

Short Report

Open Access

Phylogenetic prediction of cis-acting elements: a *cre*-like sequence in Norovirus genome?

Matías Victoria¹, Rodney Colina², Marize Pereira Miagostovich¹, José P Leite¹ and Juan Cristina^{*2}

Address: ¹Laboratório de Virologia Comparada e Ambiental, Instituto Oswaldo Cruz, FIOCRUZ, Av. Brasil 4365, Manguinhos, 21040-360 Rio de Janeiro, RJ, Brasil and ²Laboratorio de Virología Molecular. Centro de Investigaciones Nucleares, Facultad de Ciencias, Iguá 4225, 11400 Montevideo, Uruguay

Email: Matías Victoria - matvicmon@yahoo.com; Rodney Colina - rcolina@cin.edu.uy; Marize Pereira Miagostovich - marizepm@ioc.fiocruz.br; José P Leite - jogleite@ioc.fiocruz.br; Juan Cristina* - cristina@cin.edu.uy

* Corresponding author

Published: 7 September 2009

Received: 29 May 2009

BMC Research Notes 2009, 2:176 doi:10.1186/1756-0500-2-176

Accepted: 7 September 2009

This article is available from: <http://www.biomedcentral.com/1756-0500/2/176>

© 2009 Cristina et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Discrete RNA structures such as *cis*-acting replication elements (*cre*) in the coding region of RNA virus genomes create characteristic suppression of synonymous site variability (SSSV). Different phylogenetic methods have been developed to predict secondary structures in RNA viruses, for high-resolution thermodynamic scanning and for detecting SSSV. These approaches have been successfully in predicting *cis*-acting signals in different members of the family *Picornaviridae* and *Caliciviridae*. In order to gain insight into the identification of *cis*-acting signals in viruses whose mechanisms of replication are currently unknown, we performed a phylogenetic analysis of complete genome sequences from 49 Human Norovirus (NoV) strains.

Findings: The complete coding sequences of NoV ORF1 were obtained from the DDBJ database and aligned. Shannon entropy calculations and RNAalifold consensus RNA structure prediction identified a discrete, conserved, invariant sequence region with a characteristic AAACG *cre* motif at positions 240 through 291 of the RNA dependant RNA polymerase (RdRp) sequence (relative to strain [EMBL:EU794713]). This sequence region has a high probability to conform a stem-loop.

Conclusion: A new predicted stem-loop has been identified near the 5' end of the RdRp of Human NoV genome. This is the same location recently reported for *Hepatitis virus cre* stem-loop.

Findings

Internal base pairing that creates stem-loops and other RNA structures places constraints on sequence variability in bases required for structure formation in the genome of RNA viruses. For instance, the Hepatitis C virus (HCV) genome has a marked suppression of synonymous codon variability within several evolutionary conserved stem-loops in the core and NS5B coding regions that demonstrate their role in virus replication [1-3]. Discrete RNA

structures such as *cis*-acting replication elements (*cre*) in the coding region of human enteroviruses (HEVs) [4] and other viruses also create characteristic suppression of synonymous site variability (SSSV), similar to that observed in HCV [5,6]. Different phylogenetic methods have been developed to predict secondary structures in RNA viruses, like PFOLD [2,7] or Alifold [8], for high-resolution thermodynamic scanning, and like UNAFold [9] for detecting SSSV [2]. These methods have permitted to identify suita-

ble genome regions for an in-depth experimental analysis allowing establishing the role of the identified secondary RNA structures in translation or replication. This approach has permitted to raise the hypothesis that when SSSV (i.e. highly conserved synonymous sites in a RNA virus genome sequence alignment) takes place in a sequence region with a high probability of conforming a secondary structure (i.e. high probability of base pairing to generate a stable stem-loop), a cis-acting signal can be identified. This hypothesis has been successfully tested in different members of the family *Picornaviridae*, like Hepatitis A virus (HAV), Avian Encephalitis virus (AEV) and *Rhinovirus* [10,11] and in members of the family *Caliciviridae*, like *Norovirus*, *Sapovirus*, *Vesivirus* and *Lagovirus* [12].

In order to gain insight into the identification of cis-acting signals in viruses whose mechanisms of replication are currently unknown; we tested the above hypothesis for a group of 49 Human Noroviruses (NoV), for whom complete genome sequences have been recently obtained.

The complete codes of ORF1 of 49 Human NoV were obtained from the DDBJ database and aligned using the MUSCLE program [13] (for strain names and accession numbers see Additional File 1). Once aligned, the Shannon entropy at each position of the sequence dataset was calculated [14]. This permitted to measure the relative variation in each site of a sequence alignment. The results of these studies are shown in Fig. 1.

Only few, discrete, genome regions in the ORF1 of Human NoV have a Shannon entropy of zero indicating that they are invariants among all NoV sequences introduced in this analysis (Fig. 1). Interestingly, one of these discrete regions has an AAACG *cre* sequence motif [10] at position 3948 to 3952 of the alignment. These positions correspond to positions 240 through 291 of the RNA dependant RNA polymerase (RdRp) sequence (relative to strain [EU794713]) (Fig. 1).

In order to observe if this *cre* motif is situated in a sequence region that has a high probability to conform a secondary structure [8,15], we used the RNAalifold Web

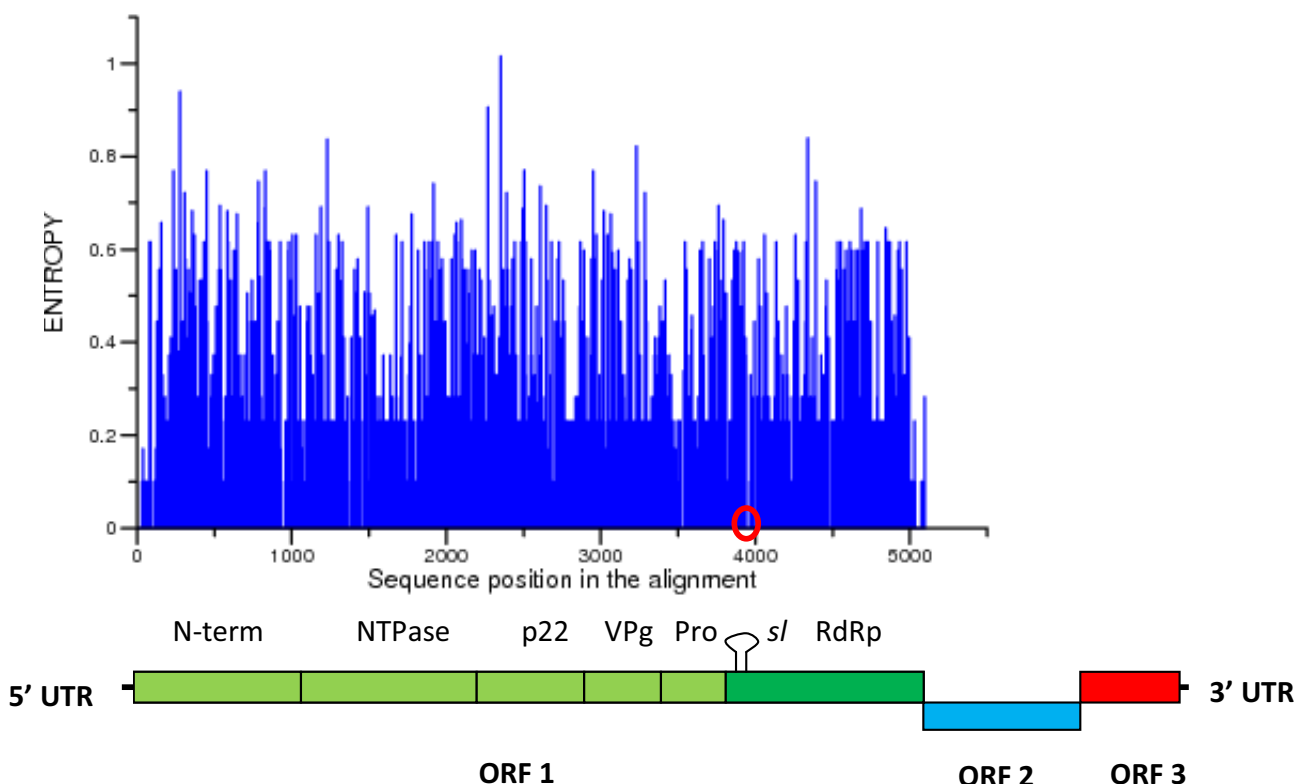


Figure 1
Shannon entropy in ORF1 of Human NoV strains. The Shannon entropy at each position of the alignment is shown. Location of positions 3922 through 3973 of the alignment (which includes the AAACG motif) in a sequence region with zero entropy is shown by a red circle. A scheme showing the position of each gene in the NoV genome is shown below the graph. The location of the stem-loop is indicated *sl*.

Server [16]. First, we obtained a graphical representation of the secondary structure in a plot of height versus position, where the height $m(k)$ is given by the number of base pairs enclosing the base at position k in the structure (i.e. loops correspond to plateaus, hairpin loops are peaks and helices to slopes). The results of these studies are shown in Fig. 2. The AAACG *cre* sequence motif is embedded in a sequence region with a very high probability to conform a stem-loop structure (see Fig. 2).

Then, the consensus RNA structures of the selected alignment region were folded using the RNAalifold program [16]. The resulting structures as well as the alignments were color-coded according to a coloring scheme for high-

lighting the mutational pattern with respect to the structure (Vienna RNA conservation coloring schema) [16]. The results of these studies are shown in Fig. 3.

Using the RNAalifold program [16] we have identified the presence of a unique and conserved stem-loop near the 5' end of the RdRp coding region for the 49 Human NoV genomes analyzed (Figs. 1 and 3). This predicted structure contains a *cre* sequence motif (Fig. 3). Interestingly, the stem-loop predicted for Human NoV is situated near the 5' end of the RdRp (Figs. 2 and 3). This is the same location recently reported for *Hepatovirus cre* stem-loop [10] (Fig. 1).

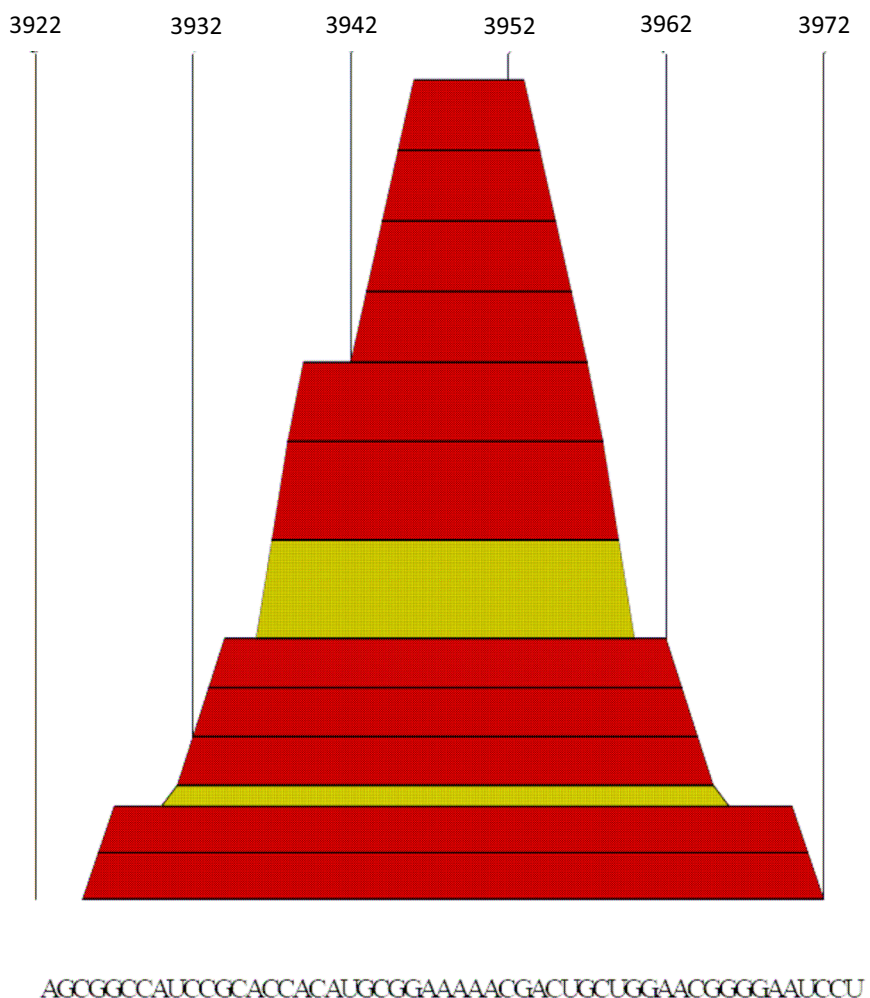


Figure 2
"Mountain plot" of a NoV ORF1 sequence region with an entropy equals to zero. A mountain plot representing a secondary structure in a plot of height versus position is shown. Sequences from positions 3922 through 3973 of the alignment are shown at the bottom of the figure. Numbers at the top of the figure show site position in the alignment. Colors correspond to the Vienna RNA conservation coloring schema [16] (see also Fig. 3). Note that the AAACG motif (underlined in red) is predicted to be located in a loop of the secondary structure.

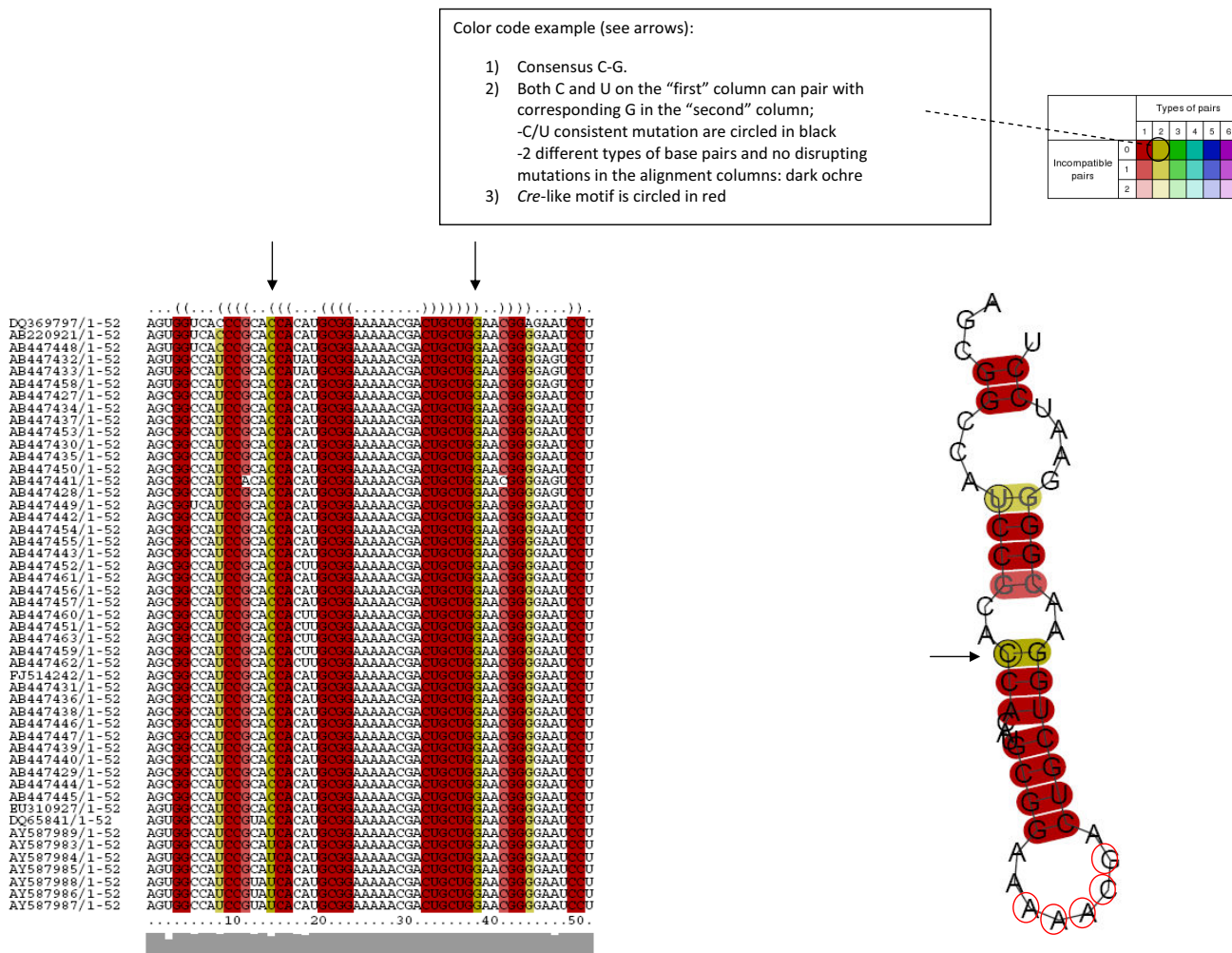


Figure 3
Conservation of a predicted secondary structure in NoV ORF1 sequences. Multiple sequence alignment across 49 Human NoV ORF1 sequences shows a consensus secondary structure from positions 3922 through 3973 of the NoV genome (positions 240 through 291 of the RdRp sequence, relative to strain [EMBL:EU794713]). The structure is shown in the dot-bracket format [16] above the alignment. Each corresponding bracket represents consensus base pairs of the alignment columns beneath. Sequences are color-coded according to consistent and compensatory mutations in the aligned sequences regarding the conserved structure (see figure text box). Pale colors indicate that a base-pair cannot be formed in some sequences of the alignment. The sequence conservation profile is shown in gray bars below the alignment. The conserved predicted secondary structure, color-coded according to the different types of base pairs in the corresponding alignment columns, is shown on the right side of the figure.

RNA structure predictions are consistent with previous analyses based on the thermodynamic folding of individual sequences [12,17,18]. Although RNA structure is clearly not the only cause of SSSV, occurring for example also in overlapping gene sequences [19], there is an impressive co-localization of the major sites of SSSV and thermodynamically predicted secondary structures [4,11,12].

NoV belong to the family *Caliciviridae*, and they are non-enveloped viruses with positive, single-stranded RNA genomes. They also share other important features with picornaviruses, like having a VPg protein covalently linked to the 5' of the genomic RNA [20]. Nevertheless, in contrast with picornaviruses, NoV express a downstream sub-genomic (sg) transcript encoding structural genes [20]. NoV are the leading cause of outbreaks of acute gastroenteritis in humans worldwide [21].

Despite the importance of these outbreaks, our understanding of the RNA structures or sequences required for NoV replication has been limited. Previous reports have identified the poly-pyrimidine tract-binding protein (PTB), poly-A binding protein (PAB) and La autoantigen to interact with the 3' untranslated region of the Norwalk virus genome [17]. Very recent studies have identified *cis*-acting signals in the 5' and 3' regions as well as at the start of the sg RNA transcript of NoV [12].

As a member of the family *Caliciviridae*, NoV are thought to replicate in a manner typical of positive-stranded RNA viruses, through the synthesis of a full-length antigenomic strand (reverse complement copy) using the viral RdRp translated initially from the RNA genome entering the cell [22]. The minus strand then acts as a template for the synthesis of full-length genomic RNA from which non-structural proteins are translated, including the RdRp. Features of the RdRp common to all positive-sense RNA viruses support this idea [23].

Although the presence of this new putative *cis*-acting signal predicted in this study was not yet investigated *in vitro* due to the lack of a standard cell culture to grow these viruses, the probability that this predicted structure will act as a functional element may open new avenues to our understanding of molecular mechanisms of NoV replication.

Extensive mutagenesis studies performed in members of the family *Picornaviridae*, like Poliovirus (PV) and Human Rhinovirus 14 (HRV-14), revealed a critical conserved AAACA/G *cre* sequence motif in the 5' half of the loop sequence that is essential for its function [10]. Similar conserved motifs are present within the loops of the *cre* elements of other picornaviruses and are important for RNA replication [10,24,25].

Paul and colleagues (2003) have shown that the PV *cre* act as the template for VPg uridylylation through a "slide-back" mechanism catalyzed by the 3D^{pol} (RdRp) [24,26,27]. The uridylylation of VPg leads to the production of VPg-pUpU, which serves as the protein primer for new RNA synthesis [28].

Interestingly, recent studies have shown that incubation of VPg with NoV 3D^{pol} (RdRp) generates VPg-poly(U) and that this uridylylated VPg can prime the replication of polyadenylated RNA [29]. In contrast, replication of antigenomic RNA was not primer dependent. Moreover, on nonpolyadenylated RNA, NoV RdRp initiated RNA synthesis *de novo* [29]. These findings clearly show that initiation of replication of the NoV genome by the RdRp requires a VPg-protein-primed initiation of replication of polyadenylated genomic RNA and a *de novo* initiation of

replication of antigenomic RNA [29]. Besides, very recent studies revealed that the NoV RdRp is a typical template-dependent RNA polymerase [30].

It is possible that the predicted stem-loop identified near the 5' end of the NoV RdRp coding region, which share a *cre*-like sequence motif with members of the family *Picornaviridae* [10], will be capable to perform the uridylylation of VPg. If that is the case, this will permit VPg to act as a primer for the synthesis of the minus strand RNA, in agreement with the results outlined above [30].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JC conceived the study. MV and JC designed, performed and revised the phylogenetic analysis. MV, RC, MPM and JPL contributed to the interpretation and discussion of the results found. JC wrote the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Origins of Norovirus strains. The table provides the strain names and accession numbers of their respective sequences of ORF1 of Norovirus strains included in this study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-2-176-S1.doc>]

Acknowledgements

We acknowledge support by CAPES (Brazil) and Universidad de la República (Uruguay) through project No. 018/08.

References

1. McMullar LK, Grakoui A, Evans MJ, Mihalik M, Puig AD, Branch AD, Feinstone SM, Rice CM: **Evidence for a functional RNA element in the hepatitis C virus core gene.** *Proc Natl Acad Sci USA* 2007, **104**:2879-2884.
2. Tuplin A, Evans DJ, Simmonds P: **Detailed mapping of RNA secondary structures in core and NS5B-encoding sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods.** *J Gen Virol* 2004, **85**:3037-3047.
3. You S, Stump DD, Branch AD, Rice CM: **A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication.** *J Virol* 2004, **78**:1352-1366.
4. Gerber K, Wimmer E, Paul AV: **Biochemical and genetic studies of the initiation of Human Rhinovirus 2 RNA replication: identification of a cis-replicating element in the coding sequence of 2A^{pro}.** *Virology* 2001, **75**:10979-10990.
5. Simmonds P, Tuplin A, Evans DJ: **Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence.** *RNA* 2004, **10**:1337-1351.
6. McKnight KL, Lemon SM: **The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication.** *RNA* 1998, **4**:1569-1584.

7. Knudsen B, Hein J: **RNA secondary structure prediction using stochastic context-free grammars and evolutionary history.** *Bioinformatics* 1999, **15**:446-454.
8. Hofacker IL, Fekete M, Stadler PF: **Secondary structure prediction for aligned RNA sequences.** *J Mol Biol* 2002, **319**:1059-1066.
9. Zuker M: **Calculating nucleic acid secondary structure.** *Curr Opin Struct Biol* 2000, **10**:303-310.
10. Yang Y, Yi M, Evans DJ, Simmonds P, Lemon SM: **Identification of a Conserved RNA Replication Element (cre) within the 3D_{pol}-coding sequence of Hepatoviruses.** *J Virol* 2008, **82**:10118-10128.
11. Tapparel C, Junier T, Gerlach D, Cordey S, Van Belle S, Perrin L, Zdobnov EM, Kaiser L: **New complete genome sequences of human rhinoviruses shed light on their phylogeny and genomic features.** *BMC Genomics* 2007, **8**:224.
12. Simmonds P, Karakasiliotis I, Bailey D, Chaudhry Y, Evans DJ: **Good-fellow: Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses.** *Nucleic Acids Res* 2008, **36**:2530-2546.
13. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792-1797.
14. Korber BT, Kuntsman K, Patterson B, Furtado M, McEvilly M, Levy R, Wolinsky S: **Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type I-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain derived sequences.** *J Virol* 1994, **68**:7467-7481.
15. Mathews DH, Sabina J, Zuker M, Turner H: **Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure.** *J Mol Biol* 1999, **288**:911-940.
16. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL: **The Vienna RNA websuite.** *Nucleic Acids Res* 2008:W70-74.
17. Gutierrez-Escolano AL, Vazquez-Ochoa M, Escobar-Herrera J, Hernandez-Acosta J: **La, PTB and PAB proteins bind to the 3' untranslated region of Norwalk virus genomic RNA.** *Biochem Biophys Res Commun* 2003, **311**:759-766.
18. Seal BS, Neill JD, Ridpath JF: **Predicted stem-loop structures and variation in nucleotide sequence of 3' noncoding regions among animal calicivirus genomes.** *Virus Genes* 1994, **8**:243-247.
19. Mizokami M, Orito E, Ohba K, Ikeo K, Lau JY, Gojbori T: **Constrained evolution with respect to gene overlap of hepatitis B virus.** *J Mol Evol* 1997, **44**:S83-S90.
20. Jiang X, Wang M, Wang K, Estes MK: **Sequence and genomic organization of Norwalk virus.** *Virology* 1993, **195**:51-61.
21. De Wit MA, Widdowson MA, Vennema H, de Bruin E, Fernandes T, Koopmans M: **Large outbreak of Norovirus: the baker who should have known better.** *J Infect* 2007, **55**:188-193.
22. Clarke IN, Lambden PR: **The molecular biology of caliciviruses.** *J Gen Virol* 1997, **78**:291-301.
23. Koonin EV: **The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses.** *J Gen Virol* 1991, **72**:2197-2206.
24. Rieder E, Paul AV, Kim DW, Van Boom JH, Wimmer E: **Genetic and biochemical studies of poliovirus cis-acting replication element cre in relation to VPg uridylylation.** *J Virol* 2000, **74**:10371-10380.
25. Yin J, Paul AV, Wimmer E, Rieder E: **Functional dissection of a poliovirus cis-acting replication element [PV-cre(2C)]: analysis of single and dual-cre viral genomes and proteins that bind specifically to PV-cre RNA.** *J Virol* 2003, **77**:5152-5166.
26. Paul AV, Yin J, Mugavero J, Rieder E, Liu Y, Wimmer E: **A "slide-back" mechanism for the initiation of protein-primer RNA synthesis by the RNA polymerase of poliovirus.** *J Biol Chem* 2003, **278**:43951-43960.
27. Paul AV, Rieder E, Kim DW, Van Boom JH, Wimmer E: **Identification of an RNA hairpin in poliovirus RNA that served as the primary template in the in vitro uridylylation of VPg.** *J Virol* 2000, **74**:10359-10370.
28. Paul AV, Van Boom JH, Filippov D, Wimmer E: **Protein primed RNA synthesis by purified poliovirus RNA polymerase.** *Nature* 1998, **393**:280-284.
29. Rohayem J, Robel I, Jager K, Scheffler U, Rudolph W: **Protein-Primed and De Novo initiation of RNA synthesis by Norovirus 3D_{pol}.** *J Virol* 2006, **80**:7060-7069.
30. Hogbom M, Jager K, Robel I, Unge T, Rohayem J: **The active form of the norovirus RNA-dependent RNA polymerase is a homodimer with cooperative activity.** *J Gen Virol* 2009, **90**:281-291.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

