Virus Attenuation by Genome-Scale Changes in Codon Pair Bias

J. Robert Coleman, *et al.*

Science **320**, 1784 (2008);
DOI: 10.1126/science.1155761

introns, although the importance of such a nonproductive pathway in the splicing reaction is unknown.

References and Notes

1. M. Company, J. Arenas, J. Abelson, *Nature* **349**, 487 (1991).
2. B. Schwer, C. H. Gross, *EMBO J.* **17**, 2086 (1998).
3. B. Schwer, T. Meszaros, *EMBO J.* **19**, 6582 (2000).
4. U. Vijayraghavan *et al.*, *EMBO J.* **5**, 1683 (1986).
5. Materials and methods are available as supporting material on *Science* Online.
6. R. M. Mayas, H. Maita, J. P. Staley, *Nat. Struct. Mol. Biol.* **13**, 482 (2006).
7. K. A. Jarrell, C. L. Peebles, R. C. Dietrich, S. L. Romiti, P. S. Perlman, *J. Biol. Chem.* **263**, 3432 (1988).
8. T. Tani, Y. Ohshima, *Nature* **337**, 87 (1989).
9. S. Augustin, M. W. Müller, R. J. Schweyen, *Nature* **343**, 383 (1990).
10. M. Mörl, C. Schmelzer, *Cell* **60**, 629 (1990).
11. D. A. Brow, *Annu. Rev. Genet.* **36**, 333 (2002).
12. C. L. Will, R. Lührmann, in *The RNA World,* R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory, New York, 2006), pp. 369–400.
13. S.-H. Kim, R.-J. Lin, *Mol. Cell. Biol.* **16**, 6810 (1996).
14. B. Schwer, C. Guthrie, *EMBO J.* **11**, 5033 (1992).
15. C. C. Query, M. M. Konarska, *Mol. Cell* **14**, 343 (2004).

16. We are grateful to T. Nilsen for critical comments and help on the manuscript. We also thank M. M. Konarska and C. Query for reading the manuscript and H. Wilson for English editing. This work was supported by a grant from Academia Sinica and National Science Council (Taiwan) grant NSC95-2321-B-001-014.

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5884/1782/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 and S2
References

11 April 2008; accepted 22 May 2008
10.1126/science.1158993

Virus Attenuation by Genome-Scale Changes in Codon Pair Bias

J. Robert Coleman,[1] Dimitris Papamichail,[2]* Steven Skiena,[2] Bruce Futcher,[1] Eckard Wimmer,[1]† Steffen Mueller[1]

As a result of the redundancy of the genetic code, adjacent pairs of amino acids can be encoded by as many as 36 different pairs of synonymous codons. A species-specific "codon pair bias" provides that some synonymous codon pairs are used more or less frequently than statistically predicted. We synthesized de novo large DNA molecules using hundreds of over- or underrepresented synonymous codon pairs to encode the poliovirus capsid protein. Underrepresented codon pairs caused decreased rates of protein translation, and polioviruses containing such amino acid–independent changes were attenuated in mice. Polioviruses thus customized were used to immunize mice and provided protective immunity after challenge. This "death by a thousand cuts" strategy could be generally applicable to attenuating many kinds of viruses.

The redundancy of the genetic code means that a typical 300–amino acid protein can be encoded in about 10^{151} ways, raising the question of to what extent the actual encoding is optimal. Actual encodings are biased to use some synonymous codons more frequently than others (the "codon bias"). For instance, in humans, the Ala codon GCC is used four times as frequently as the synonymous codon GCG. Similarly, but independently, some synonymous codon pairs are used more or less frequently than expected (the "codon pair bias") (*1*). For instance, on the basis of codon frequencies, the amino acid pair Ala-Glu is expected to be encoded by GCCGAA and GCAGAG about equally often. In fact, the codon pair GCCGAA is strongly underrepresented, even though it contains the most frequent Ala codon, such that it is used only one-seventh as often as GCAGAG (*2*) (table S1). Although it is not clear why some codon pairs are under- or overrepresented, it is possible that codon pair usage affects translation (*3*).

[1]Department of Molecular Genetics and Microbiology, Stony Brook University, Stony Brook, NY 11794, USA. [2]Department of Computer Science, Stony Brook University, Stony Brook, NY 11794, USA.

*Present address: Department of Computer Science, University of Miami, Coral Gables, FL 33124, USA.

†To whom correspondence should be addressed. E-mail: ewimmer@ms.cc.sunysb.edu

We previously reported the generation of poliovirus de novo in the absence of natural template (*4*), using reverse genetics and the ability to synthesize large DNAs. We and others recently synthesized novel polioviruses encoding precisely the same amino acid sequences as wild-type poliovirus, but using rare codons (*5*, *6*); these viruses were attenuated. Here, we used poliovirus as a model system to explore the consequences of genome-scale manipulation of codon pair bias (Fig. 1A). We call the process of designing such viruses "synthetic attenuated virus engineering" or SAVE.

We developed a computer algorithm that can recode a given amino acid sequence, but using different codon pairs, while controlling other features of the sequence such as the codon bias and the folding free energy of the RNA (*2*) (figs. S1 and S2). This algorithm was used to design two new polioviruses, PV-Min and PV-Max (Fig. 1), with a P1 region (encoding the viral capsid, 2643 nucleotides) containing under- or overrepresented codon pairs. The P1 region is suitable for such experiments because it can be deleted or substituted without affecting genome replication (*7*). Virus PV-Min was recoded to use codon pairs that are underrepresented relative to the human genome, and it contains 631 synonymous mutations. Virus PV-Max was recoded to use overrepresented codon pairs, and it contains 566 synonymous mutations. PV-Min has a codon pair bias score (CPB score) much lower than that of normal human genes (Fig. 1B). Both PV-Min and PV-Max encode precisely the same amino acid sequences as the wild type, but they use different pairwise arrangements of synonymous codons [Fig. 1B; for calculation of CPB scores, see (*2*)]. These P1 fragments were synthesized, sequenced, and incorporated into a full-length cDNA construct of poliovirus (Fig. 1C) (*2*, *8*).

In vitro transcribed RNAs of PV-Max, PV-Min, and wild-type virus were transfected into HeLa R19 cells to assess virus production (*5*, *8*). PV-Max produced 90% cytopathic effect within 24 hours after RNA transfection, similar to the transfection of wild-type RNA (*8*). The PV-Max virus generated plaques identical in size to the wild type (Fig. 1F). In contrast, the PV-Min RNA produced no visible cytopathic effect after 96 hours, and no viable virus could be isolated even after four blind passages of the supernatant from transfected cells.

We subcloned portions of the PV-Min P1 region into an otherwise wild-type virus to reduce the number of underrepresented codon pairs (Fig. 1C). These subclones yielded viruses with varying degrees of attenuation (Fig. 1, C, D, and F). Viruses containing P1 fragments X and Y were each slightly attenuated; however, when added together they yielded virus PV-MinXY, which was substantially attenuated (Fig. 1C). Virus PV-MinZ was about as attenuated as PV-MinXY. Construct PV-YZ did not yield viable virus (Fig. 1, C and D). We conclude that the inviability of PV-Min was due to the sum of defects in the various subportions.

One-step growth kinetics were examined. Like the wild-type virus, the chimeric viruses had an eclipse phase followed by exponential growth. However, as measured by plaque-forming units (PFUs), the final titer of PV-Min constructs was decreased by up to a factor of 1000 with respect to wild-type viruses (Fig. 1D). This low plaque titer could have resulted from lower production of virions (i.e., cells infected with PV-Min constructs produced fewer virus particles), or from lower specific infectivity of those virions (i.e., the particles that were produced were less efficient in establishing a plaque), or both. We examined both possibilities (*2*). When the number of viral particles produced per infected cell

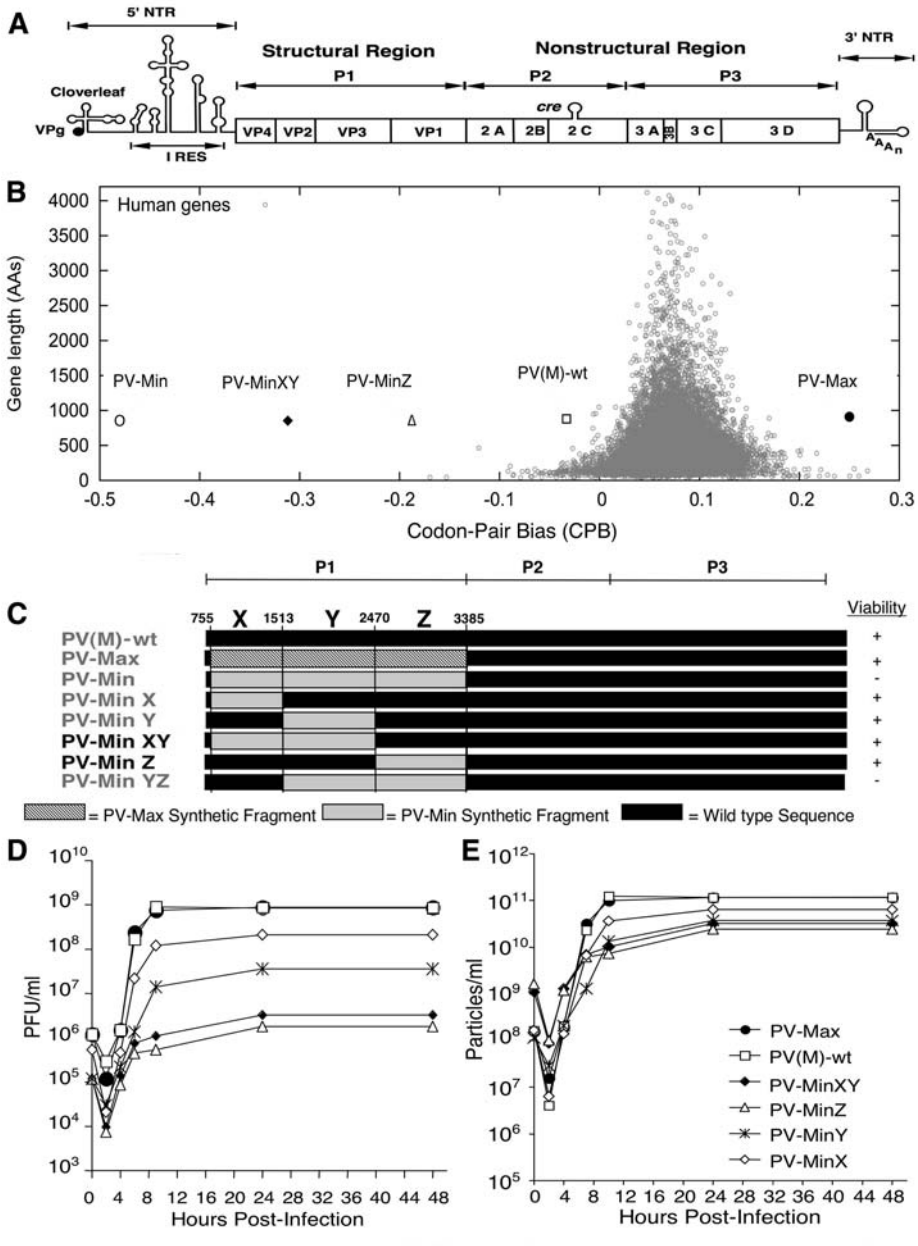

Fig. 1. (**A**) The poliovirus genome. Shown is the viral RNA with its covalently linked 5′ viral protein VPg, the 5′ nontranslated region [consisting of cloverleaf and internal ribosomal entry site (IRES)]; the long open reading frame (open box) encoding the polyprotein that is cleaved into P1 (capsid precursor), P2, and P3 (precursors to nonstructural proteins); and the 3′ nontranslated region terminated with poly(A) (*7*). The polypeptide precursors P1, P2, and P3 are processed into functional proteins by virus-encoded proteinases (*7*). *Cre* is a stem-loop structure functioning as a cis-acting replication element. (**B**) Calculated codon pair bias (CPB) score for all 14,795 annotated human genes. Each dot represents the calculated CPB score of a gene plotted against its amino acid length. Underrepresented codon pairs yield negative scores. Various poliovirus constructs are represented by symbols; PV(M)-wt is the wild-type poliovirus (CPB = −0.02). (**C**) The structures of the various chimeric, partly synthetic poliovirus constructs, and their viability on cultured cells. Nucleotide positions are shown. (**D**) A one-step growth curve with respect to PFUs. An MOI of 2 was used to infect a monolayer of HeLa R19 cells. Symbols: open squares, PV(M)-wt; solid circles, PV-Max; open diamonds, PV-Min755-1513; asterisks, PV-Min1513-2470; solid diamonds, PV-MinXY; open triangles, PV-MinZ. (**E**) As (D), but with results graphed with respect to viral particles instead of PFUs. (**F**) Plaque phenotypes of viruses after 72 hours of incubation, stained with crystal violet (plate diameter, 35 mm).

was measured, we found that cells infected with PV-MinXY or PV-MinZ produced fewer viral particles than did the wild type, but the effect was only about a factor of 10 or slightly less (Fig. 1E and Table 1). When the number of particles per PFU was measured, we found a more striking effect: PV-MinXY and PV-MinZ each required about 100 times as many viral particles as the wild type to generate a plaque (Table 1). For wild-type virus, the number of virions applied per plaque generated was about 137, whereas for PV-MinZ, the number of virions applied per plaque generated was 13,500; hence, the main defect was reduced specific infectivity of the virions. The total attenuation from both effects together was a factor of ~1000.

The heat stability of PV-MinXY and PV-MinZ was identical to that of the wild type. This observation suggests that their low specific infectivity is not a result of gross defects in the capsid (*2*) (fig. S3).

To measure the possible effect of codon pair bias on translation, we used a dicistronic reporter encoding both R-Luc and F-Luc (*2*, *5*) (Fig. 2A). Because the F-Luc reporter is translated as a fusion protein with the proteins of the P1 region, the translatability of the P1 region directly affects the amount of F-Luc protein produced. Thus, the ratio of F-Luc luminescence to R-Luc luminescence is a measure of the translatability of the various P1 encodings.

The variously encoded P1 regions were tested (Fig. 2). PV-MinXY, PV-MinZ, and PV-Min produced much less F-Luc per unit of R-Luc than did the wild-type P1 region, which strongly suggests that the underrepresented codon pairs reduced translation (Fig. 2B). The reduced translation is probably sufficient to explain the attenuated phenotype, because smaller reductions in translation caused by other methods have been observed to attenuate poliovirus; apparently, poliovirus has a fairly high threshold requirement for translation (*5*, *9*). In contrast, PV-Max P1 produced more F-Luc per unit of R-Luc than did the wild type, consistent with enhanced translation (Fig. 2B).

PV-MinXY and PV-MinZ each contain hundreds of mutations (407 and 224, respectively). If the attenuation of these viruses were due to hundreds of small defects, it should be difficult for these viruses to revert to wild-type virulence. Alternatively, if most mutations are neutral, with a small minority contributing to attenuation, then reversion should occur. To distinguish these possibilities, we serially passaged viruses PV-MinXY and PV-MinZ in HeLa R19 cells 17 and 19 times, respectively, at a multiplicity of infection (MOI) of 0.5. The titer was monitored for phenotypic reversion, and the passaged virus was sequenced. After 17 or 19 passages of PV-MinXY or PV-MinZ, respectively, no phenotypic change was detected (i.e., same titer, induction of cytopathic effect) and no nucleotide changes were seen in the synthetic region, supporting the idea that there are many small defects.

We next tested whether the synthetic viruses were also attenuated in animals. Viruses were administered to CD155 transgenic (tg) mice (which express the poliovirus receptor) via intracerebral injection (*10*), allowing direct exposure to the central nervous system, the ultimate target of poliovirus pathogenesis (*11*). PV-MinXY and PV-MinZ viruses were attenuated by a factor of 1000 (as measured by particles) or a factor of 10 (as measured by PFUs) (Table 1) (*2*). PV-Max virulence was identical to that of the wild type.

Because PV-MinZ and PV-MinXY encode exactly the same proteins as wild-type virus, they might provoke a protective immune response. Alternatively, the relatively poor translation of the mutant mRNAs might prevent such a response. To distinguish these possibilities, we administered PV-MinZ and PV-MinXY to groups of eight CD155 tg mice at a dose of 10^8 particles once a week for 3 weeks via intraperitoneal injection. Ten days after the final injection, the protective antibodies of the seven surviving mice in each group were measured via microneutralization assay, and a robust immune response was detected (fig. S3). Subsequent challenge of the immunized mice with an otherwise lethal dose of wild-type poliovirus via intramuscular injection did not lead to death or signs of paralysis or paresia; in contrast, all mock-immunized mice died.

Technology for the synthesis of large DNAs and for the redesign of living systems (*5*, *6*, *12*) allows the reengineering of viruses for specific purposes such as vaccines. We have used these approaches to generate polioviruses that use a large proportion of under- or overrepresented codon pairs. Although it has been known for many years that codon pair usage is biased (*1*), this phenomenon has previously been studied primarily by informatics (*13*, *14*). We now find that underrepresented codon pairs cause poor translation and attenuation in poliovirus. One theory for the existence of codon pair bias is that certain tRNAs interact poorly on the ribosome (*3*), and so the codon pairs causing the juxtaposition of such tRNAs are underrepresented; our translation data are consistent with this theory.

We note that attenuation is not caused by random changes in synonymous codons if those changes do not systematically reduce codon bias or codon pair bias. Previously we created the virus PV-SD, with 937 mutations in synonymous codons in the P1 region (*5*). In PV-SD, neither codon bias nor codon pair bias was changed, and that virus was not attenuated (*2*, *5*) (table S2). Here, we created virus PV-Max, which similarly contained 566 mutations in synonymous codons, and it was also not attenuated (Fig. 2). It is noteworthy that even though PV-Max contains overrepresented codons, it is not more virulent than the wild type (Table 1), possibly because evolution has already effectively optimized encoding.

The correlation between the degree of codon pair deoptimization and the degree of viral attenuation, as well as the lack of viral reversion upon passaging, are consistent with the idea that many of the 631 mutations in PV-Min cause small, additive defects. Thus, these attenuated viruses should be stable genetically, because an increase in virulence might require dozens or hundreds of reversions. This genetic stability is important; for example, the oral poliovirus vaccine (OPV) has 51 mutations, but only 5 have been shown to contribute to attenuation (*15*). Thus, OPV can (rarely) revert to neurovirulence in vaccine recipients, causing vaccine-associated paralytic poliomyelitis. Even more seriously, the vaccine strains may evolve into highly virulent circulating vaccine–derived polioviruses by mutation and recombination with related Coxsackie A viruses (*15*, *16*). Such viruses have caused small epidemics of poliomyelitis (*15*, *17*).

Finally, these results suggest that synthetic attenuated virus engineering (SAVE) could play a role in creating new vaccines for various types of viruses. By deoptimizing codon pair bias, one could systematically attenuate a virus to variable but controllable and predictable extents. This approach has four key features: (i) It produces a virus encoding precisely the same amino acid sequences as the wild-type virus, and therefore eliciting the same immune response. (ii) It is a systematic method apparently applicable to many viruses, and possibly not requiring detailed, virus-specific research. (iii) The attenuation is not subject to reversion, simply because of the sheer number of mutations. (iv) It can be combined with other attenuating changes (such as amino acid changes from adaptation of the virus to low temperatures or alternative species) or with other synthetic biology approaches to attenuation (*18*, *19*), thus taking advantage of additional modes of attenuation while providing the unique advantage of limited reversion. This "death by a thousand cuts" strategy is in contrast to existing methods of attenuation, which typically depend on a small number of mutations, and which can revert. Even for an inactivated rather than live virus approach, these features would allow a vaccine to be made from a safer starting material than the corresponding wild-type virus.

Table 1. Poliovirus specific infectivity and attenuation. PLD_{50} is the amount of virus that caused paralysis in 50% of infected mice.

Virus	Absorbance at 260 nm	Purified particles/ml*	Purified PFUs/ml	Relative specific infectivity†	PLD_{50} (particles)‡	PLD_{50} (PFUs)‡
PV(M)-wt	0.956	8.97×10^{12}	6.0×10^{10}	1.00	$10^{4.0}$	$10^{1.9}$
PV-Max	0.842	7.91×10^{12}	6.0×10^{10}	1.04	$10^{4.0}$	$10^{1.9}$
PV-MinXY	0.944	8.87×10^{12}	9.6×10^{8}	0.015	$10^{7.1}$	$10^{3.2}$
PV-MinZ	0.731	6.87×10^{12}	5.1×10^{8}	0.010	$10^{7.3}$	$10^{3.2}$

*Calculated by the formula 9.4×10^{12} particles/ml = 1 OD_{260} (*5*). †Calculated by dividing the PFU/ml of purified virus on HeLa R19 cells by particles/ml, and then normalizing to the wild-type value. In absolute terms, wild-type specific infectivity was 1/137. ‡Calculated after intracerebral injection of virus into CD155 tg mice at varying doses.

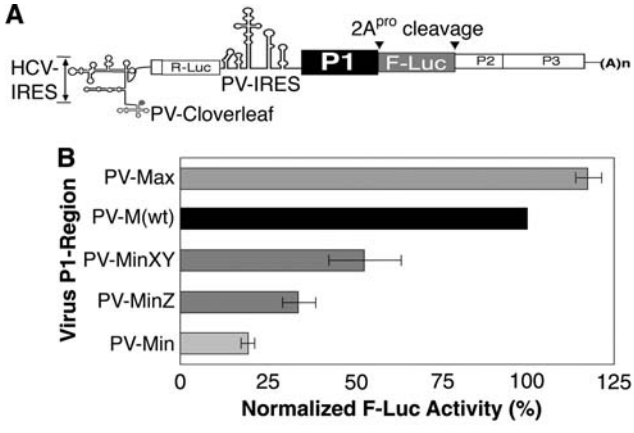

Fig. 2. Effect of altered codon pair bias on translation. (**A**) Structure of a dicistronic reporter (*5*). The first cistron uses the hepatitis C virus (HCV) IRES to initiate translation of *Renilla* luciferase (R-Luc). This first cistron provides an internal control to normalize the amount of input RNA. The second cistron uses the poliovirus IRES to initiate translation of Firefly luciferase (F-Luc). The region labeled P1 was replaced by the recoded, synthetic P1 regions of the indicated viruses. (**B**) Each dicistronic RNA was transfected, in the presence of 2 mM guanidine hydrochloride [to block replication of the transfected genome (*7*)], into HeLa R19 cells, and after 6 hours the R-Luc and F-Luc were measured. The F-Luc/R-Luc values are expressed relative to the wild type, which was set to 100%. The graph displays the average (±SD) of three independent experiments.

References and Notes

1. G. A. Gutman, G. W. Hatfield, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3699 (1989).
2. See supporting material on *Science* Online.
3. J. Buchan, *Nucleic Acids Res.* **34**, 1015 (2006).
4. J. Cello, A. Paul, E. Wimmer, *Science* **297**, 1016 (2002); published online 11 July 2002 (10.1126/science.1072266).
5. S. Mueller, D. Papamichail, J. R. Coleman, S. Skiena, E. Wimmer, *J. Virol.* **80**, 9687 (2006).
6. C. C. Burns *et al.*, *J. Virol.* **80**, 3259 (2006).
7. E. Wimmer, C. U. Hellen, X. Cao, *Annu. Rev. Genet.* **27**, 353 (1993).
8. S. van der Werf, J. Bradley, E. Wimmer, F. W. Studier, J. J. Dunn, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2330 (1986).

9. W. D. Zhao, E. Wimmer, *J. Virol.* **75**, 3719 (2001).
10. S. Koike, in *Cellular Receptors for Animal Viruses*, E. Wimmer, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1994), pp. 463–480.
11. K. Landsteiner, E. Popper, *Z. Immun. Forsch. Orig.* **2**, 377 (1909).
12. L. Chan, S. Kosuri, D. Endy, *Mol. Syst. Biol.* **1**, 10.1038/msb4100025 (2005).
13. A. Fedorov, S. Saxonov, W. Gilbert, *Nucleic Acids Res.* **30**, 1192 (2002).
14. G. Moura *et al.*, *PLoS ONE* **2**, e847 (2007).
15. O. Kew, R. Sutter, E. De Gourville, W. Dowdle, M. Pallansch, *Annu. Rev. Microbiol.* **59**, 587 (2005).
16. P. Jiang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9457 (2007).
17. O. Kew *et al.*, *Science* **296**, 356 (2002); published online 14 March 2002 (10.1126/science.1068284).
18. E. B. Flanagan, J. M. Zamparo, L. A. Ball, L. L. Rodriguez, G. W. Wertz, *J. Virol.* **75**, 6107 (2001).
19. H. Toyoda, J. Yin, S. Mueller, E. Wimmer, J. Cello, *Cancer Res.* **67**, 2857 (2007).
20. We thank H. Toyoda for lending his expertise to the mouse experiments, W. McCaig for performing serial passages of viral variants, and J. Cello and W. Karzai for comments on the manuscript. All authors declare that they have a patent pending relating to certain aspects of this work. Supported by NIH grants AI075219 and AI15122 (E.W.) and NSF grant EIA-0325123 (S.S.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5884/1784/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 and S2
References

28 January 2008; accepted 27 May 2008
10.1126/science.1155761

Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland

M. Thomas P. Gilbert,[1] Toomas Kivisild,[2] Bjarne Grønnow,[3] Pernille K. Andersen,[4] Ene Metspalu,[5] Maere Reidla,[5] Erika Tamm,[5] Erik Axelsson,[1] Anders Götherström,[6] Paula F. Campos,[1] Morten Rasmussen,[1] Mait Metspalu,[5] Thomas F. G. Higham,[7] Jean-Luc Schwenninger,[7] Roger Nathan,[7] Cees-Jan De Hoog,[8] Anders Koch,[9] Lone Nukaaraq Møller,[10]* Claus Andreasen,[11] Morten Meldgaard,[12] Richard Villems,[5] Christian Bendixen,[4] Eske Willerslev[1]†

The Paleo-Eskimo Saqqaq and Independence I cultures, documented from archaeological remains in Northern Canada and Greenland, represent the earliest human expansion into the New World's northern extremes. However, their origin and genetic relationship to later cultures are unknown. We sequenced a mitochondrial genome from a Paleo-Eskimo human by using 3400- to 4500-year-old frozen hair excavated from an early Greenlandic Saqqaq settlement. The sample is distinct from modern Native Americans and Neo-Eskimos, falling within haplogroup D2a1, a group previously observed among modern Aleuts and Siberian Sireniki Yuit. This result suggests that the earliest migrants into the New World's northern extremes derived from populations in the Bering Sea area and were not directly related to Native Americans or the later Neo-Eskimos that replaced them.

Studies into the peopling of the New World have demonstrated that a number of cultures have inhabited the geographic region encompassed by Greenland and the northern extremes of the American continent. Archaeological evidence has provided several insights into the Paleo-Eskimo cultures that ultimately migrated into Greenland, including that they comprised at least two temporally distinct extinct, and one extant, groups. Known collectively as the Paleo-Eskimos, the two waves of the extinct cultures are represented by the Independence I–Saqqaq and Pre-Dorset culture, which spanned about 3900 to 2500 ^{14}C years before present (yr B.P.), and the Independence II–Dorset cultures of circa (ca.) 2500 to 700 ^{14}C yr B.P. (*1*). Subsequently the region was peopled by the Neo-Eskimo Thule culture, a group genetically distinct from modern Native Americans and whose present-day descendants include the modern Alaskan Yupik and Inupiat and the Canadian and Greenlandic Inuit [Fig. 1; see (*2*) for subdistinction of cultures]. The genetic source of the Paleo-Eskimo cultures and their relationship to each other and to the Neo-Eskimos have not been determined. Competing theories have attributed the origin of the Paleo-Eskimos to an offshoot of the populations that gave rise to the Native American populations of North America, alternatively from the same Beringian source as the Neo-Eskimos, or even from a source population that was distinct from both the Native Americans and the Neo-Eskimos (*1*).

The majority of modern and ancient Neo-Eskimo populations have been classified with mitochondrial DNA (mtDNA) sequences of the hypervariable I (HVS1) region. Belonging mainly to haplogroups (Hgs) A2a and A2b (*3*–*7*) with a low representation (≈5%) of Hg D3 (*3*), these groups exhibit relatively low sequence variation. This supports archaeological evidence suggesting that the Thule culture originated in Alaska about 1000 years ago and spread rapidly over the next 200 years across northern Canada and Greenland (*8*).

Ancient DNA (aDNA) studies can be used to uncover genetic signatures that have been lost through population extinctions, as long as there is suitably well-preserved material for analysis. However, Paleo-Eskimo human material is scarce, despite the relative youth of the cultures and the cold preservation conditions offered by the Arctic. Thus, aDNA studies of this region have been limited to three Canadian skeletons belonging to the Dorset culture and between 800 and 1800 years old (*9*, *10*).

Archaeological excavations between 1983 and 1987 at Qeqertasussuk, Disko Bay, in western Greenland (*11*, *12*) (Fig. 1) yielded human remains from the first Paleo-Eskimo expansion into the far north of the New World. Stone tools, coupled with ^{14}C-dated twigs found among the artifacts, show that the site occupants belong both culturally and temporally to the early Saqqaq culture, 3900 to 3100 ^{14}C yr B.P. Beside four poorly preserved human long bones (*13*), the site yielded a large clump of permafrost-preserved hair that was morphologically identified as human (*2*) (fig. S1). Because hair shaft are often a suitable source of well-preserved, contaminant-free aDNA (*14*, *15*), we investigated how the Saqqaq hair clump offered insights into Paleo-Eskimo genetics.

To investigate the long-term DNA survival at the site, we first obtained 16*S* mtDNA from bowhead whale baleen (*Balaena mysticetus*) recovered from within the hair sample itself (*2*) (table S1). We then reproducibly polymerase chain reaction (PCR)–amplified, cloned, and sequenced mtDNA HVS1 sequences from the human hair and found that they appeared to derive from a single clade, Hg D2. Subsequent typing of an Hg D diagnostic C→A single-nucleotide polymorphism (SNP) at nucleotide position (n.p.) 5178 confirmed this classification (*2*) (tables S1 to S3).

[1]Center for Ancient Genetics, Department of Biology, Universitetsparken 15, DK-2100 Copenhagen, Denmark. [2]Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, Cambridge CB2 1QH, UK. [3]SILA (The Greenland Research Centre at the National Museum of Denmark), Frederikisholms Kanal 12, DK1220 Copenhagen, Denmark. [4]Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, University of Aarhus, Post Office Box 50, DK-8830 Tjele, Denmark. [5]Department of Evolutionary Biology, University of Tartu and Estonian Biocentre, Riia 23B, Tartu, 51010 Estonia. [6]Department of Evolutionary Biology, Uppsala University, Norbyvagten 18D, 74236 Uppsala, Sweden. [7]Research Laboratory for Archaeology and the History of Art, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK. [8]Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK. [9]Department of Epidemiology Research, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark. [10]Danish Centre for Experimental Parasitology, Department of Veterinary Pathobiology, Faculty of Life Sciences, University of Copenhagen, Bülowsvej 17, DK-1870 Frederiksberg C, Denmark. [11]Greenland National Museum and Archives, 3900 Nuuk, Greenland. [12]Natural History Museum of Denmark, Geologisk Museum, Øster Voldgade 5-7, DK-1350 Copenhagen, Denmark.

*Present address: Ilaqutariinnermut Pitsaaliuinermullu Aqutsisoqarfik/Family and Prevention Agency, PAARISA, Box 1160, 3900 Nuuk, Greenland.

†To whom correspondence should be addressed. E-mail: ewillerslev@bi.ku.dk