VIRUS EVOLUTION

Virus Evolution, 2019, 5(1): vez001

doi: 10.1093/ve/vez001
Rapid Communication

A reptilian endogenous foamy virus sheds light on the early evolution of retroviruses

Xiaoman Wei,[1,2,†] Yicong Chen,[1,2,†] Guangqian Duan,[2] Edward C. Holmes,[3,‡] and Jie Cui[1,*,§]

[1]Key Laboratory of Special Pathogens and Biosafety, Center for Emerging Infectious Diseases, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China, [2]University of Chinese Academy of Sciences, Beijing 100049, China and [3]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and Faculty of Medicine and Health, University of Sydney, Sydney, NSW 2006, Australia

*Corresponding author: E-mail: jiecui@wh.iov.cn

†These authors contributed equally to this work.

‡http://orcid.org/0000-0001-9596-3552

§http://orcid.org/0000-0001-8176-9951

Abstract

Endogenous retroviruses (ERVs) represent host genomic 'fossils' of ancient viruses. Foamy viruses, including those that form endogenous copies, provide strong evidence for virus-host co-divergence across the vertebrate phylogeny. Endogenous foamy viruses (EFVs) have previously been discovered in mammals, amphibians, and fish. Here we report a novel endogenous foamy virus, termed ERV-Spuma-Spu, in genome of the tuatara (*Sphenodon punctatus*), an endangered reptile species endemic to New Zealand. Phylogenetic analyses revealed that foamy viruses have likely co-diverged with their hosts over many millions of years. The discovery of ERV-Spuma-Spu fills a major gap in the fossil record of foamy viruses and provides important insights into the early evolution of retroviruses.

Key words: endogenous retroviruses; foamy virus; reptiles; evolution; tuatara

Introduction

Retroviruses (family *Retroviridae*) are viruses of major medical significance as some are associated with severe infectious disease or are oncogenic (Hayward, Cornwallis, and Jern 2015; Aiewsakun and Katzourakis 2017; Xu et al. 2018). Retroviruses are also of note because of their ability to integrate into the host germ-line, generating endogenous retroviruses (ERVs) that then exhibit Mendelian inheritance (Stoye 2012; Johnson 2015). ERVs are widely distributed in vertebrates (Hayward, Grabherr, and Jern 2013; Cui et al. 2014; Hayward, Cornwallis, and Jern 2015; Xu et al. 2018) and constitute important molecular fossils for the study of retrovirus evolution. ERVs related to all seven major retroviral genera (alpha-, beta-, delta-, epsilon-, gamma-, lenti-, and spuma-) have been described (Hayward, Cornwallis, and Jern 2015), although some of the more complex retroviruses, such as lenti-, delta- and foamy viruses, rarely appear as endogenous copies.

As well as being agents of disease, foamy viruses are of importance because they exhibit long-term virus-host co-divergence (Switzer et al. 2005). Endogenous foamy viruses (EFVs), first discovered in sloths (class Mammalia) (Katzourakis et al. 2009) also co-diverge with their hosts, and have also been

reported in primates and the Cape golden mole (Han and Worobey 2012b, 2014). The subsequent discovery of a EFV in the coelacanth genome indicated that foamy viruses may have an ancient evolutionary history (Han and Worobey 2012a), likely spanning hundreds of million years (Aiewsakun and Katzourakis 2017). Although EFVs or foamy-like elements have been reported in fish, amphibians, and mammals, to date they have not been reported in genomes of two other major classes of vertebrates—reptiles and birds (Tristem, Myles, and Hill 1995; Herniou et al. 1998; Hayward, Cornwallis, and Jern 2015; Xu et al. 2018).

Materials and methods

Genomic mining and consensus genome construction

To identify foamy viruses in reptiles, the TBLASTN program (Altschul et al. 1990) was used to screen relevant taxa from 28 reptile genomes (Supplementary Table S1) and 130 bird genomes (Supplementary Table S2) (as of October 2018) downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank). In each case, the amino acid sequences of the Pol genes of representative EFVs, endogenous foamy-like viruses, and exogenous foamy viruses were chosen as queries (Supplementary Table S3). As filters to identify significant and meaningful hits, we chose sequences with more than 30 per cent amino acid identity over a 30 per cent region, with an *e*-value set to 0.00001. Genomes that contained only single hits for EFVs were excluded as likely false positives. We extended viral flanking sequences of the hits to identify the 5′- and 3′-long terminal repeats (LTRs) using LTR finder (Xu and Wang 2007) and LTR harvest (Ellinghaus Kurtz, and Willhoeft 2008). Sequences highly similar to foamy virus proteins found in tuatara were termed 'ERV-Spuma.*n*-Spu' (in which *n* represents the number of the sequence extracted from this tuatara genome) according a recently proposed nomenclature for ERVs (Gifford et al. 2018), and aligned to generate an ERV-Spuma-Spu consensus genome (ERV-Spuma.0-Spu) (Supplementary Table S4). Conserved domains were identified using CD-Search service in NCBI (Marchler-Bauer and Bryant 2004).

Molecular dating of integration times

The ERV integration time can be estimated using the following simple relation: $T = (D/R)/2$, in which T is the integration time (million years, MY), D is the number of nucleotide differences per site between the two LTRs, and R is the genomic substitution rate (i.e. number of nucleotide substitutions per site, per year). We used the previously estimated neutral substitution rate for squamate reptiles (7.6×10^{-10} nucleotide substitutions per site, per year) (Perry et al. 2018). LTRs less than 300 bp in length were not included in this analysis. Five pairwise LTRs were used for date estimation (Supplementary Table S5).

Phylogenetic analysis

To determine the evolutionary relationship of EFVs and retroviruses, sequences of the Pol proteins were aligned using MAFFT 7.222 (Katoh and Standley 2013) and confirmed manually in MEGA7 (Kumar, Stecher, and Tamura 2016). The phylogenetic relationships among these sequences were then determined using the maximum-likelihood (ML) method in PhyML 3.1 (Guindon et al. 2010), incorporating 100 bootstrap replicates to determine node robustness. The best-fit models of amino acid substitution were determined by ProtTest 3.4.2 (Abascal, Zardoya, and Posada 2005): RtREV + Γ+I for Pol, and LG + Γ+I + F for concatenated Gag, Pol, and Env. All alignments used in the phylogenetic analyses can be found in Supplementary data sets S1 and S2.

Results and discussion

Discovery of foamy viral elements in reptile genomes

We screened all available reptilian and bird genomes by using the TBLASTN algorithm with various foamy viruses, including EFVs, as screening probes. We only considered viral hits within long genomic scaffold (>20 kilobases in length) to be *bona fide* ERVs. This genomic mining identified 118 ERV hits in tuatara (*Sphenodon punctatus*) and none in bird genomes. Hence, a total of 118 ERV hits in the tuatara genome were extracted and subjected to evolutionary analysis (Supplementary Table S6) and these ERVs were named as ERV-Spuma.*n*-Spu (where $n = 1 \sim 118$).

Genomic organization

We extracted all significant foamy viral elements and constructed a consensus genomic sequence of ERV-Spuma-Spu (Supplementary Fig. S1, Table S4), termed ERV-Spuma.0-Spu. The consensus genome harboured a pairwise LTRs and exhibits a typical spuma virus structure, encoding three main open reading frames (ORF)—*gag*, *pol*, and *env*—and one putative additional accessory gene, ORF 1 (Fig. 1). Interestingly, this accessory ORF 1 exhibits no sequence similarity to known foamy accessory genes. Notably, by searching the Conserved Domains Database (www.ncbi.nlm.nih.gov/Structure/cdd), we identified three typical foamy conserved domains for both the consensus and one of two full-length original ERV-Spuma.23-Spu: (1) Spuma virus Gag domain (pfam03276) (Winkler et al. 1997), (2) Spuma aspartic protease (A9) domain (pfam03539) which exists in all mammalian foamy virus Pol protein (Aiewsakun and Katzourakis 2017), and (3) foamy virus envelope protein domain (pfam03408) (Han and Worobey 2012a) (Supplementary Figs S2 and S3), confirming that ERV-Spuma-Spu is indeed of foamy virus origin.

Estimated integration times

To broadly estimate the integration time of ERV-Spuma-Spu, we utilized the LTR-divergence method which analyzes the degree of divergence between 5′ and 3′LTRs assuming a known rate of

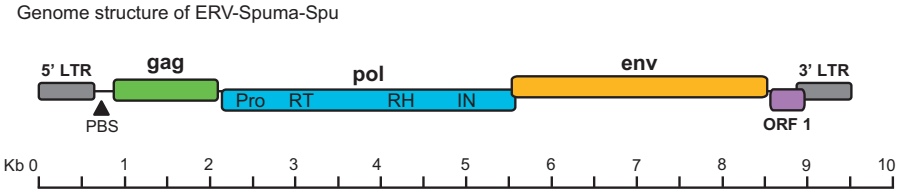

Figure 1. Genomic organizations of ERV-Spuma-Spu. LTR, long-terminal repeat; PBS, primer-binding site; Pro, aspartic protease; RT, reverse transcriptase; RH, ribonuclease H; IN, integrase.

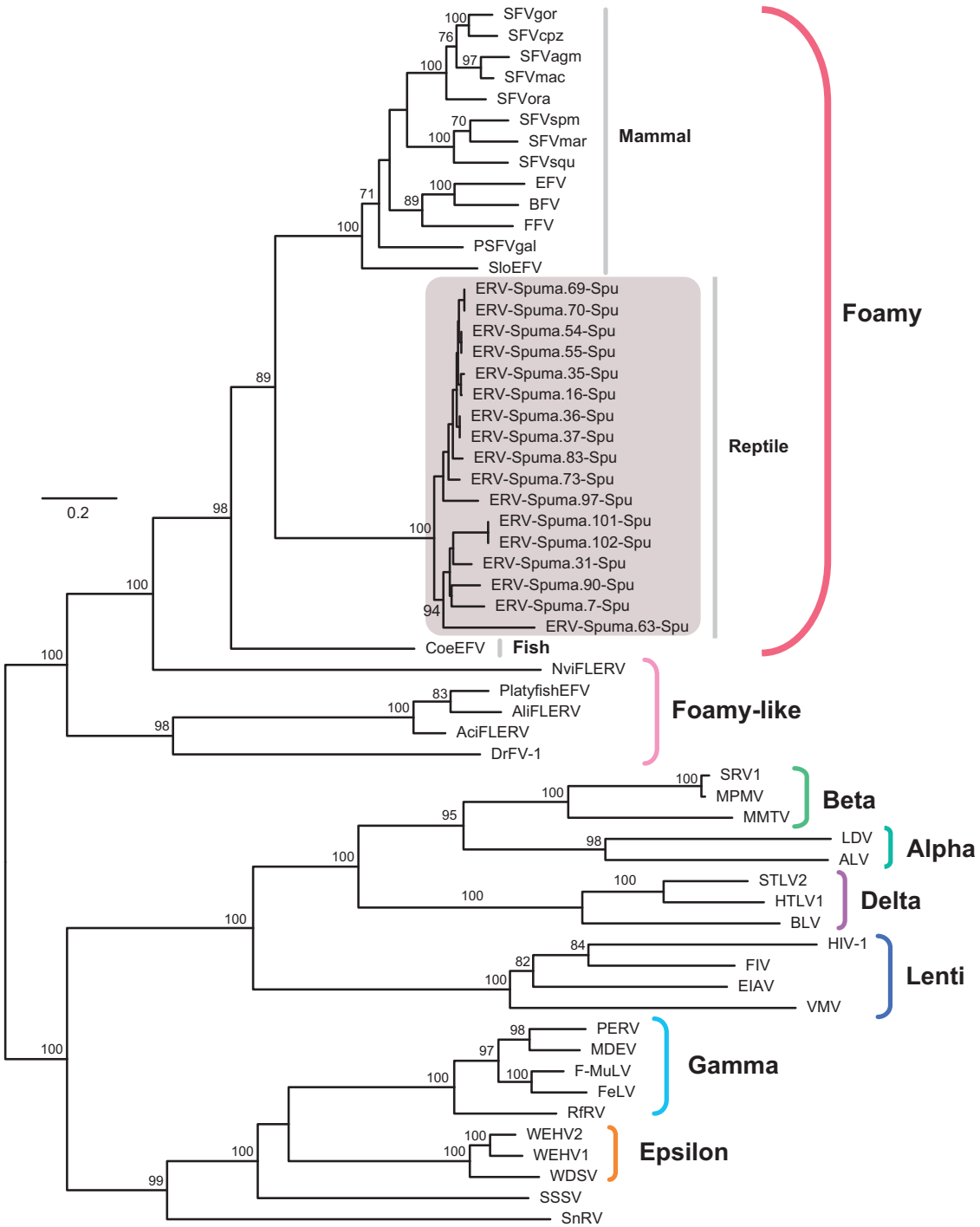

Figure 2. Phylogenetic tree of retroviruses, including ERV-Spuma-Spu, inferred using amino acid sequences of the Pol gene (490aa). The tree is midpoint rooted for clarity only. The newly identified ERV-Spuma-Spu sequences are labelled using a grey-shaded box with their accession numbers (different pol sequences in same contig are numbered in the suffix). The scale bar indicates the number of amino acid changes per site. Bootstrap values <70 per cent are not shown. The alignment of pol amino acid sequences is provided in Supplementary data set S1.

nucleotide substitution (Johnson and Coffin 1999). In total, five pairwise LTRs flanking ERV-Spuma-Spu elements were used for date estimation (Supplementary Table S5), from which we estimated an integration time of ERV-Spuma-Spu ranging from 1.3 to 35.47 MYA (million years ago). Although these dates are young relative to the age of reptiles, LTR dating may severely underestimate ERV ages (Kijima and Innan 2010; Aiewsakun and Katzourakis 2017), such that all estimates of integration time should be treated with caution.

Evolutionary relationships of ERVs-Spuma-Spu

Sequences of the Pol protein (490 amino acids in length) of ERV-Spuma-Spu were used for phylogenetic analysis. Our ML

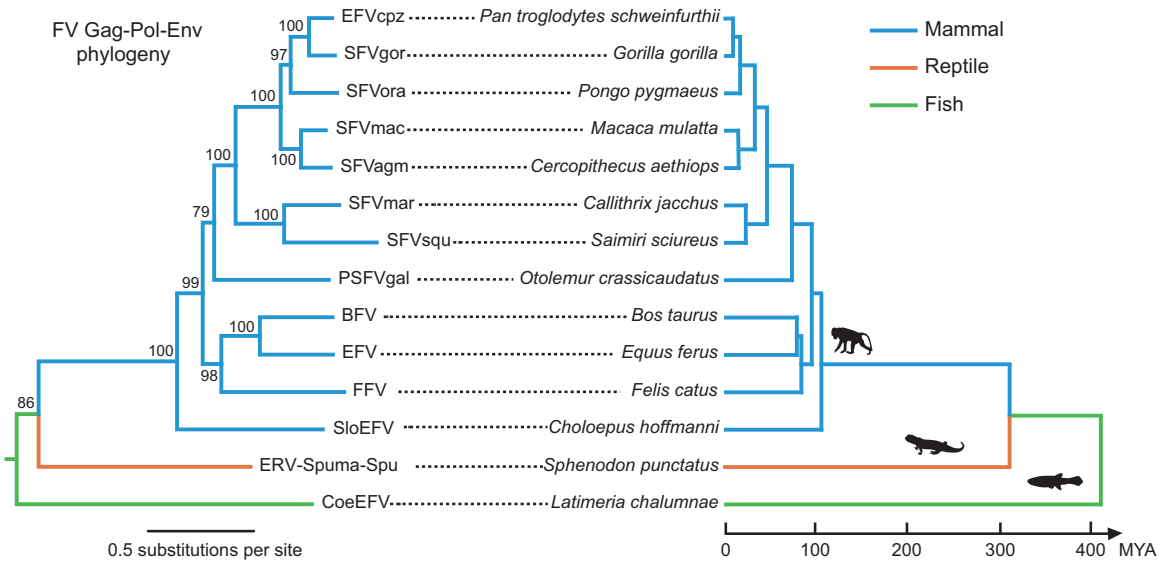

Figure 3. A simplified evolutionary relationship between foamy viruses (left) and their vertebrate hosts (right). The scale bar indicates number of amino acid changes per site in the viruses or host speciation time (million years ago, MYA). Bootstrap support values are provided at each node. The alignment of the FV Gag-Pol-Env amino acid sequences is provided in Supplementary data set S2.

phylogenetic trees revealed that the EFVs present in the tuatara genome formed a close monophyletic group within the foamy virus clade, indicative of a single origin, and with high bootstrap support (Fig. 2). The divergent phylogenetic position of ERV-Spuma-Spu is compatible with virus-host co-divergence for the entire history of the vertebrates. However, it is possible that this pattern will change with a larger sampling of taxa such that the EFV phylogeny expands.

In addition, we inferred a phylogenetic tree of FVs, EFVs, and foamy-like ERVs. This was consistent to those in previous studies (Aiewsakun and Katzourakis 2017) and revealed that the ERVs reported previously in the tuatara genome are distantly related to foamy virus and belong to class III retroviruses (Supplementary Fig. S4) (Tristem, Myles, and Hill 1995; Herniou et al. 1998). Failure to detect any ERV-Spuma-Spu-related elements in the remaining reptilian genomes suggest that the virus may be not vertically transmitted among reptiles, although this will clearly need to be reassessed with a larger sample size. The absence of EFVs in avian genomes is puzzling and clearly merits additional study.

Previous studies provided strong evidence for the co-divergence of foamy viruses and their vertebrate hosts over extended periods of evolutionary time (Katzourakis et al. 2009). An analysis of concatenated gag-pol-env protein sequences suggests that the reptilian ERV-Spuma-Spu newly described here follow the same pattern of long-term virus-host co-divergence (Fig. 3). As such, these data imply that ERV-Spuma-Spu may have diverged from the other mammalian foamy viruses along with its tuatara host more than 320 MYA (http://www.timetree.org/). The discovery of ERVs-Spuma-Spu therefore fills a major gap in our understanding of the taxonomic distribution of the foamy viruses and their evolutionary history.

Acknowledgements

J.C. is supported by National Natural Science Foundation of China (31671324) and CAS Pioneer Hundred Talents Program. E.C.H. is supported by an ARC Australian Laureate Fellowship (FL170100022).

Data availability

All the data needed to generate the conclusions made in the article are present in the article itself and/or the Supplementary data. Additional data related to this article may be requested from the authors.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Conflict of interest: None declared.

References

Abascal, F., Zardoya, R., and Posada, D. (2005) 'ProtTest: Selection of Best-fit Models of Protein Evolution', *Bioinformatics*, 21: 2104–5.

Aiewsakun, P., and Katzourakis, A. (2017) 'Marine Origin of Retroviruses in the Early Palaeozoic Era', *Nature Communications*, 8: 13954.

Altschul, S. F. et al. (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, 215: 403–10.

Cui, J. et al. (2014) 'Low Frequency of Paleoviral Infiltration across the Avian Phylogeny', *Genome Biology*, 15: 539.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008) 'LTRharvest, an Efficient and Flexible Software for De Novo Detection of LTR Retrotransposons', *BMC Bioinformatics*, 9: 18.

Guindon, S. et al. (2010) 'New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0', *Systematic Biology*, 59: 307–21.

Han, G. Z., and Worobey, M. (2012a) 'An Endogenous Foamy-Like Viral Element in the Coelacanth Genome', *PLoS Pathogens*, 8: e1002790.

——, and —— (2012b) 'An Endogenous Foamy Virus in the Aye-Aye (*Daubentonia madagascariensis*)', *Journal of Virology*, 86: 7696–8.

——, and —— (2014) 'Endogenous Viral Sequences from the Cape Golden Mole (*Chrysochloris asiatica*) reveal the Presence of Foamy Viruses in All Major Placental Mammal Clades', *PLoS One*, 9: e97931.

Hayward, A., Cornwallis, C. K., and Jern, P. (2015) 'Pan-Vertebrate Comparative Genomics Unmasks Retrovirus Macroevolution', *Proceedings of the National Academy of Sciences of the United States of America*, 112: 464–9.

——, Grabherr, M., and —— (2013) 'Broad-Scale Phylogenomics Provides Insights into Retrovirus-host Evolution', *Proceedings of the National Academy of Sciences of the United States of America*, 110: 20146–51.

Herniou, E. et al. (1998) 'Retroviral Diversity and Distribution in Vertebrates', *Journal of Virology*, 72: 5955–66.

Johnson, W. E. (2015) 'Endogenous Retroviruses in the Genomics Era', *Annual Review of Virology*, 2: 135–59.

——, and Coffin, J. M. (1999) 'Constructing Primate Phylogenies from Ancient Retrovirus Sequences', *Proceedings of the National Academy of Sciences of the United States of America*, 96: 10254–60.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Katzourakis, A. et al. (2009) 'Macroevolution of Complex Retroviruses', *Science*, 325: 1512.

Kijima, T. E., and Innan, H. (2010) 'On the Estimation of the Insertion Time of LTR Retrotransposable Elements', *Molecular Biology and Evolution*, 27: 896–904.

Kumar, S., Stecher, G., and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets', *Molecular Biology and Evolution*, 33: 1870–4.

Marchler-Bauer, A., and Bryant, S. H. (2004) 'CD-Search: Protein Domain Annotations on the Fly', *Nucleic Acids Research*, 32: W327–31.

Perry, B. W. et al. (2018) 'Molecular Adaptations for Sensing and Securing Prey and Insight into Amniote Genome Diversity from the Garter Snake Genome', *Genome Biology and Evolution*, 10: 2110–29.

Stoye, J. P. (2012) 'Studies of Endogenous Retroviruses Reveal a Continuing Evolutionary Saga', *Nature Reviews. Microbiology*, 10: 395–406.

Switzer, W. M. et al. (2005) 'Ancient Co-Speciation of Simian Foamy Viruses and Primates', *Nature*, 434: 376–80.

Tristem, M., Myles, T., and Hill, F. (1995) 'A Highly Divergent Retroviral Sequence in the Tuatara (Sphenodon)', *Virology*, 210: 206–11.

Winkler, I. et al. (1997) 'Characterization of the Genome of Feline Foamy Virus and Its Proteins Shows Distinct Features Different from Those of Primate Spumaviruses', *Journal of Virology*, 71: 6727–41.

Xu, X. et al. (2018) 'Endogenous Retroviruses of Non-Avian/Mammalian Vertebrates Illuminate Diversity and Deep History of Retroviruses', *PLoS Pathogens*, 14: e1007072.

Xu, Z., and Wang, H. (2007) 'LTR_FINDER: An Efficient Tool for the Prediction of Full-Length LTR Retrotransposons', *Nucleic Acids Research*, 35: W265–8.