chapter 1

Evolvability of Emerging Viruses

Donald S. Burke

RNA viruses are, by several orders of magnitude, the most genetically labile "life forms" (15, 23). Mutation rates for RNA viruses are typically on the order of one error per 10,000 nucleotides replicated, compared to one per 10 million nucleotides for larger DNA-based life forms like vertebrates (8) (Table 1.1). Since the average genome length of RNA viruses is only 10,000 nucleotides, and all are shorter than 40,000 nucleotides, almost all new viral RNA strands differ from their parent strand by one or more nucleotides. Indeed, the error rate of one mutation per progeny genome poises RNA viruses at the edge of "error catastrophe," error rates so fast that genetic information degenerates into replication incompetence. Viruses can also recombine with other related viruses to effectively "shuffle" newly evolved genes. Recombination serves both to hybridize highly fit variants and to replace defective and incompetent genes. This general strategy of repeated iterations of random variation, selection, and recombination between the best solutions ("most fit progeny") has been found to have widespread applications in problem solving,

Donald S. Burke, Center for Immunization Research, School of Hygiene and Public Health, Johns Hopkins University, Hampton House, Room 241, 624 North Broadway, Baltimore, MD 21205.

Pathology of Emerging Infections 2
Edited by Ann Marie Nelson and C. Robert Horsburgh, Jr.

Table 1.1 Error rates and genome sizes of RNA viruses as compared to autonomous organisms

Virus	Genome size: ν (number of nt or bp)	Error rate: $(1-q)$ (per replication round and per nt)	Error rate: $\nu(1-q)$ (per replication round and per genome)
RNA			
Bacteriophage Q_β	4,200	3×10^{-4}	1.3
Poliovirus type 1	7,400	3×10^{-4}	0.2
Vesicular stomatitis virus	11,000	1×10^{-4}	1.1
Foot and mouth disease virus	8,400	1×10^{-4}	0.8
Influenza A virus	14,000	6×10^{-5}	0.8
Sendai virus	15,000	3×10^{-5}	0.5
HIV-1 (AIDS virus)	10,000	1×10^{-4}	1.0
Avian myeloblastosis virus	7,000	5×10^{-5}	0.4
DNA			
Bacteriophage M13	6,400	7×10^{-7}	4.6×10^{-3}
Bacteriophage γ	48,500	8×10^{-8}	3.8×10^{-3}
Bacteriophage T4	166,000	2×10^{-8}	3.3×10^{-3}
Escherichia coli	4.7×10^6	7×10^{-10}	3.3×10^{-3}
Yeast (*Saccharomyces cerevisiae*)	13.8×10^6	3×10^{-10}	3.8×10^{-3}
Neurospora crassa	41.9×10^6	1×10^{-10}	4.2×10^{-3}
Human	3×10^6	$\sim 10^{-12}$	$\sim 3 \times 10^{-3}$

[a]Reprinted from reference 8 with permission from Elsevier Science.

"Genetic algorithm" strategies are widely used to rapidly calculate solutions to complex problems

even in fields far removed from biology. Such strategies are now central to many artificial intelligence systems, from database searching to machine learning. "Genetic algorithm" strategies are widely used to rapidly calculate solutions to complex problems (5, 12, 14).

This chapter will (i) review basic evolutionary theory, particularly as it relates to viruses; (ii) summarize recent viral epidemics of global significance; (iii) present a computational model of viral evolution; and (iv) offer some thoughts about how future epidemics of emerging viruses might be predicted and prevented.

Evolutionary Theory

The nucleotide sequence of a viral genome can be thought of as an information string of 10,000 bits with four alternative states (A, C, G, and U or T) per bit. The total evolutionary potential for such a system is the universe of all possible 10,000-bit strings. These can be hypothetically arranged in a "sequence space" so that each string is adjacent to its 30,000 one-step-nearest neighbors. The total dimension of this space is 4 to the 10,000th power, a number that is greater than all atoms in the universe. Obviously, many regions in this hypothetical RNA sequence space [for example, the fringe region around a pure 10,000-poly(C) sequence] are out of bounds for replication. However, many regions of RNA sequence space do permit replication, and within these there are local optima. These optima can be conceptualized as peaks on a fitness landscape in nucleotide sequence space (25, 34).

Similarly, *evolution* can be thought of as the process whereby sequence space is explored, with successful variants colonizing the fitness peaks. Because the mutation rate of RNA replication is so high, evolutionary time for exploration of sequence space for RNA-based life can be measured in weeks and years, compared to the millennia required for DNA-based life. Rephrased, evolution of RNA life occurs on a scale that can be comprehended and studied within human dimensions. The disparity between rates of RNA and DNA evolution, the difference in RNA and DNA "*evolvability*," probably accounts for the fact that most of the new emerging diseases are caused by RNA viruses; RNA-based genomes have sufficient plasticity to permit rapid host switching.

Evolution of RNA life occurs on a scale that can be comprehended and studied within human dimensions

This high evolvability of RNA may also account for the fact that almost all known arthropod-borne viruses (viruses that alternately replicate within vertebrate and arthropod cells) have RNA genomes. Although there are numerous DNA viruses of vertebrates and numerous DNA viruses of arthropods, remarkably there is only a single known DNA arbovirus (African swine fever virus) (24). It is likely that for most arthropod-borne viruses the sequence space fitness peak for growth in arthropod cells is close to, but not perfectly congruent with, that for growth in vertebrate cells, and mutation is required to trampoline back and forth through sequence space between the two host-specific optima.

Mutation alone may be insufficient to permit movement through some regions of sequence space (9, 10, 19). By definition, even single-step mutations from a local fitness optimum are less fit than their parents. Particularly in rugged fitness landscapes, genomes only slightly removed from the local optimum may be totally unfit, so that exploration of the surrounding space becomes impossible. Recombination between genomes on separated fitness optima permits such an "evolutionary broad-jumping" type of sequence space exploration; recombinant progeny may fall on previously totally unexplored fitness peaks (18). Naturally occurring recombination (or reassortment) has been closely studied in RNA viruses with segmented genomes such as influenza virus. Recombination has also recently been shown to occur commonly among HIV strains (2, 3). The role of recombination in nature is less well studied for other RNA viruses, but convincing examples have been found wherever they have been sought (1, 4).

The biological consequences of such recombination, whether by reassortment of segmented genomes or by true recombination through crossing over, may be the generation of novel variants with new epidemiological properties. For influenza, change by mutation, widely known as "drift," is of minor epidemiological significance, while change by recombination often results in a "shift" with an epidemiological impact felt on a global scale (20). It is now clear that "shifts" in influenza come about through recombination (reassortment) of RNA sequences from bird and pig influenza viruses with sequences from human viruses. Bird (and pig) influenza RNA explores certain regions of fitness space far from that occupied by human influenza RNA (21). Coinfection of a single host (the pig usually serves as the "mixing vessel" for avian and human influenza viruses) allows widely divergent sequences to coinfect the same cell and recombine (30). Many such progeny

are nonviable and never reproduce, but occasionally a new variant emerges (11, 31). The same model may be applicable to many other RNA viruses. All RNA viruses apparently can and do recombine, but the epidemiological significance of recombination is less clear. Recent studies of the nucleotide sequences of HIV strains from around the world have shown that recombination may be just as dominant an evolutionary force for this virus as it is for influenza (27, 28). Non-human primate lentiviruses may contribute to the human HIV gene pool (17).

Recombination may be just as dominant an evolutionary force for HIV as it is for influenza

Existing RNA viruses occupy only a tiny fraction of the available RNA sequence space. Clearly a dominant constraint is the ability of the proteins encoded by the viral RNA to functionally interact with host cell constituents. However, there are probably several other structural (protein/protein, protein/RNA, etc) constraints on sequence space exploration.

In addition, models from the new scientific field of "complexity" suggest that genomes in whole galaxies of sequence space may be inherently unfit not because of protein structural constraints but because of deeper constraints on the evolvability of their informational content (18). Some sequence sets may fail to evolve and remain frozen in sequence space (analogous to a solid). Others may expand through sequence space too rapidly and degenerate to chaos (analogous to a gas). If the concept of "life at the edge of chaos" has merit, perhaps only certain regions of sequence space encode the "liquid" information evolvability necessary to explore and then stably colonize new fitness peaks.

Real Viruses, Real Pandemics

In the past 30 years there have been at least seven "new" viruses that have caused global epidemics involving millions of humans (Table 1.2). All of the recent global pandemics have been of RNA viruses with an ability to recombine or reassort genetic material between viruses. For the influenza A viruses (H3N2 and H1N1) there is solid evidence that the new epidemic strains arise through mixing of genes from animal influenza viruses with genes from preexisting human influenza virus. For the retroviruses (HIV and HTLV) there is suggestive evidence that these viruses crossed the species barrier from non-human primates into humans. The recent pan-

Table 1.2 Recent viral pandemics (Ro >> 1) in human populations

Disease	Year	Location	Family	Virus
Influenza	1968	Hong Kong	*Orthomyxoviridae*	Influenza A (H3N2)
Hemorrhagic conjunctivitis	1969	Ghana	*Picornaviridae*	Enterovirus 70
Meningitis	1969	United States	*Picornaviridae*	Enterovirus 71
Hemorrhagic conjunctivitis	1970	Singapore	*Picornaviridae*	Coxsackievirus A24/variant
Influenza	1977	Russia	*Orthomyxoviridae*	Influenza A (H1N1)
AIDS	1981	United States, Zaire	*Retroviridae*	HIV-1
Leukemia, lymphoma	1982	Japan	*Retroviridae*	Human T-cell lymphotropic virus

Table 1.3 Recent localized (Ro < 1) new viral epidemics in human populations

Disease	Year	Location	Family	Virus
Neuropsychosis	1985	Germany	*Paramyxoviridae*	Borna disease virus
AIDS-like	1986	West Africa	*Retroviridae*	HIV-2
Hemorrhagic fever	1989	Venezuela	*Arenaviridae*	Guanarito virus
Influenza	1993	Netherlands	*Orthomyxoviridae*	Influenza A virus (H3N2, avian)
Pulmonary syndrome	1993	Western United States	*Bunyaviridae*	Sin Nombre virus
Hemorrhagic fever	1995	Zaire	*Filoviridae*	Ebola virus

demic picornaviridae (enterovirus 70, enterovirus 71, and coxsackievirus A24/variant) probably derived directly from preexisting human viruses, but a genetic contribution from an animal picornavirus gene pool cannot be ruled out.

Outbreaks and epidemics of new viruses in humans are continually being observed around the world. Some very recent examples are shown in Table 1.3; all are RNA viruses. As compared to the viruses causing global pandemics (above and Table 1.2), these viruses show no or only a limited capacity for human-to-human transmission. For four of the six viruses shown in Table 1.3, humans are known to become infected directly with a virus that is native to animals. In one case (Borna) an animal reservoir is suspected, and in one (Ebola) the reservoir is unknown.

New viral epidemics are also continually being observed around the world in animal populations. Table 1.4 shows some very recent examples. All but one of these recent epidemic viruses are RNA viruses; the canine parvovirus is the only example of a new epidemic animal DNA virus. Four of these viruses are thought to have arisen through interspecies transfer (canine parvovirus, lion paramyxovirus, dolphin paramyxovirus, equine paramyxovirus), while the other two are thought to have arisen through mutation of a virus already endemic in the species (pig coronavirus, chicken influenza virus).

Collectively, these new epidemics demonstrate the ability of viruses in many RNA virus families to cross species barriers where they can cause disease and become serially transmitted within the new host species.

These new epidemics demonstrate the ability of viruses in many RNA virus families to cross species barriers

Table 1.4 Some important recent viral epidemics in animal populations

Host	Disease	Year	Location	Family	Virus
Zoo	Pox/pulmonary	1973	Russia	*Poxviridae*	Cowpox (rodent) virus
Chicken	Influenza	1983	United States	*Orthomyxoviridae*	Influenza A virus (H5N2)
Pig	Respiratory	1984	Europe	*Coronaviridae*	Porcine respiratory coronavirus
Dog	Enteritis	1987	Worldwide	*Parvoviridae*	Canine parvovirus
Dolphin	Respiratory	1988	United States	*Paramyxoviridae*	Dolphin and porpoise morbillivirus
Lion	Encephalitis	1994	Tanzania	*Paramyxoviridae*	Canine distemper virus
Horse	Respiratory	1994	Australia	*Paramyxoviridae*	Equine morbillivirus
Rabbit	Hemorrhagic	1995	Australia	*Caliciviridae*	Rabbit hemorrhagic disease virus

Virtual Viruses

My colleagues at the Navy Center for Applied Research in Artificial Intelligence and I have recently been working on a computational model of viral evolution, a "virtual virus" that we call "VIV." The effort to construct a computer simulation of virus evolution was inspired by the observation that the evolutionary strategy of HIV, a diploid virus with a high mutation rate and a high recombination rate, was remarkably similar to the evolutionary computation strategies known as "genetic algorithms" used for machine learning and robotics.

Genetic algorithms (GAs) are heuristic learning (problem-solving) models based on principles drawn from natural evolution and selective breeding (5, 12). In a typical GA a *population of structures* (a population of bit strings or programs) is established that can be interpreted as a pool of candidate solutions to the given problem. *Competitive selection* is employed to allow these structures to reproduce, based on each structure's fitness as a solution to the given problem. Idealized *genetic operators*, such as mutation or recombination, are applied to the selected structures in order to create a new generation of structures. In many applications in optimization and search, these features enable the genetic algorithm to rapidly improve the average fitness of the population and to quickly identify the high-performance regions of very complex search spaces (6, 13, 29).

For the VIV computational model we constructed an artificial genetic system

For the VIV computational model we constructed an artificial genetic system (33). In this system arbitrary "nucleotide" triplets encode English letters rather than amino acids, and sequences are translated into words or groups of words rather than into polypeptides or proteins. The standard target phenotype in VIV is the words "COREPROTEIN," "POLYMERASE," and "ENVELOPE," which can be present in any order along the string. Run-on and overlapping reading frames are permitted. Fitness is assigned to each string according to the encoded spelling score. Perfect spelling of all three words is assigned a fitness of 1.0, while random gibberish is assigned a fitness score of 0.0. Although redundancies and noncoding regions are not directly scored, string brevity is rewarded with higher fitness. In a typical VIV simulation experiment, a single population of 500 random strings of lengths distributed between 100 and 500 nucleotides is permitted to evolve for 2,000 generations. Evolutionary operators such as the frequency of mutation or recombination are then systematically varied, and "adaptation curves" (plots of population fitness per generation) are analyzed.

We have drawn several conclusions from our preliminary VIV simulations, as follows: (i) the optimal point mutation rate is close to one mutation per genome per replication cycle; (ii) when added to mutation, recombination in any form speeds adaptation; (iii) homologous recombination is superior to random crossover recombination; and (iv) adaptation speed increases with homologous recombination rates up to 0.4 recombination events per replication cycle, but little more at higher recombination rates. These results suggest that HIV, a real "diploid" virus with a measured mutation rate of about one mutation per genome per replication cycle and a measured high recombination rate, may search sequence space with near-optimal efficiency (16, 26, 32, 35).

The basic VIV model can be modified to incorporate a variety of evolutionary operators, such as genome segmentation, genomic secondary structures, insertions and deletions, and feedback loops and hypercycles. Such studies are in progress.

Predicting and Preventing Viral Pandemics

It is striking how little discussion there has been on how we, the human species, might predict and prevent future pandemics

Given the considerable attention focused on emerging diseases, it is striking how little discussion there has been on how we, the human species, might predict and prevent, rather than simply detect and react to, future pandemics. To do so requires a rational approach to risk assessment. I propose three criteria to identify the set of virus families that pose the greatest risk for a new global pandemic.

The first criterion is the most obvious: recent pandemics in human history. Those viruses already proven to cause human pandemics are clearly able to do so again. These would include the *Orthomyxoviridae* (influenza), the *Lentiviridae* (HIV-1), and the *Picornaviridae* (enterovirus 70).

The second criterion is proven ability to cause major epidemics in non-human animal populations. Here a different but overlapping set of virus families is identified, including the *Orthomyxoviridae*, the *Paramyxoviridae*, and the *Coronaviridae*.

The third criterion (which may be less obvious) reflects the thesis that I have tried to build in this chapter, that intrinsic evolvability confers on a virus the potential to emerge into and to cause pandemics in human populations. Virus families with proven high mutation rates and which have genomic organizational features that foster recombination meet this criterion. These include the *Orthomyxoviridae*, *Retroviridae*, *Coronaviridae*, and *Reoviridae*.

Some of these viruses, particularly those like the *Coronaviridae* and the related *Arteriviridae*, should be considered as serious threats to human health. These are viruses with high evolvability and proven ability to cause epidemics in animal populations.

Whence and when will new viruses emerge? Of the 68 virus genera that are known to infect vertebrates, 47 are already known to infect humans. The remaining 21 virus genera are thought to infect only non-human vertebrates. Of these, 10 are large DNA viruses not likely to successfully cross species into humans; 2 are small DNA viruses; and 3 are RNA viruses known to infect only fish and fowl. This leaves six genera of RNA viruses that routinely infect other mammals but are not known to infect humans. Among these, the *Arteriviridae* (particularly simian hemorrhagic fever virus) are particularly worrisome.

Serious consideration must also be given to the geographic location of research laboratories for the study of emerging viruses (22). Two observations should guide the placement of study sites. First, viruses easily cross species boundaries between closely related host species but are less able to do so between more distantly related hosts: humans are more susceptible to viruses of monkeys than to viruses of fish. Second, for any given viral taxon

the pool of viral variation is greatest in those geographic regions where host variation is greatest. For these reasons, any effective global program to predict and prevent emerging viral pandemics should systematically study viruses in those parts of the world where the number and diversity of mammalian species, and in particular the diversity of primate species, is the greatest, such as the tropical rain forests of Africa and South America. The current complete lack of sustained United States-supported research efforts in these important regions should be corrected.

A Closing Thought

This chapter has presented a fairly abstract, conceptual approach to viral evolution and emergence. Indeed, the skeptical reader might liken it to William James's view, offered over a century ago (7):

Evolution is a change . . .

Evolution is a change
from a nohowish, untalkaboutable,
all-alikeness,
to a somehowish and in-general-talkaboutable,
not-all-alikeness,
by continuous
somethingelsifications and sticktogetherations.

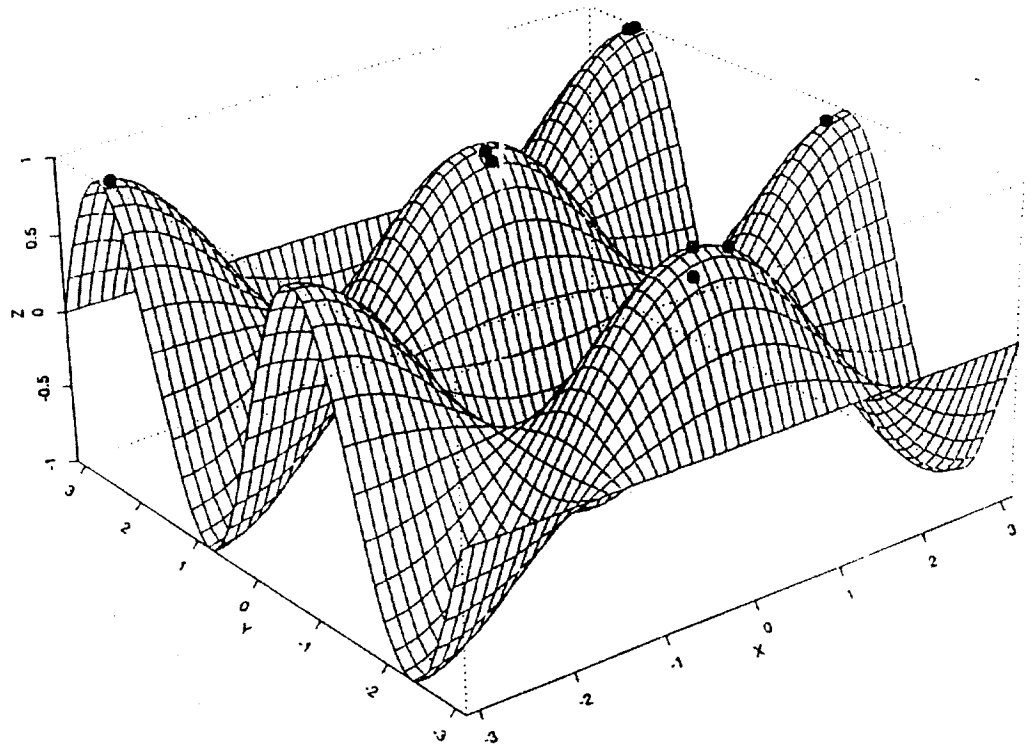


Figure 1.1 Representation of evolving bit strings in sequence space. Here the sequence space is shown only in two dimensions, the x and y axes. For a string of length L, the strings would evolve through an L-dimensional sequence space, but this is impossible to draw on a two-dimensional paper surface. Fitness is represented as the height on the z axis. In this example populations of strings are colonizing several local fitness optima.

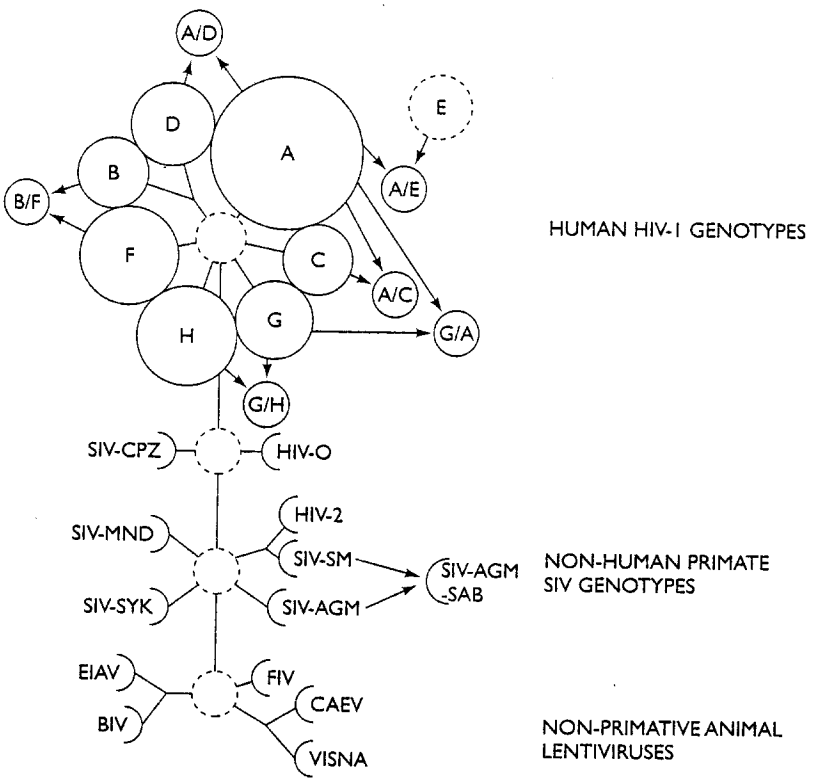
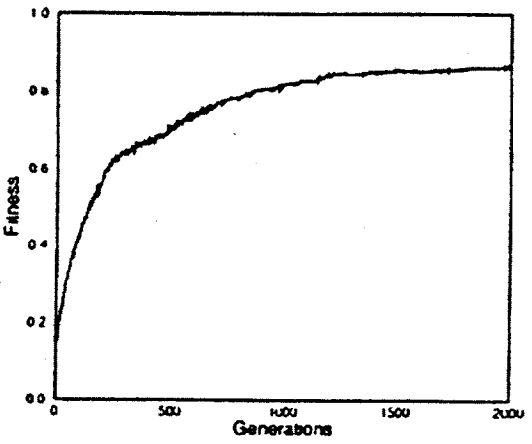


Figure 1.2 Cartoon of the lentivirus phylogenetic tree, showing the relationship between major virus groups. The dotted circles at nodes represent branch points for which relationships are not entirely certain. Viruses branching from the bottom node are the lentiviruses of horses, equine infectious anemia virus (EIAV); of cows, bovine immunodeficiency viruses (BIV); of cats, feline immunodeficiency virus (FIV); and of goats and sheep, caprine arthritis encephalitis virus (CAEV) and visna virus. Viruses from the second node are the simian immunodeficiency viruses (SIVs) of African green monkeys (AGM), mandrills (MND), Syke's monkeys (SYK), sooty mangabeys (SM), and sabeus monkeys (SB). The SIV of chimpanzees branches from the next node along with the HIV-1 outlier strains (HIV-O) and the standard genotypic variants of HIV-1, shown as letters A through H. Diameter size for each variant is roughly proportional to the genetic variation within that genotype. HIV-2 is genetically very similar to SIV-SM. Known recombinant lentiviruses are shown as the juncture of two arrows from their parent genotypes.



Experimental parameters:

- **Population size: 500**
- **Initial length: 512**
- **Length of Run: 2000 generations**
- **1-point homologous crossover**
- **Best of generation, avg over 10 runs**

Generation 0: Fitness: 0.118

```
----------_CFGCNFLSAR.----------------------------------------
------_POLFHCTWNMEPQ.----------------------------------------
--------------------------------------------------_EIEPJUHME.---
```

Generation 500: Fitness: 0.856

```
----------------------------------------_COREPROF_LTKREIN.-------
--------------------
------_POLIYMERASE.-----------------------------------------------
--------------------
----------------------_ENVEBQNLOPE.-------------------------------
----------------------
```

Generation 2000: Fitness: 0.930

```
-----------------------------------_COREPROVEJTEIN.-
---_POLYMERASE.-------------------------------------
------------------_ENVEWLOPE.----------------------
```

Figure 1.3 "Learning curves" for evolving populations of strings in the virtual virus (VIV) model. Generations (replication cycles) are shown on the *x* axis, and fitness (a value based on the spelling score and adjusted for string length; see text) is shown on the y axis. In this experiment the population sizes = 500; generations = 2,000; initial genome lengths = 100 to 500; and mutation rate = 0.003 per site per replication. Curve shows a population replicating with homologous recombination at every replication.

References

1. **Ahmen, R., C. S. Hahn, T. Somasundaram, L. Villarete, M. Matloubian, and J. H. Strauss.** 1991. Molecular basis of organ-specific selection of viral variants using chronic infection. *J. Virol.* **65**:4242–4247.
2. **Burke, D. S.** 1997. Recombination in HIV: an important viral evolutionary strategy. *Emerg. Infect. Dis.* **3**:253–259.
3. **Burke, D. S., and F. E. McCutchan.** 1996. Global distribution of human immunodeficiency virus-1 clades, p. 119–126. *In* V. T. Devita, Jr., S. Hellman, and S. A. Rosenberg (ed.), *AIDS: Biology, Diagnosis, Treatment and Prevention*, 4th ed. Lippincott-Raven Publishers, Philadelphia.
4. **Clarke, D. K., E. A. Duarte, A. Moya, S. F. Elena, E. Domingo, and J. Holland.** 1993. Bottlenecks and population passages cause profound fitness differences in RNA viruses. *J. Virol.* **67**:222–228.
5. **DeJong, K. A.** 1990. Genetic-algorithm-based learning. *In* Y. Kodratoff and R. Michalski (ed.), *Machine Learning: an Artificial Intelligence Approach*, vol. 3. Morgan Kaufmann, San Francisco.
6. **DeJong, K. A., and W. M. Spears.** 1992. A formal analysis of the role of multipoint crossover in genetic algorithms. *Ann. Math. Artificial Intelligence* **5**:1–26.
7. **Dennett, D. C.** 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, p. 393. Simon & Schuster, New York.
8. **Eigen, M.** 1993. The origin of genetic information: viruses as models. *Gene* **135**:37–47.
9. **Felsenstein, J.** 1974. The evolutionary advantage of recombination. *Genetics* **78**:737–756.
10. **Felsenstein, J., and S. Yokoyama.** 1976. The evolutionary advantage of recombination. II. Individual selection for recombination. *Genetics* **83**:845–859.
11. **Gammelin, M., A. Altmuller, U. Reinhardt, and J. Mandler.** 1990. Phylogenetic analysis of nucleoproteins suggests that human influenza A viruses emerged from a 19th-century avian ancestor. *Mol. Biol. Evol.* **7**:194–200.
12. **Goldberg, D. E.** 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Co., Reading, Mass.
13. **Grefenstette, J. J.** 1988. Credit assignment in rule discovery systems based on genetic algorithms. *Mach. Learning* **3**(2/3):225–245.
14. **Holland, J. H.** 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Mich.
15. **Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. Vande-Pol.** 1982. Rapid evolution of RNA genomes. *Science* **215**:1577–1585.
16. **Hu, W.-S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. USA* **87**:1556–1560.
17. **Jin, M. J., H. Hui, and D. L. Robertson.** 1994. Mosaic genome structure of simian immunodeficiency virus from West African green monkeys. *EMBO J.* **13**:2935–2947.
18. **Kauffman, S. A.** 1993. *The Origins of Order*, p. 114–117. Oxford University Press, New York.
19. **Kondrashov, A. S.** 1984. Deleterious mutations as an evolutionary factor. I. The advantage of recombination. *Genet. Res.* **44**:199–217.

20. **LaForce, F. M., K. L. Nichol, and N. J. Cox.** 1994. Influenza: virology, epidemiology, disease, and prevention. *Am. J. Prev. Med.* **10**(Suppl.):31–44.

21. **Ludwig, S., U. Schultz, J. Mandler, W. M. Fitch, and C. Scholtissek.** 1991. Phylogenetic relationship of the nonstructural (NS) genes of influenza A viruses. *Virology* **183**:566–577.

22. **Morse, S.** 1991. Emerging viruses: defining the rules for viral traffic. *Perspect. Biol. Med.* **34**:387–409.

23. **Morse, S., and A. Schluederberg.** 1990. Emerging viruses: the evolution of viruses and viral diseases. *J. Infect. Dis.* **162**:1–7.

24. **Murphy, F. A., C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers (ed.).** 1995. *Virus Taxonomy: Classification and Nomenclature of Viruses: Sixth Report of the International Committee on Taxonomy of Viruses*, p. 586. Springer-Verlag, New York.

25. **Provine, W. B.** 1986. *Sewall Wright and Evolutionary Biology*, p. 307–317. University of Chicago Press, Chicago.

26. **Roberts, J. D., K. Bebenek, and T. A. Kunkel.** 1988. The accuracy of reverse transcriptase from HIV-1. *Science* **242**:1171–1173.

27. **Robertson, D. L., B. H. Hahn, and P. M. Sharp.** 1995. Recombination in AIDS viruses. *J. Mol. Evol.* **40**:249–259.

28. **Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn.** 1995. Recombination in HIV-1. *Nature* **374**:124–126. (Letter.)

29. **Ross, J.** 1993. Learning Boolean function with genetic algorithms: a PAC analysis. *In* D. Whitley (ed.), *Foundations of Genetic Algorithms-2*. Morgan Kaufmann, San Francisco.

30. **Schultz, U., W. M. Fitch, S. Ludwig, J. Mandler, and C. Scholtissek.** 1991. Evolution of pig influenza viruses. *Virology* **183**:61–73.

31. **Shu, L. L., W. J. Bean, and R. G. Webster.** 1993. Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. *J. Virol.* **67**:2723–2729.

32. **Temin, H. M.** 1991. Sex and recombination in retroviruses. *Trends Genet.* **7**:71–74.

33. **Woese, C.** 1969. Models for the evolution of codon assignments. *J. Mol. Biol.* **43**:235–240.

34. **Wright, S.** 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution, p. 356. *In Proceedings of the Sixth International Congress of Genetics*.

35. **Zhang, J., and H. M. Temin.** 1994. Retrovirus recombination depends on the length of sequence identity and is not error prone. *J. Virol.* **68**:2409–2414.