura, T. Miyazawa, T. Mikami, and E. Takahashi. 2000. Predominance of canine parvovirus (CPV) in unvaccinated cat populations and emergence of new antigenic types of CPVs in cats. Virology **278:**13-9.

80. Ikeda, Y., K. Nakamura, T. Miyazawa, E. Takahashi, and M. Mochizuki. 2002. Feline host range of canine parvovirus: recent emergence of new antigenic types in cats. Emerg Infect Dis **8:**341-6.

81. Jacoby, R. O., L. J. Ball-Goodrich, D. G. Besselsen, M. D. McKisic, L. K. Riley, and A. L. Smith. 1996. Rodent parvovirus infections. Lab Anim Sci **46:**370-80.

82. Jacoby, R. O., E. A. Johnson, F. X. Paturzo, and L. Ball-Goodrich. 2000. Persistent rat virus infection in smooth muscle of euthymic and athymic rats. J Virol **74:**11841-8.

83. Jacoby, R. O., E. A. Johnson, F. X. Paturzo, D. J. Gaertner, J. L. Brandsma, and A. L. Smith. 1991. Persistent rat parvovirus infection in individually housed rats. Arch Virol **117:**193-205.

84. Jensen, K. T., J. B. Wolfinbarger, B. Aasted, and M. E. Bloom. 2000. Replication of Aleutian mink disease parvovirus in mink lymph node histocultures. J Gen Virol **81:**335-343.

85. Joo, H. S., C. R. Donaldson-Wood, and R. H. Johnson. 1976. Observations on the pathogenesis of porcine parvovirus infection. Arch Virol **51:**123-9.

86. Kasprowicz, V., A. Isa, T. Tolfvenstam, K. Jeffery, P. Bowness, and P. Klenerman. 2006. Tracking of peptide-specific CD4+ T-cell responses after an acute resolving viral infection: a study of parvovirus B19. J Virol **80:**11209-17.

87. Kaufmann, J., J. M. Buccola, W. Stead, C. Rowley, M. Wong, and C. K. Bates. 2007. Secondary symptomatic parvovirus B19 infection in a healthy adult. J Gen Intern Med **22:**877-8.

88. Kerr, J. R., M. D. Curran, J. E. Moore, D. D. Erdman, P. V. Coyle, T. Nunoue, D. Middleton, and W. P. Ferguson. 1995. Genetic diversity in the non-

structural gene of parvovirus B19 detected by single-stranded conformational polymorphism assay (SSCP) and partial nucleotide sequencing. J Virol Methods **53:**213-22.

89. Kollek, R., B. Y. Tseng, and M. Goulian. 1982. DNA polymerase requirements for parvovirus H-1 DNA replication in vitro. J Virol **41:**982-9.

90. Kunkel, T. A**.** 2004. DNA replication fidelity. J Biol Chem **279:**16895-8.

91. Lachmann, S., J. Rommeleare, and J. P. Nuesch. 2003. Novel PKCeta is required to activate replicative functions of the major nonstructural protein NS1 of minute virus of mice. J Virol **77:**8048-60.

92. Lederman, M. P., J. T.; Stout, E. R.; Bates, R. C. 1984. Virally encoded noncapsid protein associated with bovine parvovirus infection. J Virol **49:**315-318.

93. Lefrere, J. J., A. Servant-Delmas, D. Candotti, M. Mariotti, I. Thomas, Y. Brossard, F. Lefrere, R. Girot, J. P. Allain, and S. Laperche. 2005. Persistent B19 infection in immunocompetent individuals: implications for transfusion safety. Blood **106:**2890-5.

94. Li, X., and S. L. Rhode. 1991. Nonstructural protein NS2 of parvovirus H-1 is required for efficient viral protein synthesis and virus production in rat cells in vivo and in vitro. Virology **184:**117-130.

95. Lopez-Bueno, A., M. G. Mateu, and J. M. Almendral. 2003. High mutant frequency in populations of a DNA virus allows evasion from antibody therapy in an immunodeficient host. J Virol **77:**2701-2708.

96. Lopez-Bueno, A., M. P. Rubio, N. Bryant, R. McKenna, M. Agbandje-McKenna, and J. M. Almendral. 2006. Host-selected amino acid changes at the sialic acid binding pocket of the parvovirus capsid modulate cell binding affinity and determine virulence. J Virol **80:**1563-73.

97. Lopez-Bueno, A., N. Valle, J. M. Gallego, J. Perez, and J. M. Almendral. 2004. Enhanced cytoplasmic sequestration of the nuclear export receptor CRM1 by NS2 mutations developed in the host regulates parvovirus fitness. J Virol **78:**10674-84.

98. Lukashov, V. V., and J. Goudsmit. 2001. Evolutionary relationships among parvoviruses: virus-host coevolution among autonomous primate parvoviruses and links between adeno- associated and avian parvoviruses. J Virol **75:**2729-2740.

99. Macartney, L., I. A. McCandlish, H. Thompson, and H. J. Cornwell. 1984. Canine parvovirus enteritis 1: Clinical, haematological and pathological features of experimental infection. Vet Rec **115:**201-210.

100. Macartney, L., I. A. McCandlish, H. Thompson, and H. J. Cornwell. 1984. Canine parvovirus enteritis 2: Pathogenesis. Vet Rec **115:**453-460.

101. Mahy, B**.** 2006. A new pathogenic human parvovirus. Reviews in medical virology **16:**279-80.

102. Manning, A., S. J. Willey, J. E. Bell, and P. Simmonds. 2007. Comparison of tissue distribution, persistence, and molecular epidemiology of parvovirus B19 and novel human parvoviruses PARV4 and human bocavirus. J Infect Dis **195:**1345-52.

103. Martella, V., A. Cavalli, A. Pratelli, G. Bozzo, M. Camero, D. Buonavoglia, D. Narcisi, M. Tempesta, and C. Buonavoglia. 2004. A canine parvovirus mutant is spreading in Italy. J Clin Microbiol **42:**1333-6.

104. Martella, V., N. Decaro, G. Elia, and C. Buonavoglia. 2005. Surveillance activity for canine parvovirus in Italy. J Vet Med B Infect Dis Vet Public Health **52:**312-5.

105. McKenna, R., N. H. Olson, P. R. Chipman, T. S. Baker, T. F. Booth, J. Christensen, B. Aasted, J. M. Fox, M. E. Bloom, J. B. Wolfinbarger, and M. Agbandje-McKenna. 1999. Three-dimensional structure of Aleutian mink disease parvovirus: implications for disease pathogenicity. J Virol **73:**6882-91.

106. Mengeling, W. L., P. S. Paul, and T. T. Brown. 1980. Transplacental infection and embryonic death following maternal exposure to porcine parvovirus near the time of conception. Arch Virol **65:**55-62.

107. Miller, C. L., and D. J. Pintel. 2002. Interaction between parvovirus NS2 protein and nuclear export factor Crm1 is important for viral egress from the nucleus of murine cells. J Virol **76:**3257-3266.

108. Mochizuki, M., R. Harasawa, and H. Nakatani. 1993. Antigenic and genomic variabilities among recently prevalent parvoviruses of canine and feline origin in Japan. Vet Microbiol **38:**1-10.

109. Mochizuki, M., S. Konishi, M. Ajiki, and T. Akaboshi. 1989. Comparison of feline parvovirus subspecific strains using monoclonal antibodies against a feline panleukopenia virus. Nippon Juigaku Zasshi **51:**264-272.

110. Monahan, P. E., K. Jooss, and M. S. Sands. 2002. Safety of adeno-associated virus gene therapy vectors: a current evaluation. Expert Opin Drug Saf **1:**79-91.

111. Mori, J., P. Beattie, D. W. Melton, B. J. Cohen, and J. P. Clewley. 1987. Structure and mapping of the DNA of human parvovirus B19. J Gen Virol **68 (Pt 11):**2797-806.

112. Morinet, F., Tratschin, J.-D., Perol, Y., Siegl, G. 1986. Comparison of 17 isolates of the human parvovirus B19 by restriction enzyme analysis. Arch.Virol. **90: 165-172**.

113. Morita, E., A. Nakashima, H. Asao, H. Sato, and K. Sugamura. 2003. Human parvovirus B19 nonstructural protein (NS1) induces cell cycle arrest at G(1) phase. J Virol **77:**2915-21.

114. Morita, E., K. Tada, H. Chisaka, H. Asao, H. Sato, N. Yaegashi, and K. Sugamura. 2001. Human parvovirus b19 induces cell cycle arrest at g(2) phase with accumulation of mitotic cyclins. J Virol **75:**7555-63.

115. Murphy, F. A., C. M. Fauquet, M. A. Mayo, S. A. Jarvis, S. A. Ghabrial, M. D. Summers, G. P. Martelli, and D. H. L. Bishop. 1995. International Committee on Taxonomy of Viruses, International Union of Microbiological Societies. Virology Division Virus taxonomy/G: classification and nomenclature of viruses : Sixth report of the International Committee on Taxonomy of Viruses. Arch. of Virol. (Suppl.) **viii, 586.**

116. Musiani, M., M. Zerbini, G. Gentilomi, M. Plazzi, G. Gallinella, and S. Venturoli. 1995. Parvovirus B19 clearance from peripheral blood after acute infection. J Infect Dis **172:**1360-3.

117. Naeger, L. K., N. Salome, and D. J. Pintel. 1993. NS2 is required for efficient translation of viral mRNA in minute virus of mice-infected murine cells. J. Virol. **67:**1034-1043.

118. Nakamura, M., Y. Tohya, T. Miyazawa, M. Mochizuki, H. T. Phung, N. H. Nguyen, L. M. Huynh, L. T. Nguyen, P. N. Nguyen, P. V. Nguyen, N. P. Nguyen, and H. Akashi. 2004. A novel antigenic variant of canine parvovirus from a Vietnamese dog. Arch Virol **149:**2261-9.

119. Nam, H. J., B. Gurda-Whitaker, W. Y. Gan, S. Ilaria, R. McKenna, P. Mehta, R. A. Alvarez, and M. Agbandje-McKenna. 2006. Identification of the sialic acid structures recognized by minute virus of mice and the role of binding affinity in virulence adaptation. J Biol Chem **281:**25670-7.

120. Nguyen, Q. T., C. Sifer, V. Schneider, X. Allaume, A. Servant, F. Bernaudin, V. Auguste, and A. Garbarg-Chenon. 1999. Novel human erythrovirus associated with transient aplastic anemia. J Clin Microbiol **37:**2483-7.

121. Nuesch, J. P., J. Christensen, and J. Rommelaere. 2001. Initiation of minute virus of mice DNA replication is regulated at the level of origin unwinding by atypical protein kinase C phosphorylation of NS1. J Virol **75:**5730-9.

122. Nuesch, J. P., S. Lachmann, R. Corbau, and J. Rommelaere. 2003. Regulation of minute virus of mice NS1 replicative functions by atypical PKClambda in vivo. J Virol **77:**433-42.

123. Oie, K. L., G. Durrant, J. B. Wolfinbarger, D. Martin, F. Costello, S. Perryman, D. Hogan, W. J. Hadlow, and M. E. Bloom. 1996. The relationship between capsid protein (VP2) sequence and pathogenicity of Aleutian mink disease parvovirus (ADV): a possible role for raccoons in the transmission of ADV infections. J Virol **70:**852-61.

124. Olofsson, A., C. Mittelholzer, L. Treiberg Berndtsson, L. Lind, T. Mejerland, and S. Belak. 1999. Unusual, high genetic diversity of Aleutian mink disease virus. J Clin Microbiol **37:**4145-9.

125. Op De Beeck, A., and P. Caillet-Fauquet. 1997. Viruses and the cell cycle. Prog Cell Cycle Res **3:**1-19.

126. Op De Beeck, A., J. Sobczak-Thepot, H. Sirma, F. Bourgain, C. Brechot, and P. Caillet-Fauquet. 2001. NS1- and Minute Virus of Mice-Induced Cell Cycle Arrest: Involvement of p53 and p21cip1. J Virol **75:**11071-11078.

127. Opie, S. R., K. H. Warrington Jr, Jr., M. Agbandje-McKenna, S. Zolotukhin, and N. Muzyczka. 2003. Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. J Virol **77:**6995-7006.

128. Parker, J. S. L., and C. R. Parrish. 1997. Canine parvovirus host range is determined by the specific conformation of an additional region of the capsid. J. Virol. **71:**9214-9222.

129. Parrish, C. R. 1995. Pathogenesis of feline panleukopenia virus and canine parvovirus. Baillieres Clin Haematol **8:**57-71.

130. Parrish, C. R., C. Aquadro, M. L. Strassheim, J. F. Evermann, J.-Y. Sgro, and H. Mohammed. 1991. Rapid antigenic-type replacement and DNA sequence evolution of canine parvovirus. J Virol **65:**6544-6552.

131. Parrish, C. R., and L. E. Carmichael. 1983. Antigenic structure and variation of canine parvovirus type-2, feline panleukopenia virus, and mink enteritis virus. Virology **129:**401-414.

132. Parrish, C. R., J. R. Gorham, T. M. Schwartz, and L. E. Carmichael. 1984. Characterisation of antigenic variation among mink enteritis virus isolates. Am J Vet Res **45:**2591-2599.

133. Parsyan, A., E. Addo-Yobo, S. Owusu-Ofori, H. Akpene, F. Sarkodie, and J. P. Allain. 2006. Effects of transfusion on human erythrovirus B19-susceptible or -infected pediatric recipients in a genotype 3-endemic area. Transfusion **46:**1593-600.

134. Pennick, K. E., M. A. Stevenson, K. S. Latimer, B. W. Ritchie, and C. R. Gregory. 2005. Persistent viral shedding during asymptomatic Aleutian mink disease parvoviral infection in a ferret. J Vet Diagn Invest **17:**594-7.

135. Pereira, C. A., E. S. Leal, and E. L. Durigon. 2007. Selective regimen shift and demographic growth increase associated with the emergence of high-fitness variants of canine parvovirus. Infect Genet Evol **7:**399-409.

136. Perez, R., L. Francia, V. Romero, L. Maya, I. Lopez, and M. Hernandez. 2007. First detection of canine parvovirus type 2c in South America. Vet Microbiol.**124**: 147-152.

137. Porter, D. D. 1986. Aleutian disease: a persistent parvovirus infection of mink with a maximal but ineffective host humoral immune response. Prog Med Virol **33:**42-60.

138. Rubio, M. P., A. Lopez-Bueno, and J. M. Almendral. 2005. Virulent variants emerging in mice infected with the apathogenic prototype strain of the parvovirus minute virus of mice exhibit a capsid with low avidity for a primary receptor. J Virol **79:**11280-90.

139. Saikawa, T., S. Anderson, M. Momoeda, S. Kajigaya, and N. S. Young. 1993. Neutralizing linear epitopes of B19 parvovirus cluster in the VP1 unique and VP1-VP2 junction regions. J Virol **67:**3004-3009.

140. Sanabani, S., W. K. Neto, J. Pereira, and E. C. Sabino. 2006. Sequence variability of human erythroviruses present in bone marrow of Brazilian patients with various parvovirus B19-related hematological symptoms. J Clin Microbiol **44:**604-6.

141. Schwartz, D., B. Green, L. E. Carmichael, and C. R. Parrish. 2002. The canine minute virus (minute virus of canines) is a distinct parvovirus that is most similar to bovine parvovirus. Virology **302:**219-23.

142. Servant, A., S. Laperche, F. Lallemand, V. Marinho, G. De Saint Maur, J. F. Meritet, and A. Garbarg-Chenon. 2002. Genetic diversity within human erythroviruses: identification of three genotypes. J Virol **76:**9124-34.

143. Shackelton, L. A., and E. C. Holmes. 2006. Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. J Virol **80:**3666-9.

144. Shackelton, L. A., C. R. Parrish, U. Truyen, and E. C. Holmes. 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc Natl Acad Sci U S A **102:**379-384.

145. Smith, R. H., and R. M. Kotin. 2000. An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV ori DNA. J Virol **74:**3122-9.

146. Strassheim, L. S., A. Gruenberg, P. Veijalainen, J.-Y. Sgro, and C. R. Parrish. 1994. Two dominant neutralizing antigenic determinants of canine parvovirus are found on the threefold spike of the virus capsid. Virology **198:**175-184.

147. Suikkanen, S., M. Antila, A. Jaatinen, M. Vihinen-Ranta, and M. Vuento. 2003. Release of canine parvovirus from endocytic vesicles. Virology **316:**267-80.

148. Takahashi, N., N. Takada, T. Hashimoto, and T. Okamoto. 1999. Genetic heterogeneity of the immunogenic viral capsid protein region of human parvovirus B19 isolates obtained from an outbreak in a pediatric ward. FEBS Lett **450:**289-93.

149. Tattersall, P., M. Bergoin, M. E. Bloom, and e. al. 2005. Parvoviridae. In: C.M.Fauquet, M.A.Mayo, L.Maniloff, et al., eds. Virus Taxonomy. Amsterdam: Elsevier, Inc.**:**353-369.

150. Toan, N. L., A. Duechting, P. G. Kremsner, H. Song le, M. Ebinger, S. Aberle, V. Q. Binh, D. N. Duy, J. Torresi, R. Kandolf, and C. T. Bock. 2006. Phylogenetic analysis of human parvovirus B19, indicating two subgroups of genotype 1 in Vietnamese patients. J Gen Virol **87:**2941-9.

151. Toolan, H. W. 1960. Experimental production of mongoloid hamsters. Science **131:**1446-1448.

152. Tratschin, J.-D., G. K. McMaster, G. Kronauer, and G. Siegl. 1982. Canine parvovirus: relationship to wild-type and vaccine strains of feline panleukopenia virus and mink enteritis virus. J.Gen.Virol. **61:**33-41.

153. Truyen, U., J. F. Evermann, E. Vieler, and C. R. Parrish. 1996. Evolution of canine parvovirus involved loss and gain of feline host range. Virology **215:**186-189.

154. Truyen, U., K. Geissler, C. R. Parrish, W. Hermanns, and G. Siegl. 1998. No evidence for a role of modified live virus vaccines in the emergence of canine parvovirus. J Gen Virol **79:**1153-8.

155. Truyen, U., A. Gruenberg, S. F. Chang, B. Obermaier, P. Veijalainen, and C. R. Parrish. 1995. Evolution of the feline-subgroup parvoviruses and the control of canine host range in vivo. J Virol **69:**4702-4710.

156. Truyen, U., and C. R. Parrish. 2000. Epidemiology and pathology of autonomous parvoviruses. Contrib Microbiol **4:**149-62.

157. Truyen, U., G. Platzer, and C. R. Parrish. 1996. Antigenic type distribution among canine parvoviruses in dogs and cats in Germany. Vet. Record **138:**365-366.

158. Truyen, U., G. Platzer, C. R. Parrish, T. Hanichen, W. Hermanns, and O. R. Kaaden. 1994. Detection of canine parvovirus DNA in paraffin-embedded tissues by polymerase chain reaction. Zentralbl Veterinarmed [B] **41:**148-52.

159. Wan, C. H., B. A. Bauer, D. J. Pintel, and L. K. Riley. 2006. Detection of rat parvovirus type 1 and rat minute virus type 1 by polymerase chain reaction. Lab Anim **40:**63-9.

160. Wan, C. H., M. Soderlund-Venermo, D. J. Pintel, and L. K. Riley. 2002. Molecular characterization of three newly recognized rat parvoviruses. J Gen Virol **83:**2075-83.

161. Wang, D., W. Yuan, I. Davis, and C. R. Parrish. 1998. Nonstructural protein-2 and the replication of canine parvovirus. Virology **240:**273-81.

162. Wistuba, A., A. Kern, S. Weger, D. Grimm, and J. A. Kleinschmidt. 1997. Subcellular compartmentalization of adeno-associated virus type 2 assembly. J Virol **71:**1341-52.

163. Wobus, C. E., B. Hugle-Dorr, A. Girod, G. Petersen, M. Hallek, and J. A. Kleinschmidt. 2000. Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and identification of capsid domains

involved in AAV-2-cell interaction and neutralization of AAV-2 infection. J Virol **74:**9281-93.

164. Wong, S., N. S. Young, and K. E. Brown. 2003. Prevalence of parvovirus b19 in liver tissue: no association with fulminant hepatitis or hepatitis-associated aplastic anemia. J Infect Dis **187:**1581-6.

165. Wu, Z., A. Asokan, J. C. Grieger, L. Govindasamy, M. Agbandje-McKenna, and R. J. Samulski. 2006. Single amino acid changes can influence titer, heparin binding, and tissue tropism in different adeno-associated virus serotypes. J Virol **80:**11393-7.

166. Yoo, B. C., D. H. Lee, S. M. Park, J. W. Park, C. Y. Kim, H. S. Lee, J. S. Seo, K. J. Park, and W. S. Ryu. 1999. A novel parvovirus isolated from Manchurian chipmunks. Virology **253:**250-8.

167. Yoshimoto, K., S. Rosenfeld, N. Frickhofen, D. Kennedy, R. Hills, S. Kajigaya, and N. Young. 1991. A second neutralizing epitope of B19 parvovirus implicates the spike region in the immune response. J Virol **65:**7056-7060.

168. Young, N. S. 1995. B19 parvovirus. Baillieres Clin Haematol **8:**25-56.

169. Zadori, Z., J. Szelei, M.-C. Lacoste, P. Raymond, M. Allaire, I. R. Nabi, and P. Tijssen. 2001. A viral phospholipase A2 is required for parvovirus infectivity. Developmental Cell **1:**291-302.

CHAPTER 3

PHYLOGENETIC ANALYSIS OF CANINE PARVOVIRUS EMERGENCE, EVOLUTION AND DISPERSAL

From: Hoelzer, K., Shackelton, L. A., Parrish, C. R. and Holmes, E. C. Phylogenetic analysis reveals the emergence, evolution and dispersal of carnivore parvoviruses. J. Gen. Virol. 2008. 89: 2280-2289, with permission.

3.1. Summary

Canine parvovirus (CPV), first recognized as an emerging virus of dogs in 1978, resulted from a successful cross-species transmission. CPV emerged from the endemic feline panleukopenia virus (FPV), or from a closely related parvovirus of another host. Here we refine our current understanding of the evolution and population dynamics of FPV and CPV. By analyzing nearly full-length viral sequences we show that the majority of substitutions distinguishing CPV from FPV are located in the capsid protein gene, and that this gene is under positive selection in CPV, resulting in a significantly elevated rate of molecular evolution. This provides strong phylogenetic evidence for a prominent role of the viral capsid in host adaptation. In addition, an analysis of the population dynamics of more recent CPV reveals, on a global scale, a strongly spatially subdivided CPV population with little viral movement among countries and a relatively constant population size. Such limited viral migration contrasts with the global spread of the virus observed during the early phase of the CPV pandemic, but corresponds to the more endemic nature of current CPV infections.

3.2. Introduction

Efficient transmission in a new host species is a major obstacle to the successful cross-species transmission and emergence of viruses. Many emergent viruses are likely to be mal-adapted to the new host species and therefore cause only short transmission chains (18). Hence, the process of viral emergence will often require multi-stage adaptations to a new host, an idea supported by the observation that RNA viruses, which often exhibit evolutionary plasticity, are the most common agents of emerging disease (41). Indeed, there is growing evidence that although replication fidelity is

low in RNA viruses (and some ssDNA viruses), it has been fine-tuned in some cases (33, 34, 39) and might be environment-specific (30).

Canine parvovirus (CPV) represents one of the few examples where the process of cross-species viral transmission has been observed in 'real time'. CPV emerged during the 1970s as a host-range variant of an endemic parvovirus of cats, mink, or related hosts within the order Carnivora. This group of closely related viruses includes feline panleukopenia virus (FPV) and mink enteritis virus (MEV), both of which have long been endemic in their natural hosts. These viruses are globally distributed and infect a variety of domestic and feral hosts including cats, raccoons, foxes and mink (1, 25). Host range barriers restrict transmission of these viruses among dogs and related canids, in which viral replication is limited to the thymus and probably bone marrow (37).

CPV was first recognized in 1978 when it spread world-wide, but while this virus had probably circulated locally before that time, the exact time of emergence is uncertain. The first CPV identified is referred to as CPV-2 to distinguish it from the previously identified but distantly related parvovirus of dogs, minute virus of canines (canine minute virus). CPV-2 caused a severe pandemic with high mortality in dogs, but was unable to infect cats, raccoon or mink. A variant virus (referred to as CPV-2a) emerged in 1979 and replaced CPV-2 worldwide in approximately one year (27). CPV-2a differed antigenically from CPV-2 and infected both dogs and cats. Another antigenic variant (referred to as CPV-2b) with an equivalent host-range was defined by a single substitution (N426D) in the capsid protein gene and spread globally after 1984. Further mutations arose and spread widely, such as a mutation at residue 297 of the capsid protein. Currently, another mutation at the exposed capsid residue 426 (D426E), with a thus far uncharacterized impact on fitness, is spreading globally. This mutation was first described in Italy in 2000 (2), but has since been observed in other

European countries (7, 8, 20), South America (29), Asia (22) and most recently in the United States (11, 16).

Parvoviruses are small DNA viruses with a single-stranded linear genome of about 5kb. Their replication is dependent upon the host cell replication machinery, but estimates of rates of nucleotide substitution for both CPV and FPV are more comparable to those observed in RNA viruses than in other DNA viruses (35). The CPV genome contains two open reading frames (ORFs), one encoding the two non-structural proteins NS1 and NS2, and the other encoding the two structural proteins VP1 and VP2. The amino terminal (N-terminal) ends of NS1 and NS2 are identical in sequence, but the carboxy-terminal (C-terminal) end of the NS2 protein is derived by differential splicing of the mRNA, being translated in a different reading frame which overlaps that of NS1. VP1 and VP2 are splice variants and are identical in sequence except for a 143 amino acid (aa) long N-terminal stretch unique to VP1. VP2 is the major capsid protein, and accounts for approximately 90% of the viral capsid. The capsid is the major determinant of host range (14) and subject to antibody mediated selection (24). To date, most studies have focused on the evolution of the VP2 gene, with limited work on the non-structural genes.

Despite widespread vaccination of domestic carnivores, FPV and CPV have remained important pathogens of domestic and wild carnivores. In addition, both the patterns of CPV dispersal and the selection pressures acting across the viral genome are poorly understood. This non-enveloped virus causes only short infections with no persistent virus remaining after infection, but is highly resistant in the environment. The global spread of the virus could therefore occur through the movement of infected animals or mechanical vectors, although the spatial dynamics of CPV have not been examined in detail. Here we explore these ideas by performing the largest

evolutionary analysis of CPV undertaken to date, incorporating 143 sequences from 13 countries sampled over 29 years.

3.3. Materials and Methods

3.3.1. Isolate collection and sequencing.

Viral DNA was amplified by polymerase chain reaction (PCR) from clinical specimens collected at various points before or after the emergence of CPV. The DNA was either amplified directly from the clinical specimen, or after purification using the QIAamp viral DNA purification kit (Qiagen), according to the manufacturer's recommendations. PCR amplification was performed using Phusion hot start polymerase (New England Biolabs) and primers that flanked either the NS1 or VP1 genes, respectively, or that spanned the whole coding region. PCR products were either sequenced directly after purification with the QIAquick PCR purification kit, or after subcloning of the viral DNA into the bacterial vectors pJET 1.2 blunt (Fermentas) or pSMART GC HK (Lucigen Corporation) and each nucleotide was sequenced multiple times. Consensus sequences were constructed using the 'Seqman' program in the LaserGene 7.0 software package (DNASTAR). A total of 5 FPV and 6 CPV sequences that cover the complete viral coding region have been generated and were deposited in GenBank (accession numbers EU659111 to EU659121).

3.3.2. Sequence data.

The sequences of 27 near full-length genomes, 54 NS1 and 199 VP2 genes from CPV, FPV and closely related parvoviruses were either determined here or obtained from GenBank (Table 3.1). Several CPV-2c sequences (GenBank accession numbers AY3880577, AY742942, EF375479, EF375481) covered only smaller parts of the

coding region and where therefore excluded from the analysis. Similarly, not all published isolations of CPV-2c strains were apparently accompanied by GenBank submission of the viral sequences (see for example Kapil *et al.*, 2007, Hong *et al.*, 2007 or Decaro *et al.*, 2007). Of the sequences compiled, 23 near full-length sequences, 50 NS1 and 126 VP2 gene sequences contained information on the time of isolation, while the place of isolation was available for 52 NS1 and 178 VP2 gene sequences. The first 4 codons of the VP2 sequences were deleted from the VP2 gene alignment, as some of the sequences were incomplete, resulting in an alignment of 1721 nucleotides (nt). For the analysis of selection pressures acting on the VP1 unique region, a second alignment of 26 VP1 gene sequences was constructed with the intron within the VP1 unique N-terminus removed, resulting in an alignment of 2181 nt. The complete 2001 nt coding region of the NS1 gene was also analyzed, and for the analysis of selection pressures on the NS2 gene, the N-terminal 89 codons were combined with the 76 codon-long C-terminal region in the correct reading frame, resulting in an alignment of 495 nt. The near full-length sequences were trimmed to the complete NS1 and VP1 coding regions, with the introns included, resulting in a 4269 nt alignment. Sequences were aligned by hand in the Se-Al program (http://tree.bio.ed.ac.uk/software/seal/).

3.3.3. Inference of phylogenetic trees.

Maximum likelihood (ML) phylogenetic trees were estimated using PAUP* version 4.0 (36) with the best-fit model of nucleotide substitution determined using Modeltest (31). The HKY85 substitution model was the best-fit for the full-length data set and the key parameters (i.e. transition and transversion rates) were estimated from the data.

For the VP2 data sets, the GTR+I+Γ_4 substitution model was optimal. To assess support for individual nodes, bootstrap resampling values were estimated with 1000 neighbor-joining trees, again employing PAUP* (36).

3.3.4. Estimation of substitution rate and population dynamics.

Rates of nucleotide substitution per site, per year (subs/site/year) were estimated using the Bayesian Markov chain Monte Carlo (MCMC) method available in the BEAST package (http://beast.bio.ed.ac.uk/). Phylogenetic trees were estimated using the HKY85 substitution model, and either strict (constant) or relaxed (uncorrelated lognormal) molecular clocks were assumed. Four different demographic models were compared, assuming constant population size, exponential or logistic population growth or a Bayesian skyline model, which provides a piecewise depiction of changes in relative genetic diversity through time. Models were compared using Bayes Factors, with uncertainty in each parameter estimate depicted in values of the 95% highest probability density (HPD) interval. All analyses were run for sufficient time to ensure convergence as assessed using the TRACER program (http://tree.bio.ed.ac.uk/software/tracer/), with 10% of runs removed as burn-in. For those models supporting population growth, estimates of the population doubling time (λ) were obtained from the BEAST estimates of population growth (r) by using the formula $\lambda = \ln(2)/r$.

3.3.5. Estimates of site-specific selection pressures.

We used the Single Likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL) and Random Effects Likelihood (REL) algorithms available in the Datamonkey web interface of the HyPhy software package (17) to estimate the relative rates of synonymous (d_S) and nonsynonymous (d_N) substitutions per site (ratio d_N/d_S) in our

sample of CPV and FPV isolates. The HKY85 substitution model was employed in each case.

3.3.6. Analysis of migration patterns.

To determine the strength of phylogenetic clustering by country of virus isolation we employed a parsimony character mapping approach (3). Each CPV sequence was first assigned a character state reflecting its country of origin. Given the ML phylogeny for these sequences (as determined above), the minimum number of state changes needed to produce the observed distribution of country character states was estimated using parsimony (excluding ambiguous changes). To determine the expected number of changes under the null hypothesis of complete mixing among countries, the states of all isolates were randomized 1000 times. The difference between the mean number of observed and expected state changes indicates the level of geographic isolation, with statistical significance assessed by comparing the total number of observed state changes to the number expected under random mixing. All analyses were performed using PAUP* (Swofford, 2003).

Table 3.1: List of sequences analyzed here. The viral clade, sequence name and GenBank number are shown. Where applicable, information on the year and place of isolation are provided. *= sequence determined during study.

name	GenBank number	isolation date	isolation location
full-length viral sequences			
FPV			
FPLV-XJ-1.ch	EF988660	-	China
FPV-193.au_70	X55115	1970	Australia
MEV-Aba.ja_78	D00765	1978	Japan
FPV-b.us_67	M38246	1967	US
FPV-3.us_67	EU659111*	1967	US
FPV-4.us_64	EU659112*	1964	US
FPV-8a.us_89	EU659113*	1989	US
FPV-8b.us_89	EU659114*	1989	US
FPV_kai.us_06	EU659115*	2006	US
CPV-2			
CPV-pD8	NC_001539	-	-
CPV-bM38245	M38245	-	-
CPV-N.us_78	M19296	1978	US
CPV-5.us_79	EU659116*	1979	US
CPV-6.us_80	EU659117*	1980	US
CPV-2abc			
CPVB-2004.ch_104'	EF011664	2004	China
CPV-395.us_98	AY742936	1998	US
CPV-U6.ge_95	AY742935	1995	Germany
CPV-447.ge_95	AY742934	1995	Germany
CPV-339.nz_94	AY742933	1994	New Zealand
CPV-193.us_91	AY742932	1991	US
CPV-AJ564427.in	AJ564427	-	Northern India
CPV-Y1.ja_82	D26079	1982	Japan
CPV-cpv/nj01/06.ch.06	EU310373	2006	China
CPV-13.us_81	EU659118*	1981	US
CPV-410.us_100	EU659119*	2000	US
CPV-411a.us_98	EU659120*	1998	US
FPV			
FPLV-XJ-1.ch	EF988660	-	China
FPV-193.au_70	X55115	1970	Australia
MEV-Aba.ja_78	D00765	1978	Japan
FPV-b.us_67	M38246	1967	US
FPV-3.us_67	EU659111*	1967	US
FPV-4.us_64	EU659112*	1964	US
FPV-8a.us_89	EU659113*	1989	US
FPV-8b.us_89	EU659114*	1989	US
FPV_kai.us_06	EU659115*	2006	US

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
FPLV-PLI-IV.fr_68	AB000057	1968	France
FPLV-Obihiro_ja_74	AB000055	1974	Japan (Hokkaido)
FPLV-TU2_ja_75	AB000065	1975	Japan (Tokyo)
FPLV-TU4_ja_75	AB000067	1975	Japan (Tokyo)
FPLV-TU8_ja_76	AB000069	1976.5	Japan (Tokyo)
MEV-Abashiri_ja_78	D00765	1978	Japan
FPLV-TU10_ja_79	AB000062	1979	Japan (Tokyo)
FPLV-TU12_ja_79	AB000063	1979	Japan (Tokyo)
FPLV-483_ja_90	AB000048	1990	Japan (Hokkaido)
FPLV-Fukagawa_ja_93	AB000053	1993	Japan (Fukushima)
FPLV-941_ja_94	AB000049	1994	Japan (Tottori)
FPLV-AO1_ja_94	AB000051	1994	Japan (Hokkaido)
FPLV-Som1_ja_94	AB000058	1994	Japan (Nara)
FPLV-Som4_ja_94	AB000060	1995	Japan (Saitama)
CPV-2			
CPV-pD8	NC_001539	-	-
CPV-bM38245	M38245	-	-
CPV-N.us_78	M19296	1978	US
CPV-5.us_79	EU659116*	1979	US
CPV-6.us_80	EU659117*	1980	US
CPV-d.us_79	M38245	1979	US
CPV-2abc			
CPVB-2004.ch_104'	EF011664	2004	China
CPV-395.us_98	AY742936	1998	US
CPV-U6.ge_95	AY742935	1995	Germany
CPV-447.ge_95	AY742934	1995	Germany
CPV-339.nz_94	AY742933	1994	New Zealand
CPV-193.us_91	AY742932	1991	US
CPV-AJ564427.in	AJ564427	-	Northern India
CPV-Y1.ja_82	D26079	1982	Japan
CPV-cpv/nj01/06.ch.06	EU310373	2006	China
CPV-13.us_81	EU659118*	1981	US
CPV-410.us_100	EU659119*	2000	US
CPV-411a.us_98	EU659120*	1998	US
CPv-411b.us_98	EU659121*	1998	US
CPV-31.us_83	AY787929	1983	US
CPV-15.us_84	AY787926	1984	US
CPV-39.us_84	AY787930	1984	US

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
CPV-24.us_90	AY787928	1990	US
CPV-133.us_90	M74852	1990	US
CPV-402.us_98	AY742946	1998	US
CPV-407.us_98	AY742950	1999	US
CPV-412.us_98	AY742948	1998	US
CPV-431.us_103	AY742952	2003	US
CPV-435.us_103	AY742954	2003	US
CPV-436.us_104	AY742956	2004	US
NS1 sequences			
FPV			
FPLV-XJ-1.ch	EF988660	-	China
FPV-193.au_70	X55115	1970	Australia
MEV-Aba.ja_78	D00765	1978	Japan
FPV-b.us_67	M38246	1967	US
FPV-3.us_67	EU659111*	1967	US
FPV-4.us_64	EU659112*	1964	US
FPV-8a.us_89	EU659113*	1989	US
FPV-8b.us_89	EU659114*	1989	US
FPV_kai.us_06	EU659115*	2006	US
FPLV-PLI-IV.fr_68	AB000057	1968	France
FPLV-Obihiro_ja_74	AB000055	1974	Japan (Hokkaido)
FPLV-TU2_ja_75	AB000065	1975	Japan (Tokyo)
FPLV-TU4_ja_75	AB000067	1975	Japan (Tokyo)
FPLV-TU8_ja_76	AB000069	1976.5	Japan (Tokyo)
MEV-Abashiri_ja_78	D00765	1978	Japan
FPLV-TU10_ja_79	AB000062	1979	Japan (Tokyo)
FPLV-TU12_ja_79	AB000063	1979	Japan (Tokyo)
FPLV-483_ja_90	AB000048	1990	Japan (Hokkaido)
FPLV-Fukagawa_ja_93	AB000053	1993	Japan (Fukushima)
FPLV-941_ja_94	AB000049	1994	Japan (Tottori)
FPLV-AO1_ja_94	AB000051	1994	Japan (Hokkaido)
FPLV-Som1_ja_94	AB000058	1994	Japan (Nara)
FPLV-Som4_ja_94	AB000060	1995	Japan (Saitama)
CPV-2			
CPV-pD8	NC_001539	-	-
CPV-bM38245	M38245	-	-
CPV-N.us_78	M19296	1978	US
CPV-5.us_79	EU659116*	1979	US

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
CPV-6.us_80	EU659117*	1980	US
CPV-d.us_79	M38245	1979	US
CPV-2abc			
CPVB-2004.ch_104'	EF011664	2004	China
CPV-395.us_98	AY742936	1998	US
CPV-U6.ge_95	AY742935	1995	Germany
CPV-447.ge_95	AY742934	1995	Germany
CPV-339.nz_94	AY742933	1994	New Zealand
CPV-193.us_91	AY742932	1991	US
CPV-AJ564427.in	AJ564427	-	Northern India
CPV-Y1.ja_82	D26079	1982	Japan
CPV-cpv/nj01/06.ch.06	EU310373	2006	China
CPV-13.us_81	EU659118*	1981	US
CPV-410.us_100	EU659119*	2000	US
CPV-411a.us_98	EU659120*	1998	US
CPv-411b.us_98	EU659121*	1998	US
CPV-31.us_83	AY787929	1983	US
CPV-15.us_84	AY787926	1984	US
CPV-39.us_84	AY787930	1984	US
CPV-24.us_90	AY787928	1990	US
CPV-133.us_90	M74852	1990	US
CPV-402.us_98	AY742946	1998	US
CPV-407.us_98	AY742950	1999	US
CPV-412.us_98	AY742948	1998	US
CPV-431.us_103	AY742952	2003	US
CPV-435.us_103	AY742954	2003	US
FPV			
FPLV-XJ-1.ch	EF988660	-	China
FPV-193.au_70	X55115	1970	Australia
MEV-Aba.ja_78	D00765	1978	Japan
FPV-b.us_67	M38246	1967	US
FPV-3.us_67	EU659111*	1967	US
FPV-4.us_64	EU659112*	1964	US
FPV-8a.us_89	EU659113*	1989	US
FPV-8b.us_89	EU659114*	1989	US
FPV_kai.us_06	EU659115*	2006	US
FPLV-PLI-IV.fr_68	AB000057	1968	France
FPLV-Obihiro_ja_74	AB000055	1974	Japan (Hokkaido)
FPLV-TU2_ja_75	AB000065	1975	Japan (Tokyo)
FPLV-TU4_ja_75	AB000067	1975	Japan (Tokyo)
FPLV-TU8_ja_76	AB000069	1976.5	Japan (Tokyo)

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
MEV-Abashiri_ja_78	D00765	1978	Japan
FPLV-TU10_ja_79	AB000062	1979	Japan (Tokyo)
FPLV-TU12_ja_79	AB000063	1979	Japan (Tokyo)
FPLV-483_ja_90	AB000048	1990	Japan (Hokkaido)
FPLV-Fukagawa_ja_93	AB000053	1993	Japan (Fukushima)
FPLV-941_ja_94	AB000049	1994	Japan (Tottori)
FPLV-AO1_ja_94	AB000051	1994	Japan (Hokkaido)
FPLV-Som1_ja_94	AB000058	1994	Japan (Nara)
FPLV-Som4_ja_94	AB000060	1995	Japan (Saitama)
CPV-2			
CPV-pD8	NC_001539	-	-
CPV-bM38245	M38245	-	-
CPV-N.us_78	M19296	1978	US
CPV-5.us_79	EU659116*	1979	US
CPV-6.us_80	EU659117*	1980	US
CPV-d.us_79	M38245	1979	US
CPV-2abc			
CPVB-2004.ch_104'	EF011664	2004	China
CPV-395.us_98	AY742936	1998	US
CPV-U6.ge_95	AY742935	1995	Germany
CPV-447.ge_95	AY742934	1995	Germany
CPV-339.nz_94	AY742933	1994	New Zealand
CPV-193.us_91	AY742932	1991	US
CPV-AJ564427.in	AJ564427	-	Northern India
CPV-Y1.ja_82	D26079	1982	Japan
CPV-cpv/nj01/06.ch.06	EU310373	2006	China
CPV-13.us_81	EU659118*	1981	US
CPV-410.us_100	EU659119*	2000	US
CPV-411a.us_98	EU659120*	1998	US
CPv-411b.us_98	EU659121*	1998	US
CPV-31.us_83	AY787929	1983	US
CPV-15.us_84	AY787926	1984	US
CPV-39.us_84	AY787930	1984	US
CPV-24.us_90	AY787928	1990	US
CPV-133.us_90	M74852	1990	US
CPV-402.us_98	AY742946	1998	US
CPV-407.us_98	AY742950	1999	US
CPV-412.us_98	AY742948	1998	US
CPV-431.us_103	AY742952	2003	US

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
CPV-435.us_103	AY742954	2003	US
CPV-436.us_104	AY742956	2004	US
VP2 sequences			
FPV			
CPV-435.us_103	AY742954	2003	US
BFPV.fi_83	U22185	1983	Finland
FPLV/Lion/PT06.pt_106	EF418569	2006	Portugal
FPLV/Tiger/PT06.pt_106	EF418568	2006	Portugal
FPLV-483.ja_90	D88286	1990	Japan
FPLV-941.ja_94	AB000050	1994	Japan
FPLV-A04.ja_94	AB000052	1994	Japan
FPLV-ARG01.ar	EU018145	-	Argentina
FPLV-ARG02.ar	EU018144	-	Argentina
FPLV-ARG03.ar	EU018143	-	Argentina
FPLV-ARG04.ar	EU018142	-	Argentina
FPLVb-CU4	M24004	-	-
FPLV-fk.ja_93	AB000054	1993	Japan
FPLV-Gercules-Biocentr	AY665655	-	-
FPLV-JF-3	DQ099431	-	-
FPLV-obi.ja_74	AB000056	1974	Japan
FPLV-pliIV.fr_68	D88287	1968	France
FPLV-Som1.ja_94	AB000059	1994	Japan
FPLV-Som4.ja_95	AB000061	1995	Japan
FPLV-tu10.ja_79	D78584	1979	Japan
FPLV-tu12.ja_79	AB000064	1979	Japan
FPLV-tu2.ja_75	AB000066	1975	Japan
FPLV-tu4.ja_75	AB000068	1975	Japan
FPLV-tu8.ja_76.5	AB000070	1976.5	Japan
FPLV-V142	AB054225	-	-
FPLV-V208	AB054226	-	-
FPLV-V211	AB054227	-	-
FPLV-XJ-1.ch	EF988660	-	China
FPLV-ZF-5	DQ099430	-	-
FPV-193.au_70	X55115	1970	Australia
FPV-23.us_90	U22187	1990	USA
FPV-377.ge_93	U22188	1993	Germany
FPV-a.uk_62	M24002	1962	United Kingdom
FPV-T1	AF015223	-	-
MEV-a.us_73	M23999	1973	USA
MEV-Aba.ja_78	D00765	1978	Japan

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
MEV-b.us_75	M24001	1975	USA
MEV-Beregovoj-Biocentr	AY665656	-	-
MEV-Cherepanovo98	AF201477	-	-
MEV-d.uk_65	U22190	1965	United Kingdom
MEV-EF428258	EF428258	-	-
MEV-EU137663	EU137663	-	-
MEV-Rodniki-Biocentr	AY665657	-	-
RPV.us_79	M24005	1979	USA
FPV-3.us_67	EU659111*	1967	USA
FPV-4.us_64	EU659112*	1964	USA
FPV-8a.us_89	EU659113*	1989	USA
FPV-8b.us_89	EU659114*	1989	USA
FPV_kai.us_106	EU659115*	2006	USA
CPV-2			
CPV-RD87.fi_87	U22193	1987	Finland
CPV-dCornell320	M23255	-	-
CPV-128.us_78	U22186	1978	USA
CPV-RD80.fi_80	U22192	1980	Finland
CPV-pD8	NC_001539	-	-
CPV-N.us_78	M19296	1978	USA
CPV-AJ698134.in	AJ698134	-	India
CPV-M10989	M10989	-	-
CPV-5.us_79	EU659116*	1979	USA
CPV-6.us_80	EU659117*	1980	USA
CPV-2abc			
BR133.br_94	DQ340427	1994	Brazil
BR135.br_80	DQ340405	1980	Brazil
BR136.br_93	DQ340423	1993	Brazil
BR137.br_80	DQ340406	1980	Brazil
BR137.br_93	DQ340424	1993	Brazil
BR145.br_80	DQ340407	1980	Brazil
BR154.br_80	DQ340408	1980	Brazil
BR17.br_90	DQ340412	1990	Brazil
BR18.br_90	DQ340413	1990	Brazil
BR183.br_85	DQ340409	1985	Brazil
BR209.br_94	DQ340428	1994	Brazil
BR22.br_93	DQ340422	1993	Brazil
BR227.br_93	DQ340425	1993	Brazil
BR237.br_94	DQ340429	1994	Brazil
BR31.br_90	DQ340414	1990	Brazil

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
BR315.br_86	DQ340410	1986	Brazil
BR43.br_91	DQ340415	1990	Brazil
BR46.br_95	DQ340430	1994	Brazil
BR47.br_91	DQ340416	1991	Brazil
BR491.br_92	DQ340418	1992	Brazil
BR52.br_91	DQ340417	1991	Brazil
BR56.br_95	DQ340431	1995	Brazil
BR570.br_92	DQ340419	1992	Brazil
BR315.br_86	DQ340410	1986	Brazil
BR593.br_92	DQ340420	1992	Brazil
BR597.br_92	DQ340421	1992	Brazil
BR6.br_80	DQ340404	1980	Brazil
BR62.br_95	DQ340432	1995	Brazil
BR7168.br_100	DQ340433	2000	Brazil
BR8.br_90	DQ340411	1990	Brazil
BR8155.br_100	DQ340434	2000	Brazil
BR84.br_94	DQ340426	1994	Brazil
CPV-133.us_90	M74852	1990	USA
CPV-15.us_84	M24003	1984	USA
CPV-193.us_91	AY742932	1991	USA
CPV-209.vi_97	AB054219	1997	Vietnam
CPV-24.us_90	U22896	1990	USA
CPV2a-AJ564427.in	AJ564427	-	India
CPV2a-APD1.ch	EU213074	-	China
CPV2a-BD2.ch	EU213079	-	China
CPV2a-BD3.ch	EU213080	-	China
CPV2a-BD4.ch	EU213081	-	China
CPV2a-BJ004/O7.ch	EF666059	-	China, Beijing
CPV2a-BJ005/O7.ch	EF666060	-	China, Beijing
CPV2a-BJ008/O7.ch	EF666061	-	China, Beijing
CPV2a-BJ010/O7.ch	EF666062	-	China, Beijing
CPV2a-BJ015/O7.ch	EF666063	-	China, Beijing
CPV2a-BJ017/O6.ch	EF666065	-	China, Beijing
CPV2a-BJ017/O7.ch	EF666064	-	China, Beijing
CPV2a-BJ018/O7.ch	EF666066	-	China, Beijing
CPV2a-BJ020/O7.ch	EF666067	-	China, Beijing
CPV2a-BJ027/O7.ch	EF666068	-	China, Beijing
CPV2a-BJ034/O7.ch	EU145953	-	China
CPV2a-BJ042/O7.ch	EF666069	-	China, Beijing
CPV2a-BJ050/O7.ch	EU145955	-	China
CPV2a-BJ064/O7.ch	EU145956	-	China
CPV2a-BJ068/O7.ch	EU145957	-	China

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
CPV2a-BJ069/O7.ch	EU145958	-	China
CPV2a-BJ077/O7.ch	EU145959	-	China
CPV2a-BJ082/O7.ch	EU145960	-	China
CPV2a-BJ085/O7.ch	EU145961	-	China
CPV2a-HT.ch	EU213077	-	China
CPV2a-HZ0761.ch	EU213073	-	China
CPV2a-JB1.ch	EU213076	-	China
CPV2a-K001.sk	EU009200	-	South Korea
CPV2a-K014.sk	EU009201	-	South Korea
CPV2a-K015.sk	EU009202	-	South Korea
CPV2a-K022.sk	EU009203	-	South Korea
CPV2a-K026.sk	EU009204	-	South Korea
CPV2a-KT1.ch	EU213082	-	China
CPV2a-KT2.ch	EU213083	-	China
CPV2a-NJ-04.ch	EU095252	-	China, Beijing
CPV2a-Sho-nan	AB128923	-	-
CPV2a-ZD2.ch	EU213084	-	China
CPV2a-ZD3	EU213085	-	-
CPV2b-BD1.ch	EU213075	-	China
CPV2b-BJ044/O7.ch	EU145954	-	China
CPV2b-HN-3.ch	Q177497	-	China
CPV2b-K029.sk	EU009205	-	South Korea
CPV2b-K031.sk	EU009206	-	South Korea
CPV2b-ZD1.ch	EU213078	-	China
CPV2c-Pome.sk	EF599098	-	South Korea
CPV2-DQ903936.ch	DQ903936	-	China
CPV-31.us_83	M24000	1983	USA
CPV-314.ja_93	D78585	1993	Japan
CPV-339.nz_94	AY742933	1994	New Zealand
CPV-39.us_84	M74849	1984	USA
CPV-395.us_98	AY742936	1998	USA
CPV-431.us_103	AY742951	2003	USA
CPV-435.us_103	AY742953	2003	USA
CPV-436.us_104	AY742955	2004	USA
CPV-447.ge_95	AY742934	1995	Germany
CPV-46.po_94	Z46651	1994	Poland
CPV-584.it_94	AF306446	1994	Italy
CPV-616.it_95	AF306449	1995	Italy
CPV-618.it_95	AF306447	1995	Italy
CPV-632.it_96	AF306445	1996	Italy
CPV-637.it_95	AF306450	1995	Italy
CPV-677.it_98	AF306448	1998	Italy

Table 3.1 (continued)

name	GenBank number	isolation date	isolation location
CPV-695.eu_99	AF401519	1999	Europe
CPV-699.eu_100	AF393506	2000	Europe
CPVB-2004.ch_104	EF011664	2004	China
CPV-CPVKK0501	EF189717	-	-
CPV-DH326.sk	EF599097	-	South Korea
CPV-DH426.sk	EF599096	-	South Korea
CPV-Ik.ja_97	AB115504	1997	Japan
CPV-RPPV.ch_104	DQ354068	2004	China
CPV-SHZ	EU170352	-	China
CPV-T10.ta_95	U72696	1995	Taiwan
CPV-T25.ta_95	U72697	1995	Taiwan
CPV-695.eu_99	AF401519	1999	Europe
CPV-699.eu_100	AF393506	2000	Europe
CPV-T37.ta_95	U72698	1995	Taiwan
CPV-T4.ta_94	U72695	1994	Taiwan
CPV-Ta9.ta_98	AB054213	1998	Taiwan
CPV-TWN1.ta_106	EF592511	2006	Taiwan
CPV-U6.ge_95	AY742935	1995	Germany
CPV-V120.vi_97	AB054215	1997	Vietnam
CPV-v123.vi_97	AB054218	1997	Vietnam
CPV-v129.vi_97	AB054216	1997	Vietnam
CPV-V139.vi_97	AB054222	1997	Vietnam
CPV-V154.vi_97	AB054217	1997	Vietnam
CPV-V203.vi_97	AB054224	1997	Vietnam
CPV-V204	AB054221	-	-
CPV-v217.vi_97	AB054220	1997	Vietnam
CPV-w42.it_95	AF306444	1995	Italy
CPV-Y1.ja_82	D26079	1982	Japan
HCM-18.ho.vi_102	AB120722	2002	Vietnam
HCM-23.ho.vi_102	AB120723	2002	Vietnam
HCM-6.ho.vi_102	AB120720	2002	Vietnam
HCM-8.ho.vi_102	AB120721	2002	Vietnam
HNI-1-18.ha.vi_102	AB120728	2002	Vietnam
HNI-2-13.ha.vi_102	AB120724	2002	Vietnam
HNI-3-11.ha.vi_102	AB120726	2002	Vietnam
HNI-3-4.ha.vi_102	AB120725	2002	Vietnam
HNI-4-1.ha.vi_102	AB120727	2002	Vietnam
LCPV-t1.ta_95	AB054214	1995	Taiwan
LCPV-V140.vi_97	AB054223	1997	Vietnam
Taichung.ta_103	AY869724	2003	Taiwan
CPV-13.us_81	EU659118*	1981	USA
CPV-410.us_100	EU659119*	2000	USA

Table 3.1 (continued)

CPV-411a.us_98	EU659120*	1998	USA
CPV-411b.us_98	EU659121*	1998	USA
CPV-cpv/nj01/06.ch_106	EU310373	2006	China

3.4. Results

3.4.1. Phylogenetic analysis of the full-length genome.

We analyzed 27 near full-length genomes of carnivore parvoviruses, several of which were determined during the course of this study. The FPV sequences cover a time period of 42 years (1964 to 2006) and were collected in various geographical regions. The oldest CPV isolate was collected in 1978, the year in which CPV was first described, and the CPV sequences cover a total of 29 years. The topology of the ML tree of these data (Figure 3.1) was similar to that of trees inferred from the VP2 region, with the longest internal branch separating the FPV and CPV clades. 16 nucleotide substitutions separated the FPV and CPV clades, while 7 substitutions separated the CPV-2 from the CPV-2a clade. The majority of these substitutions (11 between FPV and CPV and 5 between CPV-2 and CPV-2a) were located in the capsid protein region (data not shown).

3.4.2. Substitution rates and population dynamics.

Bayesian MCMC estimates of rates of nucleotide substitution were inferred for the full-length genome, and separately for the NS1 and VP2 genes, either analyzing all carnivore parvoviruses simultaneously or subdividing the sequences into FPV and CPV (i.e. all CPV-2, CPV-2a and derived viruses). For the VP2 sequence, two additional data sets were analyzed, one consisting of the CPV-2a sequences, and the other comprising all FPV and most CPV sequences but excluding the original CPV-2 sequences.

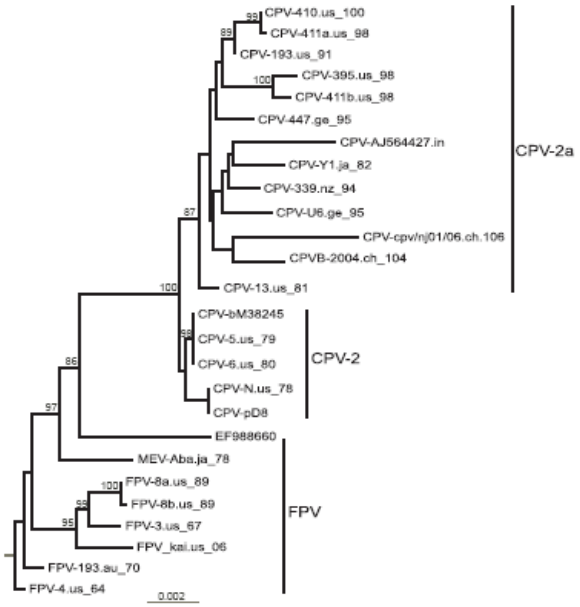


Figure 3.1: Phylogenetic tree of 27 full-length carnivore parvovirus sequences, rooted with the oldest FPV sequence (FPV-4.us.64). Horizontal branch lengths are drawn to scale (nucleotide substitutions per site), and bootstrap support (>85%) is indicated above the respective branches. The three recognized clades, FPV and related viruses, CPV-2 and CPV-2a and derived viruses, are indicated.

Very similar substitution rates were obtained for individual data sets under all molecular clock and demographic models, reflecting the robustness of this analysis (all results available from the authors on request). However, estimates of the substitution rate from the capsid protein gene were strikingly different between FPV and CPV (Table 3.2). The mean substitution rate for FPV was 8.2×10^{-5} subs/site/year (95% HPD = $3.6 \times 10^{-5} – 1.3 \times 10^{-4}$ subs/site/year), while that of CPV was significantly higher at 2.2×10^{-4} subs/site/year (95% HPD = 1.7×10^{-4} - 2.7×10^{-4} subs/site/year). Estimates of substitution rate for both FPV and CPV based on the NS1 gene were similar (6.6×10^{-5} subs/site/year and 1.0×10^{-4} subs/site/year, respectively), with overlapping HPD values.

Table 3.2: Nucleotide substitution rates, population dynamics and times to common ancestry for CPV and FPV. - = not applicable; const. = constant population size; log. = logistic population growth; BSP = Bayesian skyline model; Mean substitution rate = subs/site/year; HPD = 95% highest probability density interval; Time to Most Recent Common Ancestor (TMRCA) = age since most recent sequence included in analysis (i.e. 2006).

	VP2					NS1			full-length
	FPV	**CPV**	**CPV-2a**	**no CPV-2**	**all**	**FPV**	**CPV**	**all**	
demographic model	const.	log.	log.	const.	BSP	const.	const.	const	const.
sequence length (bp)	1721	1721	1721	1721	1721	2007	2007	2007	4269
number of sequences	30	96	90	120	126	22	27	49	23
date range of sequences	1962-2006	1978-2004	1980-2004	1962-2006	1962-2006	1963-2006	1978-2004	1963 - 2006	1964-2006
Mean substitution rate	8.2×10^{-5}	2.2×10^{-4}	2.1×10^{-4}	1.6×10^{-4}	1.4×10^{-4}	6.4×10^{-5}	1.0×10^{-4}	6.2×10^{-5}	1.2×10^{-4}
HPD substitution rate |	3.6×10^{-5} - 1.3×10^{-4}	1.7×10^{-4} - 2.7×10^{-4}	1.6×10^{-4} - 2.7×10^{-4}	1.1×10^{-4} - 2.0×10^{-4}	1.1×10^{-4} - 1.7×10^{-4}	1.7×10^{-5} - 1.2×10^{-4}	2.2×10^{-5} - 2.0×10^{-4}	3.4×10^{-5} - 9.3×10^{-5}	7.1×10^{-5} - 1.7×10^{-5}
Time to Most Recent Common Ancestor (years)	98	32	28	71	105	128	85	127	71
HPD age	53-169	29-37	26-32	50-100	100-117	50-260	30-192	65-215	47-103
mean growth rate (1/years)	-	0.2	0.4	-	-	-	-	-	-
HPD growth rate (years)	-	0.1	0.1	-	-	-	-	-	-
	-	0.3	0.4	-	-	-	-	-	-
Mean epidemic doubling time (λ)	-	3.7	1.8	-	-	-	-	-	-

Using the same approach we were able to estimate the Time to the Most Recent Common Ancestor (TMRCA) of both FPV and CPV. The estimated TMRCA for FPV was 98 years from the most recent analyzed isolate collected in 2006 (95% HPD = 53 – 168 years) while that for CPV was 32 years (95% HPD = 29 – 37 years) which closely corresponds with the first CPV case recognized in 1978.

The mean TMRCA estimate for CPV-2a (28 years) is lower than that for all CPVs, again matching the first observed CPV-2a collected in 1979. The TMRCA estimates based on the NS1 gene place the emergence of FPV, and particularly CPV, considerably earlier than those based on the capsid protein gene (128 years for FPV and 85 years for CPV). However, the HPD values are very wide (and overlap with those inferred from VP2), revealing the inherent statistical uncertainty in this estimate (although more recent dates are observed under other demographic models tested – results not shown).

The population dynamics of all CPVs and the CPV-2a subset were best described by logistic population growth in which a rapid growth phase is followed by a phase of increasingly slower growth, while a constant population size provided a better description of both the FPV data set and the data set containing all sequences except the CPV-2 strain. The complete data set, comprising all VP2 sequences, was best described by the more complex Bayesian skyline model, reflecting the inclusion of viruses with differing epidemiological dynamics. Notably, the mean epidemic doubling time for CPV-2a and derived viruses (1.8 years) is approximately half that for all CPVs combined (3.7 years), revealing the initially rapid spread of the CPV-2a variant (although, again, the 95% HPD values overlap). However, these estimates of growth rate should be treated with caution due to the potentially confounding effect of population subdivision. Both FPV and CPV NS1 sequences were best described by models of constant population size.

3.4. 3. Selection pressures in CPV and FPV.

An analysis of selection pressures, manifest as d_N/d_S ratios, across both viruses and genes revealed the sporadic action of positive selection, with mean d_N/d_S values of 0.15 for NS1, 0.53 for NS2, 0.15 for the VP1 unique region and 0.12 for VP2 (Table 2.3). Notably, the short NS2 protein contained four positively selected sites, all within the 76 aa long C-terminal region where the encoded sequence overlaps that of NS1 in a different reading frame.

Hence, the 'positively selected' sites in this region are difficult to interpret. While the possibility of positive selection cannot be ruled out, the results might represent false-positives caused by synonymous mutations in the alternate reading frame. In contrast, only a single NS1 site, at codon 597, appeared to be under positive selection. This amino acid is located in the C-terminal region of NS1, in the region where NS2 is encoded in the different reading frame, and overlaps with residues 106 and 107 of NS2, again suggesting that this result might be false-positive. No positively selected sites were detected in the VP1 unique region.

As expected given the capsid's known role in host range control, four surface exposed capsid residues, at VP2 positions 101 (d_N/d_S= infinity; d_N= 2.24), 300 (d_N/d_S = infinity; d_N= 3.0), 324 (d_N/d_S = infinity; d_N= 4.43) and 426 (d_N/d_S = infinity; d_N= 4.56), were under positive selection across all sequences (Figure 3.2). Among the FPV sequences, only VP2 codon 562 in the VP2 gene appears to be under positive selection (d_N/d_S = infinity; d_N = 13.05). Among the CPV sequences, VP2 residues 87 (d_N/d_S = infinity; d_N= 5.03) and 426 (d_N/d_S = infinity; d_N= 4.55) were under positive selection.

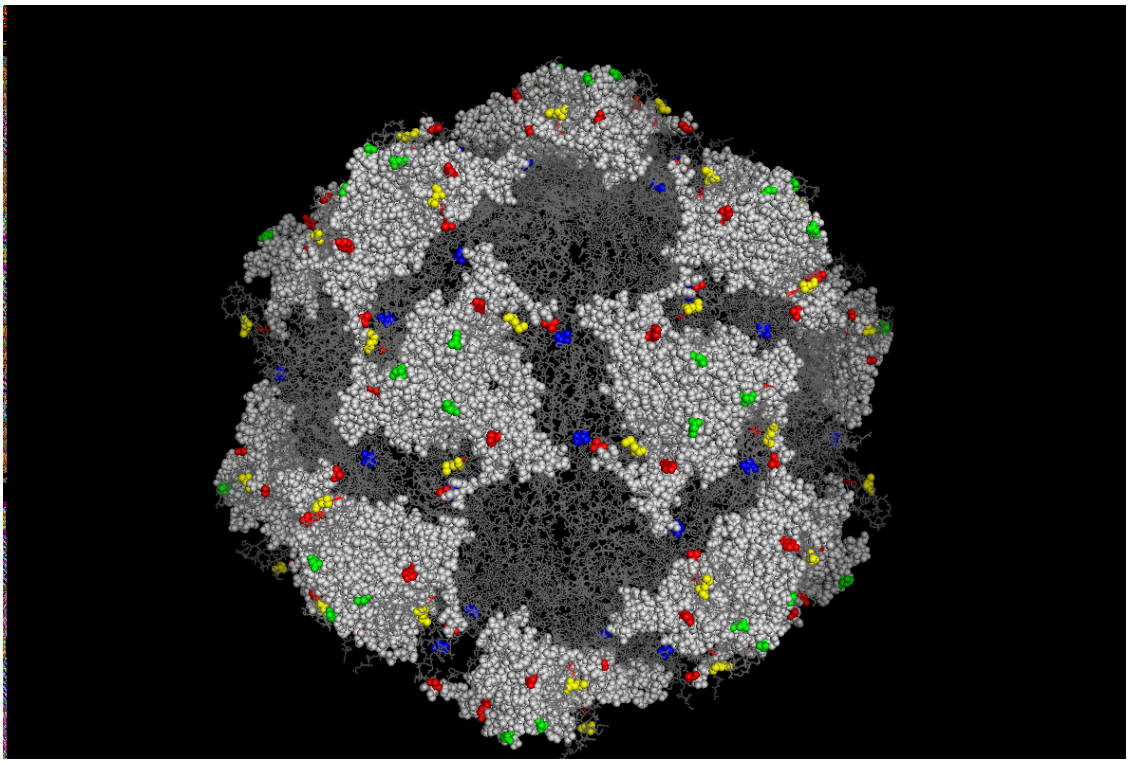


Figure 3.2: The positively selected capsid protein residues are mapped onto the crystal structure of the CPV half capsid. Red= residues positively selected among FPV and CPV; blue= residue positively selected only among FPV sequences; yellow= residues positively selected only among CPV sequences; green= residue 426.

3.4.4. Migration patterns of CPV.

The differences between the observed and expected numbers of migration events involving CPV isolates sampled from different countries are shown in Table 3.4, with negative values indicative of population subdivision and positive values characterizing directional migration between the specified geographic regions. Overall, we detected a strong signal of population subdivision, with only a few instances of migration between specific localities. Indeed, only 12 of the 182 possible directional migration pathways showed signs of CPV migration, and none had strong signals, suggesting only sporadic geographical movement of these viruses.

Table 3.3: Analysis of selection pressures in each genomic region. P-values or, in the case of REL, Bayes factors, are reported. P-values < 0.1 and Bayes factors > 100 are shown in bold. - =none detected;∞ = "infinity"; *=residues were mapped to the NS2 gene of minute virus of mice using a previously published alignment (40) and the functions listed correspond to those known in that virus. ; †= only detected in the full-length data set; ‡= location and function of the NS1 and NS2 mutations were inferred from the known functions of homologous regions in minute virus of mice; TfR = Transferrin receptor; #= sequence contained too many sequences for analysis.

gene		mean d_N/d_S	mutation	d_N	d_S	d_N/d_S	SLAC	FEL	IFEL	REL	location‡	potential function‡
NS1*		0.15	L597H	10.47	0	∞	0.23	**0.089**	0.114	77	C-terminus	P38 promoter transactivation
NS2*		0.53	T94R	2.42	0.91	2.66	0.445	0.25	0.195	**331**	C- terminus	CRM1 inter-action
			N107R	2.42	2.12	1.14	0.602	0.76	0.26	**136**	C- terminus	CRM1 interaction
			D151N	2.42	0.92	2.63	0.615	0.296	**0.09**	**301**	C-terminus	14-3-3 family inter-action
			E160K	2.42	0.87	2.78	0.52	0.38	1	**325**	C-terminus	unclear
VP1		0.15	-	-	-	-	-	-	-	-	-	-
VP2	**all**	0.12	I101T†	2.24	0	∞	0.35	0.10	0.34	**5617†**	3fold spike	possibly TfR binding§
			A300G	3.0	0	∞	**0.089**	0.189	0.361	6†	3fold spike	TfR binding, antigenic change§
			Y324I†	4.43	0	∞	0.506	0.54	0.286	**504 †**	3fold spike	likely TfR binding
			N426D/E	4.56	0	∞	**0.043**	0.255	0.124	6†	surface	antigenic escape
	FPV	0.11	V562L	13.05	0	∞	0.408	0.33	**0.097**	4704	capsid surface	unclear
	CPV	0.11	M87L	5.03	0	∞	0.465	0.219	**0.069**	#	shoulder	TfR binding; possibly antigenic§
			N426D/E	4.55	0	∞	**0.045**	0.251	0.104	#	surface	antigenic escape

Conversely, 95 of the 182 possible pathways were negative, indicative of population subdivision, sometimes strongly so. Interestingly, China represents a far weaker source of viruses for other regions than expected given the number of available sequences from this country. Similarly, Brazil is also a weaker source population than expected by chance (especially for movements to China and Taiwan), but appears to be a source for movement to Europe (Italy and Poland) and potentially the United States. Overall, the strongest evidence of viral movement was from the United States to Vietnam (observed-expected = 1.54) and from Vietnam to China (observed-expected = 1.36), but the signal was weak.

When we mapped the positively selected sites onto the phylogenetic tree depicted in Figure 3.3 (data not shown), the VP2 N426D mutation appeared to be co-circulating in several countries, such as South Korea, the United States and Taiwan, but was absent in all but four sequences collected in China (CPV2b-BD1.ch, CPV2b-BK044/07.ch,CPV2b-HN-3.ch and CPV2b-ZD1.ch), of which the first three clustered with sequences from Vietnam. Interestingly, the N426D mutation was not monophyletic, indicating that it has arisen independently in these countries. Only a single Brazilian sequence (BR185.br_85) had the N426D mutation, and this sequence had signatures of migration, since it clustered with the European and U.S. isolates. instead of the other Brazilian isolates. Geographic structure was even more striking in the case of the Y324I mutation. This was only present in sequences from China and South Korea and again was not monophyletic. Conversely, the VP2 mutations at codons 87, 101 and 300 were present in all but five CPV-2a sequences.

Table 3.4: Migration analysis of CPV, based on VP2 gene. Observed minus expected migration pathways are shown below, with directional movements from country indicated in row to country shown in column. Likely migration events (positive values) are indicated by shading. Country codes: A, India; B, Brazil; C, China; F, Finland; G, Germany; I, Italy; J, Japan; N, New Zealand; O, outgroup (1964 FPV isolate, UK); P, Poland; S, South Korea; T, Taiwan; U, USA; V, Vietnam.

To/from	A	B	C	F	G	I	J	N	O	P	S	T	U	V	Σ
A	NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	−0.15	NA	−2.71	−0.13	−0.31	**1.10**	−0.46	−0.14	−0.15	**0.86**	−1.25	−1.04	**0.07**	−2.08	−6.37
C	−0.26	−3.79	NA	−0.27	−0.48	−1.68	**0.24**	−0.28	−0.25	−0.24	−2.29	−1.85	−3.49	−3.67	−18.3
F	0	0	0	NA	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	NA	0	0	0	0	0	0	0	0	0	0
I	−0.01	−0.08	−0.06	0	−0.01	NA	−0.02	−0.01	−0.01	0	−0.04	−0.04	−0.06	−0.06	−0.39
J	0	0	−0.01	0	0	0	NA	0	0	0	0	0	0	0	−0.01
N	0	0	0	0	0	0	0	NA	0	0	0	0	0	0	0
O	0	**1**	0	0	0	0	0	0	NA	0	0	0	0	0	1
P	0	0	0	0	0	0	0	0	0	NA	0	0	0	0	0
S	−0.01	−0.13	−0.12	−0.01	−0.02	−0.05	0.02	−0.01	−0.01	−0.01	NA	−0.07	−0.1	−0.12	−0.67
T	0	−0.08	**0.92**	−0.01	−0.01	−0.04	**0.98**	−0.01	−0.01	−0.04	−0.05	NA	−0.08	−0.05	**1.58**
U	−0.03	−0.52	−0.48	−0.04	**0.95**	**0.80**	−0.09	−0.03	−0.03	−0.02	−0.29	−0.24	NA	**1.54**	**1.53**
V	−0.04	−0.65	**1.36**	−0.03	−0.09	−0.27	−0.1	−0.05	−0.04	−0.04	−0.33	**0.73**	−0.53	NA	−0.08
Σ	−0.49	−4.25	−1.11	−0.49	**0.03**	−0.14	**0.54**	−0.51	−0.48	**0.54**	−4.25	−2.5	−4.20	−4.45	

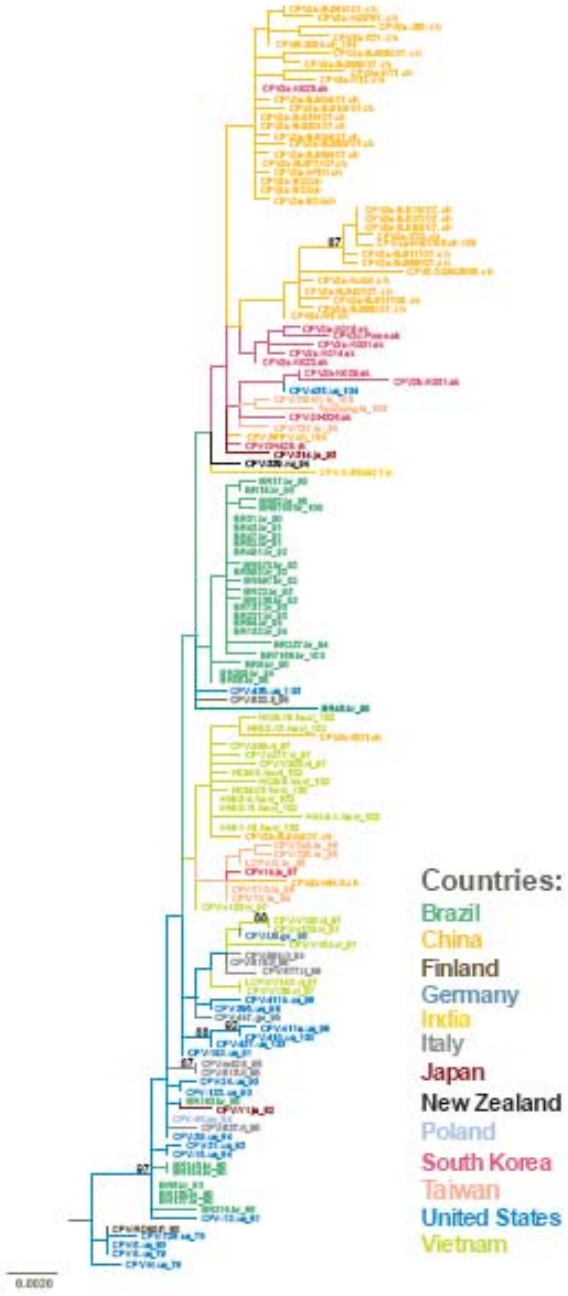


Figure 3.3: Phylogenetic tree of 152 CPV VP2 sequences used for the analysis of migration patterns. Horizontal branch lengths are drawn to scale (nucleotide substitutions per site), and bootstrap support (>85%) is indicated above their respective branches. The country of isolation is indicated by color-coding, as indicated in the key.

Three of the Vietnamese sequences collected in 1997 from feral cats (CPV-V139.vi_97, CPV-V203.vi_97 and LCPV_V140.vi_97) and one sequence from South Korea (CPV-2c-Pome) harbored a G300D mutation (15) and one Chinese sequence isolated from a red panda in 2004 (CPV-RPPV.ch_104) possessed a G300V mutation (32). Finally, only one sequence in the data set (HNI-4-1.ha.vi_102) contained the VP2 D426E mutation. As such, no inference can be made from this dataset regarding the global movement of this mutation.

3.5. Discussion

This study provides novel insights into the population dynamics of the emerging CPV virus, contrasting the evolutionary pressures acting on CPV and FPV, and among the genes of CPV. The emergence of CPV was characterized by strong selection for specific mutations within VP2 but not NS1. Optimal receptor binding and antigenic escape were the likely driving forces, as indicated by the location and function of the positively selected sites in the capsid protein (located at VP2 codons 101, 300, 324, 426).

This study is the first to systematically explore CPV dynamics on a global scale. Notably, we found the CPV population to be spatially structured, with very limited gene flow between clusters. This is in contrast with a previous study which reported strong temporal, but mild or absent spatial clustering, among CPV isolates from Brazil, the U.S. and Europe (Pereira *et al.*, 2000), but is consistent with other

reports of geographic clustering of isolates from Vietnam, Japan/Taiwan and the United States (9). Three of the mutations at positively selected sites (at VP2 residues 87,101 and 300) arose soon after CPV emerged. The mutations at residues 87 and 101 have arisen once and all but five of the CPV-2a viruses harbor the A300G mutation. The remaining CPV-2a viruses were isolated from feral feline hosts and show alternative substitutions of residue 300, suggesting a role of this residue in the feline adaptation of CPV. We have previously shown that an Asp at residue 300 is selected during passage of CPV-2 in feline cells in culture, and also that this prevents the infection of dog cells (Parrish and Carmichael, 1986; Parker and Parrish, 1997). The two positively selected mutations at VP2 residues 324 and 426 arose later, and apparently independently, in different geographic locations. Thus the early period of CPV emergence was characterized by global viral spread and global strain replacement, while later mutations likely arose through parallel emergence of the same mutations in separate geographic regions. Indeed, by far the strongest phylogeographic signal in these data is that of population subdivision, with relatively little movement of viruses between geographic areas. In addition, the majority of migrations that do occur are between countries in close geographic proximity, such as Taiwan or Vietnam and China, or between countries with close economic connections such as the United States and Germany or Italy, indicating that CPV is likely spread among countries through the movement of mechanical vectors or, in some cases, infected animals. The movement of viruses from the United States to Vietnam appears surprising, but all of these sequences were isolated in 1997, only 3 years after the embargo against Vietnam was lifted, indicating a potential role of the developing trade relations between the countries.

While recombination can potentially impact phylogenetic relationships, we found no clear signal of recombination among the FPV or CPV isolates studied here

(unpublished data; available from the authors on request). Indeed, the presence and role of recombination among FPV or CPV sequences is uncertain. While phylogenetic evidence of recombination has been detected for related parvoviruses of mink (Aleutian mink disease virus), swine (porcine parvovirus) and rodents (several rodent parvoviruses including LuIII) (Shackelton et al, 2007), there is no clear evidence for the process among the FPV- and CPV-related parvoviruses studied to date. It is therefore unlikely that the analyses reported here were strongly impacted by recombination.

Our molecular clock estimates of emergence times for CPV-2 and CPV-2a are notable in that they are very close to the dates when the viruses were first detected in nature. In particular, we estimate that CPV-2 emerged around 1974 (i.e. 32 years before the most recent analyzed sequence collected in 2006), only four years before it spread globally in 1978. This is compatible with serological evidence that the virus circulated in Europe a few years before the CPV pandemic (38). Similarly, our estimate for the emergence date of CPV-2a is approximately four years later than CPV-2, which corresponds closely with the first detection of CPV-2a in nature in 1979. Notably, the pattern of logistic population growth for CPV-2 and CPV-2a indicates the initially rapid spread of the virus among animals. This corresponds to the expected population dynamics of a newly emerging pathogen, and has also been reported by previous studies (28, 35). In marked contrast, FPV evolution is characterized by constant population sizes, as expected for an endemic pathogen.

The change in evolutionary dynamics associated with the process of emergence is also apparent in the significantly higher rate of substitution in the capsid protein gene of CPV compared to the FPV sequences, with mean rates of 1.2 and 0.4 substitutions per genome per year, respectively. Although these rates are comparable to those inferred previously (12, 28, 35), this is the first time a statistically significant

difference between FPV and CPV has been revealed. In theory, the elevated rate of evolutionary change in CPV might be due to either lower replication fidelity in dogs or increased positive selection after the transmission to the new host. Since these viruses are all replicated by host cell DNA polymerases, significantly different fidelities appear unlikely. Moreover, substitution rates inferred from the NS1 gene were comparable between FPV and CPV sequences, and similar to that inferred for FPV based on the capsid gene, indicating that the elevated substitution rate observed among the CPV sequences was most likely due to strong positive selection, probably driven by host adaptation and immune escape.

There is clear evidence for strong positive selection on the capsid protein gene, and differences among the host species again appear to be of importance in this regard. Four VP2 residues - 101, 300, 324 and 426 - are subject to positive selection among all carnivore parvoviruses. Residues 101 and 300 are located on a surface-exposed region of the viral capsid (Figure 3.2), interact with the transferrin receptor (which is the receptor mediating viral entry), and partially determine the re-acquired feline host range of CPV-2a. The substitution at residue 324 has not been previously characterized, but is adjacent to residue 323, a known host-range mutation, which together with residue 93 controls canine TfR binding and the canine host range (13). The 324 mutation appears to be circulating only in China and South Korea (data not shown), and has appeared a number of times independently. Although clear information on time of sampling is available for only a few of the sequences containing this mutation, most appear to have been collected in 2006 or 2007, indicating that this mutation probably arose recently. Mutations at aa 426 change the antigenic profile of the virus (23, 26). This codon has undergone two mutations since CPV emerged, first from asparagine to aspartate, and more recently to glutamate, indicating complicated selection dynamics acting on this residue. The exact

phenotypic consequences of the mutations are not clear, and the N426D mutation has been present in some CPV populations for decades without replacing the wild-type (Parrish *et al*., 1991).

Mapping the codon 426 variants onto the phylogenetic tree in Figure 3.2 revealed both the Asn and Asp amino acids were co-circulating in the United States, Taiwan, Vietnam and Europe, while the Asp mutation was at very low frequency in China and Brazil (data not shown; available from the authors on request). Interestingly, the Asp codon-containing sequences from those countries showed signatures of migration and clustered with sequences from other countries that carried the respective mutation. The D426E mutation has recently been isolated from several countries around the world, but only a single VP2 sequence of known origin that harbored this mutation was available, preventing more detailed analysis. One additional sequence containing the D426E mutation (GenBank accession number AF401519) was excluded from the analysis since the place of isolation could not be clearly determined. Among the CPV sequences, residues 87 and 426 are under strong positive selection. Residue 87 is also surface exposed and, although its specific functions are unknown, appears to be associated with the re-gained feline host range of CPV-2a, together with residue 300.

Less positive selection was observed in FPV, and only VP2 codon 562 appeared to be under positive selection, although relatively strongly. Adaptive evolution at this site has not been described previously. However, two residues in close spatial proximity, 564 and 568, partially determine the *in vivo* host range of FPV (Truyen et. al, 1994a).

Finally, our study is notable in that we detected putatively positively selected sites in the NS1 (codon 597, overlapping codon 107 in NS2) and NS2 (codons 94, 107,151 and 160) genes, although these are potentially false positives due to the

overlapping reading frames. The role of NS2 in the CPV life-cycle is unclear and knock-out mutants appear to behave similarly to wild-type both in tissue culture and in dog infections (40). However, in the related rodent parvoviruses, NS2 is required for efficient translation and capsid assembly in cells from natural host species but not in cells from other hosts (4-6, 10, 19, 21). A possible role for NS2 in host adaptation therefore merits further study.

3.6. Acknowledgements

Leland E. Carmichael and Edward Dubovi provided clinical samples, and Wendy S. Weichert and Nell Bond provided excellent technical support. Christian Nelson provided valuable help with the three-dimensional representation of the positively selected capsid residues. This work was supported by NIH grants GM080533-01 to E.C.H and AI028385 to CRP.

REFERENCES

1. Barker, I. K., R. C. Povey, and D. R. Voigt. 1983. Response of mink, skunk, red fox and raccoon to inoculation with mink virus enteritis, feline panleukopenia and canine parvovirus and prevalence of antibody in wild carnivores in Ontario. Can J Comp Med **47**: 188-197.

2. Buonavoglia, C., V. Martella, A. Pratelli, M. Tempesta, A. Cavalli, D. Buonavoglia, G. Bozzo, G. Elia, N. Decaro, and L. Carmichael. 2001. Evidence for evolution of canine parvovirus type 2 in Italy. J Gen Virol **82:**3021-5.

3. Carrington, C. V., J. E. Foster, O. G. Pybus, S. N. Bennett, and E. C. Holmes. 2005. Invasion and maintenance of dengue virus type 2 and type 4 in the Americas. J Virol **79:**14680-7.

4. Choi, E. Y., A. E. Newman, L. Burger, and D. Pintel. 2005. Replication of minute virus of mice DNA is critically dependent on accumulated levels of NS2. J Virol **79:**12375-81.

5. Cotmore, S. F., A. M. D'Abramo, Jr., L. F. Carbonell, J. Bratton, and P. Tattersall. 1997. The NS2 polypeptide of parvovirus MVM is required for capsid assembly in murine cells. Virology **231:**267-280.

6. D'Abramo, A. M., Jr., A. A. Ali, F. Wang, S. F. Cotmore, and P. Tattersall. 2005. Host range mutants of minute virus of mice with a single VP2 amino acid change require additional silent mutations that regulate NS2 accumulation. Virology **340:**143-54.

7. Decaro, N., C. Desario, D. D. Addie, V. Martella, M. J. Vieira, G. Elia, A. Zicola, C. Davis, G. Thompson, E. Thiry, U. Truyen, and C. Buonavoglia. 2007. The study molecular epidemiology of canine parvovirus, Europe. Emerg Infect Dis **13:**1222-4.

8. Decaro, N., V. Martella, C. Desario, A. L. Bellacicco, M. Camero, L. Manna, D. d'Aloja, and C. Buonavoglia. 2006. First detection of canine parvovirus type 2c in pups with haemorrhagic enteritis in Spain. J Vet Med B Infect Dis Vet Public Health **53:**468-72.

9. Doki, M., K. Fujita, R. Miura, M. Yoneda, Y. Ishikawa, A. Taneno, and C. Kai. 2006. Sequence analysis of VP2 gene of canine parvovirus isolated from domestic dogs in Japan in 1999 and 2000. Comparative immunology, microbiology and infectious diseases **29:**199-206.

10. Eichwald, V., L. Daeffler, M. Klein, J. Rommelaere, and N. Salome. 2002. The NS2 proteins of parvovirus minute virus of mice are required for efficient nuclear egress of progeny virions in mouse cells. J Virol **76:**10307-19.

11. Hong, C., N. Decaro, C. Desario, P. Tanner, M. C. Pardo, S. Sanchez, C. Buonavoglia, and J. T. Saliki. 2007. Occurrence of canine parvovirus type 2c in the United States. J Vet Diagn Invest **19:**535-9.

12. Horiuchi, M., Y. Yamaguchi, T. Gojobori, M. Mochizuki, H. Nagasawa, Y. Toyoda, N. Ishiguro, and M. Shinagawa. 1998. Differences in the evolutionary pattern of feline panleukopenia virus and canine parvovirus. Virology **249:**440-52.

13. Hueffer, K., L. Govindasamy, M. Agbandje-McKenna, and C. R. Parrish. 2003. Combinations of two capsid regions controlling canine host range determine canine transferrin receptor binding by canine and feline parvoviruses. J Virol **77:**10099-10105.

14. Hueffer, K., J. S. Parker, W. S. Weichert, R. E. Geisel, J. Y. Sgro, and C. R. Parrish. 2003. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. J. Virol. **77:**1718-1726.

15. Ikeda, Y., M. Mochizuki, R. Naito, K. Nakamura, T. Miyazawa, T. Mikami, and E. Takahashi. 2000. Predominance of canine parvovirus (CPV) in unvaccinated cat populations and emergence of new antigenic types of CPVs in cats. Virology **278:**13-9.

16. Kapil, S., E. Cooper, C. Lamm, B. Murray, G. Rezabek, L. Johnston, 3rd, G. Campbell, and B. Johnson. 2007. Canine parvovirus types 2c and 2b circulating in North American dogs in 2006 and 2007. J Clin Microbiol **45:**4044-7.

17. Kosakovsky Pond, S. L., and Frost, S.D.W. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments Bioinformatics 2005 **21:**2531-2533.

18. Kuiken, T., E. C. Holmes, J. McCauley, G. F. Rimmelzwaan, C. S. Williams, and B. T. Grenfell. 2006. Host species barriers to influenza virus infections. Science **312:**394-7.

19. Li, X., and Rhode, S.L. (1991). 1991. Nonstructural protein NS2 of parvovirus H-1 is required for efficient viral protein synthesis and virus production in rat cells in vivo and in vitro. Virology **184:**117-130

20. Martella, V., A. Cavalli, A. Pratelli, G. Bozzo, M. Camero, D. Buonavoglia, D. Narcisi, M. Tempesta, and C. Buonavoglia. 2004. A canine parvovirus mutant is spreading in Italy. J Clin Microbiol **42:**1333-6.

21. Naeger, L. K., N. Salome, and D. J. Pintel. 1993. NS2 is required for efficient translation of viral mRNA in minute virus of mice-infected murine cells. J Virol **67:**1034-1043.

22. Nakamura, M., Y. Tohya, T. Miyazawa, M. Mochizuki, H. T. Phung, N. H. Nguyen, L. M. Huynh, L. T. Nguyen, P. N. Nguyen, P. V. Nguyen, N. P. Nguyen, and H. Akashi. 2004. A novel antigenic variant of canine parvovirus from a Vietnamese dog. Arch Virol **149:**2261-9.

23. Nakamura, M., Y. Tohya, T. Miyazawa, M. Mochizuki, H. T. T. Phung, N. H. Nguyen, L. M. T. Huynh, L. T. Nguyen, P. N. Nguyen, P. V. Nguyen, N. P. T. Nguyen, and H. Akashi. 2004. A novel antigenic variant of Canine parvovirus from a Vietnamese dog. Arch Virol **149:**2261-2269.

24. Nelson, C. D., L. M. Palermo, S. L. Hafenstein, and C. R. Parrish. 2007. Different mechanisms of antibody-mediated neutralization of parvoviruses revealed using the Fab fragments of monoclonal antibodies. Virology **361:**283-93.

25. Parrish, C. R. 1990. Emergence, natural history, and variation of canine, mink, and feline parvoviruses. Adv Virus Res **38:**403-450.

26. Parrish, C. R., C. F. Aquadro, M. L. Strassheim, J. F. Evermann, J. Y. Sgro, and H. O. Mohammed. 1991. Rapid antigenic-type replacement and DNA sequence evolution of canine parvovirus. J. Virol. **65:**6544-6552.

27. Parrish, C. R., P. Have, W. J. Foreyt, J. F. Evermann, M. Senda, and L. E. Carmichael. 1988. The global spread and replacement of canine parvovirus strains. J Gen Virol 69 **(Pt 5):**1111-6.

28. Pereira, C. A., E. S. Leal, and E. L. Durigon. 2006. Selective regimen shift and demographic growth increase associated with the emergence of high-fitness variants of canine parvovirus. Infect Genet Evol **7**: 399-409.

29. Perez, R., L. Francia, V. Romero, L. Maya, I. Lopez, and M. Hernandez. 2007. First detection of canine parvovirus type 2c in South America. Vet Microbiol **124:**147-52.

30. Pita, J. S., J. R. de Miranda, W. L. Schneider, and M. J. Roossinck. 2007. Environment determines fidelity for an RNA virus replicase. J Virol **81:**9072-7.

31. Posada, D., Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14:**817-818.

32. Qin, Q., I. K. Loeffler, M. Li, K. Tian, and F. Wei. 2006. Sequence analysis of a canine parvovirus isolated from a red panda (Ailurus fulgens) in China.Virus Genes **34**: 299-302.

33. Sanjuan, R., J. M. Cuevas, V. Furio, E. C. Holmes, and A. Moya. 2007. Selection for robustness in mutagenized RNA viruses. PLoS Genet **3:**e93.

34. Sanjuan, R., J. M. Cuevas, A. Moya, and S. F. Elena. 2005. Epistasis and the adaptability of an RNA virus. Genetics **170:**1001-8.

35. Shackelton, L. A., C. R. Parrish, U. Truyen, and E. C. Holmes. 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc Natl Acad Sci U S A **102:**379-384.

36. Swofford, D. L. 1993. PAUP - phylogentic analysis with parsimony. Center for Biodiversity, Illinois Natural History Survey, Champaign, Ill.

37. Truyen, U., and C. R. Parrish. 1992. Canine and feline host ranges of canine parvovirus and feline panleukopenia virus: distinct host cell tropisms of each virus in vitro and in vivo. J Virol **66:**5399-5408.

38. Truyen, U., G. Platzer, C. R. Parrish, T. Hanichen, W. Hermanns, and O. R. Kaaden. 1994. Detection of canine parvovirus DNA in paraffin-embedded tissues by polymerase chain reaction. Zentralbl Veterinarmed [B] **41:**148-52.

39. Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature **439:**344-8.

40. Wang, D., W. Yuan, I. Davis, and C. R. Parrish. 1998. Nonstructural protein-2 and the replication of canine parvovirus. Virology **240:**273-81.

41. Woolhouse, M. E., L. H. Taylor, and D. T. Haydon. 2001. Population biology of multihost pathogens. Science **292:**1109-12.

CHAPTER 4

WITHIN-HOST DYNAMICS OF EMERGING AND ENDEMIC PARVOVIRUSES OF DOGS AND CATS

From: Hoelzer, K., Shackelton, L.A., Holmes, E.C. and Parrish, C.R. Within-host genetic diversity of endemic and emerging parvoviruses of dogs and cats. J. Virol. 2008. 82: 11096-11105, with permission.

4.1. Summary

Viral emergence can result from the adaptation of endemic pathogens to new or altered host environments, a process that is strongly influenced by the underlying sequence diversity. To determine the extent and structure of intra-host genetic diversity in a recently emerged single-stranded DNA virus, we analyzed viral population structures during natural infections of animals with canine parvovirus (CPV) or its ancestor feline panleukopenia virus (FPV). We compared infections that occurred shortly after CPV emerged with more recent infections, and examined the population structure of CPV after experimental cross-species transmission to cats. Infections with CPV and FPV showed limited genetic diversity regardless of the analyzed host tissue or year of isolation. Co-infections with genetically distinct viral strains were detected in some cases, and rearranged genomes were seen in both FPV and CPV. The sporadic presence of some sequences with multiple mutations suggested the occurrence of either particularly error-prone viral replication or co-infection by distantly related strains. Finally, some potentially organ-specific host effects were seen during experimental cross-species transmission, with many of the mutations located in the non-structural protein NS2. These included residues with evidence of positive selection at the population level, which is compatible with a role of this protein in host adaptation.

4.2. Introduction

Emerging viruses that gain new host ranges are major threats to human and animal health. However, the evolutionary mechanisms that allow viruses to infect new hosts

and establish self-sustaining transmission chains are complex and far from understood. In most cases studied to date, several mutations together determine the new host range, but where, when, and how these mutations arise has largely remained enigmatic. Many cases of viral emergence probably represent multi-stage adaptations to an altered host environment, where the initial emerging virus is poorly adapted to the recipient host and causes only inefficient transmission. Further mutations are then required for complete adaptation to the new host. The success of the new pathogen is therefore likely to be highly influenced by the genetic diversity of the viral population, with more variable viral populations being particularly prone to generate successful emerging infections.

The extent of genetic diversity present within viral populations is largely determined by a balance between erroneous replication (measured as mutation rate) and purifying selection (which is commonly expressed in the substitution rate, which is the number of fixed mutations/ base pair/year), with particularly high error rates in RNA viruses that replicate using RNA-dependent RNA polymerases (15), or, in the case of retroviruses, reverse trancriptases. Although large double-stranded DNA viruses possess mutation rates far lower than those seen in RNA viruses, some small single-stranded (ss) DNA viruses appear to both mutate (14, 39) and have substitution rates closer to RNA viruses than to dsDNA viruses (reviewed in (16)). The parvoviruses represent a particularly well-studied case, with the mean substitution rate for the canine parvovirus (CPV) capsid protein gene at $\sim 1 \times 10^{-4}$ substitutions/nucleotide/year (38, 42). However, the extent and structure of within-host variability in ssDNA viruses, including co-infection as a potential source of genetic diversity, has rarely been analyzed.

Various phenomena such as recombination, non-random codon usage and host effects such as hypermutation mediated by “Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B” (APOBEC3G) might also affect standing genetic diversity, although their role in the evolution of ssDNA viruses is largely unclear. In addition, some rapidly evolving viruses such as HIV undergo tissue-specific variation (24, 47). Tissue-specific effects may also occur during infections with parvoviruses, but this has not been analyzed previously.

CPV emerged in the 1970s as a new virus of dogs, being derived from either feline panleukopenia virus (FPV) or a very closely related virus of another host (Figure 4.1). After circulating undetected in dogs in Europe or Eurasia for a few years, the virus spread globally in 1978 (35). This virus is referred to as CPV type-2 (CPV-2) to distinguish it from the distantly related minute virus of canines. In 1979 a variant strain of CPV-2 (named CPV-type2a (CPV-2a)) emerged and replaced CPV-2 globally by the end of 1980 (37). The emerging strain differed antigenically and readily infected cats, while CPV-2 did not replicate in felines (44). In 1984 an antigenic variant of CPV-2a with apparently identical host range, which is referred to as CPV-2b, arose. CPV-2a and CPV-2b are currently co-circulating in the global dog population, but their relative frequencies appear to vary between geographic regions and are potentially also subject to temporal fluctuation (reviewed in (20)). Natural infections of cats and wild felines with CPV have been reported (4, 18, 31), but FPV has remained the more prevalent parvovirus causing disease in cats.

Both FPV and CPV cause acute infections of young animals, resulting in the development of long lived protective immunity (7). Infection occurs through the oro-nasal route and initial replication takes place in the lymphatic tissues of the oro-pharynx (Figure 4.4A). After a primary viremia, the virus begins to replicate in the thymus, spleen and bone marrow 2 to 3 days after infection. The virus then becomes

disseminated throughout the animal, and replication is seen in the gut- associated lymphoid tissues, particularly the Peyer's patches. The virus subsequently replicates in the rapidly dividing epithelial cells within the crypts of the small intestinal villi, and is shed at high levels in the feces 4 to 7 days post infection (6, 7, 29, 33). The viruses appear to be cleared less than 2 weeks post infection with no residual replicating virus remaining (reviewed in (33).

The CPV genome (depicted in Figures 4.2 (below the consensus sequences) and 4.3 (top panel)), contains two open reading frames (ORFs), one of which encodes the two nonstructural proteins NS1 and NS2, with the other encoding the two structural proteins VP1 and VP2 (40). The amino terminal ends of NS1 and NS2 overlap and are identical in sequence, but the carboxy terminal domain of NS2 is derived by differential splicing, fusing it to a different reading frame which overlaps that of NS1. The sequence of VP2 is completely contained within that of VP1, which has an additional 143 amino acid (aa) long N-terminal sequence. VP2 makes up 90% of the viral capsid, which is the major determinant of canine and feline host ranges (8, 22, 32). Since the feline host range of CPV-2a was not due to back-mutations to the FPV sequences, some of the newly arisen mutations that characterized CPV-2a were likely of compensatory nature(43, 45).

Here we analyze the DNA sequence diversity and population structures of CPV and FPV within samples collected during natural and experimental infections. We define both the sequence variation generated *de novo* and that which likely results from co-infection. FPV samples from various tissues or feces were collected over a 43 year period, with the oldest sample originating from 1963. The earliest CPV sample was collected in 1979, the second year of the pandemic in dogs. We inoculated cats with CPV collected from naturally infected dogs, and analyzed the developing

population structures of the virus in different feline organs, comparing the sequence diversity in these populations to that present in the inoculum.

4.3. Materials and Methods

4.3.1. Virus samples.

The viruses examined are listed in Table 4.1. All samples were collected as original clinical specimens (feces or tissues) from infected animals, and were stored at -80°C. Diluted samples were either added directly to polymerase chain reaction (PCR) reactions without any prior preparation, or were purified using the QIAamp DNA Mini kit (Qiagen, Valencia, CA) according to the manufacturer's recommendation before amplification.

4.3.2. PCR, cloning and sequencing.

PCR analysis was performed using the high fidelity Phusion Hot Start DNA polymerase (New England Biolabs, Beverly, MA) and primer pairs that either flanked the region between nts. 118 and 4837 in the viral genome (compared to CPV-2 reference sequence, GenBank accession number: M38245.1), or that amplified that region as two overlapping fragments (nts. 118 to 2419 and nts. 2033 to 4837 (reverse primer 1) or nts. 2033 and 4680 (reverse primer 2), all primer sequences available from the authors upon request).

Table 4. 1. Characterization of FPV and CPV isolates from virus samples analyzed from natural infections. The host species, year of isolation and tissue type are indicated, and the number of clones analyzed for each gene fragment is shown. GenBank accession numbers for respective consensus sequences are provided and re-arranged genomes are described where applicable.

Virus	GenBank accession	Host	Year of collection	Sample tissue	Number of clones		Re-arranged genomes (VP2)	
					NS1	VP2	Deleted region (nt.)	Inserted region (nt)
FPV-4.us.64	EU659112	cat	1964	spleen	7	5	-	-
FPV-3.us.67	EU659111	cat	1967	kidney	39	38	1981-3010 2525-2889 2710-3090 2689-3336 2843-3311	1549-1980 - - - -
CPV-5.us.79	EU659116	dog	1979	spleen	4	11	-	-
CPV-6.us.80	EU659117	dog	1980	feces	7	8	-	-
CPV-13.us.81	EU659118	dog	1981	feces	37	21	2383-3587	3702-3990
FPV-8.us.89a* FPV-8.us.89b*	EU659113 EU659114	mountain lion	1989	tissue	7	7	-	-
CPV-411.us.98a* CPV-411.us.98b*	EU659120 EU659121	dog	1998	feces	8	17	-	-
CPV-410.us.100	EU659119	dog	2000	feces	41	27	-	-
FPV- kai.us.106	EU659115	cat	2006	small intestine	40		3153-3420	2764-3154

PCR products were purified by agrose gel purification using the Qiaquick gel extraction kit (Qiagen) and were either prepared for cloning into the pSMART GC HK vector according to the manufacturers recommendation (Lucigen Corporation, Middelton,WI), or were used without further preparation for cloning into the pJET1.2 vector (Fermentas, Glen Burnie, MD) using the manufactuerer's blunt-end cloning protocol. Inserts were sequenced using primer sets that covered the entire insert in both directions (primer sequences available from the authors upon request).

4.3.3. Quantitative PCR analysis.

A TaqMan real-time PCR assay for quantification of parvovirus genomes was established, targeting a conserved 60 nt region located between nts. 1039 and 1114 of the CPV genome (*qPCR forward primer*: AAATGAAACCAGAAACCGTTGAA; *qPCR reverse primer*: TCCCGCGCTTTGTTTCC). A TaqMan minor groove binding probe (MGB, ABI) was designed to bind to the amplified region (*TaqMan MGB probe*: ACAGTGACGACAGCAC). The linearized sequence of a CPV-2b genome, cloned into a plasmid, was used as an external standard for quantification of copy numbers.

4.3.4. Sequence analysis.

Isolate consensus sequences have been published previously (19), and are available as GenBank accession numbers EU659111 to EU659121. Mutations detected in the course of this study were compared to all publically available FPV and CPV sequences covering the respective genome region (alignments were identical to those published previously (19)).

Some nucleotide sequences contained deletions or insertions, which were further analyzed using BLAST (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). These

sequences were aligned separately to the consensus sequence. Gene rearrangements were only confirmed if the exact boundaries of the deletion or duplication event could be inferred from the sequence information. To confirm that those re-arranged sequences were present in the original samples before PCR amplification, we chose one of the gene rearrangements, detected in the FPV-kai.us.106 sample, and designed PCR primer pairs for which one primer each binds within the duplicated gene region (sequence available from the authors upon request).

4.3.5. Experimental cat infections.

All animal infections were approved by the Cornell University Animal Care and Use Committee. The original virus-containing specimens (not tissue culture passaged) were resuspended in phosphate-buffered saline (pH 7.2) and sterile-filtered through a 0.22 micron filter prior to inoculation. The viral titers in the CPV isolates CPV-13.us.81 and CPV-410.us.100 were determined as $TCID_{50}$ in Nordon Laboratories feline kidney (NLFK) cells as described previously (44) and the viral genome copy numbers were determined by quantitative PCR. Groups of two ~5 week old parvovirus seronegative kittens (Liberty Research, Waverly, NY) were inoculated through the oro-nasal route with 5×10^5 $TCID_{50}$ of CPV-13.us.81 or with 3×10^5 $TCID_{50}$ of CPV-410.us.100. Those contained 5.5×10^{14} or 3.4×10^{14} viral copy numbers respectively as determined by qPCR. The kittens were monitored daily for clinical symptoms and viral shedding was assessed by PCR analysis of rectal swabs. On days 0 and 6 post inoculation, serum samples were collected to confirm the absence of anti-viral antibodies in hemagglutination inhibition (HI) tests as described elsewhere (34). On day 6 (for the CPV13 virus challenged kittens) or 8 (for the CPV410 virus challenged kittens) after inoculation the kittens were euthanized, then thymus, bone marrow and

fecal samples were collected and examined for viral DNA by PCR. PCR products (where these were generated) were cloned and sequenced as described above.

4.4. Results

Overall we obtained approximately 900,000 nucleotides of sequence, including 190 clones covering the NS1/2 gene region and 172 covering the VP1/2 gene region (Table 4. 1).

4.4.1. Analysis of natural FPV infections.

We analyzed viral genetic diversity in clinical FPV samples collected over a 43 year period (Figure 4.1), representing a broad variety of host tissues (Table 4.1). Most mutations that distinguished isolate consensus sequences were synonymous (Figure 4.2A). No mutations were detected in the NS1 carboxy-terminus where NS2 is encoded in a different reading frame and only 1 nonsynonymous mutation was observed among the 12 mutations in the VP2 protein gene (8.3%).

The population structures within individual FPV samples were very homogenous and comparable across samples (Figure 4.3A). Eleven of 90 sequences (12.2%) that covered the capsid protein gene harbored a mutation, with 3 of those being nonsynonymous. The VP2 Ile 101 to Thr change was present in 3 of 7 sequences from the same sample. Six of 93 (6.5%) NS1-covering sequences harbored one or more mutations each, of which 64% were nonsynonymous. One mutation at

nucleotide position 2166 resulted in nonsynonymous mutations in both the NS1 and C-terminal NS2 reading frames (Figure 4.3A).

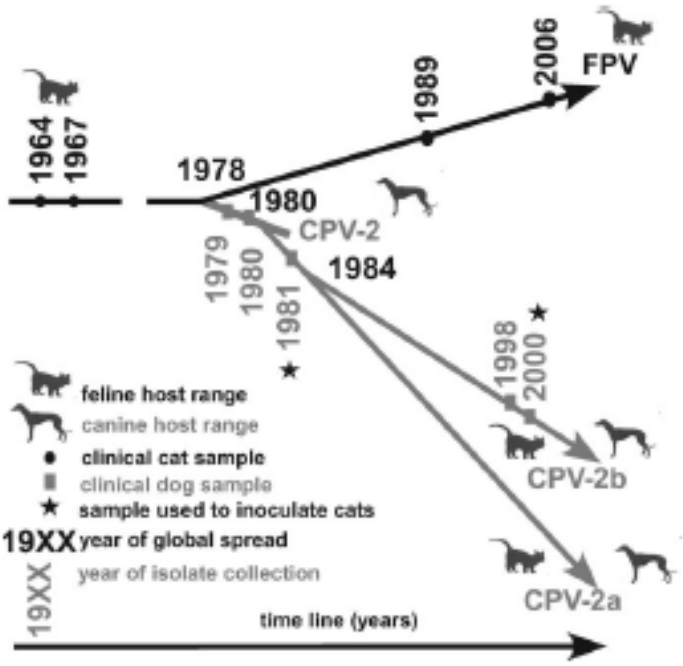


Figure 4.1. Overview of natural virus samples analyzed. The host from which the sample was collected, the virus type, and year of collection are indicated. The likely natural host range and year of global spread are indicated. Viruses used for inoculation of kittens are marked.

In most FPV samples, one mutation was detected every 4-6 × 10^{-5} nucleotides and the majority of mutations were generally located in the NS1 coding region (Table 4.3). About half of the mutations detected in viral genomes from infected feline hosts can also be found in FPV GenBank sequences, such as 10 of 20 mutations in regions with sufficient sequence data available in GenBank.

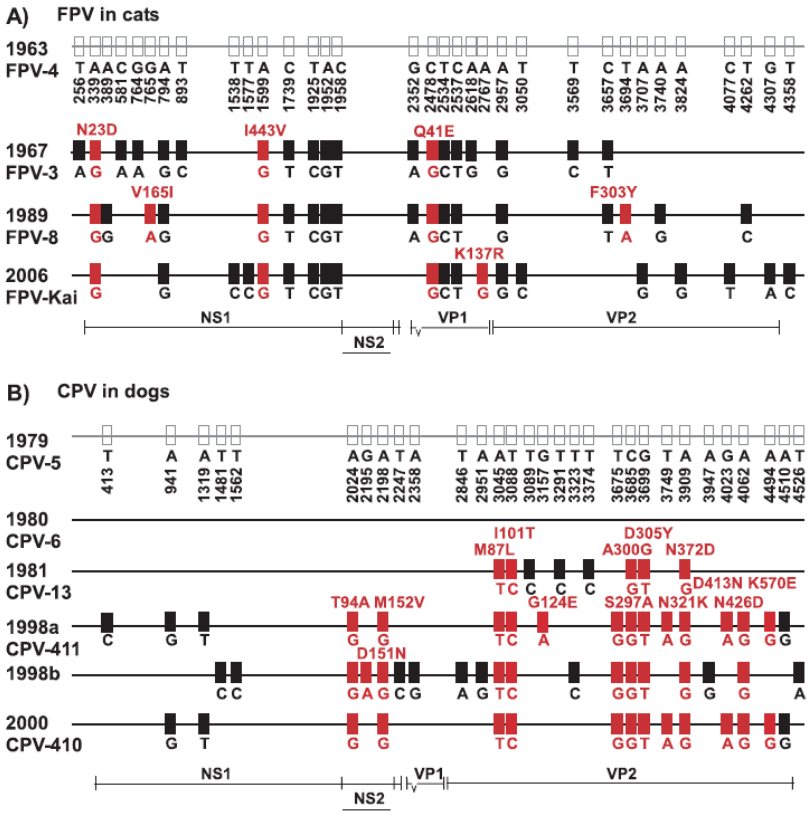


Figure 4.2. Mutations in the consensus sequences. The isolate consensus sequences from clinical samples of FPV (A) or CPV (B) infections were compared, using the oldest FPV or CPV samples as reference sequence in each case. The nucleotide position is indicated on the top, and the character state of individual nucleotides is indicated below each sequence. Changes in amino acid sequence are characterized above the sequence where appropriate. Synonymous changes are indicated by black boxes, nonsynonymous changes by red boxes, while the location in the viral genome can be inferred by referring to the genome map shown below the sequences.

Of the 12 nonsynonymous mutations, 4 (33%) were also present in GenBank FPV sequences (Table 4.2) indicating a likely relatively high associated fitness of these mutations.

Moreover, in the cases of FPV-8.us.89 and CPV-411.us.98 the possibility of co-infections is underlined by the presence of these respective mutations in the global FPV population.

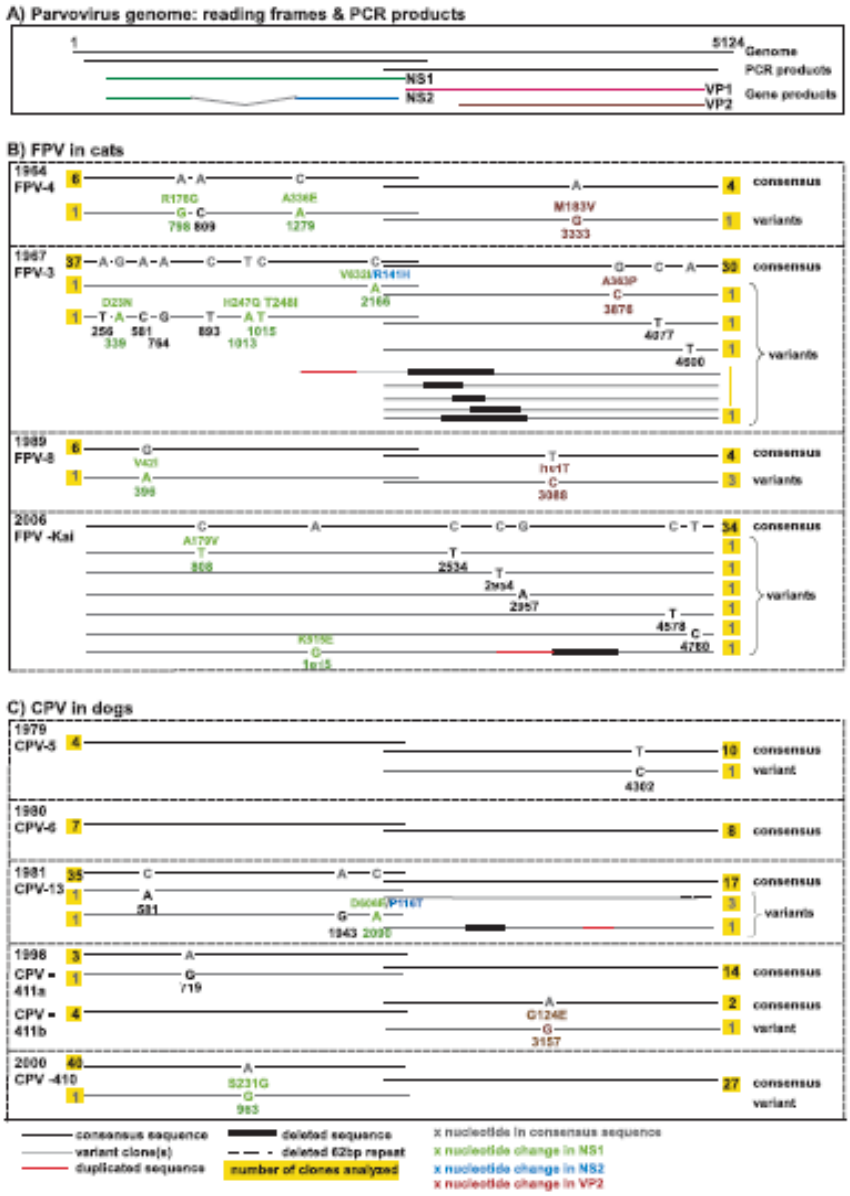


Figure 4.3. Intra-host diversity in FPV and CPV samples. (A) The gene regions covered by PCR amplification and a corresponding translation map of the parvovirus genome are indicated. B and C). Divergent viral sequences detected in naturally FPV (B) or CPV (C) infected animals are shown. The location of mutations in the parvovirus genome and the type of nucleotide substitution are indicated for each divergent sequence, the number of individual clones analyzed for each sample and the gene product affected by nonsynonymous mutations are indicated and the year of isolation is identified.

Table 4.2. Characterization of within-host mutations detected in individual FPV or CPV sequences. The location of single mutations detected in (A) FPV or (B) CPV samples. Nonsynonymous mutations are further characterized, and mutations present in other GenBank sequences are indicated by shading. (C) Since the sample CPV-411.us.98 most likely contained two co-infecting virus strains we analyzed the two virus populations (1998a and 1998b) separately and then compared the two consensus sequences present in sample CPV-411.

Virus	Virus type	Year of isolation	nt location	Protein region	Amino acid change	in nature	Potential function
FPV-4	FPV	1964	798	NS1	R176G	no	DNA binding[a]
			809	NS1	-	no	-
			1279	NS1	A336E	no	helicase [a]
			3333	VP2	M183V	no	unclear
FPV-3		1967	256	NS1	-	*	-
			339	NS1	D23N	yes	DNA binding[a]
			581	NS1	-	yes	-
			764	NS1	-	yes	-
			893	NS1	-	yes	-
			1013	NS1	H247Q	yes	DNA binding [a]
			1015	NS1	T248I	yes	DNA binding [a]
			2166	NS1/NS2	V632I (NS1) R141H (NS2)	no	VP2 phosphorylation[a]
			3876	VP2	A363P	no	helicase[a]
			4077	VP2	-	yes	-
			4600	VP2	-	*	-
FPV-8	FPV	1989	396	NS1	V42L	no	DNA binding[a]
			3088	VP2	I101T	yes	TfR binding
FPV-Kai	FPV	2006	808	NS1	A179V	no	DNA binding[a]
			1815	NS1	K515E	no	VP2 phosphorylation[a]
			2534	VP1	-	yes	-
			2954	VP2	-	no	-
			2957	VP2	-	yes	-
			4578	VP2	-	*	-
			4760	VP2	-	*	-

Table 4.2 (continued)

Virus	Virus type	Year of isolation	nt location	Protein region	Amino acid change	in nature	Potential function
CPV-5	CPV-2	1979	4302	VP2	-	no	unclear
CPV-6	CPV-2	1980	-	-	-	-	-
CPV-13	CPV-2a		581	NS1	-	yes	-
		1981	1943	NS1	-	no	-
			2090	NS1/NS2	D606E (NS1)/ P116T (NS2)	no	Intra-cellular traffickin g[a]
CPV-411a	CPV-2b	1998a	719	NS1	-	no	-
CPV-411b	CPV-2b	1998b	3157	VP2	G124E	no	-
CPV-410	CPV-2b	2000	963	NS1	S231G	no	unclear
CPV-411	CPV-2b	1998a/	413	NS1	-	yes	-
		1998b	941	NS1	-	yes	-
			1319	NS1	-	yes	-
			1481	NS1	-	yes	-
			1562	NS1	-	yes	-
			2195	NS1/ NS2	D151N (NS2)	yes	VP2 phosphor ylationa
			2247	VP1	-	yes	-
			2358	VP2	-	yes	-
			2846	VP2	-	yes	
			2951	VP2	-	yes	-
			3157	VP2	G124E	no	unclear
			3323	VP2	-	yes	-
			3749	VP2	N321K	yes	possibly TfR binding
			3947	VP2	-	yes	-
			4023	VP2	-	yes	-
			4494	VP2	K570E	yes	unclear
			4510	VP2	-	yes	-
			4526	VP2	-	*	-

Finally, genomic rearrangements were detected within the capsid protein gene regions of FPV-3.us.67 and FPV-kai.us.106 (Table 4.1) (with the FPV-kai.us.106 clones spanning the complete viral coding region between nts. 140 and 4815) and were verified as described above (data not shown).

4.4.2. Analysis of natural CPV infections.

Clinical CPV samples were isolated over a 22 year period, starting the year after CPV spread worldwide (Figure 4.1). Most samples were in feces, except for CPV-5.us.79 which was a sample from the spleen (Table 4.1). The sequences were assigned to the previously defined CPV-2 (CPV-5.us.79 and CPV-6.us.80), CPV-2a (CPV-13.us.81) and CPV-2b (CPV-410.us. 00 and CPV-411.is.98) antigenic types (Figure 4. 1) (35,37). Most mutations that distinguished the isolate consensus sequences clustered in the capsid protein gene (Figure 4.2B). Eleven of 20 (55%) mutations in this gene were nonsynonymous while genome-wide 14 nonsynonymous mutations were found among the 30 mutations (46.7%). Three mutations were located in the carboxy terminal region of NS1 where NS2 is encoded in a different reading frame, changing residues 641 of NS1 or 94 and 152 of NS2.

The CPV-411.us.98 sample, collected from a puppy infected during a shelter outbreak, harbored two clearly distinct viral populations (Figure 4.3B). Both strains were the CPV-2b type and all but one of the distinguishing mutations were also detected in GenBank sequences (Table 4.2), indicating that this likely represented a co-infection by two different viruses. The sample also contained a single capsid protein gene clone that was of CPV-2 type, possibly a remnant of vaccine virus.

Table 4.3. Analysis of heterogeneity in the viral population**.** The total number of mutations detected in each sample is indicated. In cases where co-infection was suspected (i.e. samples FPV-8.us.89 and CPV-411.us.98), the data was analyzed both by looking at the two viral populations separately and disregarding mutations distinguishing the potential subpopulations (A) and by disregarding the potential of co-infection and counting all mutations detected within an infected animal (B). Comparable estimates were then calculated from a previously published study of heterogeneity in a CPV infected cat. The number of mutations within an animal equals the total number of mutations detected, thus counting mutations that were present in more than one sequence multiple times. The average number of mutations/ nucleotide equals the total number of mutations detected in the respective sample, divided by the number of nucleotides analyzed.

Sample	Mutations within animal	Non-synonymous fraction	Average number mutations/ nucleotide	Mutations in NS1 & NS2		Mutations in VP1 & VP2	
				number	fraction	number	fraction
A) Disregarding mutations that distinguish putatively co-infecting strains							
FPV-4.us.64	4	0.75	1.4×10^{-4}	3	0.75	1	0.25
FPV-3.us.67	11	0.45	5.8×10^{-5}	8	0.73	3	0.27
FPV-8.us.89a	1	1.00	3.8×10^{-5}	1	1.0	-	-
FPV-8.us.89b	0	-	-	-	-	-	-
FPV-kai.us.06	7	0.29	3.7×10^{-5}	2	0.29	5	0.71
CPV-5.us.79	1	0	2.6×10^{-5}	0	0	1	1.00
CPV-6.us.80	0	-	-	-	-	-	-
CPV-13.us.81	3	0.33	2.2×10^{-5}	3	1.00	0	0
CPV-411.us.98a	1	0	2.3×10^{-5}	1	1.00	0	0
CPV-411.us.98b	1	1.00	3.6×10^{-5}	0	0	1	1.00
CPV-410.us.00	1	1.00	6.1×10^{-6}	1	1.00	0	0
Kitten 1	18	0.56	1.8×10^{-4}	17	0.94	1	0.06
Kitten 2	0	-	-	-	-	-	-
B) Accounting also for mutations that distinguish putatively co-infecting strains							
FPV-8.us.89*	4	1.00	1.2×10^{-4}	1	0.25	3	**0.75**
CPV-411.us98*	20	0.3	3.2×10^{-4}	7	0.35	13	**0.65**
C) Comparison to previously published studies							
CPV cat (4)	19	0.63	7.8×10^{-4}	-	-	-	-

However, since the possibility of cross-contamination in this case could not be ruled out, the sequence was excluded from further analyses.

In most CPV samples, one mutation was detected every 2-4 × 10^{-5} nucleotides, and mutations appeared to be somewhat more frequent in the NS1 than VP2 ORF (Table 4.3). Four of the 97 NS1 clones (4.1%) and two of the 83 VP2 clones (2.4%) harbored mutations from the consensus sequences, with individual clones harboring 1 or 2 mutations. Only one of the 7 mutations detected within CPV-infected dogs (14.3%) was also detected in sequences from GenBank (Table 4.2). We found evidence of one gene duplication and deletion in the VP2 gene fragment of CPV-13.us.81 (Table 4.1), and 3 of 21 CPV-13.us.81 sequences differed in the frequencies of a repeated 62 nt sequence located in the 3' end of the viral genome (Figure 4.3B), a phenomenon described previously for CPVs and FPVs (36).

4.4.3. Analysis of experimental cat infections.

Only low levels of viral replication and little evidence of clinical disease were seen in the two kittens inoculated with CPV-13.us.81. Kitten 1 developed mild clinical signs on day 4 after inoculation and shed low amounts of parvovirus in the feces the following day, while Kitten 2 showed mild clinical signs and viral shedding on day 6 post inoculation. Neither of the kittens inoculated with CPV-410.us.100 showed clinical signs, and no clear evidence of viral fecal shedding was detected by conventional or qPCR (data not shown). We analyzed a total of 29 NS1 and 22 VP2 spanning clones isolated from different tissues of the two CPV-13.us.81 inoculated animals (Figure 4.4B). In general, all viral populations were highly homogeneous.

Table 4.4. Mutations and re-arrangements detected in individual CPV sequences after experimental passage through cats. (A) Single mutations and (B) gene re-arrangements in samples collected after experimental infection with CPV-13.us.81. The functional impact of non-synonymous mutations is indicated where known or suspected. No virus could be recovered after experimental inoculation of cats with CPV-410.us.00. Mutations also observed in GenBank sequences are indicated by shading.

Cat #	*Organ of isolation*	*Gene region*	*nt location*	*Amino acid change*	*Observed in nature*	*Functional implication*
1	feces	NS1	941	-	yes	-
1	feces	NS1	1319	-	yes	-
1	feces	NS1/ 2	2024	T94A (NS2)	yes	Intra-cellular trafficking [a]
1	feces	NS1/ 2	2198	M152V (NS2)	yes	Interaction with members of the 14-3-3 protein family [a]
1	feces	VP2	2798	start codon	no	VP2 start codon
1	bone marrow	NS1/ 2	1855	D528V	yes	VP2 phosphorylation[a]

Cat #	**Organ of isolation**	Gene region	**Deleted region (nt)**	**Detected in inoculum**
1	thymus	VP2	2134-2875	no
			2274-3342	no
			2676-3770	no

Mutations were detected in the bone marrow (VP2 Asp528 to Val)) and feces of Kitten 1, with the later affecting the start codon of VP2. Two distinct sequences were recovered from the feces of Kitten 1, with 4 of the 8 clones identical to the challenge virus (Figure 4.3B). The other 4 clones harbored the same set of 4 mutations in the NS1 region, two synonymous changes in the NS1 carboxy terminal sequence and

nonsynonymous in the NS2 ORF (Thr94 to Ala and Met152 to Val) and two synonymous and located in the NS1 amino terminus. All mutations except the one in the VP2 gene of Kitten 1 were also present in GenBank sequences (Table 4.4).

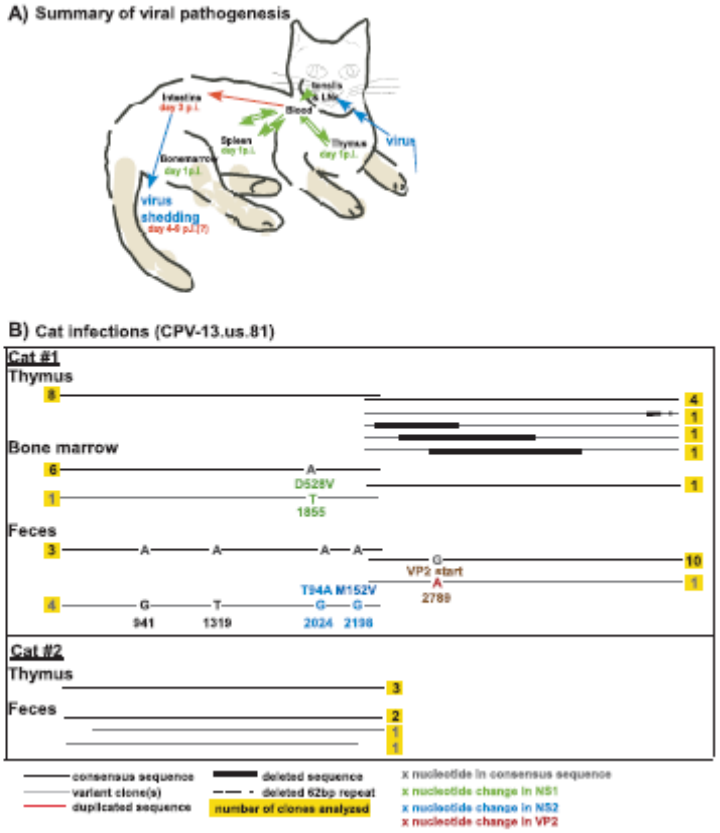


Figure 4.4. Analysis of the sequence recovered from experimental CPV-13.us.81 infection of cats. (A) Summary of the expected pathogenesis of FPV in cats, including the viral location at various days post infection (p.i.) and organs infected. (B) Characterization of viral sequences recovered from different organs, including the type of nucleotide substitution and the attained protein (for nonsynonymous changes) of each infected cats.

Gene re-arrangements were detected in sequences from the thymus of Kitten 1 and both forms of 62 nt repeat arrangements (Table 4.4), which were present in the challenge virus, were detected after experimental passage in cats.

4.5. Discussion

The emergence of CPV provides a valuable opportunity to examine the mechanisms of cross-species transmission, allowing a comparison of the viruses under endemic and emerging conditions. Here we define the levels of variation among the viral sequences in different hosts, comparing the long-adapted FPV in cats, CPV in dogs at various times after its global spread in 1978, as well CPV sequences during single experimental passages in cats. We compared the degree of sequence diversity in naturally infected dogs and cats, distinguishing between newly arising variation and heterogeneity due to co-infection, and analyzed the effect of host switching on the degree of sequence heterogeneity detected.

4.5.1. Intra-host sequence diversity.

In general, we observed low levels of sequence variation during natural infections by either FPV or CPV, and also during the experimental cross-species transfer of CPV to cats. Only 6.3% of viral sequences cloned from natural infections harbored any mutations. This contrasts with most previous studies of intra-host population structure during parvovirus infections, but has also been suggested by one other recent study of CPV in dogs (3). Battalani et al. (2006) analyzed individual virus genomes isolated from a single CPV-infected cat and detected high levels of sequence diversity within a 1745 nt fragment of the VP2 gene (4), with 10 distinct sequences observed among 14 analyzed viral clones (71%). Two antigenically distinct CPV variants (CPV type-2a and a variant with the VP2 D426E substitution, referred to as CPV type-2c) were isolated from this single animal, indicating that at least part of the observed variation was likely due to co- or superinfection rather than newly arising mutations.

4.5.2. Evidence for co-infections.

Notably, our study provided strong evidence for multiple infections. In particular, 3 of 7 clones analyzed for FPV-8.us.89 contained the same change (VP2 residue Ile101 to Thr) in the capsid protein gene, while being otherwise identical to the consensus sequence. This particular mutation is circulating in the FPV (Table 4.2) and CPV (19) population, thereby making a co-infection event a plausible explanation. The CPV-411.us.98 isolate presented particularly strong evidence of co-infection. This sample contained two genetically distinct CPV-2b viruses (CPV-411a.us.98 and CPV-411b.us.98), which are distinguished by 6 and 12 mutations in the NS1 or VP1 gene, respectively. All but one of the mutations differentiating those two viral strains are present in GenBank sequences, again supporting the idea of co-infection since the mutations are clearly circulating in the global CPV population. If we assumed co-infection as the source of some heterogeneity, the average number of mutations per nucleotide was in the range of 2×10^{-5} - 6×10^{-6} in all but one of the analyzed samples (the estimate for kitten 1 was excluded here as the artificial host switching might impact the developing population structure and the estimates might therefore not be comparable). Conversely, the estimates were considerably higher for the questionable samples if it was assumed that all variation arose by de-novo mutation. These higher estimates were in the range estimated based on the report of Batallani et al. (3), so that similar reasoning might explain the high diversity reported by these authors. Co-infections have been described for vaccine and field strains of CPV (13) and for several related parvovirus families, including the human parvovirus B19 (5). Co-infections with multiple parvovirus strains may thus occur frequently, potentially facilitating recombination (41).

4.5.3. Gene rearrangements.

We detected several gene duplication and deletion events among the FPV and CPV genomes examined. Such rearranged sequences have long been recognized among the parvoviruses (1, 10, 21), and likely result from template switching of the polymerase during replication. This process, which is also thought to be responsible for non-homologous and homologous recombination among RNA viruses (23, 28), may be facilitated by the complex secondary structures of the viral ssDNA genome. Recombination among CPV genomes was suggested as the source of genomes containing various combinations of mutations after extended tissue culture passage (2).

4.5.4. Effect of host species, organ or sampling time after CPV emergence.

The viral population structures in all isolates were generally characterized by low levels of sequence diversity, regardless of their host species, organ or time of isolation, or of the specific viral strain. On average, one mutation was detected every 10^4 -10^5 nucleotides, which is comparable to the annual substitution rate estimates of these viruses obtained on a population level (43, 19). Since variant CPV sequences arise very readily during tissue culture passage (2), strong purifying selection appears to quickly purge most arising mutants during natural infections, resulting in the low degree of sequence diversity within infected hosts. We detected no observable differences in intra-host population structure between CPV isolates collected during the first wave of spread (from 1979 and 1980) compared to those collected after the virus had been circulating in dogs for longer times, suggesting that dynamics of intra-host mutation and selection did not change markedly during this time period.

To better determine the effects of a single cross-species transmission event on the virus, we examined experimental CPV infections of cats, which revealed only

limited viral replication in the susceptible cats. The CPV sequences recovered from the cats showed low levels of heterogeneity, but parts of the viral capsid protein gene were deleted in some cases, indicative of defective genomes. In addition, one virus isolated from the feces of Cat 1 harbored a likely lethal mutation that altered the start codon of VP2. Mutations were detected in the NS1/2 regions of viruses isolated from bone marrow and feces, with approximately 50% of viruses in the feces of cat 1 harboring the same set of 4 mutations. Two of the changes (at residues 94 and 152 of NS2) mapped to the gene region where NS2 is derived from a different reading frame overlapping the NS1 reading frame and have been identified as potentially positively selected on a population level (19). Since these represent 2 of 4 apparently linked mutations observed after artificial host switching, a role of some or all of these residues in host adaptation is likely. Little is known about the function of NS2 in CPV, and NS2 knock-out mutants showed no obvious differences in replication in cell culture or dogs (46). In the rodent parvoviruses MVM and LuIII, NS2 appears to be required for efficient translation, capsid assembly, and nuclear transport. In these viruses host-specific effects modulate the functions of NS2 (9, 11, 12, 17, 25, 30). Interestingly, upon infection of severe combined immunodeficient (SCID) mice with MVM in the presence of polyclonal anti-capsid antibodies, the viral population harbored nonsynonymous changes in the NS2 C-terminus which likely affected CRM1 binding (27). In this case nonsynonymous mutations were absent from the capsid protein genes (27), although such mutations arose readily during infections in the absence of polyclonal antibodies (26). The role of NS2 in host adaptation of CPV therefore merits further study.

4.5.5. Distribution of mutations in the genomes.

The majority of mutations involved in the emergence of CPV -2 clustered in the capsid protein ORF, and on a population level this genome region has continued to evolve more rapidly than the non-structural ORF, at least among the CPV sequences (19). Among the FPV sequences, on the contrary, these differences between the genome regions were not observable (19). Interestingly, the mutations observed within individual infected animals here generally showed no clear clustering in any specific genome region, though mutations were somewhat more likely to occur in the non-structural ORF. Particularly the scarcity of mutations detected in the capsid protein region after experimental cross-species transmission to cats appears surprising and highlights the limited effect of positive selection within a single infected animal.

4.6. Acknowledgements

We thank Virginia Scarpino, Wendy Weichert and Melanie Ho for technical support. This work was in part supported by National Institutes of Health Grants GM080533 to ECH and AI028385 to CRP. KH is supported by a graduate assistantship from the College of Veterinary Medicine at Cornell.

REFERENCES

1. Allen, J. M., D. J. Debelak, T. C. Reynolds, and A. D. Miller. 1997. Identification and elimination of replication-competent adeno-associated virus (AAV) that can arise by nonhomologous recombination during AAV vector production. J Virol **71:**6816-6822.

2. Badgett, M. R., A. Auer, L. E. Carmichael, C. R. Parrish, and J. J. Bull. 2002. Evolutionary dynamics of viral attenuation. J Virol **76:**10524-10529.

3. Battilani, M., L. Gallina, F. Vaccari, and L. Morganti. 2007. Co-infection with multiple variants of canine parvovirus type 2 (CPV-2). Vet Res Commun **31 Suppl 1:**209-12.

4. Battilani, M., A. Scagliarini, S. Ciulli, L. Morganti, and S. Prosperi. 2006. High genetic diversity of the VP2 gene of a canine parvovirus strain detected in a domestic cat. Virology **352:**22-6.

5. Candotti, D., N. Etiz, A. Parsyan, and J. P. Allain. 2004. Identification and characterization of persistent human erythrovirus infection in blood donor samples. J Virol **78:**12169-78.

6. Carman, P. S., and R. C. Povey. 1985. Pathogenesis of canine parvovirus-2 in dogs - histopathology and antigen identification in tissues. Res Vet Sci **38:**141-150.

7. Carman, P. S., and R. C. Povey. 1985. Pathogenesis of canine parvovirus-2 in dogs: hematology, serology and virus recovery. Res Vet Sci **38:**134-140.

8. Chang, S. F., J. Y. Sgro, and C. R. Parrish. 1992. Multiple amino acids in the capsid structure of canine parvovirus coordinately determine the canine host range and specific antigenic and hemagglutination properties. J Virol **66:**6858-6567.

9. Choi, E. Y., A. E. Newman, L. Burger, and D. Pintel. 2005. Replication of minute virus of mice DNA is critically dependent on accumulated levels of NS2. J Virol **79:**12375-81.

10. Clement, N., B. Avalosse, K. El Bakkouri, T. Velu, and A. Brandenburger. 2001. Cloning and Sequencing of Defective Particles Derived from the Autonomous Parvovirus Minute Virus of Mice for the Construction of Vectors with Minimal cis-Acting Sequences. J Virol **75:**1284-1293.

11. Cotmore, S. F., A. M. D'Abramo, Jr., L. F. Carbonell, J. Bratton, and P. Tattersall. 1997. The NS2 polypeptide of parvovirus MVM is required for capsid assembly in murine cells. Virology **231:**267-280.

12. D'Abramo, A. M., Jr., A. A. Ali, F. Wang, S. F. Cotmore, and P. Tattersall. 2005. Host range mutants of minute virus of mice with a single VP2 amino acid change require additional silent mutations that regulate NS2 accumulation. Virology **340:**143-54.

13. Decaro, N., C. Desario, G. Elia, M. Campolo, A. Lorusso, V. Mari, V. Martella, and C. Buonavoglia. 2007. Occurrence of severe gastroenteritis in pups after canine parvovirus vaccine administration: a clinical and laboratory diagnostic dilemma. Vaccine **25:**1161-6.

14. Drake, J. W. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci(USA) **88:**7160-7164.

15. Drake, J. W. 1999. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Ann N Y Acad Sci **870:**100-107.

16. Duffy, S., L. A. Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet **9:**267-276.

17. Eichwald, V., L. Daeffler, M. Klein, J. Rommelaere, and N. Salome. 2002. The NS2 proteins of parvovirus minute virus of mice are required for efficient nuclear egress of progeny virions in mouse cells. J Virol **76:**10307-10319.

18. Gamoh, K., M. Senda, Y. Inoue, and O. Itoh. 2005. Efficacy of an inactivated feline panleucopenia virus vaccine against a canine parvovirus isolated from a domestic cat. The Vet Rec **157:**285-287.

19. Hoelzer, K., L. A. Shackelton, C. R. Parrish, and E. C. Holmes. 2008. Phylogenetic analysis reveals the emergence, evolution and dispersal of carnivore parvoviruses. J Gen Virol **in press**.

20. Hoelzer, K. and C.R. Parrish **.** 2008. Evolution and variation of the parvoviruses. *In* E. Domingo, Holland, J. and Parrish, C.R. (ed.), Origin and Evolution of Viruses, 2nd Ed. Elsevier, London.

21. Hogan, A., and E. A. Faust. 1986. Nonhomologous recombination in the parvovirus chromosome: role for a CTATTTCT motif. Mol. Cell Biol. **6:**3005-3009.

22. Hueffer, K., L. Govindasamy, M. Agbandje-McKenna, and C. R. Parrish. 2003. Combinations of two capsid regions controlling canine host range

determine canine transferrin receptor binding by canine and feline parvoviruses. J Virol **77:**10099-10105.

23. Lai, M. M. C. 1992. RNA recombination in animal and plant viruses. Microbiol Reviews **56:** 61-79.

24. Lemey, P., A. Rambaut, and O. G. Pybus. 2006. HIV evolutionary dynamics within and among hosts. AIDS reviews **8:**125-140.

25. Li, X., and Rhode, S.L. 1991. Nonstructural protein NS2 of parvovirus H-1 is required for efficient viral protein synthesis and virus production in rat cells in vivo and in vitro. Virology **184:**117-130

26. Lopez-Bueno, A., M. G. Mateu, and J. M. Almendral. 2003. High mutant frequency in populations of a DNA virus allows evasion from antibody therapy in an immunodeficient host. J Virol **77:**2701-2708.

27. Lopez-Bueno, A., N. Valle, J. M. Gallego, J. Perez, and J. M. Almendral. 2004. Enhanced cytoplasmic sequestration of the nuclear export receptor CRM1 by NS2 mutations developed in the host regulates parvovirus fitness. J Virol **78:**10674-10684.

28. Ma, H., R. Xu, H. Cheng, H. S. Kuo, M. During, and R. H. Fang. 2003. Gene transfer into human keloid tissue with adeno-associated virus vector. J Trauma **54:**569-573.

29. Meunier, P. C., B. J. Cooper, M. J. G. Appel, M. E. Lanieu, and D. O. Slauson. 1985. Pathogenesis of canine parvovirus enteritis: sequential virus distribution and passive immunization studies. Vet Pathol **22:**617-624.

30. Naeger, L. K., N. Salome, and D. J. Pintel. 1993. NS2 is required for efficient translation of viral mRNA in minute virus of mice-infected murine cells. J Virol **67:**1034-1043.

31. Nakamura, M., Y. Tohya, T. Miyazawa, M. Mochizuki, H. T. Phung, N. H. Nguyen, L. M. Huynh, L. T. Nguyen, P. N. Nguyen, P. V. Nguyen, N. P. Nguyen, and H. Akashi. 2004. A novel antigenic variant of canine parvovirus from a Vietnamese dog. Arch Virol **149:**2261-2269.

32. Parrish, C. R. 1999. Host range relationships and the evolution of canine parvovirus. Vet Microbiol **69:**29-40.

33. Parrish, C. R. 1995. Pathogenesis of feline panleukopenia virus and canine parvovirus. Baillieres Clin Haematol **8:**57-71.

34. Parrish, C. R., and L.E. Carmichael. 1983. Antigenic structure and variation of canine parvovirus, feline panleukopenia virus, and mink enteritis virus. Virology **129:**401-414.

35. Parrish, C. R., C. F. Aquadro, M. L. Strassheim, J. F. Evermann, J. Y. Sgro, and H. O. Mohammed. 1991. Rapid antigenic-type replacement and DNA sequence evolution of canine parvovirus. J Virol **65:**6544-6552.

36. Parrish, C. R., Aquadro, C.F., and L.E. Carmichael. 1988. Canine host range and a specific epitope map along with variant sequences in the capsid protein gene of canine parvovirus and related feline, mink and raccoon parvoviruses. Virology **166:**293-307.

37. Parrish, C. R., P. H. O'Connell, J. F. Evermann, and L. E. Carmichael. 1985. Natural variation of canine parvovirus. Science **230:**1046-1048.

38. Pereira, C. A., E. S. Leal, and E. L. Durigon. 2007. Selective regimen shift and demographic growth increase associated with the emergence of high-fitness variants of canine parvovirus. Infect Genet Evol **7:**399-409.

39. Raney, J. L., R. R. Delongchamp, and C. R. Valentine. 2004. Spontaneous mutant frequency and mutation spectrum for gene A of phiX174 grown in E. coli. Environ Mol Mutagen **44:**119-127.

40. Reed, A. P., E. V. Jones, and T. J. Miller. 1988. Nucleotide sequence and genome organization of canine parvovirus. J Virol **62:**266-276.

41. Shackelton, L. A., K. Hoelzer, C. R. Parrish, and E. C. Holmes. 2007. Comparative analysis reveals frequent recombination in the parvoviruses. J Gen Virol **88:**3294-3301.

42. Shackelton, L. A., C. R. Parrish, U. Truyen, and E. C. Holmes. 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc Natl Acad Sci USA **102:**379-384.

43. Truyen, U., M. Agbandje, and C. R. Parrish. 1994. Characterization of the feline host range and a specific epitope of feline panleukopenia virus. Virology **200:**494-503.

44. Truyen, U., and C. R. Parrish. 1992. Canine and feline host ranges of canine parvovirus and feline panleukopenia virus: distinct host cell tropisms of each virus in vitro and in vivo. J Virol **66:**5399-5408.

45. Truyen, U. W. E., J. F. Evermann, E. Vieler, and C. R. Parrish. 1996. Evolution of Canine Parvovirus Involved Loss and Gain of Feline Host Range. Virology **215:**186-189.

46. Wang, D., W. Yuan, I. Davis, and C. R. Parrish. 1998. Nonstructural protein-2 and the replication of canine parvovirus. Virology **240:**273-281.

47. Yamaguchi, Y., and T. Gojobori. 1997. Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. Proc Natl Acad Sci USA **94:**1264-1269.

CHAPTER 5

INTRAHOST DYNAMICS OF CANINE INFLUENZA VIRUS IN NATURALLY AND EXPERIMENTALLY INFECTED DOGS

Dr. Pablo Murcia was instrumental for generating the presented sequence data and Drs. Baillie and Wood were also intimately involved in the deep amplicon sequencing. The experimental dog infections were performed in close collaboration with Drs. Osterrieder and Metzger.

5.1. Summary

Canine Influenza virus (CIV) is a novel pathogen of dogs. The virus emerged around 2004 following a host jump from an H3N8 equine influenza virus. CIV is still spreading in certain US dog populations, even though some evidence suggests relatively low virus transmissibility. Little is known about the factors facilitating this successful cross-species transmission or the subsequent evolution, spread and potentially continuing host adaptation of CIV. Here we address these questions by analyzing the intra-host dynamics of CIV in dogs over the course of infection.

Virus populations in experimentally infected dogs were characterized by low levels of genetic diversity, which decreased over time. Intriguingly, where a sample for day 4 post inoculation was available, genetic diversity appeared to increase slightly on this day, after reaching a nadir on day 3 post inoculation. Most detected mutations had low prevalence (largely singletons), and most mutations present at higher prevalence were pre-existing in the challenge virus. Mutations were often not detected on more than one subsequent day and variable sites were not strongly correlated between animals or between days. However, several mutations that were not detected in the challenge virus seemed to be maintained in the infected animals over several days, and many mutations that reached higher frequency were located in known antigenic sites.

5.2. Introduction

Host-jumping viruses can represent important threats to human and animal health. Numerous devastating emerging human pathogens such as the human immunodeficiency virus, Influenza A virus, or SARS coronavirus and several common human diseases such as mumps, measles and pertussis have been shown or are believed to have emerged after cross-species transmission from an animal host (4,

7, 14, 19, 21). However, despite the profound evolutionary importance of cross-species transmission, the evolutionary mechanisms underpinning this process remain opaque. In general, the adaptation of a pathogen to a new host is believed to be a sequential process, where the initial virus may be maladapted to the new host and only inefficiently transmitted. Consequently, further adaptation is required before the novel pathogen can cause wide-spread epidemics (1, 11). During this adaptation phase, the virus may only circulate locally and causes predominantly short, self-limiting transmission chains. Many outbreaks of novel diseases end at this stage - but the factors determining the success or failure of these host-transferred viruses are unclear (21).

The propensity for a virus to generate genetic variation, as well as the fitness distribution of mutations within individual hosts, is therefore essential to understanding the process of cross-species transmission and emergence. In general, the shape of a viral population within an infected host represents a balance between diversity generated through error-prone replication and how rapidly it is removed by purifying selection. The shapes of these viral populations are likely host-specific and might be strongly impacted by host-associated factors such as immune status or age. However, the within-host population structure of most newly emerged viruses, and the relative importance of intra-host dynamics on disease emergence and subsequent adaptation, are unknown.

Canine influenza virus (CIV) was first recognized as a new pathogen of dogs in January 2004, when it caused an outbreak of respiratory disease in racing greyhounds in Florida (3). Serological testing of archival samples later revealed that the virus had been circulating in Florida racetracks since at least 2000, and CIV

seroprevalence was high in racing greyhounds and shelter dogs by mid-2004 (3). By August 2004, the disease had also been detected in dogs at greyhound race tracks in Texas, Alabama, Arkansas, West Virginia and Kansas (3). CIV continued to spread across the US, increasingly affecting non-racing dogs (12, 22). By October 2008, a national syndromic serosurvey had analyzed CIVantibody-positive samples from 26 different states and the District of Columbia (Cornell University College of Veterinary Medicine, http://diaglab.vet.cornell.edu/issues/civ-stat2.asp). This survey included dogs presenting with clinical signs that were compatible with CIV infection, but that were unassociated with greyhound racetracks. Between 2005 and 2008, > 6000 serological samples had been analyzed though this program. On average, 17% of samples tested positive for CIV, with the ratio ranging from 1.4% to 63% depending on the state. High percentages of positives were detected in samples from Colorado (62.6%), Delaware (56.9%) and Wyoming (43.4%), while samples from Arizona (27.3%), Illinois (19.6%), Massachusetts (23.4%), New York (23.3%), Pennsylvania (16.7%) and Washington DC (21.8%) showed lower levels of positives. Intriguingly, few or no positives were detected in samples from Georgia (n=62), Maryland (n=64), Minnesota (n=68), North Carolina (n=69), Ohio (n=183), Texas (n=119), Virginia (n=106), Washington State (n=130) and Wisconsin (n=73), indicating a very heterogeneous, patchy distribution of the virus, and likely very local viral spread.

Phylogenetic analysis suggests that CIV emerged through direct transfer of an intact equine H3N8 influenza strain, suggesting a probably negligible role of reassortment in this emergence event (3). Moreover, all CIV sequences analyzed to date are monophyletic, indicating that a single cross-species transmission event has occurred (3, 22). In the hemagglutinin (HA) gene (H3) a set of five amino acid changes appears to distinguish the currently sequenced CIV isolates from their equine

ancestors, while another three amino acid changes in this gene distinguish the more recent from the older CIV isolates, thus potentially representing the result of further host adaptation or antigenic escape (12). Intriguingly, the CIV isolates are antigenically similar to contemporary EIV isolates, while qualitative differences between older and more recent CIV isolates my exist (12). However, to date there is limited data available for CIV and many questions regarding the emergence of CIV and its subsequent spread in the North American dog population have remained unanswered.

Here we analyzed, for the first time, the viral intra-host sequence variation and temporal dynamics of CIV during infections of dogs. To address the potential impact of our experimental inoculation, we compared the dynamics observed during our experimental infections with those observed in the field by analyzing a selected sample collected from a naturally infected dog.

5.3. Materials and Methods

5.3. 1. Experimental dog infections.

The dog infections and analysis of antibody titers, clinical symptoms and viral shedding have been described previously (15) and are summarized below for the dogs analyzed in this study (Tables 5.1). In short, 12 purpose-bred intact beagle bitches (Marshall Farms), approximately 8 weeks of age, were placed in group housing prior to challenge inoculation and were not separated based on group affiliation. Individual dogs were identified by ear tattoos and none of the dogs had detectable antibodies to CIV, as determined by HA Inhibition assay, prior to vaccination.

The dogs were randomly divided into three groups of four. The allocation of individual dogs remained obscured for the duration of the experiment and the evaluation of serological data. Dogs were vaccinated subcutaneously (SC) and group 1 received 2.4 X 10^6 PFU of a recombinant equine herpesvirus expressing the viral HA (rH_EIV) (15), while group 2 received a commercial equine H3N8 subunit vaccine. Group 3 received only virus resuspension buffer (negative control). The dogs received a booster vaccination of 4.1 X 10^6 PFU of rH_EIV, the commercial subunit vaccine or resuspension buffer 4 weeks later. All dogs were challenged three weeks after booster vaccination with 1x10^6 PFU A/canine/PA/10915-07 which had been amplified by growth in embryonated eggs. For the inoculation, 2 ml of virus-containing allantoic fluid were placed in a custom-engineered nebulizer and administered with flow-through oxygen to each individual dog for ~10 minutes.

Blood was drawn from the cephalic vein on days -56 (initial vaccination), -35, -21 (booster vaccination), -14, and -7 before challenge. In addition, blood was drawn on days 0, 8, and 15 post challenge. Two nasal swabs were taken from each dog on days 0 to 10 post inoculation (p.i.) using one 3" non-sterile foam-tipped swab (Fisher) and one TX754B Mini Alpha swab (ITW Texwipe). The swabs were placed immediately in 1 ml of transport media (PBS supplemented with 10% glycerol, 4% penicillin/streptomycin/amphotericin B, 10% FBS, and 0.05% enrofloxacin (Baytril®)) on ice.

Physical examinations were performed 2 days prior to challenge, on the day of challenge (day 0) and from days 1 to 8, 10 and 15 post challenge and have been described previously (15). Briefly, observations of the activity level, demeanor, heart

rate, respiratory rate, rectal temperature, and overall general appearance were performed and subsequently combined. A fever was defined as a rectal temperature above 39.5°C (i.e. 103.1 degrees Fahrenheit). Clinical signs were scored using 5 categories, with 1 = absence of symptoms; 2 = mild clinical signs or mildly depressed animals; 3 = moderate clinical signs or moderately depressed animals; 4 = severe clinical signs or depressed animals with clear clinical signs and 5 = severe clinical signs and clearly depressed animals.

All sera were tested by use of standard microtiter haemagglutination inhibition (HI) and an inhibition enzyme-linked immunosorbent (ELISA) assay and results have been reported previously (15).

Two naïve dogs were selected based on the clinical manifestation (Table 5.1), rectal temperature (Table 5.1) and antibody titers (Table 5.2), and CIV sequences isolated from these animals were prepared for sequencing as described below.

Table 5.1: Summary of clinical data for experimentally infected dogs, previously published by (15). Days analyzed during this study are indicated by bolding. Pyrexia was defined as a temperature above 103.1 degrees Fahrenheit. Code for clinical signs: 1= no symptoms; 2= mild clinical signs or mildly depressed; 3= moderate clinical signs or moderately depressed; 4= severe clinical signs or depressed with clear clinical signs; 5= severe clinical signs and clearly depressed.

measure	dog #	Immune status	days post inoculation				
			day 1	day 2	day 3	day 4	day 5
temperature	1	naïve	101.3	**104.0**	**103.2**	**103.5**	102.6
	2		101.8	**103.8**	**102.4**	103.3	102.6
clinical signs	1	naïve	2	**2**	**3**	**3**	4
	2		2	**3**	**2**	2	1

Table 5.2: Changes in anti-influenza antibody (HA inhibition) titers of naïve and vaccinated dogs after mock vaccination (day-53) and after inoculation (day 0), previously published by (15).

dog #	Immune status	days post vaccination			
		day (-)14	day 0	day 8	day 15
1	naïve	8	8	128	1024
2		8	8	128	256

5.3.2. Naturally infected dog sample.

A nasal swab from a naturally infected, clinically sick and client-owned animal was submitted to the Cornell Diagnostic lab for routine CIV testing. Aliquots of nasal swabs from CIV-positive dogs were stored at -80C° until viral RNA isolation.

5.3.3. RNA isolation, RT-PCR analysis, cloning and sequencing.

Viral RNA from experimentally or naturally infected dogs was isolated using the QIAamp viral RNA mini kit (QIAGEN, Valencia, CA) according to the manufacturer's recommendation, starting with 140 µl nasal swab material, and eluting in a total volume of 70 µl of buffer AVL (QIAGEN, Valencia, CA). Viral RNA was stored at -80°C until reverse transcription (RT). Viral RNA was reversed transcribed using two-step reverse transcription. The circular DNA (cDNA) was generated using Superscript III reverse transcriptase (INVITROGEN, Carlsbad, CA) according to the manufacture's recommendation, starting from 5 µl viral RNA . Primer Bm-HA1 (5' TATTCGTCTCAGGGAGCAAAAGCAGGGG3') (previously published by (8)) was used to generate the cDNA and RT was performed at 55°C for 90 min, followed by incubation at 70°C for 10 min. The second strand was generated using Platinum *Pfx* DNA polymerase (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol, starting from 5 µl cDNA. Primers Bm-HA1 and EHA1007rw (5'TTGGGGCATTTTCCATATGT3') were used for the PCR amplification, which

was performed during 40 cycles (94°C for 30 sec, 55°C for 1 min and 68°C for 1 min), followed by a final extension at 68°C for 10 min. Primers spanned the region between nt -43 (upstream of the HA1 start codon) and nt 965, generating a fragment of 1008 nts.

PCR products were purified by agrose gel purification using the Qiaquick gel extraction kit (QIAGENE) and were cloned into pCR4Blunt-TOPO (Invitrogen) according to the manufacturer's protocol. Cloned constructs were transformed into electrocompetent E. coli hosts (One Shot TOP10 cells, Invitrogen) and plated with ampicillin or kanamycin selection. Amplicons were sequenced using commercial primers M13 forward and M13 (20) reverse. To test whether sequences generated by Sanger and Cornell University were comparable, three samples were partially sequenced at both places, and the phylogenetic relationship between the sequences generated at either place was compared (data not shown).

5.3.4. Evolutionary analysis.

Between 22 and 88 HA1 sequences from each sample, and on each day (Table 5.5), were determined as described above and aligned by hand in the Se-Al program (http://tree.bio.ed.ac.uk/software/seal/). Sequence alignments were trimmed to cover amino acids 1 to 316 of the nascent HA1, thus including the signal peptide. To map the amino-acids onto the inferred sequence of the mature HA, alignments were further trimmed to correspond to the mature HA1 (all alignments available from the authors on request). To compare the evolutionary dynamics within infected animals to those observed on a population level, we constructed alignments of 165 equine sequences and 6 CIV sequences, both covering the corresponding region (Tables 5.3 and 5.4), from the Influenza Virus Resource

(http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). These alignments were trimmed using Se-Al, and analyzed as described below. Due to the very limited number of CIV sequences available in GenBank, no population-level analysis of CIV was performed.

Unrooted maximum likelihood (ML) phylogenetic trees were estimated for the challenge virus, the alignment of EIV sequences and for sequences collected from each analyzed animal using PAUP* version 4.0 (18) with the best-fitting model of nucleotide substitution determined using Modeltest (13) (all parameter values available from the authors on request). The mean pairwise genetic diversity for each sample was subsequently calculated from the uncorrected pairwise distance matrix between taxa (available from the authors upon request).

We used the Single Likelihood Ancestor Counting (SLAC) algorithm available in the Datamonkey web interface of the HyPhy software package (10) to estimate the relative rates of synonymous (d_S) and nonsynonymous (d_N) substitutions per site (ratio d_N/d_S) in our challenge virus, the sequences collected from individual infected animals, the EIV consensus sequences, and the EIV alignment (again employing the best-fitting substitution model from Modeltest).

Mutations detected within individual infected animals were characterized further as to their frequency and presence in other samples, and were mapped to inferred amino acid sequences of the mature HA1. Mutations present at a frequency > 1 were subsequently compared to the alignments of 165 equine H3N8 and 6 CIV sequences (Tables 5.3 and 5.4) to assess their presence in the canine or equine H3N8 population.

Table 5.3: Overview of H3N8 equine influenza virus sequences used to assess the population-level evolution of equine influenza virus.

Genbank accession	Country of isolation	Year of isolation	virus name
AB360549	Japan	2007	A/equine/Ibaraki/1/07(H3N8)
AB369862	Japan	2007	A/equine/Kanazawa/1/2007(H3N8)
AB435160	Japan	2008	A/equine/Tokyo/3/2008(H3N8)
AB435161	Japan	2008	A/equine/Hyogo/1/2008(H3N8)
AB436910	Mongolia	2008	A/equine/Mongolia/1/2008(H3N8)
AF197241	USA	1998	A/equine/Kentucky/1/1998(H3N8)
AF197242	USA	1994	A/equine/Florida/1/94(H3N8)
AF197243	Canada	1990	A/equine/Saskatoon/1/1990(H3N8)
AF197244	Argentina	1995	A/equine/Argentina/1/95(H3N8)
AF197245	Argentina	1994	A/equine/Argentina/2/94(H3N8)
AF197246	Argentina	1996	A/equine/Argentina/1/96(H3N8)
AF197247	USA	1995	A/equine/Kentucky/9/1995(H3N8)
AF197248	USA	1996	A/equine/Kentucky/1/1996(H3N8)
AF197249	USA	1997	A/equine/Kentucky/1/1997(H3N8)
AJ223192	United Kingdom	1989	A/equine/Lichfield/89(H3N8)
AJ223193	United Kingdom	1989	A/equine/Newmarket-Bob Champion/89(H3N8)
AJ223194	Germany	1989	A/equine/Berlin/3/89(H3N8)
AJ223195	Germany	1989	A/equine/Berlin/4/89(H3N8)
AJ223196	Germany	1991	A/equine/Berlin/1/91(H3N8)
AJ223197	Germany	1991	A/equine/Berlin/2/91(H3N8)
AX018718	n/a	n/a	Equine influenza virus H3N8
AY048077	Argentina	1997	A/equine/Argentina/97(H3N8)
AY048078	Argentina	1999	A/equine/Argentina/99-2(H3N8)
AY048079	Argentina	1999	A/equine/Argentina/99-3(H3N8)
AY048080	Argentina	1999	A/equine/Argentina/99-4(H3N8)
AY048081	Argentina	1999	A/equine/Argentina/99(H3N8)
AY273167	USA	1999	A/equine/New York/99(H3N8)
AY273168	USA	2000	A/equine/Oklahoma/2000(H3N8)
AY383755	Chile	1985	A/equine/Santiago/1/1985(H3N8)
AY855341	USA	2002	A/equine/Kentucky/5/2002(H3N8)
AY919314	Sweden	2001	A/equi 2/Gotland/01(H3N8)
BD244629	n/a	n/a	Equine influenza virus H3N8
BD244630	n/a	n/a	Equine influenza virus H3N8
BD244631	n/a	n/a	Equine influenza virus H3N8
BD244632	n/a	n/a	Equine influenza virus H3N8
CS287777	n/a	n/a	Equine influenza virus H3N8
CY018869	USA	1991	A/equine/Idaho/37875/1991(H3N8)
CY028796	USA	1978	A/equine/Kentucky/1/1978(H3N8)
CY028804	USA	1980	A/equine/Kentucky/4/1980(H3N8)
CY028812	USA	1980	A/equine/California/1/1980(H3N8)

Table 5.3 (continued)

Genbank accession	Country of isolation	Year of isolation	virus name
CY028820	USA	1981	A/equine/Kentucky/2/1981(H3N8)
CY028828	USA	1981	A/equine/Kentucky/1/1981(H3N8)
CY028836	USA	1963	A/equine/Miami/1/1963(H3N8)
CY028844	USA	1976	A/equine/Kentucky/pass_the_pepper1/1976(H3N8)
CY028852	USA	1976	A/equine/Kentucky/bitter_boredom5/1976(H3N8)
CY028860	USA	1981	A/equine/Georgia/1/1981(H3N8)
CY028868	USA	1981	A/equine/Georgia/3/1981(H3N8)
CY028876	USA	1981	A/equine/Georgia/9/1981(H3N8)
CY028884	USA	1981	A/equine/Georgia/10/1981(H3N8)
CY028892	USA	1981	A/equine/Georgia/13/1981(H3N8)
CY028900	USA	1981	A/equine/Kentucky/magnificent_genius1/1981(H3N8)
CY028908	USA	1982	A/equine/California/103/1982(H3N8)
CY028916	USA	1983	A/equine/New York/VR-297/1983(H3N8)
CY030077	USA	1981	A/equine/Kentucky/3/1981(H3N8)
CY030085	USA	1982	A/equine/California/83/1982(H3N8)
CY030093	USA	1986	A/equine/Kentucky/1/1986(H3N8)
CY030101	USA	1987	A/equine/Kentucky/1/1987(H3N8)
CY030109	USA	1988	A/equine/Kentucky/692/1988(H3N8)
CY030117	USA	1988	A/equine/Kentucky/694/1988(H3N8)
CY030125	USA	1988	A/equine/Kentucky/698/1988(H3N8)
CY030133	USA	1990	A/equine/Kentucky/1277/1990(H3N8)
CY030141	USA	1991	A/equine/Texas/39655/1991(H3N8)
CY030149	USA	1992	A/equine/Kentucky/1/1992(H3N8)
CY030157	USA	1991	A/equine/Alaska/29759/1991(H3N8)
CY030165	USA	1986	A/equine/Tennessee/5/1986(H3N8)
CY030173	USA	1991	A/equine/Kentucky/1/1991(H3N8)
CY030181	USA	1994	A/equine/Kentucky/8/1994(H3N8)
CY030189	USA	1983	A/equine/New York/1/1983(H3N8)
CY030743	USA	1981	A/equine/Kentucky/Rosie100/1981(H3N8)
CY030751	USA	1997	A/equine/California/4537/1997(H3N8)
CY030759	USA	2002	A/equine/California/8560/2002(H3N8)
CY031538	USA	1987	A/equine/Kentucky/2/1987(H3N8)
CY032221	USA	1986	A/equine/Kentucky/2/1986(H3N8)
CY032229	USA	1986	A/equine/Kentucky/3/1986(H3N8)
CY032293	Brazil	1963	A/equine/Sao Paulo/6/1963(H3N8)
CY032301	Argentina	1985	A/equine/Cordoba/18/1985(H3N8)
CY032309	USA	1985	A/equine/Santa Fe/1/1985(H3N8)
CY032317	United Kingdom	1989	A/equine/Sussex/1/1989(H3N8)
CY032325	United Kingdom	1989	A/equine/Rook/93753/1989(H3N8)

Table 5.3 (continued)

Genbank accession	Country of isolation	Year of isolation	virus name
CY032333	Italy	1991	A/equine/Italy/824/1991(H3N8)
CY032341	Italy	1992	A/equine/Italy/1199/1992(H3N8)
CY032349	Austria	1992	A/equine/Austria/421/1992(H3N8)
CY032357	Switzerland	1993	A/equine/Switzerland/173/1993(H3N8)
CY032365	Italy	1991	A/equine/Rome/5/1991(H3N8)
CY032373	Italy	1991	A/equine/Italy/1062/1991(H3N8)
CY032381	Switzerland	1979	A/equine/Switzerland/1118/1979(H3N8)
CY032389	Romania	1980	A/equine/Romania/1/1980(H3N8)
CY032397	Brazil	1969	A/equine/Sao Paulo/1/1969(H3N8)
CY032405	France	1979	A/equine/Fontainbleu/1/1979(H3N8)
CY032413	Germany	1989	A/equine/Berlin/1/1989(H3N8)
CY032421	Uruguay	1963	A/equine/Uruguay/1/1963(H3N8)
CY032937	USA	1980	A/equine/Kentucky/2/1980(H3N8)
CY032945	Algeria	1972	A/equine/Algiers/1/1972(H3N8)
CY032953	South Africa	1986	A/equine/Johannesburg/1/1986(H3N8)
CY033481	n/a	1978	A/equine/Kascakew/1/1978(H3N8)
CY034934	Japan	1971	A/equine/Sachiyama/1/1971(H3N8)
D30677	United Kingdom	1979	A/equine/NewMarket/D64/79(H3N8)
D30678	United Kingdom	1989	A/equine/Sussex/93753/89(H3N8)
D30679	France	1989	A/equine/Yvelines/2136/1989(H3N8)
D30680	USA	1991	A/equine/Alaska/1/1991(H3N8)
D30681	USA	1991	A/equine/Arundel/12369/91(H3N8)
D30682	Hong Kong	1992	A/equine/Hong Kong/1/92(H3N8)
D30683	USA	1992	A/equine/Kentucky/1/1992(H3N8)
D30684	Italy	1991	A/equine/Rome/5/1991(H3N8)
D30685	United Kingdom	1992	A/equine/Lambourn/22778/92(H3N8)
D30686	Spain	1993	A/equine/LaPlata/1/1993(H3N8)
DQ124189	USA	1999	A/equine/New York/1/1999(H3N8)
DQ124191	USA	2002	A/equine/Kentucky/5/2002(H3N8)
DQ124192	USA	2003	A/equine/Ohio/1/2003(H3N8)
DQ124193	USA	2003	A/equine/Massachussetts/213/2003(H3N8)
DQ124194	USA	2003	A/equine/New York/452/2003(H3N8)
DQ124195	USA	2003	A/equine/California/191/2003(H3N8)
DQ222913	USA	2003	A/equine/Wisconsin/1/03(H3N8)
EF117330	Italy	2005	A/equine/Bari/2005(H3N8)
EF536318	Italy	2003	A/equine/Rome/1/2003(H3N8)
EF536319	Italy	2003	A/equine/Rome/2/2003(H3N8)
EF536320	Italy	2004	A/equine/Rome/2004(H3N8)
EF536321	Italy	1999	A/equine/Brescia/1999(H3N8)
EF541428	Germany	2000	A/equine/Berlin/00(H3N8)
EF541429	Germany	2002	A/equine/Berlin/13/02(H3N8)

Table 5.3 (continued)

Genbank accession	Country of isolation	Year of isolation	virus name
EF541430	Germany	2002	A/equine/Berlin/14/02(H3N8)
EF541431	United Kingdom	2000	A/equine/Leicestershire/2/2000(H3N8)
EF541432	United Kingdom	2004	A/equine/Aboyne/410355/04(H3N8)
EF541433	United Kingdom	2004	A/equine/Stoke-on-Trent/410956/04(H3N8)
EF541434	United Kingdom	2004	A/equine/Arundel/410749/04(H3N8)
EF541435	United Kingdom	2001	A/equine/Cheltenham/1/01(H3N8)
EF541436	United Kingdom	2000	A/equine/Leicestershire/1/2000(H3N8)
EF541437	United Kingdom	2005	A/equine/Lanark/1/2005(H3N8)
EF541438	United Kingdom	2005	A/equine/Wales/1/2005(H3N8)
EF541439	United Kingdom	2005	A/equine/Essex/1/2005(H3N8)
EF541440	United Kingdom	2005	A/equine/Essex/2/2005(H3N8)
EF541441	United Kingdom	2005	A/equine/Essex/3/2005(H3N8)
EF541442	Kingdom	2005	A/equine/Aboyne/1/2005(H3N8)
EF541443	Italy	1991	A/equine/Italy/788/91(H3N8)
EU503046	Argentina	2001	A/equine/Argentina/1/2001(H3N8)
EU855741	Canada	2003	A/equine/Guelph/G03-55399/2003(H3N8)
EU855742	Canada	2004	A/equine/Guelph/G04-54701/2004(H3N8)
EU855743	Canada	2003	A/equine/Guelph/G03-0250/2003(H3N8)
EU855744	Canada	2006	A/equine/Guelph/06-28865/2006(H3N8)
EU926631	Chile	2006	A/equine/Lonquen/1/2006(H3N8)
FJ209731	Egypt	2008	A/equine/Egypt/6066NAMRU3-VSVRI/2008(H3N8)
FJ605181	Greece	2007	A/equine/Athens/02/2007(H3N8)
FJ605182	Greece	2007	A/equine/Athens/04/2007(H3N8)
L27597	Hong Kong	1992	A/equine/Hong Kong/1/92(H3N8)
L39913	Argentina	1993	A/equine/Argentina/1/93(H3N8)
L39914	USA	1994	A/equine/Kentucky/1/1994(H3N8)
L39915	USA	1990	A/equine/Kentucky/1/1990(H3N8)
L39916	USA	1993	A/equine/Florida/1/93(H3N8)

Table 5.3 (continued)

Genbank accession	Country of isolation	Year of isolation	virus name
L39917	USA	1992	A/equine/Kentucky/1/1992(H3N8)
L39918	USA	1991	A/equine/Kentucky/1/1991(H3N8)
M24718	Uruguay	1963	A/equine/Uruguay/1/1963(H3N8)
M24719	USA	1963	A/equine/Miami/1/1963(H3N8)
M24720	Japan	1971	A/equine/Tokyo/1971(H3N8)
M24721	Algeria	1972	A/equine/Algiers/1972(H3N8)
M24722	United Kingdom	1976	A/equine/New Market/1976(H3N8)
M24723	France	1976	A/equine/Fontainebleau/1976(H3N8)
M24724	Romania	1980	A/equine/Romania/1980(H3N8)
M24725	Chile	1985	A/equine/Santiago/1/1985(H3N8)
M24726	USA	1985	A/equine/Tennessee/5/1985(H3N8)
M24727	USA	1986	A/equine/Kentucky/2/1986(H3N8)
M24728	USA	1987	A/equine/Kentucky/1/1987(H3N8)
M29257	USA	1963	A/equine/Miami/1963(H3N8)
M65018	China	1989	A/equine/Jilin/1/1989(H3N8)
M73773	France	1976	A/equine/France/1/1976(H3N8)
S64310	Sweden	1991	Equine influenza virus H3N8
S77429	Sweden	1979	Equine influenza virus H3N8
U58195	USA	1981	A/equine/Kentucky/1/1981(H3N8)
X68437	United Kingdom	1989	A/equine/Suffolk/89(H3N8)
X85085	USA	1991	A/eq/Arundel/91/(H3N8)
X85086	Netherlands	1989	A/eq/Ella/89/(H3N8)
X85087	United Kingdom	1992	A/eq/Lambourn/92/(H3N8)
X85088	United Kingdom	1993	A/eq/Newmarket/93/(H3N8)
X85089	United Kingdom	1993	A/eq/Newmarket/93/(H3N8)
X85090	United Kingdom	1989	A/eq/Sussex/89/(H3N8)
X95637	Nigeria	1991	A/equine/Ibadan/6/91(H3N8)
X95638	Nigeria	1991	A/equine/Ibadan/9/91(H3N8)
Y14053	Sweden	1988	A/equi 2/Skara/88(H3N8)
Y14054	Sweden	1988	A/equi 2/Solvalla/79(H3N8)
Y14055	Sweden	1984	A/equi 2/Aby/84(H3N8)
Y14056	Sweden	1990	A/equi 2/Visingso/90(H3N8)
Y14057	Sweden	1993	A/equi 2/Avesta/93(H3N8)
Y14058	Sweden	1994	A/equi 2/Soderala/94(H3N8)
Y14059	Sweden	1996	A/equi 2/Alvdalen/96(H3N8)
Y14060	Sweden	1996	A/equi 2/Bollnas/96(H3N8)

Table 5.4: Overview of canine influenza sequences used to assess the population-level evolution of canine H3N8 influenza virus.

Genbank accession	Country of isolation	Year of isolation	virus name
DQ124157	USA	2003	A/canine/Florida/242/2003(H3N8)
DQ124190	USA	2004	A/canine/Florida/43/2004(H3N8)
DQ124196	USA	2004	A/canine/Texas/1/2004(H3N8)
DQ146419	USA	2005	A/canine/Iowa/13628/2005(H3N8)
EU402407	USA	2005	A/dog/Jacksonville/C3/2005(H3N8)
EU402408	USA	2005	A/dog/Miami/E3/2005(H3N8)

5.4. Results

Overall, we obtained >400,000 bases of sequence from naturally and experimentally CIV infected dogs as well as the challenge virus inoculum. Consecutive samples, collected on days 2, 3 and in one case 4 p.i., were analyzed and compared to each other and to the challenge virus. Due to low viral loads on day 1 and after day 3 or 4 p.i. (15), we restricted our analysis to those days on which the viral populations were sufficiently large (Table 5.5). By day 8 p.i., all inoculated dogs had produced detectable antibody titers, which in nearly all animals further increased by day 15, indicating that the virus challenge had elicited an immune response (see Table 5.2). All dogs showed pyrexia and mild to moderate clinical signs on day 2 p.i., which in most cases lasted until day 4 p.i. (see Table 5.1). In both naïve animals temperatures transiently decreased on day 3 before rising again on day 4 p.i. Similar dynamics were observed for the viral titers during that same time period, indicating that the viral population size in naïve animals potentially temporarily decreased on day 3 p.i (15).

5.4.1. Analysis of intra-host dynamics after experimental challenge.

Samples collected from naïve dogs on day 2 p.i. were similar in genetic diversity to the challenge virus, with the mean estimate for the challenge virus equaling 0.0014 and that for dogs 1 and 2 equaling 0.0017 and 0.0012 respectively. Samples collected on later days showed lower levels of diversity, with the estimates for day 3 equaling 0.0005 and 0.0008 for dogs 1 and 2, respectively (Table 5.5). Intriguingly, the mean pairwise genetic diversity appeared to follow similar dynamics as described above for viral titers. For dog 1 where data for all three days are available, the mean pairwise genetic diversity decreased on day 3 p.i. from 0.0014 to 0.0005 and then increased again to 0.0010 on day 4 (Table 5.5).

Importantly, we detected far weaker purifying selection – such that d_N/d_S ratios were close to the random expectation of 1.0 – in both the challenge virus and the naïve dogs than in our sample of EIV GenBank sequences (Table 5.5); the mean d_N/d_S ratio was 0.277 for the EIV sample and 0.694 for the challenge virus. Intriguingly, the d_N/d_S ratio increased on day 3, from 0.73 to 1.63 in dog 1 and from 1.13 to 1.41 in dog 2, and decreased again to 0.77 for dog 1 on day 4.

Neither insertions nor deletions were detected in any of the naïve dog samples or in the challenge virus, but stop codons were detected in both the challenge virus (amino acid (aa) 176 of the mature HA) and dog 1 on day 3 p.i. (aa 143 of the mature HA), suggesting that these particular strains were defective.

Table 5.5: Mutations and selection pressures detected within the challenge virus sample, in individual infected animals or among EIV GenBank sequences.

Animal	Day	No. sequences	Mean pairwise diversity	d_N/d_S	No. amino acid variants (% of total)	stop codons	InDels	Variable sites		Number of mutations present at frequency >1
								absolute number	% of total sites	
Challenge virus		88	0.0014	0.694	33 (37.5)	K191 stop	0	22	7.0	8
1	2	62	0.0017	0.723	20 (32.3)	0	0	20	6.3	4
1	3	70	0.0005	1.632	15 (21.4)	S158stop	0	10	3.2	1
1	4	72	0.0010	0.767	25 (34.7)	0	0	17	5.4	3
2	2	64	0.0012	1.131	19 (29.7)	0	0	17	5.4	3
2	3	22	0.0008	1.414	7 (31.8)	0	0	4	1.3	2
natural infection		57	0.0009	0.340	13 (22.8)	G131stop	1 (47)	7	2.2	2
Equine influenza		165		0.277	123 (74.6)	0	8	-	-	-

Intriguingly, while the challenge virus contained several sequences that shared the same mutations, these sequence clusters had nearly completely disappeared in both naïve animals by day 3 p.i. (Figures 5.1 and 5.2). In all samples, the majority of sequences were identical to the challenge virus consensus sequence, indicating that there was not a major bottleneck at transmission. Most mutant sequences were characterized by a single mutation, which was commonly only detected once (i.e. 'singeltons'; Table 5.6). Mutations present at higher frequency were generally pre-formed in the challenge virus (Table 5.7). In some cases, different mutations were detected at the same site in different samples (Table 5.8), but the location of mutations was generally not identical in different samples and mutations did not cluster in specific regions of HA1 (Figure 5.3). Overall, only 34% of mutant sites were detected more than once, which included mutations pre-formed in the challenge virus and apparently retained for some time in the infected dogs.

Of the 7 sites that were shared between different samples from infected naïve dogs, 6 (86%) were repeatedly sampled from the same animal, indicating that those mutations had probably at least temporarily remained at low level in the viral population. 8 out of 12 sites (66.7%) that were present at frequency >1 in at least one sample were also polymorphic in the EIV population, even though the amino acids present in the EIV population sometimes differed from those detected in the naïve dogs (Table 5.7). Several of these mutations were located in antigenic sites B, C,D or E, and the same was true for mutations detected in more than one sample (Table 5.8). Unfortunately, the very limited number of CIV sequences available in GenBank precluded a meaningful analysis of variable sites in the CIV population.

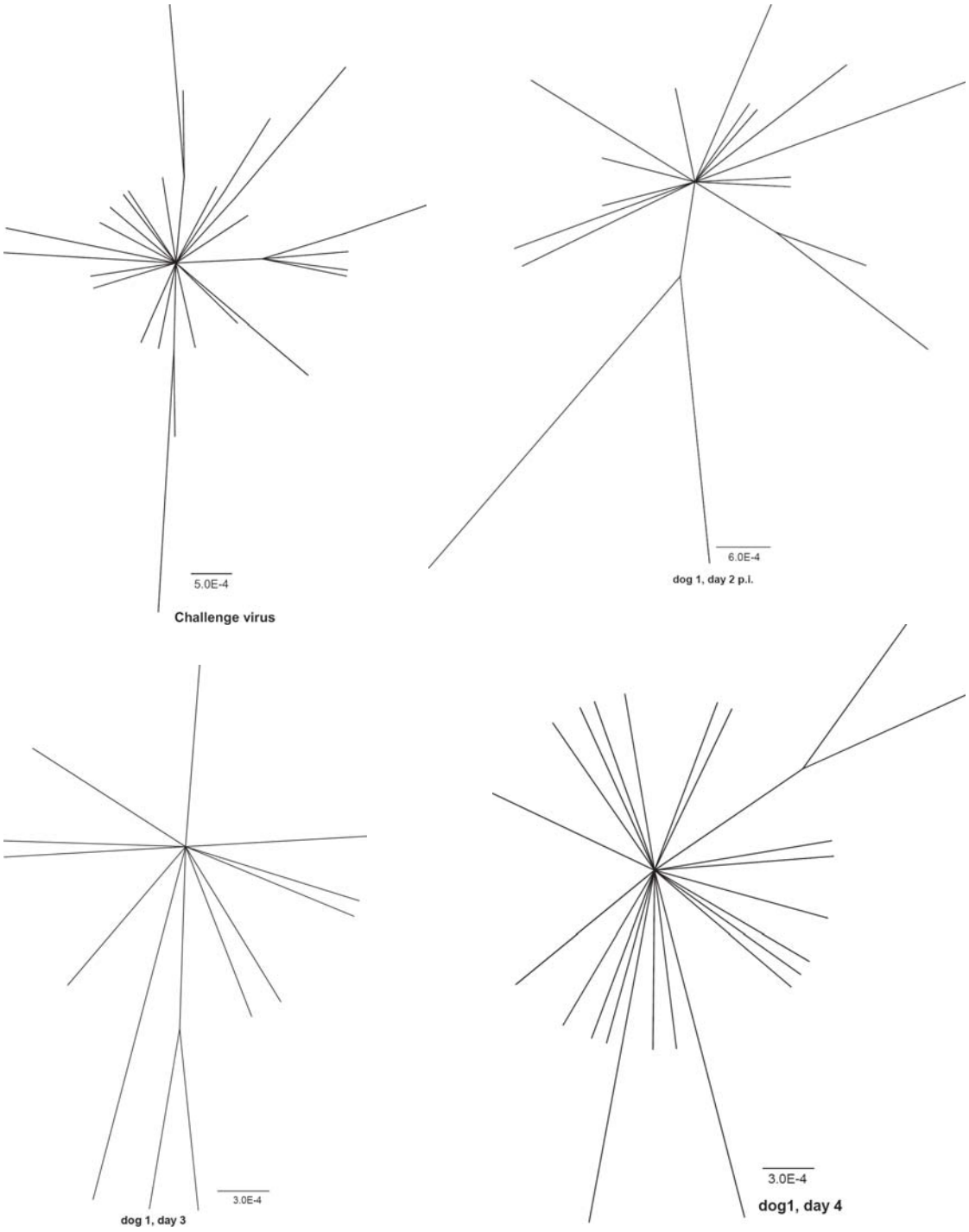


Figure 5.1. Phylogenetic relationship between CIV sequences isolated from the challenge virus or dog 1 on days 2, 3 or 4 post inoculation. Branch lengths are drawn to scale.

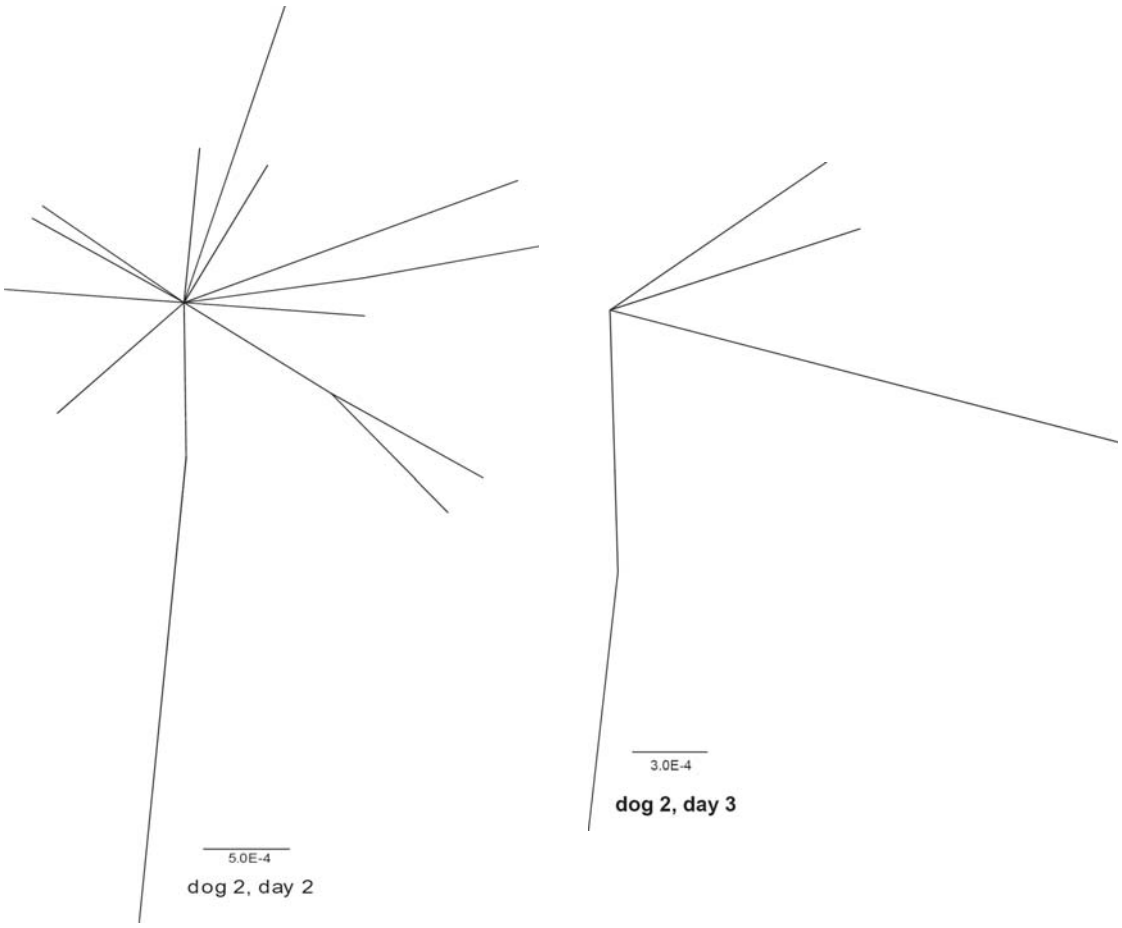


Figure 5.2. Phylogenetic relationship between CIV sequences isolated from dog 2, collected on days 2 and 3 post inoculation. Branch lengths are drawn to scale.

Table 5.6. Location of mutations detected in challenge virus and naturally infected animals. Mutations with frequency >1 are indicated in bold.

sample	Location of mutations	frequency
Challenge virus	signal peptide aa 6	1
	signal peptide aa 7	1
	30	**2**
	31	**2**
	32	1
	44	1
	77	1
	104	1
	123	1
	126	1
	155	1
	156	**2**
	165	1
	176	1
	185	1
	218	**6**
	220	**10**
	225	1
	248	**5**
	290	1
	291	**3**
	301	1
dog1, day 2	29	1
	46	1
	56	**2**
	68	1
	83	1
	120	1
	161	1
	187	1
	195	1
	212	1
	218	**7**
	220	**16**
	271	1
	281	1
	287	**2**
	288	1
	289	1
	290	1
	292	1
	301	1
dog1, day 3	80	1
	83	1

Table 5.6 (continued)

sample	Location of mutations	frequency
dog1, day 3 (continued)	143	1
	156	1
	186	1
	219	1
	220	**4**
	242	1
	253	1
	285	1
dog1, day 4	signal peptide aa 9	1
	19	**3**
	44	1
	49	1
	56	1
	76	1
	83	**2**
	88	1
	92	1
	126	1
	195	1
	220	**9**
	232	1
	242	1
	254	1
	277	1
	287	1
dog2, day 2	signal peptide aa 1	1
	32	1
	38	1
	49	1
	50	1
	83	1
	116	1
	118	**4**
	160	1
	210	1
	218	**6**
	220	**11**
	226	1
	231	1
	274	1
	296	1
	300	1
dog2, day 3	41	1

Table 5.6 (continued)

sample	Location of mutations	frequency
dog2, day 3 (continued)	83	**4**
	156	1
	220	1
A1100	15	1
	116	1
	126	**2**
	135	1
	175	1
	247	1
	299	**18**

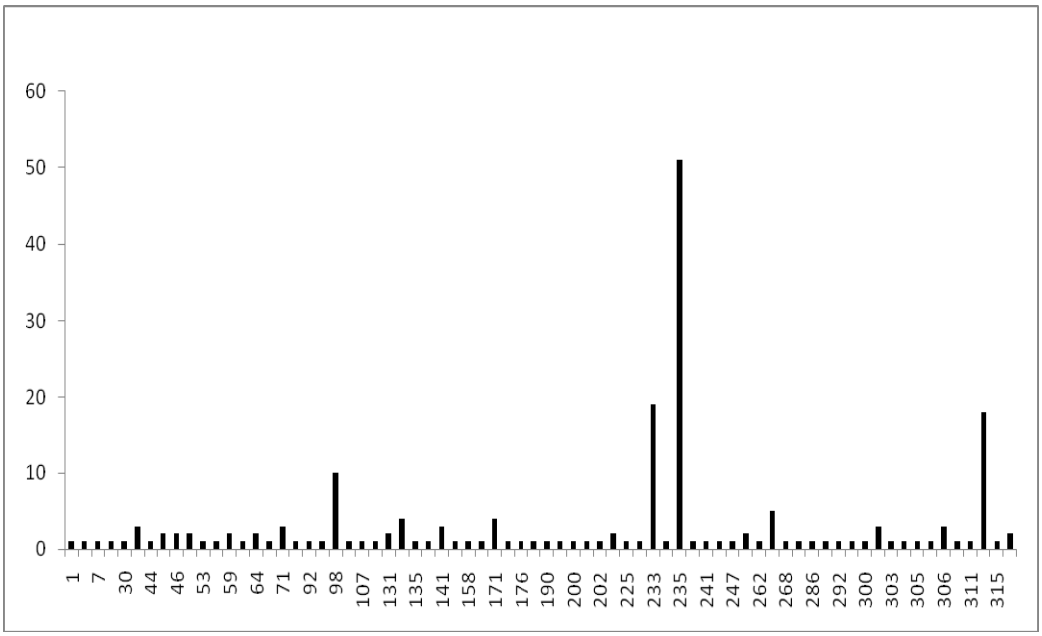


Figure 5.3 : Frequency with which mutations were detected at specific sites in HA1. Mutations are mapped onto the nascent form of the HA protein.

Table 5.7. Characterization of mutations present at frequency > 1. Mutations were mapped on the inferred structure of the mature HA1 protein. Antigenic properties were inferred based on the studies by Wilson et al. and Smith et al. (17, 20). "Cluster-difference" mutation refers to the study by Smith et al., identifying sites distinguishing antigenic clusters. Only position 136 was detected as positively selected among the EIV isolates using the SLAC algorithm (dN/dS = 6.27, p=0.028).

sample	Location of mutations	Frequency of mutations	Function and location of mutations	mutation in EIV (amino acid)	mutation in CIV (amino acid)
Challenge virus	S30N	2	-	yes (S,T)	no
	D31G	2	-	no	no
	K156E	2	site B; cluster-difference	yes (K,E)	no
	G218E	6	site D, polymorphic in CIV, 217 is cluster-difference	yes (G, R)	yes (G,E)
	R220S	10	site D	no	no
	N248T	5	site D	yes (N, K)	no
	S270S	2	clade-forming in CIV	no	no
	N290K	3	-	yes (N,S)	no
dog 1,day 2	**Y56H**	**2**	near C-site, 54 cluster-difference	yes (Y,H)	no
	G218E	7	see above	see above	see above
	R220S	16	see above	see above	see above
	S287G	**2**	**-**	**no**	no
dog 1, day 3	R220S	4	see above	see above	see above
dog 1, day 4	**A19T**	**3**	-	yes (A,H)	no
	S83N	**2**	site E; cluster-difference; clade-forming in CIV	yes (N,K,S)	no
	R220S	9	see above	see above	see above
dog 2, day 2	**V118M**	**4**	site D, clade-defining in CIV	no(L)	yes (L,V)
	E218G	6	see above	see above	see above
	R220S	11	see above	see above	see above
dog 2, day 3	**S83N**	**4**	see above	see above	see above
naturally infected	T126A	2	near site A; 124 is cluster-difference	yes (T,I,A)	no
	E299K	18	-	no (K)	no

Table 5.8: Characterization of mutations detected in more than one sample from experimentally infected dogs. Only site 218 is variable in the alignment of CIV sequences.

location of mutation detected with frequency > 1	number of samples	days of sampling	mutations	repeated in animal	function and location	variable in EIV (amino acid)
32	2	0;2	D32G	no	-	yes (D,G)
44	2	0;3	Q44H	no	site C	no
49	2	2;3	G49V	no	site C	yes (G,E)
56	2	2;4	Y56H	yes		yes (Y,H)
83	5	2;3;4	S83N	yes	site E, pos. selection	yes (N,K,S)
126	2	0;4	T126A, T126I	no		yes (T,I,A)
156	3	0;3	K156E	no	site B, pos. selection	no
218	3	0;2	G218E	no	site D	yes (G, R)
195	2	2;4	Y195C	yes	site B	no
220	5	0;2;3;4	R220S	yes	site D	no
242	2	2;4	I242S	yes	site D	yes (I,V,T)
287	2	2;4	S287G, S287C	yes	-	no
290	2	0;2	N290K	no	-	yes (N,S)
301	2	0;2	T301I; T301P	no	-	no

Many of those mutations that appeared to be generated *de novo* in the infected animals and that were sampled at frequency > 1 in at least one sample were also maintained in these dogs. Mutations Y56H and S287G, for instance, were detected in 2 sequences (3.2%) each in dog 1 on day 2, and in one sequence each on day 4 in the same dog (Table 5.6). Mutation N83S is particularly noteworthy since this mutation was present in 2 (2.8%) of the sequences sampled from dog 1 on day 4 and in 4 sequences (18.2%) sampled from dog 2 on day 3. Moreover, except for the challenge virus, mutation N83S was present at least at frequency one in all samples analyzed (Table 5.6), indicating that it was probably not pre-formed in the challenge virus but was maintained in the dogs. Of those 5 *de novo* generated mutations in that category (Y56H, S287G, A19T, S83N and V118M), 3 were also polymorphic in the EIV sample. Mutation V118M is particularly intriguing since that site does not vary in EIV and the mutation L118V appears to differentiate older from more recent CIV isolates, suggesting that it represents a host-specific adaptation (12).

5.4.2. Analysis of intra-host dynamics following natural infection.

The sample collected from a naturally infected dog showed a generally similar population structure to the samples from experimentally infected dogs described above. Again, the majority of sequences were identical to the consensus sequence while most mutant sequences differed by only one mutation (Figure 5.4). One cluster of sequences was present in the sample, characterized by a mutation at position 299 (Table 5.7). The mean pairwise genetic diversity in the sample was with 0.0009 comparable to that observed in some naïve dogs (Table 5.5). However, the mean d_N/d_S ratio in this sample was with 0.34 considerably lower than in any of the samples from naïve dogs and comparable to that observed in the EIV population level analysis,

potentially indicating stronger purifying selection in this naturally infected dog. A single deletion, in frame, was present in one of the sequences from this naturally infected dog (at aa 47 of the mature HA), and a stop codon was detected in one of the other sequences (at aa 131 of the mature HA). 7 variable sites were detected in this sample, two of which (116 and 126) were also detected in the naïve dogs. One of two mutations present at frequency > 1 (T126A) is polymorphic in the EIV sample, and was also detected in the naïve dogs (Table 5.6). Interestingly, this mutation is located in close proximity to antigenic site A (Table 5.7).

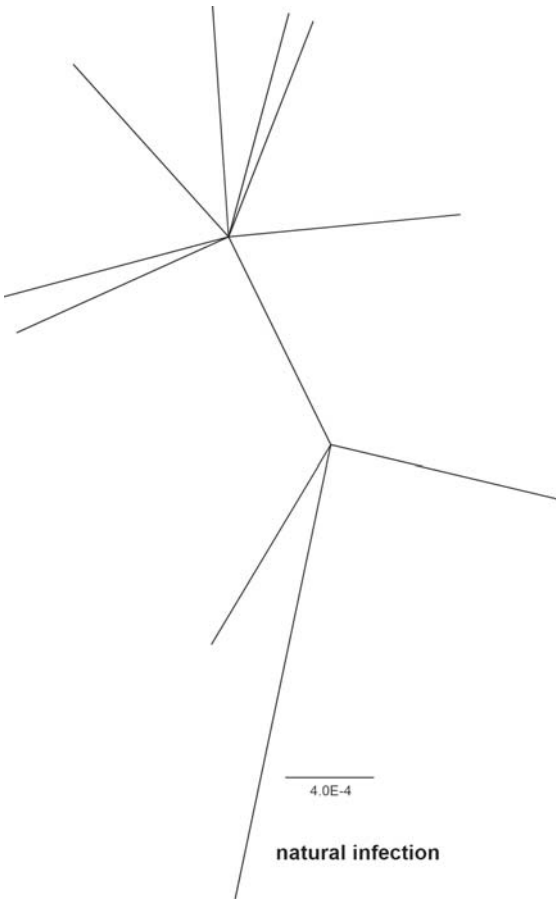


Figure 5.4. Phylogenetic relationship between CIV sequences isolated from a naturally infected animal. Branch lengths are drawn to scale.

5.5. Discussion

The emergence of CIV through transfer of a complete EIV sequence from horses to dogs represents a valuable and rare opportunity to study the evolution of a newly emerged virus in a novel host population. Here, we define the intra-host dynamics of CIV over the course of infection, contrasting the dynamics in experimentally infected animals to those in naturally infected dogs.

5.5.1. Sequence diversity and population structure in naïve animals.

In general, the evolutionary dynamics observed in the naïve dogs, characterized by a high turnover of mutations, the predominance of singletons and a consensus sequence that remains unaltered over the course of infection, are characteristic for influenza infections of mammalian hosts. The viral populations detected in both naïve animals on day 2 p.i. were fairly comparable to those detected in the challenge virus, potentially indicating a relatively small transmission bottleneck. This is further supported by the observation that many mutations detected in the infected animals were already present in the challenge virus, and therefore were likely transmitted during the inoculation and remained in the population at very low levels, even though it cannot be formally disproven that these mutations arose *de novo* in the infected dogs. The size and nature of transmission bottlenecks during natural infections are unknown for most viruses, including influenza, but artificially induced population bottlenecks have been shown to exert an important effect on the pathogenicity and evolution of RNA viruses (see for example (2, 5, 6)). The route of inoculation used in our experiment and the relatively large amount of virus used for inoculation might have inflated the number of viruses that are able to pass among hosts. However, the

spatial distribution of CIV appears patchy, and CIV is predominantly being maintained and transmitted in populations with high population densities such as animal shelters or racetracks. This might suggest a low efficiency of transmission and a large number of viruses necessary to establish infections, but further field studies are needed to address this question.

An important observation from our study was that weaker purifying selection was observation within than among hosts, manifest as inflated d_N/d_S ratios. A similar observation has been made in dengue virus (9) and suggests that the intra-host samples contain transient deleterious mutations that have yet to be removed by purifying selection. However, the absence of viruses containing stop codons and/or indels indicates that lethal mutations are purged quickly. It is also possible that the high d_N/d_S ratios contain some signal of positive selection. In particular, since the challenge virus was amplified in eggs, the inoculum virus may have been adapted to the chicken such that back-adaptation occurs in dogs. Indeed, since most of the mutations that appear to have been generated *de novo* in the naïve dogs fall at known antigenic sites, some positive selection for antigenic escape is possible. All dogs analyzed mounted strong immune responses by day 8 post inoculation, again indicating a potential role for immune selection. However, given the comparably low frequency of most mutations that were not found in the challenge virus, and since even on day 4 p.i. most virus sequences were identical to the challenge virus consensus sequence, a predominant role of immune escape appears unlikely.

The strong, temporary increase in d_N/d_S ratio and decrease in genetic diversity observed on day 3 p.i. is also compatible with the role of innate or early adaptive immune responses on that day. A temporary predominance of the immune responses

could explain the drop in viral titers and body temperature observed in the naïve infected animals on that day. These dynamics, though not well understood, appear to be common for several influenza infections in mammalian hosts and are likely an effect of the commencing immune response. The fact that two apparently de-novo generated mutations (at aa 19 and 83) appear on day 4 p.i. again lends to the idea of a second bottleneck event on day 3 p.i.

The S83N mutation is particularly intriguing since it seems to arise *de novo* in both animals (on days 3 or 4 respectively), falls in antigenic site E, and has previously been characterized as defining differences in antigenic clusters (17). This site appears to be variable in EIV (with aa N, K and S present in the EIV population) but is conserved among all CIV GenBank sequences available to date, with the mutation of N83S representing one of 5 conserved mutations that distinguish the CIV from contemporary EIV isolates (12). Since this site reverted independently in some sequences from both animals, the N83S mutation might be deleterious or slightly deleterious in these animals. Epistatic interaction or hitchhiking with another mutation elsewhere in the genome frequently complicates the evolutionary dynamics of emergence (16) and might conceivably be involved in the emergence of CIV. Moreover, since the mutation at aa 83 emerged later during the infection and appears to affect a highly antigenic site, a role of immune escape appears possible.

Mutation V118M is also noteworthy as this mutation seems to have been generated *de novo* in animal 7151 and reached considerably high frequency in this animal on day 2 but was not detected again on day 3 (although this might be due to the limited number of sequences available for this sample). Residue 118 is located in antigenic site D and a role in antigenic escape therefore appears likely. Moreover, a

L118V mutation appears to distinguish the early from more recent CIV consensus sequences (12) and the second mutation at this site therefore appears surprising.

5.5.2. Sequence diversity and population structure after natural infection.

The population structure observed in the naturally infected dog appears comparable to that observed in naïve animals and to influenza infections of mammalian hosts in general, with most mutations present as singletons, and only one or a few mutations in any sequence that differ from the consensus sequence. However, the sample was characterized by a markedly lower d_N/d_S ratio than in the experimentally infected dogs, which was comparable to the inter-host estimate obtained based on the EIV sequences. Intriguingly, this sample stemmed from a client-owned and clinically sick dog, while the experimentally infected dogs showed only mild, transient clinical signs. Experimental inoculations of healthy, well-fed animals under laboratory conditions frequently fail to produce the severe clinical signs observed in the field, and for some pathogens it has been shown that stress or starvation can enhance the likelihood of clinical disease. The clinically ill dog analyzed in this study may have differed from the experimentally infected animals with respect to immune status, elicited immune response or viral load, and unfortunately no serological information and no detailed case history is available for this animal. The fact that both Indels and stop codons were readily detected in this sample appears surprising as this would support the absence of strong purifying selection on the viruses in this animal, which appears to contradict the apparently strong purifying selection that seems to be shaping this viral population.

In conclusion, our studies of CIV intra-host diversity show that viral populations appear to be dynamic and vary over time. Mutations seem to arise readily in the infected animals, but never reach high frequency and appear to be frequently lost on subsequent days. Immune selection might be one driving force of CIV intra-host dynamics, but the intra-host dynamics appear to be mainly dominated by genetic drift, with considerably weaker purifying selection acting within a host than on the population-level.

REFERENCES

1. Antia, R., R. R. Regoes, J. C. Koella, and C. T. Bergstrom. 2003. The role of evolution in the emergence of infectious diseases. Nature **426:**658-61.

2. Bergstrom, C. T., P. McElhany, and L. A. Real. 1999. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. Proc Natl Acad Sci USA **96:**5095-100.

3. Crawford, P. C., E. J. Dubovi, W. L. Castleman, I. Stephenson, E. P. Gibbs, L. Chen, C. Smith, R. C. Hill, P. Ferro, J. Pompey, R. A. Bright, M. J. Medina, C. M. Johnson, C. W. Olsen, N. J. Cox, A. I. Klimov, J. M. Katz, and R. O. Donis. 2005. Transmission of equine influenza virus to dogs. Science **310:**482-5.

4. Cui, J., N. Han, D. Streicker, G. Li, X. Tang, Z. Shi, Z. Hu, G. Zhao, A. Fontanet, Y. Guan, L. Wang, G. Jones, H. E. Field, P. Daszak, and S. Zhang. 2007. Evolutionary relationships between bat coronaviruses and their hosts. Emerg Infect Dis **13:**1526-32.

5. Domingo, E., and J. J. Holland. 1997. RNA virus mutations and fitness for survival. Annu Rev Microbiol **51:**151-78.

6. Elena, S. F., R. Sanjuan, A. V. Borderia, and P. E. Turner. 2001. Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. Infect Genet Evol **1:**41-8.

7. Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature **397:**436-41.

8. Hoffmann, E., J. Stech, Y. Guan, R. G. Webster, and D. R. Perez. 2001. Universal primer set for the full-length amplification of all influenza A viruses. Arch Virol **146:**2275-89.

9. Holmes, E. C. 2003. Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. J Virol **77:**11296-8.

10. Kosakovsky Pond, S. L., and Frost, S.D.W. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments Bioinformatics **2005 21:**2531-2533.

11. Kuiken, T., E. C. Holmes, J. McCauley, G. F. Rimmelzwaan, C. S. Williams, and B. T. Grenfell. 2006. Host species barriers to influenza virus infections. Science **312:**394-7.

12. Payungporn, S., P. C. Crawford, T. S. Kouo, L. M. Chen, J. Pompey, W. L. Castleman, E. J. Dubovi, J. M. Katz, and R. O. Donis. 2008. Influenza A virus (H3N8) in dogs with respiratory disease, Florida. Emerg Infect Dis **14:**902-8.

13. Posada, D., Crandall, K.A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14:**817-818.

14. Reid, A. H., T. G. Fanning, J. V. Hultin, and J. K. Taubenberger. 1999. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. Proc Natl Acad Sci USA **96:**1651-6.

15. Rosas, C., G. R. Van de Walle, S. M. Metzger, K. Hoelzer, E. J. Dubovi, S. G. Kim, C. R. Parrish, and N. Osterrieder. 2008. Evaluation of a vectored equine herpesvirus type 1 (EHV-1) vaccine expressing H3 haemagglutinin in the protection of dogs against canine influenza. Vaccine **26:**2335-43.

16. Sanjuan, R., A. Moya, and S. F. Elena. 2004. The contribution of epistasis to the architecture of fitness in an RNA virus. Proc Natl Acad Sci U S A **101:**15376-9.

17. Smith, D. J., A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. Osterhaus, and R. A. Fouchier. 2004. Mapping the antigenic and genetic evolution of influenza virus. Science **305:**371-6.

18. Swofford, D. L**.** 1993. PAUP - phylogentic analysis with parsimony. Center for Biodiversity, Illinois Natural History Survey, Champaign, Ill.

19. Vana, G., and K. M. Westover. 2008. Origin of the 1918 Spanish influenza virus: a comparative genomic analysis. Mol Phylogenet Evol **47:**1100-10.

20. Wilson, I. A., J. J. Skehel, and D. C. Wiley. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature **289:**366-73.

21. Wolfe, N. D., C. P. Dunavan, and J. Diamond. 2007. Origins of major human infectious diseases. Nature **447:**279-83.

22. Yoon, K. J., V. L. Cooper, K. J. Schwartz, K. M. Harmon, W. I. Kim, B. H. Janke, J. Strohbehn, D. Butts, and J. Troutman. 2005. Influenza virus infection in racing greyhounds. Emerg Infect Dis **11:**1974-6.

CHAPTER 6

PRESENCE AND ROLE OF CYTOSINE METHYLATION IN DNA VIRUSES

From: Hoelzer, K., Shackelton, L.A. and Parrish, C.R.
Presence and role of cytosine methylation in DNA viruses.
Nucl. Acids Res. 2008. 36: 2825-2837, with permission.

6. 1. Abstract

Nucleotide composition varies greatly among DNA viruses, yet the evolutionary pressures and biological mechanisms driving these patterns are unclear. One of the most striking discrepancies lies in the frequency of CpG (the dinucleotide CG, linked by a phosphate group), which is underrepresented in most small DNA viruses (those with genomes below 10kb) but not in larger DNA viruses. Cytosine methylation might be partially responsible, but research on this topic has focused on a few virus groups. For several viruses that integrate their genome into the host genome, the methylation status during this stage has been studied extensively, and the relationship between methylation and viral-induced tumor formation has been examined carefully. However, for actively replicating viruses—particularly small DNA viruses—the methylation status of CpG motifs is rarely known and the effects on the viral life cycle are obscure. In vertebrate host genomes, most cytosines at CpG sites are methylated, which in vertebrates acts to regulate gene expression and facilitates the recognition of unmethylated, potentially pathogen-associated DNA. Here we briefly introduce cytosine methylation before reviewing what is currently known about CpG methylation in DNA viruses.

6. 2. Introduction

6.2. 1. CpG underrepresentation in vertebrate genomes.

The denotation ‘CpG’ is shorthand for the occurrence of a cytosine linked, through a phosphate bond, to a guanine. CpGs are underrepresented in most eukaryote genomes, but the frequency varies widely among species and is negatively correlated with the presence and extent of cytosine methylation in the genome (see, for example, (1,2)). In vertebrate genomes, CpGs are present at one-third to one-fourth of the expected frequency, yet the reasons are disputed (1-4). Cytosine and guanine tend to have

higher stacking energies than adenine and thymine, so structural constrains may be important in CpG avoidance (see, for example, (5)). In many species, the proportion of tRNAs containing CpG in their anticodons is lower than that of tRNAs with other dinucleotides; therefore, the transcription efficiency might be higher for codons not containing CpGs (*e.g.,* reviewed in (6)). Another explanation may lie in the fact that unmethylated CpGs can stimulate innate immune responses, potentially resulting in autoimmune reactions (reviewed in (7-9)); therefore, large numbers of CpGs may be detrimental if not all are methylated. Finally, methylated cytosines have a tendency for spontaneous deamination, which may also account for the CpG depletion (see for example (10-12)). While the deamination of unmethylated cytosines leads to uracil and can be corrected by cellular DNA repair machinery, the transition of methylated cytosines to thymines is irreversible, leading to elevated mutation frequencies in highly methylated genomes.

In general, vertebrate genomic CpGs are highly methylated. Sixty to 90% of genomic CpGs are thought to be in a methylated state (reviewed in (13,14)), but both CpG frequency and methylation patterns can vary widely across a single vertebrate genome. Notably, some regions are CpG-enriched yet practically devoid of methylation. These sequence stretches, termed 'CpG islands', are >500bp in length and comprise ~1% of total genomic DNA (*e.g.* the human genome contains more than 29,000 such islands, with estimates reaching as high as 45,000 islands per haploid genome (15,16)). The islands are often associated with 5' promoter regions of housekeeping genes (for reviews see, among others, (15,17,18)).

6.2.2. Vertebrate CpG methyltransferases.

Vertebrate genomes are methylated by DNA methyltransferases (DNMTs) which convert cytosine to 5-methylcytosine (5Me-cytosine). DNMTs are functionally divided into 'de novo' (DNMT3a and DNMT3b) and 'maintenance' (DNMT1) methyltransferases (Table 6.1). Their catalytic domains appear highly conserved across species and S-adenosyl methionine (SAM) appears to function as the only methyl donor (reviews of mammalian DNA methyltransferases can be found in (19-22)).

Table 6.1: Overview of vertebrate DNA methyltransferases and their functions. The enzymatic functions attributed to known DNMTs are summarized below, and some key references are provided. Some simplifications were made for the purpose of clarity, and the reader is referred to specific reviews of DNMTs, indicated in the text, for more detail. Cases where the functional role has been proposed, but not yet established conclusively, are indicated by question marks.

DNMT	Functional role	reference
De-novo DNMTs		
DNMT 3a	Embryonic development, methylation of CpG sites, meiosis induction in sperm (?)	(110-112)
DNMT 3b	Embryonic development, spermatogenesis (?)	(110,113)
DNMT 3L	Maternal genomic imprinting, silencing of retrotransposons in spermatogonial stem cells	(114)
Maintantance DNMTs		
DNMT 1	Cellular maintenance methylation, maintenance of imprinting, silencing of mobile elements during genomic demethylation, contribution to histone deacetylases	(115-118)
DNMT 2	Unclear. Methylation in Drosophila melanogaster (?)	(119)

DNMT expression levels in non-transformed tissues have rarely been studied explicitly. The available data indicate the translation of all three DNMTs in the majority of tissues, but translation levels seem to vary between tissues, cell differentiation levels, host developmental stages and potentially host species.

Robertson et. al (23) analyzed DNMT mRNA levels in various human tissues and found DNMT1, 3a and 3b expressed in nearly all analyzed fetal and adult tissues. In adult tissues, DNMT1 appeared to be generally more highly expressed than DNMT3a and 3b, and DNMT1 mRNA was detected in all analyzed tissues except the small intestine. DNMT3a and 3b mRNAs were detected in large quantities in heart, skeletal muscle, thymus, kidney, liver and peripheral blood mononuclear cells, but were present at low levels in all analyzed tissues. DNMT1 and 3a seemed to be expressed at high levels in all analyzed fetal tissues, while DNMT 3b was expressed at high levels only in fetal liver. The authors proposed fetal hepatic hematopoiesis as the reason for the observed high levels of DNMT3b in the fetal liver. In bovine fetuses, Golding et al. (24) found DNMT3b highly expressed in rumen, kidney, testes and lung while the highest mRNA levels were again detected in the liver. MRNAs of all three DNMTs were again detected in all of the adult bovine tissues analyzed, but highest levels were detected in kidney, brain and testes. Mizuno et al. (25) detected mRNAs for all three DNMTs in human neutrophils, monocytes, T-lymphocytes, bone marrow cells and CD34 positive immune cells, but mRNA levels varied between cell types. DNMT1 appeared to be expressed at high levels in all cell types except neutrophiles, while DNMT3a mRNA levels in neutrophiles and T lymphocytes appeared high. DNMT3b mRNA levels were low in differentiated cells such as neutrophiles, but high in bone marrow and particularly CD34 positive cells.

Methylation patterns are generally stable, sustained by DNMT1 and are inherited by both daughter DNA molecules during mitosis. However, during meiosis both parental genomes are demethylated prior to fertilization, and cellular methylation patterns become established de-novo during embryogenesis. This process usually starts during implantation and is finished by the end of gastrulation (recently reviewed in (26)). The development of site-specific methylation patterns is crucial for normal

embryonic development (see, for example, (27-29)) and is responsible for genomic imprinting and for X chromosome silencing (30,31).

6.2.3. Methylation-induced gene silencing.

CpG methylation acts to suppress transcription in several ways (reviewed, for example, in (32)). It can directly prevent the binding of transcription factors to the promoter regions of genes, most likely due to steric hindrance. Alternatively, a number of proteins, known as methyl-CpG binding (MeCP) proteins, can selectively bind methylated CpG sites. One of these, Kaiso, recognizes methylated CpG sites through a zinc-finger motif, while another subgroup shares a conserved methyl-CpG-binding domain (MBD). The MBD proteins MBD 1, 2 and MeCP2, as well as Kaiso, all function as transcriptional repressors. MBD 3, however, associates with the nucleosome remodeling and histone deacetylation (NuRD) complex, a co-repressor complex containing histone deacetylases. Deacetylation of histones increases their affinity for DNA, leading to the formation of inactive heterochromatin (reviewed in (30,33,34)).

6.2.4. Methylation-induced silencing of foreign DNA.

Methylation is involved in the inactivation of integrated foreign DNA, such as retrotransposons, proviral sequences, and other transposable elements (TEs) (33). These 'parasitic' DNAs constitute a quantitatively significant portion of mammalian genomes, comprising, for example, ~40% of the human genome. They are notably GC rich and, within mammalian genomes, are hypermethylated at CpG motifs. In several studies the inhibition of de-novo or maintenance methylation was associated with increased transcription levels of TEs, indicating that effective TE silencing requires

both de-novo and maintenance methylation, likely catalyzed by DNMT3L and DNMT1, respectively (reviewed by (35,36)).

6.2.5. Methylation and innate immunity.

CpG methylation also plays an important role in immune surveillance and the detection of pathogens, as unmethylated CpGs are a signature of bacterial and other pathogen-associated DNA. CpG-containing oligodeoxynucleotides (ODNs), even as short as six nucleotides in length, can be potent stimulators of the vertebrate immune system, resulting in the activation of a wide variety of immune cells such as B cells, NK cells, monocytes/macrophages and dendritic cells (DCs). Stimulation with CpG ODNs directly activates plasmacytoid dendritic cells (PDCs) and B cells. The PDCs secrete large amounts of IFN-α and IL-12 which subsequently stimulate monocytes, myeloid DCs, T cells and type-1 NK cells (37-39). One pathway through which unmethylated CpGs trigger innate immune responses is by binding a member of the Toll-like receptor (TLR) family, TLR9 (see, as one example, (40)). TLR9 is expressed at high levels on subsets of human and mouse DCs and gene homologues have been found in many other mammalian species. The degree to which particular CpG sequences stimulate immune responses depends upon their flanking nucleotide sequences, and motifs which confer maximal stimulation are at least partially species-specific (reviewed in (41)). For example, ODNs containing the motif GACGTT very efficiently stimulate murine and rabbit B cells, but only weakly stimulate human B cells, which are most effectively stimulated by GTCGTT (42). Some similarities in recognition sequences among related species appear likely. For example, GTCGTT motifs are highly immune-stimulatory in a variety of vertebrates including humans, dogs, cats, cows, chicken and non-human primates (41). ODNs that are thymine rich at the 3' end and contain a TpC dinucleotide at their 5' end appear to be generally

potent immune stimulators, while those that contain CpG motifs towards the 3' end appear to be less immunogenic (41). Generalities, however, may be misleading in some cases, as resulting immune effects are thought to be highly cell-type and species specific (43-45).

CpG dinucleotides in vertebrate genomes are most commonly preceded by a C and followed by a G (46), a signature which does not appear to be immune-stimulatory. Motifs such as TTCCGA, on the other hand, appear to be potent stimulators of the human immune system and may play a role in autoimmune diseases such as lupus erythematosus. Other CpG motifs, for example those containing poly(G) sequences, are thought to inhibit NF-κB induced immune stimulation (47). The relatively low occurrence of stimulatory motifs in mammalian DNA might explain the lower immunogenicity of unmethylated vertebrate DNA compared to bacterial DNA.

6.2.6. Stimulatory motifs in DNA virus genomes.

The genomes of some, but not all, viruses are also lacking many such stimulatory sequences. For example, likely stimulatory sequences are highly underrepresented in the genomes of adenovirus serotypes 2 and 5, but not in serotype 12. Adenovirus type 2 and 5 can cause persistent infections, their DNA is not immune-stimulatory and it might even suppress immune stimulation by DNA from other sources. Type 12, however, does not cause persistent infections, and its DNA appears highly immune-stimulatory. Experimental evidence for a correlation between the stimulatory potential of the viral DNA and the number of supposedly stimulatory motifs has been presented (48).

Table 6.2 summarizes the frequencies of certain putatively stimulatory and non-stimulatory motifs in genomes from several members of different viral families.

Table 6.2. Frequency of TLR9 stimulatory and non-stimulatory/inhibitory sequences. GpC and CpG contents, as well as individual nucleotide and motif frequencies, were determined from the publicly available reference sequences (RefSeq) in Genbank. (MatLab script available from the authors upon request.) Sequence length, frequencies and percentages of individual nucleotides, GpC and CpG content, as well as the frequencies of potentially stimulating and non-stimulating motifs, are provided below, along with RefSeq accession numbers. Expected values were calculated for each motif and are shown in brackets following the observed number of motifs in the sequence. The expected values were rounded to the next integer. Potentially stimulating motifs (*i.e.* GACGTT, GTCGCT, GTCGTT) were selected from motifs that Rankin *et al.* ((120)) found to be stimulatory in some domestic animals. Motifs that are most likely non-stimulatory (*i.e.* GTCGTT, CGCGCG, CCCGCG) and that might be inhibitory are those described in (41). Potentially stimulatory sequences are those described by Rankin *et al.*(120); values represent the total number of the respective motif present in the genome. The number in parentheses indicates the number of motifs expected in the sequence, given the individual nucleotide composition. This was calculated using the following formula: E(UVWXYZ)= 10^-12* sequence length*(%U*%V*%W*%X*%Y*%Z), where E(UVWXYZ)= expected value of the motif UVWXYZ; U,V,W,X,Y,Z = nucleotides of which the motif consists, and %U = % of viral sequence consisting of base U. Expected values were rounded to the nearest integer.
Likely non-stimulatory and potentially inhibitory sequences are those described by Krieg; values again represent the total number of the respective motif present in the genome, and expected values were calculated as describe above.
'Total stimulatory' or 'total non-stimulatory' values represents the sum of all putatively stimulatory or non-stimulatory motifs in the sequence.
'Ratio of stimulatory/non-stimulatory sequences' represent the fraction of the total number of stimulatory divided by the total number of non-stimulatory sequences.
'CpG' or 'GpC' represents the total number of the respective dinucleotide CpG or GpC in the viral DNA sequence.

Virus	Host	Base frequencies							Stimulatory				Non-stimulatory				Ratio stimulatory-/nonstimulatory	Sequence length (nts)	Refseq number
		A (%)	C (%)	G (%)	T (%)	CpG	GpC]	CpG/GpC	GACGTT	GTCGTT	GTCGCT	Total stimulatory	TTCCGA	GCGCGC	CCCGCG	Total non-stimulatory			
	Small ss DNA viruses																		
FPV	Cat	1816 (35.44)	835 (16.3)	1036 (20.2)	1437 (28.04)	76 (169)	167 (169)	0.455 (1)	0 (1)	0 (1)	0 (1)	0 (3)	0 (1)	2 (0)	0 (0)	2 (1)	– (3)	5124	M38246 #
CPV	Dog	1923 (35.92)	839 (15.76)	1054 (19.8)	1507 (28.3)	74 (166)	161 (166)	0.46 (1)	0 (1)	0 (1)	0 (1)	0 (3)	0 (1)	1 (0)	0 (0)	1	– (3)	5323	NC_001539
AAV1	Human	1121 (23.76)	1393 (29.52)	1273 (26.98)	931 (19.73)	316 (376)	383 (276)	0.825 (1)	1 (1)	0 (1)	0 (1)	1 (3)	0 (1)	5 (2)	1 (3)	6 (6)	0.167 (0.5)	4718	NC_002077
AAV2	Human	1198 (25.6)	1262 (26.97)	1255 (28.82)	964 (20.6)	266 (364)	304 (364)	0.875 (1)	2 (1)	0 (1)	0 (1)	2 (3)	1 (1)	4 (2)	0 (2)	5 (5)	0.400 (0.6)	4679	NC_001401
AAV4	Human	1137 (23.85)	1405 (29.47)	1333 (27.96)	892 (18.7)	325 (392)	368 (392)	0.883 (1)	0 (1)	0 (1)	0 (1)	0 (3)	1 (1)	4 (3)	1 (3)	6 (7)	– (0.429)	4767	NC_001829
AAV5	Human	1167 (25.14)	1322 (28.48)	1238 (26.67)	915 (19.7)	287 (352)	320 (352)	0.897 (1)	0 (1)	0 (1)	0 (1)	0 (3)	1 (1)	1 (2)	1 (2)	3 (5)	– (0.6)	4642	NC_006152
	Small ds DNA viruses																		
SV40	Human	1518 (28.95)	1100 (20.98)	1039 (19.82)	1586 (30.25)	27 (218)	266 (218)	0.102 (1)	0 (1)	0 (1)	0 (1)	0 (3)	0 (1)	0 (1)	0 (0)	0 (0)	– (3)	5243	NC_001669
JC polyoma-virus	human	1527 (29.77)	1035 (20.18)	1040 (20.27)	1528 (29.79)	16 (210)	255 210)	0.063 (1)	0 (1)	0 (1)	0 (1)	0 (3)	0 (1)	0 (0)	0 (0)	0 (1)	– (3)	5130	NC_001699
Murine polyoma-virus	Mouse	1396 (26.36)	1270 (23.98)	1233 (23.29)	1398 (26.4)	94 (295)	289 (295)	0.325 (1)	0 (1)	0 (1)	1 (1)	1 (3)	0 (1)	0 (1)	0 (1)	0 (3)	– (0.5)	5295	NC_001515
	Large ds DNA viruses																		
HSV1	Human	24240 (15.92)	51458 (33.79)	52513 (34.49)	24050 (15.79)	17917 (17745)	16465 (17745)	1.088 (1)	31 (24)	37 (24)	49 (52)	117 (100)	24 (24)	236 (241)	297 (236)	557 (501)	0.21 (0.1996)	152261	NC_001806
Gallid herpes-virus 1	Chicken	38272 (25.74)	35807 (24.08)	35798 (24.08)	38810 (26.1)	8666 (8622)	9401 (8622)	0.922 (1)	25 (37)	22 (37)	31 (34)	78 (108)	45 (36)	74 (29)	40 (29)	159 (94)	0.49 (1.148)	148687	NC_006623
Human adeno-virus D	Human	7988 (22.76)	9877 (28.14)	9981 (28.44)	7250 (20.66)	2410 (2808)	3034 (2808)	0.794 (1)	1 (7)	2 (7)	8 (10)	11 (24)	2 (8)	36 (18)	22 (18)	60 (40)	0.183 (0.6)	35096	NC_002067
Simian adeno-virus 1	Non-human primate	7939 (23.04)	9677 (28.09)	9343 (27.11)	7491 (21.75)	2327 (2623)	2703 (2623)	0.861 (1)	3 (8)	7 (7)	10 (10)	20 (25)	6 (8)	11 (15)	19 (16)	36 (39)	0.556 (0.64)	34450	NC_006879
Fowl adeno-virus A	chicken	10284 (23.48)	12087 (27.59)	11699 (26.71)	9734 (22.22)	3332 (3228)	3109 (3228)	1.072 (1)	11 (10)	10 (10)	9 (12)	30 (32)	15 (11)	27 (18)	19 (18)	61 (47)	0.492 (0.68)	43804	NC_001720
Fowl adeno-virus D	Chicken	10485 (23.27)	12305 (27.3)	11931 (26.48)	10342 (22.95)	3857 (3258)	3178 (3258)	1.214 (1)	10 (11)	11 (10)	11 (12)	32 (33)	17 (11)	15 (17)	32 (18)	64 (46)	0.5 (0.72)	45063	NC_000899
Frog adeno-virus	Frog	7827 (29.9)	4693 (17.94)	5210 (19.91)	8433 (32.23)	509 (1011)	1017 (1011)	0.500 (1)	2 (6)	3 (7)	3 (4)	8 (17)	0 (6)	3 (1)	0 (1)	3 (8)	2.667 (2.1)	26163	NC_002501
Frog virus 3	Frog	24269 (22.92)	28815 (27.2)	29488 (27.84)	23331 (22.03)	5633 (8020)	6342 (8020)	0.888 (1)	31 (25)	14 (24)	11 (29)	56 (78)	10 (24)	7 (46)	25 (44)	42 (114)	1.333 (0.68)	105903	NC_005946

General conclusions are difficult to draw since the biological effects of individual motifs likely differ between virus families and host species. Moreover, the limited genome size and CpG underrepresentation complicates inference for small DNA viruses. However, certain patterns appear noticeable even without using formal statistical inference. The ratio of total stimulatory to total non-stimulatory motifs is higher than expected based on the viral nucleotide composition for Frog adenovirus, Frog virus 3 and HSV1, while it is lower than expected for all other viruses for which this value can be calculated. Putatively stimulatory motifs are rarely present more frequently than expected based on the viral nucleotide composition, with the exception of Herpes simplex virus 1 (HSV1) and Frog virus 3. Putatively non-stimulatory motifs, on the other hand, are present at higher frequency than expected in several cases, as seen for several adeno- and herpesviruses. The driving forces determining this nucleotide composition are difficult to disentangle. HSV1 induces mainly subclinical, persistent infections but HSV1 DNA has been shown to efficiently stimulate TLR9 responses (49). The mammalian genomes examined contain a functional TLR9 ortholog, but chickens lack an orthologous TLR9 gene, and the same is likely true for most if not all avian species. Unfortunately, data on other avian species is thus far limited (50) and it is not known, in most cases, whether amphibian species have a functional TLR9. Several other factors likely influence presence and distribution of CpG containing motifs, but many of these remains to be understood.

6.3. DNA viruses and cytosine methylation

To understand the impact of cytosine methylation on the viral life cycle and the evolution of base composition, the particularities of each virus will need to be considered. Differences will inevitably exist between actively replicating viral DNA and that which is integrated into the host genome. The type of viral persistence will

also be of importance. The integration of adeno- or polyomavirus DNA into the host genome is usually a terminal process since the viruses cannot liberate their genomes and are therefore no longer infectious. The evolutionary roles of methylation in these cases will likely differ from that in other viruses, such as Herpesviruses, which can liberate their genome after periods of latency. But differences may also exist between large and small viruses- with many larger viruses encoding their own replication machinery and additional proteins which modify host cell processes and immune responses. The susceptibility of the viral genome to methylation and immune recognition will also be affected by other factors, such as the location of replication within the cell and the specific intracellular trafficking route (Figure 6.1).

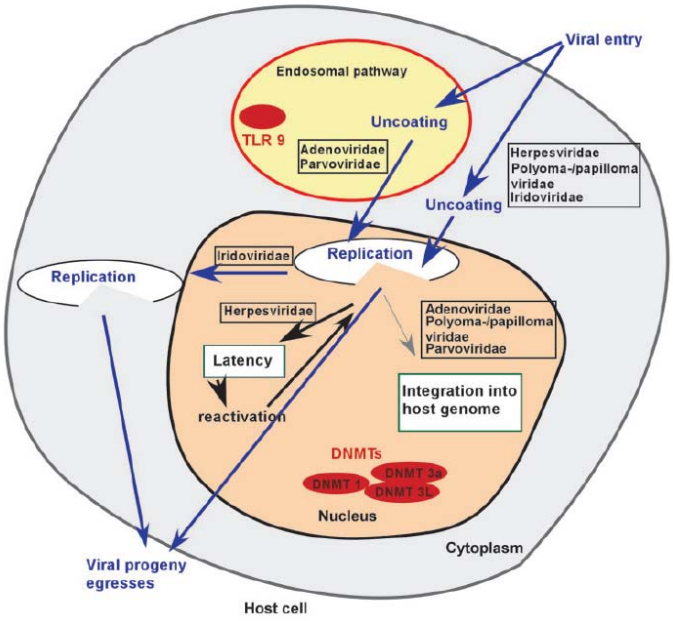


Figure 6.1: Comparison of DNA virus infection pathways. Major intracellular trafficking routes and characteristics of replication are shown for the DNA virus families discussed. The importance of methylation and immune recognition is indicated. Simplifications and generalizations were made for the purpose of clarity.

Viruses that integrate into the host genome and the effects of methylation on their life cycles have long been of particular interest. Many studies have focused on

adenoviruses, but comprehensive knowledge of the methylation status of other viruses, such as polyomaviruses and herpesviruses, has also been obtained (Table 6.3).

Early studies of viral methylation focused on DNA from polyomavirus-infected cells (51). The early methods (*e.g.* quantification of methyl-H^3 incorporation into the genome) used to measure methylation were limited in their sensitivity and did not show the genomic distribution of methylated sites. Technical advances later made it possible to study cytosine methylation in a more detailed and site-specific manner. However, the various roles of methylation during persistent viral infections, in the silencing of viral genomes, and in immune evasion and tumor formation are still being uncovered.

The impact of methylation during active viral replication is generally incompletely understood, and for many small DNA viruses it is currently not known whether the viral genome is methylated during active replication.

6.3.1. Adenovirus DNA methylation – summarizing technical advances in viral methylation studies

Early studies of adenovirus methylation relied on chromatographic and radioactive techniques to distinguish methylated from unmethylated bases. The techniques were complicated by limited sensitivity, the difficulty of isolating pure viral fractions, and the inability to discriminate between host and integrated viral DNA. The investigators nevertheless carefully compared actively replicating and integrated viral DNA and their conclusions have stood the test of time. Genomic DNA from adenovirus type 2 and 12 infected cells was heavily methylated, more than cellular DNA from uninfected control cells (52). Whether the integration of viral DNA altered methylation patterns in the host genome or whether the observed differences were due exclusively to the integrated viral DNA was not determined.

Table 6.3: Overview of GC content, CpG frequency and methylation status of small and large DNA viruses. The GC content, CpG content and, where known, methylation status is shown for viral families/subfamilies. Where applicable, a distinction is made between active replication and latency. If latent, the state of the genome (*i.e.* integrated or episomal) is specified. Preferred host species are indicated along with any known effect the virus has on host cell methylation. Where relevant, representative references are provided.

Virus	Genome size[a] (kb)	GC frequency[a]	CpG content[a] (ρCPG)	Methylation status during active replication		Methylation status during latency			Host species	Effect on host methylation
				Replicating	Reference[b]	Integrated	Episomal	Reference[b]		
Large dsDNA viruses										
Adenoviridae	28–45	0.3–0.65	0.5–1.13	Un/hypomethylated	(121)	Methylated	–	(121)	Mammal, bird	DNMT upregulation
Alpha-herpesvirinae	130–150	0.4–0.71	0.9–1.17	Un/hypomethylated	(67)	–	Un/hypo-mehtylated	(89)	Mammal, bird	DNMT upregulation
Beta-herpesvirinae	140–240	0.4–0.67	1.0–1.25	Unknown	N/A	–	Unknown	N/A	Mammal	Unknown
Gamma-herpesvirinae (122)	110–185	0.3–0.61	0.3–0.66	Un/hypomethylated	(66)	–	Methylated	(65)	Mammal	DNMT upregulation
Ranid herpesvirus[c]	220–230	0.5–0.55	0.8–0.95	Methylated	(70)	–	Unknown	N/A	Amphibian	Viral 5-cytosine methyltransferases?
Iridoviridae	140–383	0.2–0.56	0.5–0.84	Methylated	(74)	–	–	–	Amphibian, fish	Viral 5-cytosine methyltransferases?
Poxviridae	130–375	0.2–0.64	0.8–1.23	Unknown	N/A	–	–	–	Mammal bird invertebrate	Unknown
Small dsDNA viruses										
Papilloma-viridae	7–8	0.4–0.54	0.1–0.57	Partially methylated	(60)	Methylated	–	(123)	Mammal	DNMT upregulation
Polyoma-viridae	5	0.4–0.48	0.05–0.78	Un/hypomethylated	(51)	Methylated	–	(51)	Mammal, bird	DNMT upregulation
Small ssDNA viruses										
Autonomous Parvoviridae	4–6	0.3–0.5	0.3–0.71	Unknown	N/A	–	–	–	Mammal	Unknown
Dependo-virinae	4–6	0.4–0.58	0.6–1.03	Unknown	N/A	–	–	–	Mammal	Unknown
Circoviridae	2	0.5–0.57	0.4–0.87	Unknown	N/A	–	–	–	Mammal, bird	Unknown
Anellovirus	4	0.5	0.67	Unknown	N/A	–	–	–	Human	Unknown

However, actively replicating adenovirus type 2 and 12 DNA appeared to have few or no methylated bases (52).

More sensitive analysis was allowed by pairs of methylation sensitive and insensitive restriction endonucleases. In particular, the isoschizomers HpaII and MspI have been used frequently in studies of DNA methylation . Both enzymes recognize the motif CCGG, but the catalytic activity of HpaII is inhibited by CpG methylation while the activity of MspI is not affected. The new technique verified the previous results with actively replicating adenovirus type 2 and 12 DNA showing no detectable methylated CCGG sites and the restriction patterns were consistent among viral DNA from different cells (53). Inference from restriction enzyme analyses, however, remained limited to those sites recognizable by restriction enzyme pairs. Several subsequent studies focused on adenovirus DNA and the absence of methylated sites could later be conclusively established through bisulfate sequencing (51-53). Bisulfate converts unmethylated cytosines to thymines and thus allows the sensitive detection of specific methylated sites by sequencing the genome prior to and after treatment. The absence of methylated sites in the replicating adenovirus 2 genome has been shown using this technique (54-56).

Many studies of adenovirus DNA have focused on integrated viral genomes or genomic segments, largely because of their oncogenic potential. Integrated adenoviral DNA is known to be hypermethylated, but the regulatory events which determine the methylation status of replicating or integrated sequences are not well understood. A recombinant between adenovirus 12 and host cell DNA (referred to as SYREC) is unmethylated over the whole genome during active replication including the host-derived sequence, which is methylated in the chromosome. This led to speculations about the existence of adenovirus-encoded de-methylation proteins, or a role for the virus-encoded replication machinery in methylation avoidance (reviewed recently in

(57)). However, no specific mechanism has been defined, and the speed of viral replication or compartmentalization of this process within the host cell have also been proposed as factors contributing to the lack of methylation.

6.3.2. Methylation status of polyoma- and papillomaviruses

Studies of papovavirus methylation also focused on integrated sequences, mostly of human papillomavirus strains and SV40 polyomavirus. Polyomavirus and papillomavirus genomes are CpG-depleted (Table 6.3) and any CpG motifs present are clustered within certain regions of the viral genome (58), complicating the analysis of methylation patterns. The early studies, however, concluded that there was little methylation in replicating polyomaviruses (51), with the possible exception of a CpG site near the early promoter (reviewed in (59)). Several other studies concurred and found several human papillomavirus strains to be hypomethylated during active replication, again with methylated sites clustering in specific regions of the viral genome. The integrated viral DNA, on the contrary, appeared heavily methylated. However, distinguishing between actively replicating and integrated DNA appears difficult in some cases (60-62).

6.3.3. Methylation status of herpesvirus genomes

Herpesviruses establish latent infections without integrating into the host genome, and reactivate to reestablish active replication. During latency, the circularized viral genome remains quiescent in an episomal state, and can be replicated by the host cell replication machinery. Active (*e.g.* lytic) viral replication differs in several respects. Production of viral progeny inevitably results in cell lysis, utilizes the viral-encoded replication machinery and involves a separate origin of replication. (*e.g.* in the case of Epstein-Barr Virus (EBV) *oriLyt* functions as the origin of replication in the lytic

cycle while *oriP* is responsible for replication of episomal EBV DNA (63)) The methylation status during lytic DNA replication is likely different from that during latency, but most studies have focused on latent infections and the differences have rarely been studied explicitly.

Differences might also exist among herpesvirus subfamilies, as CpG motifs are highly underrepresented in gammaherpesvirus genomes, but not in alpha- or betaherpesvirus genomes (64). The reasons for these differences remain elusive. Gammaherpesviruses cause predominantly persistent, lymphoproliferative diseases. Studies of some gammaherpesviruses, such as EBV and Kaposi's sarcoma-associated herpesvirus (KSHV), indicate that methylation serves an intricate regulatory role during the viral life cycle (see below). The EBV genome appears to be hypomethylated or unmethylated during lytic infection, but highly methylated during latency, and appears to become demethylated during reactivation (65). There is some evidence for methylation during active replication of another gammaherpesvirus, herpesvirus saimiri, since Kaschka-Dierich *et al.* (1981) found both the linear and circular viral DNA to be heavily methylated. (66). The alphaherpesvirus herpes simplex virus (HSV) appears unmethylated during active replication and latency (67-69) while two frog herpesviruses (ranid herpesvirus 1 and 2) appear to be methylated during replication. The ranid herpesvirus genomes also contain putative DNA cytosine-5 methyltransferases sequences located in ORFs 86 and 120 (70).

Methylation plays a pivotal role in the life cycle of EBV (reviewed in (71,72)), where methylation of specific genes appears to be involved in the transition from lytic to latent infection. Most EBV nuclear antigens (EBNAs) are expressed only for short periods during the lytic infection, after which their transcription is silenced through hypermethylation of the viral promoters Cp and Wp. Production of the indispensable EBNA1 is then achieved by expression from an alternative "TATA-less" promoter

Qp. Reinitiation of lytic infection can occur spontaneously or may be triggered by events such as immunoglobulin crosslinking of the host cell and cell proliferation. Reactivation is mediated through the expression of the immediate early genes Zta and Rta, which are regulated by the promoters Zp and Rp, the latter of which is hypermethylated during latency. While Rp hypermethylation is a major regulator of latent infection, Zp hypermethylation is dispensable. EBV therefore seems to use methylation-induced gene silencing as an immune evasion strategy. While a role for methylation in the EBV lifecycle is widely accepted, its importance for other herpesviruses is still disputed.

6.3.4. Methylation status of iridiviruses and ascoviruses

The genomes of iridoviruses appear to be heavily methylated during active replication and methylation might again have a regulatory role. Iridoviruses are large DNA viruses that infect fish, amphibians and reptiles. Initial stages of replication occur in the nucleus, after which the nascent viral genomes are transported into the cytoplasm where the second stage of replication occurs (reviewed by (73)). Willis and Granoff (74) analyzed the iridovirus frog virus 3 (FV3) by radioactive labeling/restriction enzyme digestion and found the viral DNA to be heavily methylated, with an estimated 20% of cytosines methylated. Time-course experiments later indicated changing methylation patterns over the course of infection. Willis *et al.* (75) analyzed nuclear and cytoplasmic extracts collected at various times post-infection (p.i.) and found methylated viral DNA after ~6-7h p.i. This DNA appeared to be located in the cytoplasm, while nuclear DNA, collected at earlier time points, appeared to be hypomethylated. The authors also provided evidence against demethylation of parental viral genomes upon host cell infection. Further, Schetter *et al.* (76) showed that nascent FV3 DNA in the nucleus is unmethylated or hypomethylated at early times p.i,

but later becomes hypermethylated (after ~6h p.i. in their assay). Cytoplasmic viral DNA, on the other hand, appears to be methylated at all times. The authors reported DNA methyltransferase activity in nuclear extracts of FV3 infected cells, beginning at a time p.i. which coincided with the appearance of methylated viral genomes. This activity was absent from cytoplasmic extracts, indicating FV3 might regulate methylation of its genome. It was later discovered that the FV3 genome encodes a putative DNA cytosine-5 methyltransferase sequence, which bears similarities to other known methyltransferases (77). The putative methyltransferase, however, is more homologous to prokaryotic enzymes than to eukaryotic methyltransferases. While transcription (from an early promoter) and translation of this FV3 protein have been verified, direct evidence for its methylation activity is still limited.

Another iridovirus, fish lymphocystis disease virus (FLDV), was also found to have a highly methylated genome, even though its GC content is considerably lower than that of FV3 (78,79). FLDV also contains a putative 5-cytosine methyltransferase (53.3% identical to that of FV3) (80) but evidence for its function is, thus far, scarce. Besides the frog herpesviruses and iridoviruses, ascoviruses, which infect insects and are closely related to the iridoviruses, are also heavily methylated during replication (81).

6.3.5. Methylation status of small DNA viruses

In contrast to most large DNA viruses, CpG is underrepresented in the majority of small DNA viruses (64,82,83). It is possible that cytosine methylation is partly responsible for this underrepresentation. Small DNA viruses encode only a minimal number of proteins and use the host cell replication machinery for replication. Low CpG frequencies may have been selected to avoid methylation by host methyltransferases, to maximize translation efficiencies, or as a means to reducing

CpG mediated immune responses. However, apart from the small polyomaviruses and papillomaviruses described above, there appears to have been few close examinations of the methylation status of small DNA viruses (particularly those that do not integrate), such as autonomous parvoviruses, circoviruses and anelloviruses. The roles and effects of methylation might be different from those seen in integrating and/or large DNA viruses, but further research is needed to understand the specific effects of methylation on small, non-integrating DNA viruses.

Adeno-associated parvovirus (AAV) integrates into the host genome, and has been examined primarily in the context of guanine methylation. Productive infection with an AAV normally requires co-infection with a helper virus (usually an adenovirus or, less frequently, a herpesvirus). In the absence of a helper virus, AAVs may remain in the cell as persistent dsDNA, or they may be stably integrated into the host genome. Upon infection by a helpervirus, the integrated AAV DNA can be rescued and undergo active replication. The CpG contents of AAVs more closely resemble those of their helper viruses than those of the closely related autonomous parvoviruses, which show strong CpG suppression. The mechanisms of integration have been extensively studied in AAV serotype 2 (AAV2), which predominantly integrates in a specific region of chromosome 19 (locus 19q13.3t). Integration is directed by AAV-encoded Rep proteins, which likely recognize guanine residues within GCTC repeating motifs (located in the AAV terminal hairpin). In vitro methylation of these guanine residues has been shown to effectively inhibit DNA-protein interactions and likely interferes with integration in vivo (84). Similar results were obtained when examining the integration of AAV4 (which utilizes a slightly different recognition motif) into African green monkey cells (85). Nevertheless, it is not known what role, if any, cytosine-5 methylation plays in the natural AAV life

cycle, if AAV methylation influences integration, or if guanines or cytosines are methylated during active replication.

6.4. Methylation and integrated or latent viral DNA

Methylation of integrated viral DNA has been studied extensively and the topic is extensively covered in the scientific literature. Areas of interest have included changes in methylation patterns during integration, the effects of viral methylation on host cells and the correlation between virally-induced methylation and tumor formation. Excellent reviews have been published on these topics and the reader is referred to these articles, indicated in the following discussion, for additional detail. Here we briefly summarize what is known about methylation during integration and latency and the regulatory elements involved, contrasting the level of understanding to what is known about methylation during active replication.

As already mentioned above, most DNA viruses such as adenoviruses (57,59,86-88) and polyomaviruses (83) appear specifically methylated during latency or when in an integrated form. In contrast, the alphaherpesvirus HSV does not appear to be methylated during latency (89), but the reasons for this difference are unclear. The processes by which integrating viruses are de-novo methylated are only partly understood, but it is clear that the extent and pattern of de-novo methylation depend upon the time after integration and site in the host genome (90), the integrated sequences (90,91,92 ,93,94), and possibly the cell type (90); (reviewed in (57)). De-novo methylation tends to be initiated within confined genomic regions and then spreads throughout the viral genome (90,93,95), (reviewed in (87)). For example, in a study examining the methylation pattern of adenovirus 12 over the course of integration into hamster cells, methylation began in the center of the viral DNA

(between map units 30-75) and subsequently spread outward in both directions. Some regions of the viral genome remained unmethylated throughout the study, namely the right terminal repeat and regions near the left terminus (95). It is not known which factors initiate methylation of the genome, nor the determinants of the sequential methylation pattern.

Methylation patterns appear to be site-specific. In many viruses there seems to be an inverse correlation between viral gene expression and degree of methylation, with late viral genes being generally more susceptible to methylation effects than early genes (see, for example, (53,91)). Promoter regions of integrated viruses are frequently hypomethylated. For example, promoter regions in adenovirus 12 are hypomethylated at CCGG sites while the viral genome is otherwise heavily methylated (reviewed in (57,59,96)). Methylation of viral genes, particularly at promoter sequences, appears to be a reversible event (reviewed in (96)) but the mechanism by which viral genomes are demethylated remains to be determined. Demethylation has also been shown to occur after in vitro methylation of several viral DNAs. For example, SV40 DNA recovered 20h after microinjection into cells retained the artificial methylation pattern under non-permissive culture conditions, while allowing early gene transcription. The DNA, however, became demethylated after viral replication (97). Loss of methylation has also been observed in adenoviruses, in the case of the SYRECs mentioned above (86,98) which are methylated when covalently linked to the host genome, but unmethylated during active replication. They require the presence of fully functional adenovirus helper for efficient lytic replication, which led some to postulate a role for the viral transcription machinery in the loss of methylation. Unfortunately experimental data in support of this hypothesis is so far scarce.

Infection with integrating (and even some non-integrating) viruses can induce alterations in the methylation patterns of host DNA, a potential factor in the development of malignant tumors. For example, herpesviruses KSHV and EBV, polyomaviruses BK and SV40, papillomavirus HPV, several adenoviruses and hepatitis B virus encode proteins that activate/upregulate DNMT1, 3a, or 3b. This results in the hypermethylation and, therefore, the downregulation of a number of cellular genes. One of the cellular genes frequently hypermethylated in these infections is the cell cycle regulating tumor suppressor gene p16INK4a, which is commonly hypermethylated in many types of cancer (reviewed in (99)). Differentiation of host cell tumors and the level of genome methylation of integrated viruses generally appear to be positively correlated, as seen for various papillomaviruses such as Shope papillomavirus (100) and human papillomavirus (HPV) (101,102), the herpesvirus EBV (reviewed in (72,103)), and various adenoviruses (reviewed in (96)).

6.5. Methylation and viral gene expression

For iridoviruses, gene expression at functional levels can be maintained despite genome hypermethylation. Methylated sites, however, are not distributed randomly across the genome and methylation levels differ among CpG motifs. FLDV, for example, has a considerably higher level of methylation at CCGG than CGCG motifs (79). The physiological significance of this, however, is still disputed. In vitro studies of the related FV3 showed that specific methylation (using HpaII methyltransferase) of CCGG motifs in a late promoter, L1140, abolished its function. However, the methylation of all CpG sites by the indiscriminate SssI methyltransferase or the specific methylation of GCGC motifs by HhaI methyltransferase had no dramatic effect. This appears to be a rare example where, counterintuitively, complete

methylation of a late promoter does not abolish its function (methylation frequently abolishes the function of late viral promoters, while early promoters tend to be less susceptible to the repressive effect of cytosine methylation). The underlying mechanism has not been elucidated, but steric hindrance in asymmetrically methylated DNA might play a role (104). It seems possible that methylation of CCGG motifs results in complex secondary structures in the promoter region, whereas indiscriminant CpG methylation or methylation of GCGC motifs balances these secondary structures. A similar phenomenon was seen in adenovirus type 2, where methylation of CCGG sites within the (late) E2a promoter inhibited the transcription of viral DNA, while methylation of GCGC motifs had no repressive effect (105,106). The transcription activity of the polyomavirus SV40 was found not drastically decreased when in vitro methylated with indiscriminant DNMTs from rat liver (107,108). In contrast, upon specific methylation of HpaII sites (CCGG-specific), transcription activity was reduced markedly (reviewed in (59)). However, the interpretation of these latter results is complicated by the possibility of incomplete methylation by the rat liver extract, as rat liver DNA is a competitive inhibitor of adenovirus methylation (55) and low level contaminants might likewise have inhibited complete methylation of SV40.

In vitro methylation studies involving human papillomavirus type 16 (HPV16) showed that indiscriminate SssI methylation significantly decreased the transcriptional activity of the long control region (LCR), which is devoid of ORFs but contains several *cis*-acting regulatory elements. Enhancers located in the LCR are responsive to both cellular and viral factors, such as the viral E2 gene products. Transcriptional silencing was measured using a LCR-containing reporter plasmid and inhibition was again believed to be steric, specifically due to interference with the binding of transcription regulator E2. Notably, hypomethylation of HPV16 E2 binding sites has been observed in highly differentiated, but not in undifferentiated, tumor cells (101).

However, direct comparison of the HPV16 and SV40 studies is difficult. The effects of indiscriminate methylation on early and late viral gene products in the genomic context are hard to infer from this study.

The time of transcription appears to impact the effects of methylation in several viruses. For example, the in vitro HpaII (CCGG-specific) methylation of SV40 early genes (*i.e.* those coding for the large T-antigen) does not lead to transcriptional repression, and the in vitro methylated sites are subsequently lost during active replication (108). Late viral gene expression, however, can be efficiently inhibited by in vitro methylation of HpaII sites within the 5' part of this region (109). Site-specific methylation might be involved in the transition to latency in some cases (as in the case of EBV). On the other hand, methylation does not seem to be involved in the switch from early to late gene expression in HSV (89). Likewise, the major late promoter of adenovirus 2, which induces the switch from early to late gene expression, is unmethylated in non-integrated viruses (56), suggesting methylation does not play a major role in the regulation of early versus late gene expression for this virus.

6.6. Concluding Summary

A number of studies show that CpG methylation can greatly affect the life cycles of DNA viruses, but its exact role in natural infections remains unclear. Those effects likely differ between different viruses and are dependent upon many factors, such as the stage in the viral life cycle, the host species and infected tissue, the flanking nucleotide motifs, the genes in question and the genomic location. Where it has been studied, major differences have been seen for some viruses between the actively replicating and the latent or integrated viral DNA; the former is often unmethylated or hypomethylated, while the later often shows specific and regulated methylation.

Differences may also exist between large and small DNA viruses, potentially because of their rate and site of replication and differences in their ability to manipulate the host responses to unmethylated DNA. Thus, it is necessary to not only examine individual viruses, but to examine their dynamic methylation status and the effects of any viral methylation on both the virus and the host cell. The various ways in which mutational bias, codon usage and host selections (*e.g.* through TLR9-dependent responses) influence CpG frequency and methylation status need to be defined. It will be especially important to determine the unique roles and regulators of methylation, and their divergent effects on the genome evolution of small and large DNA viruses during active or latent infection, and in packaged, integrated or episomal DNA. A deeper knowledge of the relationships among the virus, the host cell and methylation (and de-methylation) machinery will provide essential insights into determinants of methylation status, gene expression, replicative behavior, as well as activation of pattern recognition receptors and immune responses.

REFERENCES

1. Bird, A.P. (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.*, **8**, 1499-1504.

2. Schorderet, D.F. and Gartler, S.M. (1992) Analysis of CpG suppression in methylated and nonmethylated species. *Proc. Natl. Acad. Sci. USA*, **89**, 957-961.

3. Swartz, M.N., Trautner, T.A. and Kornberg, A. (1962) Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J Biol. Chem.*, **237**, 1961-1967.

4. Burge, C., Campbell, A.M. and Karlin, S. (1992) Over- and under-representation of short oligonucleotides in DNA sequences. *Proc. Natl. Acad. Sci. USA*, **89**, 1358-1362.

5. Sponer, J., Gabb, H.A., Leszczynski, J. and Hobza, P. (1997) Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study. *Biophy. J.*, **73**, 76-87.

6. Andersson, S.G. and Kurland, C.G. (1990) Codon preferences in free-living microorganisms. *Microbiological reviews*, **54**, 198-210.

7. Krieg, A.M. (2002) A role for Toll in autoimmunity. *Nature immunology*, **3**, 423-424.

8. Bachmann, M.F. and Kopf, M. (2001) On the role of the innate immunity in autoimmune disease. *J. Exp. Med.*, **193**, F47-50.

9. Janeway, C.A., Jr. and Medzhitov, R. (2002) Innate immune recognition. *Ann. Rev. Immunol.*, **20**, 197-216.

10. Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978) Molecular basis of base substitution hotspots in Escherichia coli. *Nature*, **274**, 775-780.

11. Sved, J. and Bird, A. (1990) The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc. Natl. Acad. Sci. U S A*, **87**, 4692-4696.

12. Fryxell, K.J. and Zuckerkandl, E. (2000) Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol. Biol. Evol.* , **17**, 1371-1383.

13. Kass, S.U., Pruss, D. and Wolffe, A.P. (1997) How does DNA methylation repress transcription? *Trends Genet*, **13**, 444-449.

14. Ng, H.H. and Bird, A. (1999) DNA methylation and chromatin modification. *Curr. Opin. Geneti. Dev.*, **9**, 158-163.

15. Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA*, **90**, 11995-11999.

16. McClelland, M. and Ivarie, R. (1982) Asymmetrical distribution of CpG in an 'average' mammalian gene. *Nucleic Acids Res.*, **10**, 7865-7877.

17. Takai, D. and Jones, P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA*, **99**, 3740-3745.

18. Caiafa, P. and Zampieri, M. (2005) DNA methylation and chromatin structure: the puzzling CpG islands. *J. Cell. Biochem.*, **94**, 257-265.

19. Pradhan, S. and Esteve, P.O. (2003) Mammalian DNA (cytosine-5) methyltransferases and their expression. *Clini. Immunol. (Orlando, Florida)*, **109**, 6-16.

20. Hermann, A., Gowher, H. and Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell. Mol. Life Sci.*, **61**, 2571-2587.

21. Bird, A.P. and Wolffe, A.P. (1999) Methylation-induced repression--belts, braces, and chromatin. *Cell*, **99**, 451-454.

22. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6-21.

23. Robertson, K.D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F.A. and Jones, P.A. (1999) The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res.*, **27**, 2291-2298.

24. Golding, M.C. and Westhusin, M.E. (2003) Analysis of DNA (cytosine 5) methyltransferase mRNA sequence and expression in bovine preimplantation embryos, fetal and adult tissues. *Gene Expression Patterns*, **3**, 551-558.

25. Mizuno, S., Chijiwa, T., Okamura, T., Akashi, K., Fukumaki, Y., Niho, Y. and Sasaki, H. (2001) Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia. *Blood*, **97**, 1172-1179.

26. Salozhin, S.V., Prokhorchuk, E.B. and Georgiev, G.P. (2005) Methylation of DNA--one of the major epigenetic markers. *Biochemistry*, **70**, 525-532.

27. Reik, W., Dean, W. and Walter, J. (2001) Epigenetic reprogramming in mammalian development. *Science (Washington DC, U.S.)*, **293**, 1089-1093.

28. Fulka, H., Mrazek, M., Tepla, O. and Fulka, J., Jr. (2004) DNA methylation pattern in human zygotes and developing embryos. *Reproduction (Cambridge, England)*, **128**, 703-708.

29. Aranyi, T. and Paldi, A. (2006) The constant variation: DNA methylation changes during preimplantation development. *FEBS L.*, **580**, 6521-6526.

30. Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33 Suppl**, 245-254.

31. Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Ann. Rev. Biochem.*, **74**, 481-514.

32. Attwood, J.T., Yung, R.L. and Richardson, B.C. (2002) DNA methylation and the regulation of gene transcription. *Cell. Mol. Life Sci.*, **59**, 241-257.

33. Singal, R. and Ginder, G.D. (1999) DNA methylation. *Blood*, **93**, 4059-4070.

34. Hendrich, B. and Tweedie, S. (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.*, **19**, 269-277.

35. Slotkin, R.K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.*, **8**, 272-285.

36. Robertson, K.D. and Wolffe, A.P. (2000) DNA methylation in health and disease. *Nat. Rev. Genet.*, **1**, 11-19.

37. Rothenfusser, S., Hornung, V., Krug, A., Towarowski, A., Krieg, A.M., Endres, S. and Hartmann, G. (2001) Distinct CpG oligonucleotide sequences activate human gamma delta T cells via interferon-alpha/-beta. *Eur. J. Immunol.*, **31**, 3525-3534.

38. Kojima, Y., Xin, K.-Q., Ooki, T., Hamajima, K., Oikawa, T., Shinoda, K., Ozaki, T., Hoshino, Y., Jounai, N., Nakazawa, M. *et al.* (2002) Adjuvant effect of multi-CpG motifs on an HIV-1 DNA vaccine. *Vaccine*, **20**, 2857-2865.

39. Krieg, A.M., Yi, A.K., Matson, S., Waldschmidt, T.J., Bishop, G.A., Teasdale, R., Koretzky, G.A. and Klinman, D.M. (1995) CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature*, **374**, 546-549.

40. Krug, A., Towarowski, A., Britsch, S., Rothenfusser, S., Hornung, V., Bals, R., Giese, T., Engelmann, H., Endres, S., Krieg, A.M. *et al.* (2001) Toll-like receptor expression reveals CpG DNA as a unique microbial stimulus for plasmacytoid dendritic cells which synergizes with CD40 ligand to induce high amounts of IL-12. *Eur. J. Immunol*, **31**, 3026-3037.

41. Krieg, A.M. (2002) CpG motifs in bacterial DNA and their immune effects. *Ann. Rev. Immunol.*, **20**, 709-760.

42. Hartmann, G. and Krieg, A.M. (2000) Mechanism and function of a newly identified CpG DNA motif in human primary B cells. *J. Immunol.*, **164**, 944-953.

43. Mutwiri, G., Pontarollo, R., Babiuk, S., Griebel, P., van Drunen Littel-van den Hurk, S., Mena, A., Tsang, C., Alcon, V., Nichani, A., Ioannou, X. *et al.* (2003) Biological activity of immunostimulatory CpG DNA motifs in domestic animals. *Veterinary immunology and immunopathology*, **91**, 89-103.

44. Iho, S., Yamamoto, T., Takahashi, T. and Yamamoto, S. (1999) Oligodeoxynucleotides containing palindrome sequences with internal 5'-CpG-3' act directly on human NK and activated T cells to induce IFN-gamma production in vitro. *J. Immunol.*, **163**, 3642-3652.

45. Kanellos, T.S., Sylvester, I.D., Butler, V.L., Ambali, A.G., Partidos, C.D., Hamblin, A.S. and Russell, P.H. (1999) Mammalian granulocyte-macrophage colony-stimulating factor and some CpG motifs have an effect on the immunogenicity of DNA and subunit vaccines in fish. *Immunology*, **96**, 507-510.

46. Han, J., Hsu, C., Zhu, Z., Longshore, J.W. and Finley, W.H. (1994) Over-representation of the disease associated (CAG) and (CGG) repeats in the human genome. *Nucleic Acids Res.*, **22**, 1735-1740.

47. Lenert, P., Stunz, L., Yi, A.K., Krieg, A.M. and Ashman, R.F. (2001) CpG stimulation of primary mouse B cells is blocked by inhibitory oligodeoxyribonucleotides at a site proximal to NF-kappaB activation. *Antisense & nucleic acid drug development*, **11**, 247-256.

48. Krieg, A.M., Wu, T., Weeratna, R., Efler, S.M., Love-Homan, L., Yang, L., Yi, A.K., Short, D. and Davis, H.L. (1998) Sequence motifs in adenoviral

DNA block immune activation by stimulatory CpG motifs. *Proc. Natl. Acad. Sci. USA*, **95**, 12631-12636.

49. Krug, A., Luker, G.D., Barchet, W., Leib, D.A., Akira, S. and Colonna, M. (2004) Herpes simplex virus type 1 activates murine natural interferon-producing cells through toll-like receptor 9. *Blood*, **103**, 1433-1437.

50. Roach, J.C., Glusman, G., Rowen, L., Kaur, A., Purcell, M.K., Smith, K.D., Hood, L.E. and Aderem, A. (2005) The evolution of vertebrate Toll-like receptors. *Proc. Natl. Acad. Sci. USA*, **102**, 9577-9582.

51. Winocour, E., Kaye, A.M. and Stollar, V. (1965) Synthesis and transmethylation of DNA in polyoma-infected cultures. *Virology*, **27**, 156-169.

52. Gunthert, U., Schweiger, M., Stupp, M. and Doerfler, W. (1976) DNA Methylation in Adenovirus, Adenovirus-Transformed Cells, and Host Cells. *Proc. Natl. Acad. Sci. USA*, **73**, 3923-3927.

53. Vardimon, L., Neumann, R., Kuhlmann, I., Sutter, D. and Doerfler, W. (1980) DNA methylation and viral gene expression in adenovirus-transformed and -infected cells. *Nucleic Acids Res.*, **8**, 2461-2473.

54. Wienhues, U. and Doerfler, W. (1985) Lack of evidence for methylation of parental and newly synthesized adenovirus type 2 DNA in productive infections. *J. Virol.*, **56**, 320-324.

55. von Acken, U., Simon, D., Grunert, F., Doring, H.P. and Kroger, H. (1979) Methylation of viral DNA in vivo and in vitro. *Virology*, **99**, 152-157.

56. Kammer, C. and Doerfler, W. (1995) Genomic sequencing reveals absence of DNA methylation in the major late promoter of adenovirus type 2 DNA in the virion and in productively infected cells. *FEBS L.*, **362**, 301-305.

57. Doerfler, W. (2005) On the biological significance of DNA methylation. *Biochemistry*, **70**, 505-524.

58. Soeda, E., Arrand, J.R., Smolar, N. and Griffin, B.E. (1979) Sequence from early region of polyoma virus DNA containing viral replication origin and encoding small, middle and (part of) large T antigens. *Cell*, **17**, 357-370.

59. Doerfler, W. (1981) DNA methylation--a regulatory signal in eukaryotic gene expression. *The J. Gen. Virol.*, **57**, 1-20.

60. Danos, O., Katinka, M. and Yaniv, M. (1980) Molecular cloning, refined physical map and heterogeneity of methylation sites of papilloma virus type 1a DNA. *Eur. J. Biochem. / FEBS*, **109**, 457-461.

61. Danos, O., Katinka, M. and Yaniv, M. (1982) Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. *The EMBO journal*, **1**, 231-236.

62. Burnett, T.S. and Sleeman, J.P. (1984) Uneven distribution of methylation sites within the human papillomavirus la genome: possible relevance to viral gene expression. *Nucleic Acids Res.*, **12**, 8847-8860.

63. Hammerschmidt, W. and Sugden, B. (1988) Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein-Barr virus. *Cell*, **55**, 427-433.

64. Takacs, M., Segesdi, J., Balog, K., Mezei, M., Toth, G. and Minarovits, J. (2001) Relative deficiency in CpG dinucleotides is a widespread but not unique feature of Gammaherpesvirinae genomes. *Acta microbiologica et immunologica Hungarica*, **48**, 349-357.

65. Szyf, M., Eliasson, L., Mann, V., Klein, G. and Razin, A. (1985) Cellular and viral DNA hypomethylation associated with induction of Epstein-Barr virus lytic cycle. *Proc. Natl. Acad. Sci. USA*, **82**, 8090-8094.

66. Kaschka-Dierich, C., Bauer, I., Fleckenstein, B. and Desrosiers, R.C. (1981) Episomal and nonepisomal herpesvirus DNA in lymphoid tumor cell lines. *Haematology and blood transfusion*, **26**, 197-203.

67. Low, M., Hay, J. and Keir, H.M. (1969) DNA of herpes simplex virus is not a substrate for methylation in vivo. *J. Mol. Biol.*, **46**, 205-207.

68. Lundberg, P., Welander, P., Han, X. and Cantin, E. (2003) Herpes simplex virus type 1 DNA is immunostimulatory in vitro and in vivo. *J. Virol.*, **77**, 11158-11169.

69. Kubat, N.J., Tran, R K, McAnay, P and Bloom, D C. (2004) Specific histone tail modification and NOT DNA methylation is a determinant of Herpes Simplex Virus Type 1 latent gene expression. *J. Virol.*, **78**, 1139-1149.

70. Davison, A.J., Cunningham, C., Sauerbier, W. and McKinnell, R.G. (2006) Genome sequences of two frog herpesviruses. *J. Gen. Virol.*, **87**, 3509-3514.

71. Ambinder, R.F., Robertson, K.D. and Tao, Q. (1999) DNA methylation and the Epstein-Barr virus. *Seminars in cancer biology*, **9**, 369-375.

72. Tao, Q. and Robertson, K.D. (2003) Stealth technology: how Epstein-Barr virus utilizes DNA methylation to cloak itself from immune detection. *Clinical immunology (Orlando, Florida)*, **109**, 53-63.

73. Chinchar, V.G. (2002) Ranaviruses (family Iridoviridae): emerging cold-blooded killers. *Arch. Virol.*, **147**, 447-470.

74. Willis, D.B. and Granoff, A. (1980) Frog virus 3 DNA is heavily methylated at CpG sequences. *Virology*, **107**, 250-257.

75. Willis, D.B., Goorha, R. and Granoff, A. (1984) DNA methyltransferase induced by frog virus 3. *J. Virol.*, **49**, 86-91.

76. Schetter, C., Grunemann, B., Holker, I. and Doerfler, W. (1993) Patterns of frog virus 3 DNA methylation and DNA methyltransferase activity in nuclei of infected cells. *J. Virol.*, **67**, 6973-6978.

77. Kaur, K., Rohozinski, J. and Goorha, R. (1995) Identification and characterization of the frog virus 3 DNA methyltransferase gene. *J. Gen. Virol.*, **76 (Pt 8)**, 1937-1943.

78. Darai, G., Anders, K., Koch, H.G., Delius, H., Gelderblom, H., Samalecos, C. and Flugel, R.M. (1983) Analysis of the genome of fish lymphocystis disease virus isolated directly from epidermal tumours of pleuronectes. *Virology*, **126**, 466-479.

79. Wagner, H., Simon, D., Werner, E., Gelderblom, H., Darai, C. and Flugel, R.M. (1985) Methylation pattern of fish lymphocystis disease virus DNA. *J. Virol.*, **53**, 1005-1007.

80. Tidona, C.A., Schnitzler, P., Kehm, R. and Darai, G. (1996) Identification of the gene encoding the DNA (cytosine-5) methyltransferase of lymphocystis disease virus. *Virus genes*, **12**, 219-229.

81. Bigot, Y., Stasiak, K., Rouleux-Bonnin, F. and Federici, B.A. (2000) Characterization of repetitive DNA regions and methylated DNA in ascovirus genomes. *J. Gen. Virol.*, **81**, 3073-3082.

82. Karlin, S., Ladunga, I. and Blaisdell, B.E. (1994) Heterogeneity of genomes: measures and values. *Proc. Natl. Acad. Sci. U S A*, **91**, 12837-12841.

83. Shackelton, L.A., Parrish, C.R. and Holmes, E.C. (2006) Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J. Mol. Evol.*, **62**, 551-563.

84. Weitzman, M.D., Kyostio, S.R., Kotin, R.M. and Owens, R.A. (1994) Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. U S A*, **91**, 5808-5812.

85. Amiss, T.J., McCarty, D.M., Skulimowski, A. and Samulski, R.J. (2003) Identification and characterization of an adeno-associated virus integration site in CV-1 cells from the African green monkey. *J. Virol.*, **77**, 1904-1915.

86. Doerfler, W., Kruczek, I., Eick, D., Vardimon, L. and Kron, B. (1983) DNA methylation and gene activity: the adenovirus system as a model. *Cold Spring Harbor symposia on quantitative biology*, **47 Pt 2**, 593-603.

87. Doerfler, W. (1993) Adenoviral DNA integration and changes in DNA methylation patterns: a different view of insertional mutagenesis. *Progress in nucleic acid research and molecular biology*, **46**, 1-36.

88. Doerfler, W. (1996) A new concept in (adenoviral) oncogenesis: integration of foreign DNA and its consequences. *Biochimica et biophysica acta*, **1288**, F79-99.

89. Kubat, N.J., Tran, R.K., McAnany, P. and Bloom, D.C. (2004) Specific histone tail modification and not DNA methylation is a determinant of herpes simplex virus type 1 latent gene expression. *J. Virol.*, **78**, 1139-1149.

90. Orend, G., Knoblauch, M., Kammer, C., Tjia, S.T., Schmitz, B., Linkwitz, A., Meyer, G., Maas, J. and Doerfler, W. (1995) The initiation of de novo methylation of foreign DNA integrated into a mammalian genome is not exclusively targeted by nucleotide sequence. *J. Virol.*, **69**, 1226-1242.

91. Sutter, D. and Doerfler, W. (1980) Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. *Proc. Natl. Acad. Sci. U S A*, **77**, 253-256.

92. Kruczek, I. and Doerfler, W. (1982) The unmethylated state of the promoter/leader and 5'-regions of integrated adenovirus genes correlates with gene expression. *The EMBO journal*, **1**, 409-414.

93. Kuhlmann, I. and Doerfler, W. (1983) Loss of viral genomes from hamster tumor cells and nonrandom alterations in patterns of methylation of integrated adenovirus type 12 DNA. *J. Virol.*, **47**, 631-636.

94. Hertz, J.M., Schell, G. and Doerfler, W. (1999) Factors affecting de novo methylation of foreign DNA in mouse embryonic stem cells. *J. Biol. Chem.*, **274**, 24232-24240.

95. Orend, G., Kuhlmann, I. and Doerfler, W. (1991) Spreading of DNA methylation across integrated foreign (adenovirus type 12) genomes in mammalian cells. *J. Virol.*, **65**, 4301-4308.

96. Doerfler, W. (1991) Patterns of DNA methylation--evolutionary vestiges of foreign DNA inactivation as a host defense mechanism. A proposal. *Biological chemistry Hoppe-Seyler*, **372**, 557-564.

97. Graessmann, M., Graessmann, A., Wagner, H., Werner, E. and Simon, D. (1983) Complete DNA methylation does not prevent polyoma and simian virus 40 virus early gene expression. *Proc. Natl. Acad. Sci. U S A*, **80**, 6470-6474.

98. Deuring, R., Klotz, G. and Doerfler, W. (1981) An unusual symmetric recombinant between adenovirus type 12 DNA and human cell DNA. *Proc. Natl. Acad. Sci. U S A*, **78**, 3142-3146.

99. Flanagan, J.M. (2007) Host epigenetic modifications by oncogenic viruses. *British journal of cancer*, **96**, 183-188.

100. Sugawara, K., Fujinaga, K., Yamashita, T. and Ito, Y. (1983) Integration and methylation of shope papilloma virus DNA in the transplantable Vx2 and Vx7 rabbit carcinomas. *Virology*, **131**, 88-99.

101. Kim, K., Garner-Hamrick, P.A., Fisher, C., Lee, D. and Lambert, P.F. (2003) Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection. *J. Virol.*, **77**, 12450-12459.

102. Badal, S., Badal, V., Calleja-Macias, I.E., Kalantari, M., Chuang, L.S., Li, B.F. and Bernard, H.U. (2004) The human papillomavirus-18 genome is efficiently targeted by cellular DNA methylation. *Virology*, **324**, 483-492.

103. Verma, M. (2003) Viral genes and methylation. *Annals of the New York Academy of Sciences*, **983**, 170-180.

104. Munnes, M., Schetter, C., Holker, I. and Doerfler, W. (1995) A fully 5'-CG-3' but not a 5'-CCGG-3' methylated late frog virus 3 promoter retains activity. *J. Virol.*, **69**, 2240-2247.

105. Vardimon, L., Kuhlmann, I., Doerfler, W. and Cedar, H. (1981) Methylation of adenovirus genes in transformed cells and in vitro: influence on the regulation of gene expression? *Eur. J. Cell Biol.*, **25**, 13-15.

106. Vardimon, L., Gunthert, U. and Doerfler, W. (1982) In vitro methylation of the BsuRI (5'-GGCC-3') sites in the E2a region of adenovirus type 2 DNA does

not affect expression in Xenopus laevis oocytes. *Molecular and cellular biology*, **2**, 1574-1580.

107. Simon, D., Grunert, F., von Acken, U., Doring, H.P. and Kroger, H. (1978) DNA-methylase from regenerating rat liver: purification and characterisation. *Nucleic Acids Res.*, **5**, 2153-2167.

108. Subramanian, K.N. (1982) Effect of in vitro methylation at CpG sites on gene expression in a genome functioning autonomously in a vertebrate host. *Nucleic Acids Res.*, **10**, 3475-3486.

109. Fradin, A., Manley, J.L. and Prives, C.L. (1982) Methylation of simian virus 40 Hpa II site affects late, but not early, viral gene expression. *Proc. Natl. Acad. Sci. U S A*, **79**, 5142-5146.

110. Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247-257.

111. Oka, M., Rodic, N., Graddy, J., Chang, L.-J. and Terada, N. (2006) CpG Sites Preferentially Methylated by Dnmt3a in Vivo. *J. Biol. Chem.*, **281**, 9901-9908.

112. Yaman, R. and Grandjean, V. (2006) Timing of entry of meiosis depends on a mark generated by DNA methyltransferase 3a in testis. *Mol. Reprod. Dev.*, **73**, 390-397.

113. Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W.W., Okumura, K. and Li, E. (1999) Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene*, **236**, 87-95.

114. Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. and Bestor, T.H. (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science*, **294**, 2536-2539.

115. Bestor, T.H. and Ingram, V.M. (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl. Acad. Sci. U S A*, **80**, 5559-5563.

116. Li, E., Beard, C. and Jaenisch, R. (1993) Role for DNA methylation in genomic imprinting. *Nature*, **366**, 362-365.

117. Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335-340.

118. Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.*, **19**, 187-191.

119. Kunert, N., Marhold, J., Stanke, J., Stach, D. and Lyko, F. (2003) A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development (Cambridge, England)*, **130**, 5083-5090.

120. Rankin, R., Pontarollo, R., Ioannou, X., Krieg, A.M., Hecker, R., Babiuk, L.A. and van Drunen Littel-van den Hurk, S. (2001) CpG motif identification for veterinary and laboratory species demonstrates that sequence recognition is highly conserved. *Antisense & nucleic acid drug development*, **11**, 333-340.

121. Gunthert, U., Schweiger, M., Stupp, M. and Doerfler, W. (1976) DNA methylation in adenovirus, adenovirus-transformed cells, and host cells. *Proc. Natl. Acad. Sci. U S A*, **73**, 3923-3927.

122. Perlmann, C., Saemundsen, A.K. and Klein, G. (1982) A fraction of Epstein--Barr virus virion DNA is methylated in and around the EcoRI-J fragment. *Virology*, **123**, 217-221.

123. Badal, V., Chuang, L.S., Tan, E.H., Badal, S., Villa, L.L., Wheeler, C.M., Li, B.F. and Bernard, H.U. (2003) CpG methylation of human papillomavirus type 16 DNA in cervical cancer cell lines and in clinical specimens: genomic hypomethylation correlates with carcinogenic progression. *J. Virol.*, **77**, 6227-6234.

CHAPTER 7

SYNONYMOUS RECODING OF THE PARVOVIRUS CAPSID PROTEIN GENE REDUCES VIRAL FITNESS

Dr. Laura A. Shackelton generated the recoded gene sequence studied here.

7.1. Abstract

The evolution of viral genomes is subject to a variety of selective pressures, some of which act on gene products while others are directly linked to the DNA sequence composition. For example, the frequency of cytosine-guanine (CpG) dinucleotides (which are generally underrepresented in many vertebrate and invertebrate species) varies greatly among DNA viruses. This dinucleotide is strongly suppressed in most viruses with genomes of <10,000 bases, but rarely suppressed in viruses with larger genomes. CpG motifs represent mutation hot-spots, can trigger innate immune responses and are a key component of numerous regulatory responses. The biological relevance of CpG bias in DNA viruses, however, remains elusive. Here we addressed the biological role of CpG avoidance for the case of a small autonomous parvovirus: canine parvovirus (CPV). We focused on the capsid protein gene because this genomic region exhibits the strongest underrepresentation of CpGs, and increased the CpG composition in this region by nearly 20-fold while keeping the coding sequence unaltered. Because this recoded virus failed to produce viable progeny we constructed two additional viruses with smaller portions of the capsid protein gene replaced. All three recoded gene constructs successfully expressed proteins at approximately equivalent levels to the wild-type virus sequence when expressed from mammalian expression vectors. However, only the virus with the shortest recoded sequence produced viral progeny - and at a markedly lower level than wild-type virus. The genome of the wild type virus was unmethylated after replication in tissue culture, suggesting a minor role of CpG methylation. The synonymous recoding strongly decreased viral fitness in what appears to be a translation-independent manner, indicating that translational silencing was most likely not a major driver of CpG avoidance in this case.

7.2. Background

The genome composition and evolution of most organisms is affected by nucleotide-composition biases, dinucleotide frequencies, and non-random synonymous codon usage. Codon usage is generally species specific (14), but some trends such as the scarcity of codons containing the dinucleotides 'CG' or 'TA' appear to be nearly universal (22). In higher vertebrates, codon choice may also affect tissue-specific gene expression (see for example (35, 38)). The strength of codon bias and the choice of synonymous codons varies between species and also between individual genes in a genome (reviewed for example in (17, 29)). Translational optimization appears to be the major determinant of codon bias in highly expressed genes of prokaryotes and lower eukaryotes, with well-studied examples including *Escherichia coli* and *Saccaromyces cerevisiae* (e.g. (1, 4, 12, 13, 41)). In contrast, mutational pressures and selection on GC content seem important in determining codon usage in mammalian genes (recently reviewed in (7)).

Non-random synonymous codon usage and dinucleotide-composition bias in viruses is less well understood but some evidence points towards adaptation to host translation as an important evolutionary driver (e.g. (20, 27, 49)). For example in influenza A, the asteroviridae and potentially the irioviridae, nucleotide composition and codon usage appear to correspond well with those of the host species and have been proposed as one mechanism of viral adaptation after cross-species transmission (15, 43, 44). Replacement of preferred with non-preferred codons in a poliovirus resulted in reduced virulence and replicative fitness (6, 30), also indicating some codon optimization. However, in other virus families such as the adenoviridae, no

association of viral codon-usage biases with those of their host species can be detected (9). Some viruses (such as papillomaviridae (49) and maybe hepatnaviridae (34)) seem strategically to choose codons that are relatively rarely used by their hosts to express some of their viral genes, perhaps to restrict the level of protein production and reduce their immunogenicity (recently reviewed in (48)).

Nucleotide composition and dinucleotide patterns within viral genomes vary considerably among virus families, and occasionally even between subgroups (21, 39). The CpG dinucleotide (i.e. a cytosine linked to a guanidine through a phosphate bond) represents one of the most striking examples; in most small DNA viruses (genomes of <10,000 bases) this dinucleotide is reduced by $\geq$ 75% compared to the frequencies predicted from their overall base compositions (39). In contrast, the genomes of larger DNA viruses generally do not show a suppression of CpGs. The small parvoviruses are particularly intriguing, because CpGs are highly underrepresented in most of their members; the exception are the adeno-associated viruses (AAVs) which replicate in cells co-infected with large adeno- or herpesviruses (39). The AAVs appear comparable in nucleotide composition to their helperviruses.

Numerous reasons for the various nucleotide compositional discrepancies have been proposed: the ability of the larger viruses to encode viral polymerases; differences in the tolerance of mutations; control or avoidance of immune reactions; and the avoidance of gene silencing (reviewed for example in (19)). CpG methylation (or the lack thereof) might be involved in any of these scenarios. In eukaryotic cells, methylation patterns are used to discriminate between methylated host and largely unmethylated pathogen DNA, and non-methylated CpGs can trigger innate immune responses. Cytosine methylation also regulates differential gene expression and can

control replication and error correction (reviewed recently for example in (10, 37, 45)). Methylation can inactivate transposable elements (reviewed in (40)) and (particularly in promoter regions) can lead to transcriptional silencing (37). Methylated CpGs are also prone to mutation due to spontaneous deamination which cannot be corrected by the cellular repair machinery (8, 11). Thus, methylation can affect viral evolution in several ways. The DNA of many animal viruses appears to be unmethylated during active replication, but for several small DNA virus families including the parvoviridae, the nature and any role of the CpG methylation is not well understood (reviewed in (18)).

Canine parvovirus (CPV) is a member of the autonomous parvoviruses which emerged in the mid-1970s from the endemic feline panleukopenia virus (FPV) or a closely related carnivore parvovirus. This group of viruses has been endemic for a long time, and members infect various domestic and wild hosts including many felids, raccoons, foxes and mink (2). Transmission of FPV between those host species appears to occur readily while host-range barriers restricted transmission to dogs until CPV- a strain adapted to dogs - emerged (42). Infections of cats with more recent strains of CPV have been reported, but dogs and their wild relatives represent the major host species of CPV.

Here we report the synonymous recoding of part of the capsid protein gene of CPV to favor CpG-containing codons. We analyzed the observable effects on viral fitness and gene expression, and examined the potential interplay between the recoded sequence and CpG methylation. Additionally, we showed for the first time that the CpGs of the parvovirus genome are not methylated during active replication.

7.3 Methods

7.3.1. Computational analysis of nucleotide composition and codon usage.

Viral nucleotide and dinucleotide compositions as well as sequence-specific codon usages were analyzed using LaserGene (DNASTAR, version 7). Analysis of the effective number of codons (ENC), frequency of optimal codons (Fop) and GC content of the third codon position (GC3) was performed using Codon W (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=codonw). Codon usages of vertebrate species were reported from the literature (*Homo sapiens*) or for domestic cat (*Felis catus*), domestic dog (*Canis familiaris*), grey wolf (*Canis lupis*) and American mink (*Neovison vison*) obtained from the codon usage database (31).

7.3.2. Viruses and recoding.

CPV wild-type (wt) virus (CPV-wt), of antigenic type CPV-2 (GenBank reference number: M38245.1), was produced from a previously described infectious plasmid clone of the viral genome (33). A partially recoded VP2 gene sequence (between nts. 3008 and 4526 of sequence M38245.1) was constructed using the program Se-Al (http://tree.bio.ed.ac.uk/software/seal/) by converting CpN to CpG at all sites where this introduced synonymous changes. The recoded VP2 gene was synthesized and cloned into CPV_2wt using the BsrGI and PacI sites, generating the recoded virus (CPV-Rec). Two viruses with smaller portions of the recoded sequence were created (CPV-1/3Rec and CPV-2/3Rec) by swapping the 645bp PflMI-SpeI fragment between the CPV-wt and CPV-Rec. Virus stocks were generated in Norden Laboratory Feline

Kidney (NLFK) cells by lipofectamine transfection (Invitrogen, Carlsbad, CA) of 5µg plasmid DNA containing the full length genomes, and cells were subsequently subcultured for 2 to 5 days. The cultures were frozen and thawed, then a 1:2 dilution was used to inoculate fresh NLFK cultures to measure the ability of any virus generated to infect cells.

7.3.3. Expression vector generation.

A mammalian expression plasmid containing the CPV-2 VP1/2 gene fragment in pcDNA neo (CPV-Exp-wt) has been described previously (47). The BsrGI- PacI fragment in this construct was replaced by the corresponding fragments of CPV-Rec, CPV-1/3Rec or CPV-2/3Rec, resulting in expression vectors CPV-Exp-Rec, CPV-Exp-1/3Rec, and CPV-Exp-2/3Rec respectively. Plasmid DNAs (5 µg amounts) were transfected into NLFK cells using lipofectamine and cultured for 2 to 5 days.

7.3.4. Immunofluorescence assays.

Cells were either seeded on glass coverslips 2 days after transfection with infectious plasmids as described above, or inoculated with cell culture passages 24 h after cells were seeded onto the coverslips. After 2-5 days cells were fixed in 4% paraformaldehyde (PFA) for 10 min, then permeabilized in phosphate-buffered saline (PBS) with 0.5% bovine serum albumin (BSA), 0.2% Triton X100. Cells were then stained for infection with a Cy-2 labeled anti-NS1 antibody (CE-10.Cy2), or stained for assembled capsids using an Alexa594 labeled anti-capsid antibody (MAb8.Alexa594). Assembly of viral capsids from expression vectors was also determined by staining with MAb8.Alexa594.

For the determination of transfection or infection efficiencies, cells were transfected with infectious plasmids or inoculated 24 h after seeding onto glass

coverslips and fixed 2 days post transfection or infection. Cells were then analyzed for the presence of viral capsids by staining cells with MAb8.Alexa594 and nuclear DNA was detected with 4'-6-Diamidino-2-phenylindole (DAPI). A total of 300 cells were counted per coverslip and the reported infection ratios represent averages of 3 or more coverslips. All statistical analyses were performed in Microsoft Office Excel 2007.

7.3.5. Western blot analysis.

NLFK cells transfected with the expression vectors were washed and lysed in SDS sample buffer (10% SDS, 50 mM Tris.HCl and 2-mercaptoethanol). After boiling for 5 min samples were run on 8% denaturing SDS polyacrylamide gels, transferred to a nylon membrane and detected with the polyclonal rabbit anti-CPV antibody. Bound antibody was detected with a horseradish peroxidase conjugated goat-anti mouse antibody (Jackson ImmunoResearch Laboratories, West Grove, PA), and analyzed with the SuperSignal substrate (Pierce Chemical Company, Rockford, IL).

7.3.6. Restriction fragment length polymorphism (RFLP) analysis.

Viral RF DNA was isolated from infected or transfected NLFK cultures as described previously. As controls, approximately full-length viral genomic DNA was recovered from bacterial plasmid CPV-wt using endonucleases SphI and XmaI (NEB, Ispwich, MA), with the bacterial DNA serving as negative control. To obtain positive controls, this DNA was subsequently in-vitro methylated using MssI methylase (NEB, Ipswich, MA) according to the manufacturer's recommendations. All methylation reactions were performed for at least 6 h to ascertain sufficient methylation. Restriction analysis was performed by digestion with isoschizomer pairs HpaII and MspI or HhaI and CfoI for 3 h. Fragments were separated by agrose gel electrophoresis in a 1% TAE

gel, and DNA bands were visualized under UV light after staining with ethidium bromide.

7.4. Results

7.4.1. Analysis of CpG frequencies and recoding of part of the viral genome.

We analyzed nucleotide composition, codon usage and CpG content in the ~5.1 kb ssDNA genome of CPV. We found CpGs to be strongly suppressed - particularly in the capsid protein gene, with only 12 of the 75 CpG sites located within the ~ 1700 nt long capsid gene VP2 (Table 7.1). The parvovirus genomes are characterized by a considerably lower GC content (Table 7.1) and a lower effective number of codons (ENC) than their hosts (Table 7. 2). CPV and its ancestor FPV are very similar in codon usage, as would be expected from their close phylogenetic relationship (Table 7.2). The codon preferences of their mammalian host species (i.e. cats, dogs, mink, and wolves) are also very similar, but clearly distinct from that of humans (Table 7.3). Intriguingly, the codon usage of CPV and FPV generally does not correspond to that of their hosts.

We targeted a 1518 nt. long sequence in the capsid protein region for recoding, and introduced 198 additional CpGs through synonymous recoding (Table 7.1). We further constructed 2 intermediate viruses with portions of that capsid protein gene region replaced (Table 7.1; Figure 7.1), introducing 144 or 65 additional CpGs, respectively. By doing so we strongly increased the frequency of some CpG-containing low frequency codons, while leaving the usage of other codons largely unaltered (Table 7.3). Overall, the effective number of codons (ENC) increased from 36 to 46, with particularly the GC3 value increasing from 0.16 to 0.3 (Table 7.2),

indicating a tendency towards the use of lower frequency codons in the recoded viruses. In general, the codon usage and GC content of the recoded viruses shifted to that of the host species.

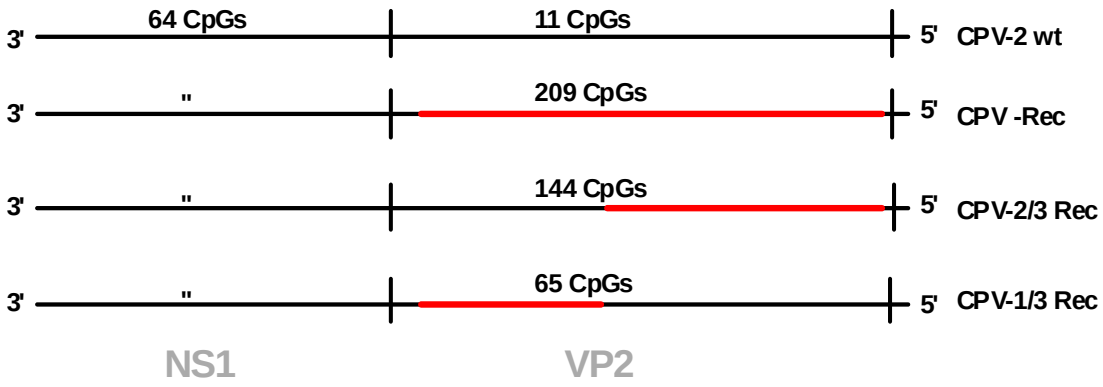


Figure 7.1. Schematic outline of the CPV-2 wt and recoded genomes, indicating the number of CpGs present in the genomic regions and the relative location of the recoded sequences.

7.4.2. Replication fitness of the recoded viruses.

After transfection of the viruses containing recoded portions in full length plasmid clones we detected expression of the non-structural protein NS1 and the capsid genes in all cases (Figure 7.2). Expression from CPV-Rec and CPV-2/3Rec viruses was very weak and no infectious progeny were detected after the culture supernatants were used to inoculate new cultures (Figure 7.3). After transfection or infection with virus CPV-1/3Rec we saw clear protein expression (Figures 7.2 and 7.3), but the infection rates were clearly lower than for virus CPV-2wt (Table 7.4.), indicating a strong decrease in transfection and infection efficiency associated with the recoded virus.

Table 7.1. Characterization of nucleotide and dinucleotide composition in Feline Panleukopenia virus (FPV), CPV-2wt and the recoded CPV genomes. The GC content in the coding region of the vertebrate host genomes is as follows: domestic cat (*Felis catus)*: 52.74%; domestic dog (*Canis familiaris)*: 53.16%; American mink (*Neovison vison)*: 52.14%; grey wolf (*Canis lupus)*: 56.75%.

	FPV	CPV-2wt	CPV-Rec	CPV-2/3Rec	CPV-1/3Rec
G&C – content (%)	36.5	36.6	40.5	39.3	37.7
CpG dinucleotide frequency in genome	76	75	273	208	129
CpG dinucleotide frequency in VP2	11	11	209	144	65
GpC frequency in genome	167	167	204	193	176
GpC frequency in VP2	53	52	89	78	61
Base frequency(%) (A, G, T, C)	35,20,29,16	35,20,29,16	34,22,26,18	34,22,26,18	35,21,27,17

Table 7.2. Summary statistics for human (from van Hemert et al, 2007), FPV, CPV _2wt, recoded viruses and their hosts. ENC = effective number of codons; FoP = frequency of optimal codons; GC3 = GC content of the third codon position . n/a = not available.

	Homo sapiens	Felis catus	Canis familiaris	FPV	CPV_2wt	CPV_Rec	CPV-2/3Rec	CPV-1/3Rec
ENC	55	52	n/a	36	36	44	42	39
FoP	n/a	n/a	n/a	0.3	0.22	0.24	0.24	0.22
GC3	0.57	0.63	0.62	0.16	0.16	0.31	0.26	0.20

Table 7.3. Codon usage in humans (from (28)), domestic cats (*Felis catus*), American mink (*Neovison vison*), grey wolves (*Canis lupis*), domestic dogs (*Canis familiaris*), Feline Panleukopenia virus (FPV), CPV-2 wt and recoded genomes (complete coding region covering NS1 and VP2, introns and UTRs removed), with codons containing CpG or GpC motifs indicated by shading.

codon	amino acid	Homo sapiens	Vertebrate hosts				Parvoviruses		Recoded viruses		
			Felis catus	Neovison vison	Canis lupus	Canis familiaris	FPV	CPV_2wt	CPV_ Rec	CPV_ 2/3Rec	CPV_ 1/3Rec
gca	A	63	19	18	15	20	56	59	40	46	53
gcc		20	45	45	53	44	0	8	12	10	10
gcg		1	**12**	**11**	**10**	**11**	**8**	**7**	**33**	**24**	**15**
gct		16	24	27	22	25	36	26	15	20	22
gga	G	50	25	25	10	24	37	39	30	32	37
ggc		2	**35**	**38**	**52**	**35**	**6**	**7**	**21**	**19**	**10**
ggg		32	25	23	22	25	15	12	11	11	12
ggt		15	15	14	16	16	42	41	38	38	41
aca	T	55	24	26	28	26	46	47	26	28	44
acc		13	40	41	47	39	2	7	18	15	9
acg		6	**16**	**14**	**10**	**13**	**5**	**3**	**32**	**24**	**12**
act		26	21	18	14	22	46	43	24	33	35
aga	R	89	22	22	14	20	91	82	82	82	82
agg		8	24	19	16	21	0	2	2	2	2
cga		1	9	10	8	11	0	2	2	2	2
cgc		0	18	17	29	20	0	9	12	12	9
cgg		1	20	23	24	21	5	4	2	2	4
cgt		1	7	23	8	7	5	2	0	0	2

Table 7.3 (continued)

codon	Amino acid	Homo sapiens	Felis catus	Neovison vison	Canis lupus	Canis familiaris	FPV	CPV_2wt	CPV_ Rec	CPV_ 2/3Rec	CPV_ 1/3Rec
cca	P	36	24	25	21	25	74	69	33	45	56
ccc		27	38	37	46	35	0	0	14	11	3
ccg		5	**13**	**9**	**12**	**12**	**0**	**5**	**38**	**28**	**15**
cct		32	26	29	21	27	26	27	15	16	26
agc	S	9	25	24	27	25	3	4	9	6	7
agt		37	13	15	11	14	27	34.	30	33	31
tca		31	12	13	11	13	20	19	11	16	14
tcc		10	25	25	34	24	0	1	11	7	6
tcg		2	6	5	4	6	0	6	10	7	9
tct		11	19	18	14	18	50	36	29	31	33
tgc	C	19	57	56	61	58	25	5	5	5	5
tgt		81	43	44	39	42	75	95	95	95	95
caa	Q	61	28	24	20	25	89	93	93	93	93
cag		39	72	76	80	75	11	7	7	7	7
aac	N	29	59	57	69	57	21	29	39	36	32
aat		71	41	43	31	43	79	71	61	64	68
gac	D	50	58	61	69	57	12	33	42	33	36
gat		50	42	39	31	43	88	67	58	67	64
gaa	E	77	43	41	29	40	83	87	87	87	87
gag		23	57	59	71	60	17	13	13	13	13
cac	H	46	63	60	76	61	9	19	32	29	23
cat		54	37	40	24	39	91	81	68	71	77
ata	I	53	16	12	16	15	21	28	25	25	28
atc		13	53	55	59	53	4	7	21	18	10
att		34	32	33	26	32	75	65	54	56	62

Table 7.3 (continued)

codon	Amino acid	Homo sapiens	Felis catus	Neovison vison	Canis lupus	Canis familiaris	FPV	CPV_2wt	CPV_ Rec	CPV_ 2/3Rec	CPV_ 1/3Rec
cta	L	16	6	6	5	6	15	13	12	12	13
ctc		12	21	24	26	22	0	2	3	3	2
ctg		33	43	40	53	43	3	3	3	3	3
ctt		3	11	13	8	12	9	23	23	23	23
tta		32	6	5	3	6	56	41	41	41	41
ttg		3	13	12	6	12	18	18	18	18	18
aaa	K	87	43	38	33	40	95	80	80	80	80
aag		13	57	63	67	60	5	20	20	20	20
atg	M	100	100	100	100	100	100	100	100	100	100
ttc	F	34	59	59	80	59	7	5	21	20	7
ttt		66	49	41	20	41	93	95	79	80.	93
tgg	W	100	100	100		100	100	100	100	100	100
tac	Y	29	61	61	63	60	0	16	28	28	16
tat		71	39	39	37	40	100	84	72	72	4
gta	V	73	10	10	10	10	45	33	30	30	32
gtc		11	28	25	27	27	3	6	16	10	12
gtg		3	47	50	58	48	7	16	15	16	15
gtt		13	16	15	6	14	45	46	39	44	41

Table 7.4. Transfection and infection efficiencies of CPV-2 wt and CPV1/3-Rec, measured 48h after transfection or infection.

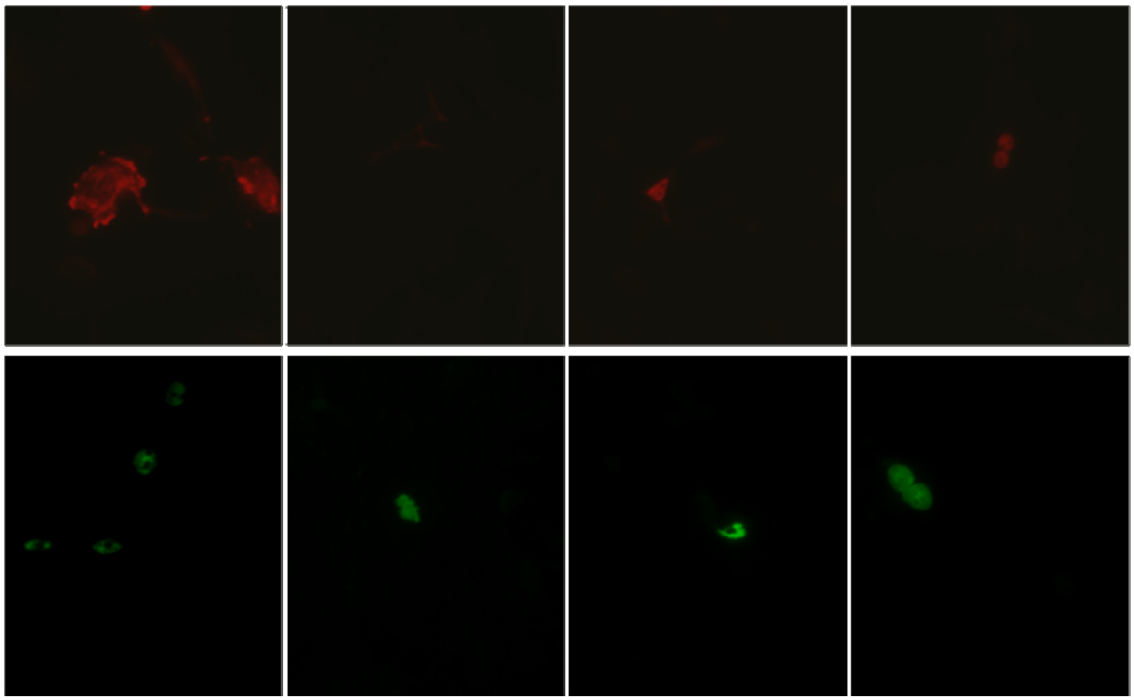
virus	**# cells counted**	**Transfection efficacy [%]; (Standard Deviation)**	**Infection efficacy [%]; (Standard Deviation)**
CPV-2wt	300	7.5 (0.58)	21 (1.3)
CPV-1/3Rec	300	4.2 (1.3)	8.8 (0.69)

7.4.3. Expression of recoded genes from mammalian expression vectors.

All recoded capsid-protein sequences allowed successful expression of the capsid proteins from a plasmid under the control of the CMV promoter. The expression levels were comparable to the wild-type sequence as determined by Western-blot analysis (Figure 7. 5) and by immunofluorescence staining of assembled capsids (Figure 7.4). All expression vectors expressed both capsid proteins VP1 and VP2 which assembled capsids, indicating that the recoding had not strongly affected the ability of the gene sequence to be translated.

7.4.4. Methylation status analysis of the wild type virus genome.

Restriction fragment-length polymorphism (RFLP) analysis of viral replicative form (RF) DNA with methylation-sensitive and - insensitive isoschizomeres showed the absence of methylation at HpaII (motif:CCGG) and Hha I (motif: GCGC) sites from the CPV-2wt genome (Figures 7.6. and 7.7.). As controls we analyzed non-methylated (neg. control) and in-vitro methylated (pos. control) CPV-2wt ds DNA recovered from the bacterial plasmid. As predicted, the methylated DNA was resistant to HpaII or HhaI digestion, but not to MspI digestion, and restriction patterns of the latter were similar to those obtained from RF DNA. In vitro methylated DNA was partially resistant to CfoI digestion (Figure 7.7.), indicating that overhanging CpG methylation blocked digestion with this enzyme, while the RF DNA digested with this enzyme appeared completely digested. In conclusion, these results strongly indicate the absence of methylation from these sites in the CPV-2wt genome.

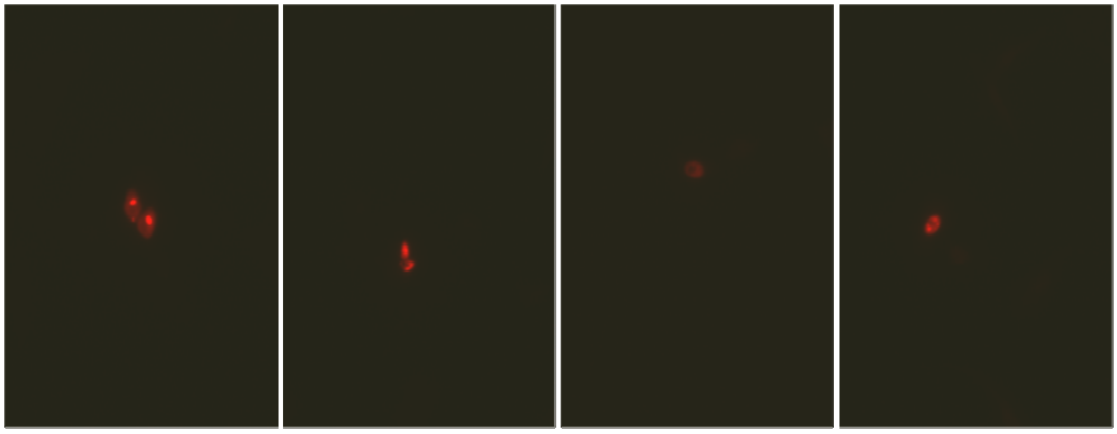


CPV-2wt CPV - Rec CPV2/3 - Rec CPV 1/3 - Rec

Figure 7.2. Characterization of recoded genomes - expression of viral proteins and capsid assembly after transfection. Green = NS1 expression; Red = capsid expression.

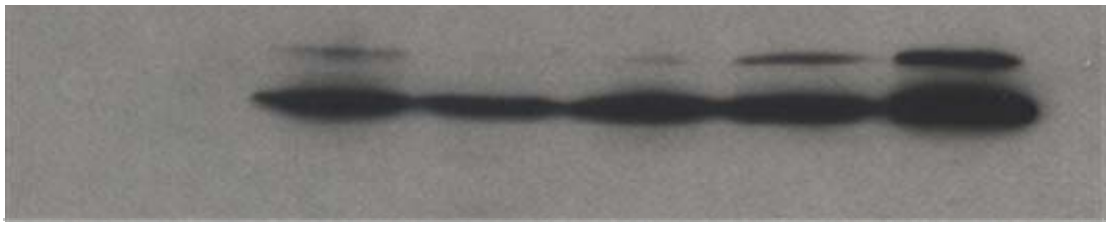
CPV-2wt CPV - Rec CPV2/3 - Rec CPV 1/3 - Rec

Figure 7.3. Characterization of recoded genomes - presence of assembled capsids after inoculating cultures with tissue-culture supernatants.



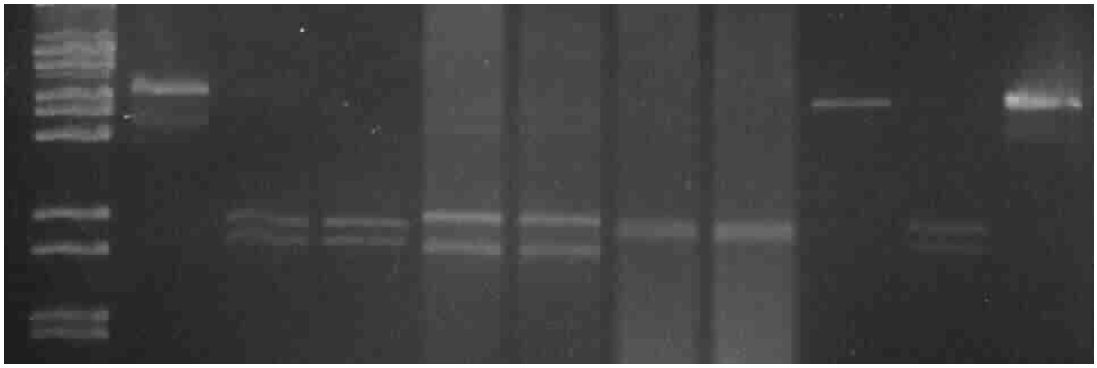
CPV-Exp CPV-Exp-Rec CPV-Exp-2/3Rec CPV-Exp-1/3Rec

Figure 7.4. Characterization of recoded genomes: expression and assembly of viral capsids from expression vectors.



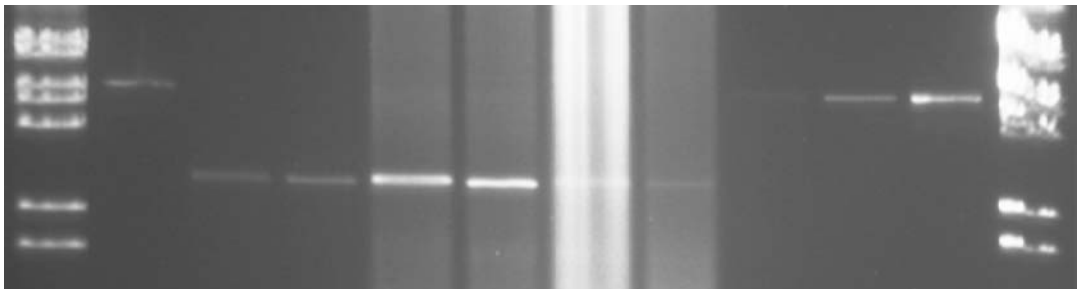
neg. control CPV-Exp-1/3Rec CPV-Exp-2/3Rec CPV-Exp-Rec CPV-Exp CPV-2wt

Figure 7.5. Characterization of recoded genomes: Western-blot analysis of expression of viral capsid proteins VP1 and VP2 from expression vectors. Top band = VP1, bottom band = VP2.



1 2a 2b 3a 3b 4a 4b 5a 5b 1

Figure 7. 6. Restriction enzyme polymorphism analysis of CPV-2wt RF DNA by digestion with HpaII (i.e. CpG methylation sensitive) & MspI (methylation insensitive) restriction enzymes. 1= uncut viral DNA liberated from CPV-2wt; 2= neg. control; 3 = viral RF DNA from CPV-2wt transfected cultures; 4= viral RF DNA from CPV-2 wt inoculated cultures; 5 = positive control; a= digested using HpaII; b= digested using MspI.



1 2a 2b 3a 3b 4a 4b 5a 5b 1

Figure 7. 7. Restriction enzyme polymorphism analysis of CPV-2wt RF DNA by digestion with HhaI (i.e. CpG methylation sensitive) & Cfo I (i.e. partially methylation sensitive) restriction enzymes. 1= uncut viral DNA liberated from CPV-2wt; 2= neg. control; 3 = viral RF DNA from CPV-2wt transfected cultures; 4= viral RF DNA from CPV-2wt inoculated cultures; 5 = positive control; a= digested using HhaI; b= digested using CfoI.

7.5. Discussion

The genomes of most parvoviruses show highly suppressed CpG composition, but the underlying reasons are not fully understood (21, 39). Here we added many additional CpGs to the capsid protein gene of CPV to examine the functional roles of that suppression, and also analyzed the potential relation to cytosine methylation. We showed that synonymous recoding strongly reduced viral fitness but not viral gene expression, and further showed the absence of detectable cytosine methylation from CPV-2wt. A significant role of CpG methylation through genome inactivation or mutational pressures therefore appears unlikely.

The engineering of a viable virus with altered codon usage in a ~400 base region of the capsid protein gene (CPV1/3-Rec) indicates the tolerance of the CPV genome towards at least moderate recoding. However, the lower transfection and infection efficiencies compared to wild type indicate a fitness cost associated even with this moderate amount of synonymous recoding. Because all recoded capsid proteins were expressed at reasonable levels from the expression plasmid, and initial viral replication and capsid assembly was seen after transfection of all full length plasmids, a block in viral packaging might be a plausible explanation for at least some of the observed fitness cost, particularly for the viruses containing the larger recoded region. Studies of genome replication and packaging in a closely related parvovirus (the minute virus of mice (MVM)), show that changes in the sequences near the right-hand end of the viral genome which increase the frequency of Gs interfere with the packaging of the viral DNA, most likely through interference with the folded DNA structures on capsid insertion (5).

The RFLP-based analysis of methylation status limits our analysis to those sites recognized by the restriction enzymes used. Although this limits the certainty with which conclusions about the complete absence of cytosine methylation can be made, HpaII sites appear to be of particular importance for the methylation-induced inactivation of promoters (recently reviewed in (16)). Moreover, HpaII and HhaI sites constitute a considerable amount of the CpG sites present in CPV-2wt (16) and appear very representative of all CpG sites present in the CPV genome. The absence of methylation at these sites therefore certainly indicates the scarcity, if not complete absence, of CpG sites in the CPV genome. The absence of detectable methylation of the CPV-2wt genomes raises the potential of CpG suppression being selected to avoide immune stimulation. It is unclear what role (if any) immune recognition has on the genome composition of CPV. Parvoviruses traffic through the endosomal pathway during cell entry and therefore would co-localize with TLR-9 in cells that express that receptor (16, 26, 32, 36, 46). Internalized herpesvirus DNA can be recognized through the TLR-9 pathway and a similar effect for DNA in CPV capsids (or perhaps free viral DNA released from lysed cells) appears likely (25). However, the stimulatory potential of CpGs is determined by the flanking nucleotides and the host species affected (3, 23, 24), and little is known about the functional roles of TLR-9 in dogs. This topic and in particular the innate immune responses to CPV DNA during infections of canine cells therefore merits further study.

We detected efficient protein expression from the wild type and recoded sequences but subtle impacts on translational efficiencies might have been missed in the analysis. However, our observation of reduced replicative fitness but relatively unaltered protein expression after recoding corresponds well to previous studies of poliovirus, where codon deoptimization led to reduced replicative fitness and pathogenicity but relatively unaltered protein expression (6, 30). In the case of polio,

codon deoptimization has been proposed as a potential means to improve vaccine safety because deoptimization greatly decreases viral pathogenicity.

Our study represents the first applied study of codon usage and cytosine methylation in autonomous parvoviruses. We have shown that CpG underrepresentation is related to viral fitness in a (most likely) translation-independent manner, and that cytosine methylation does not appear to be a major driver of CpG underrepresentation in these genomes. Although further studies are needed to elucidate the true importance and role of CpG underrepresentation in the autonomous parvovirus genome, our study proves the importance of CpG underrepresentation for viral fitness and raises the question of a potential role of immune evasion as major driver for CpG avoidance in the CPV genome.

REFERENCES

1. Akashi, H., and A. Eyre-Walker. 1998. Translational selection and molecular evolution. Curr Opin Genet Dev **8:**688-93.

2. Barker, I. K., R. C. Povey, and D. R. Voigt. 1983. Response of mink, skunk, red fox and raccoon to inoculation with mink virus enteritis, feline panleukopenia and canine parvovirus and prevalence of antibody in wild carnivores in Ontario. Can J Comp Med **47:** 188-197.

3. Bauer, S., Kirschnig, C J, Haeckler, H, Redecker, V, Hausmann, S, Akira, S, Wagner, H and Lipford, G B. 2001. Human TLR9 confers resopnsiveness to bacterial DNA via specie-specific CpG motif recognition. Proc Natl Acad Sci USA **98:**9237 - 9242.

4. Bennetzen, J. L., and B. D. Hall. 1982. Codon selection in yeast. J Biol Chem **257:**3026-31.

5. Burnett, E., S. F. Cotmore, and P. Tattersall. 2006. Segregation of a single outboard left-end origin is essential for the viability of parvovirus minute virus of mice. J Virol **80:**10879-83.

6. Burns, C. C., J. Shaw, R. Campagnoli, J. Jorba, A. Vincent, J. Quay, and O. Kew. 2006. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J Virol **80:**3259-72.

7. Chamary, J. V., J. L. Parmley, and L. D. Hurst. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet **7:**98-108.

8. Coulondre, C., J. H. Miller, P. J. Farabaugh, and W. Gilbert. 1978. Molecular basis of base substitution hotspots in Escherichia coli. Nature **274:**775-80.

9. Das, S., S. Paul, and C. Dutta. 2006. Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydropathy. Virus Res **117:**227-36.

10. Doerfler, W. 2005. On the Biological Significance of DNA Methylation. Bichemestry (Moscow) **70:**505-524.

11. Fryxell, K. J., and E. Zuckerkandl. 2000. Cytosine deamination plays a primary role in the evolution of mammalian isochores. Mol Biol Evol **17:**1371-83.

12. Gouy, M., and C. Gautier. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res **10:**7055-74.

13. Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier. 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res **9:**43-74.

14. Grantham, R., C. Gautier, M. Gouy, R. Mercier, and A. Pave. 1980. Codon catalog usage and the genome hypothesis. Nucleic Acids Res **8:**49-62.

15. Greenbaum, B. D., A. J. Levine, G. Bhanot, and R. Rabadan. 2008. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS Pathog **4:**1000079.

16. Harbison, C. E., J. A. Chiorini, and C. R. Parrish. 2008. The parvovirus capsid odyssey: from the cell surface to the nucleus. Trends Microbiol **16:**208-14.

17. Hershberg, R., and D. A. Petrov. 2008. Selection on codon bias. Annu Rev Genet **42:**287-99.

18. Hoelzer, K., L. A. Shackelton, and C. R. Parrish. 2008. Presence and role of cytosine methylation in DNA viruses of animals. Nucleic Acids Res **36:**2825-37.

19. Hoelzer, K., L. A. Shackelton, and C. R. Parrish. 2008. Presence and role of cytosine methylation in DNA viruses of animals. Nucl. Acids Res.**:**gkn121.

20. Jenkins, G. M., and E. C. Holmes. 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Research **92:**1-7.

21. Karlin, S., W. Doerfler, and L. R. Cardon. 1994. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? J. Virol. **68:**2889-2897.

22. Karlin, S., and J. Mrazek. 1996. What drives codon choices in human genes? J Mol Biol **262:**459-72.

23. Krieg, A. M. 2002. CoG Motifs in Bacterial DNA and their Immune Effects. Annual Review of Immunology **20:**709-760.

24. Krieg, A. M. 2001. Now I know my CpGs. Trends in Microbiology **9:**249-252.

25. Krug, A., G. D. Luker, W. Barchet, D. A. Leib, S. Akira, and M. Colonna. 2004. Herpes simplex virus type 1 activates murine natural interferon-producing cells through toll-like receptor 9. Blood **103:**1433-7.

26. Leifer, C. A., J. C. Brooks, K. Hoelzer, J. Lopez, M. N. Kennedy, A. Mazzoni, and D. M. Segal. 2006. Cytoplasmic targeting motifs control localization of toll-like receptor 9. J Biol Chem **281:**35585-92.

27. Lucks, J. B., D. R. Nelson, G. R. Kudla, and J. B. Plotkin. 2008. Genome landscapes and bacteriophage codon usage. PLoS Comput Biol **4:**e1000001.

28. Meintjes, P. L., and A. G. Rodrigo. 2005. Evolution of relative synonymous codon usage in Human Immunodeficiency Virus type-1. J Bioinform Comput Biol **3:**157-68.

29. Mooers, A. O., and E. C. Holmes. 2000. The evolution of base composition and phylogenetic inference. Trends Ecol Evol **15:**365-369.

30. Mueller, S., D. Papamichail, J. R. Coleman, S. Skiena, and E. Wimmer. 2006. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol **80:**9687-96.

31. Nakamura, Y., T. Gojobori, and T. Ikemura. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res **28:**292.

32. Parker, J. S., and C. R. Parrish. 2000. Cellular uptake and infection by canine parvovirus involves rapid dynamin-regulated clathrin-mediated endocytosis, followed by slower intracellular trafficking. J Virol **74:**1919-30.

33. Parrish, C. R., C. F. Aquadro, M. L. Strassheim, J. F. Evermann, J. Y. Sgro, and H. O. Mohammed. 1991. Rapid antigenic-type replacement and DNA sequence evolution of canine parvovirus. J Virol **65:**6544-52.

34. Pinto, R. M., L. Aragones, M. I. Costafreda, E. Ribes, and A. Bosch. 2007. Codon usage and replicative strategies of hepatitis A virus. Virus Res **127:**158-63.

35. Plotkin, J. B., H. Robins, and A. J. Levine. 2004. Tissue-specific codon usage and the expression of human genes. Proc Natl Acad Sci USA **101:**12588-91.

36. Ros, C., C. J. Burckhardt, and C. Kempf. 2002. Cytoplasmic trafficking of minute virus of mice: low-pH requirement, routing to late endosomes, and proteasome interaction. J Virol **76:**12634-45.

37. Salozhin, S. V., E. B. Prokhorchuk, and G. P. Georgiev. 2005. Methylation of DNA - One of the Major Epigenetic Markers. Biochemistry (Moscow) **70:**525-532.

38. Semon, M., J. R. Lobry, and L. Duret. 2006. No evidence for tissue-specific adaptation of synonymous codon usage in humans. Mol Biol Evol **23:**523-9.

39. Shackelton, L., C. Parrish, and E. Holmes. 2006. Evolutionary Basis of Codon Usage and Nucleotide Composition Bias in Vertebrate DNA Viruses. J Molec Evol **62:**551-563.

40. Slotkin, R. K., and R. Martienssen. 2007. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet **8:**272-285.

41. Stoletzki, N., and A. Eyre-Walker. 2007. Synonymous codon usage in Escherichia coli: selection for translational accuracy. Mol Biol Evol **24:**374-81.

42. Truyen, U., and C. R. Parrish. 1992. Canine and feline host ranges of canine parvovirus and feline panleukopenia virus: distinct host cell tropisms of each virus in vitro and in vivo. J Virol **66:**5399-5408.

43. Tsai, C. T., C. H. Lin, and C. Y. Chang. 2007. Analysis of codon usage bias and base compositional constraints in iridovirus genomes. Virus Res **126:**196-206.

44. van Hemert, F. J., B. Berkhout, and V. V. Lukashov. 2007. Host-related nucleotide composition and codon usage as driving forces in the recent evolution of the Astroviridae. Virology **361:**447-54.

45. Vanyushin, B. F. 2005. Enzymatic DNA methylation is an epigenetic control for Genetic Functions of the Cell. Biochemistry (Moscow) **70:**488-499.

46. Wagner, H. 2002. Interactions between bacterial CpG-DNA and TLR-9 bridge innate and adaptive immunity. Current Opinion Microbiol **5:**62-69.

47. Yuan, W., and C. R. Parrish. 2001. Canine parvovirus capsid assembly and differences in mammalian and insect cells. Virology **279:**546-57.

48. Zheng, Z. M., and C. C. Baker. 2006. Papillomavirus genome structure, expression, and post-transcriptional regulation. Front Biosci **11:**2286-302.

49. Zhou, J., W. J. Liu, S. W. Peng, X. Y. Sun, and I. Frazer. 1999. Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. J Virol **73:**4972-82.

CHAPTER 8

SUMMARY AND CONCLUSIONS

In conclusion, the scientific studies presented in the scope of this thesis explore some of the molecular and evolutionary determinants of viral host switching. Because the success of a cross-species transmission event is strongly affected by host- and pathogen-associated factors, we centered our studies in these areas. We however acknowledge that numerous other aspects such as those related to ecology, politics, the environment, socio-economic drivers or geology simultaneously influence disease emergences.

As we have shown, emerging diseases can differ significantly in their core properties and no ‘one-size-fits-all’ approaches to understanding disease emergences appear promising. When we analyzed the intra-host dynamics in two recently emerged viruses of dogs - canine parvovirus (CPV) and canine influenza virus (CIV) - we detected striking differences in their evolutionary dynamics, underlining the dangers inherent in drawing general conclusions. However, in both cases the intra-host dynamics of the emerged virus appeared to be similar to those observed for the ancestral virus in the donor host, indicating that the host switch had not greatly affected the evolutionary dynamics within infected hosts. Moreover, when we artificially transferred CPV between hosts we did not detect marked differences in evolutionary dynamics; this further supports the hypothesis that many host changes do not strongly affect viral intra-host dynamics.

Conversely, however, our study of CPV and its ancestor feline panleukopenia virus (FPV) on a population level showed marked differences in their evolutionary dynamics, which reinforced the discrepancy between evolutionary forces on a population versus an intra-host level. We further detected changes in evolution and spread from the initial emerged CPV strain (CPV-2) to more recent CPV isolates

(CPV-2a etc.), signifying how starkly evolutionary dynamics can change in the time after a host-switching event.

The interactions between viruses and their hosts are fine-tuned and likely determined on many different levels, several of which are still incompletely understood. We addressed some of these more subtle levels of host adaptation by addressing the impact of non-random synonymous codon usage in CPV. We clearly showed that synonymous recoding starkly decreased viral fitness in a translation-independent manner and provide evidence that cytosine-methylation is not a major driver of the observed codon bias. While further studies are needed to prove this hypothesis, in this case a role of codon bias in immune evasion appears likely. This highlights the potential for additional, more subtle barriers to successful host switching.

In conclusion, this thesis presents one of the first integrative studies of the molecular and evolutionary determinants of viral host jumping. While many questions remain open and several aspects of cross-species transmission have been deliberately excluded from this analysis for simplification purposes, some common characteristics of successful host jumps have emerged and the thesis represents a first tentative step towards understanding disease emergence through host jumping.