Palindromic nucleotide substitutions: A new software for *pestivirus* genotyping

Massimo Giangaspero[1*], Claudio Apicella[2], Ryô Harasawa[1]

[1]Department of Veterinary Microbiology, School of Veterinary Medicine, Faculty of Agriculture, Iwate University, Ueda 3-18-8, Morioka, Iwate 020-8550, Japan
[2]Department of Public Health, Food Safety and National Boards for Health Protection, Ministry of Health, Via Ribotta 5, 00166 Rome, Italy

The genus *Pestivirus* of the family *Flaviviridae* is represented by four established species; Bovine viral diarrhea virus 1 (BVDV-1), Bovine viral diarrhea virus 2 (BVDV-2), Border disease virus (BDV) and Classical swine fever virus (CSFV) and a tentative "Giraffe" species. The palindromic nucleotide substitutions (PNS) in the 5' untranslated region (UTR) of *Pestivirus* RNA have been described as a new, simple and practical method for genotyping. The preparation of specific software was considered for an easy access of the users to the method. In the present study, the new software, named PNS, was prepared and the application on the genotyping procedures with the keys for *Pestivirus* identification as specific PNS at genus, species and genotype level, respectively, was evaluated on five hundred-tirty-four sequences. The software is freely available at www.pns-software.com

Keywords: Genotypes, palindromic nucleotide substitutions, *Pestivirus*, software.

INTRODUCTION

The genus *Pestivirus* of the family *Flaviviridae* is represented by four established species; *Bovine viral diarrhea virus 1* (BVDV-1), *Bovine viral diarrhea virus 2* (BVDV-2), *Border disease virus* (BDV) and *Classical swine fever virus* (CSFV) and a tentative "Giraffe" species (King *et al*., 2012). The *Pestivirus* genome, single-stranded, positive polarity RNA, is composed by a sequence of about 12,500 nucleotides. It can be divided into three regions: a 5'-untranslated region (UTR), a single large open reading frame encoding region, and a 3'-UTR. The 5'-UTR is highly conserved among all members within the genus *Pestivirus*, thus being useful for the characterization of species or genotypes. The primary structure analysis, by sequence alignment and construction of phylogenetic trees, is the most common method for the classification of *Pestivirus* strains. Genetyx-Mac, DNASIS, Clustal X(Thompson *et al*., 1997) are among the software used for typing virus strains based on sequence alignment. In reality, the genomic sequence is tri-dimensional. The reproduction of the third structure is highly problematic. However, it is relatively easy to predict the secondary structure, according to the most probable nucleotide binding, with lowest folding energies. The secondary structure of the 5'-UTR can be divided into four domains, A-D, with domain D encompassing the 3' two thirds of the UTR predicted to fold into a complex palindromic stem-loop structure, including an internal ribosome entry site (IRES), as observed in poliovirus (Deng and Brock, 1993; Harasawa, 1994), thus representing critical regions of the 5´-UTR, responsible for translational, transcriptional and replicational events in pestiviruses.

The *Pestivirus* genome has a relatively long 5'-UTR upstream of the polyprotein open reading frame. Although the nucleotide sequence of the 5'-UTR is well conserved among the members of the *Pestivirus* genus, the 5'-UTR has been shown to contain at least three variable loci. The nucleotide substitutions in these variable loci are particularly important because the 5'-UTR of positive-sense RNA viruses generally includes regulatory motifs, which are indispensable to viral survival. Therefore, random mutations at the 5'-UTR have a high probability of incompatibility with viral survival. Thus stable nucleotide variations at this level assume high importance in terms of virus evolutionary history.

Table 1. Summary of *Pestivirus* strains (n 533) evaluated according to the Palindromic nucleotide substitution (PNS) method at the 5' untranslated region of RNA.

Species	Number of strains	Host	Geographical origin
BVDV-1	274	Cattle, Sheep, Pig, Deer, Roe deer, Human, Contaminant	Argentina, Austria, Belgium, Brazil, Canada, China, Belgium, France, Germany, India, Ireland, Italy, Japan, New Zealand, Slovakia, South Africa, Spain, Sweden, Switzerland, UK, USA
BVDV-2	77	Cattle Sheep Contaminant	Argentina, Austria, Belgium, Brazil, Canada, France, Germany, Italy, Japan, Netherland, NewZealand, Slovakia, Tunisia, UK, USA
BVDV-3*	3	Cattle	Brazil, Thailand
BDV	131	Sheep, Pyrenean chamois, Cattle, Pig, Reindeer, Wisent	Australia, France, Germany, Japan, New Zealand, Spain, Switzerland, Tunisia, Turkey, UK, USA
BDV-2*	3	Sheep, Goat	Italy
CSFV	43	Pig, Sheep	China, France, Garmany, Honduras, Italy, Japan, Malaysia, Netherlands, Poland, Russia, Spain, Switzerland, USA
Pronghorn*	1	Pronghorn	USA
Giraffe*	1	Giraffe	Kenya
Bungowannah*	1	Pig	Australia

* Tentative species

Nucleotide sequences at the three variable loci, V1, V2 and V3, in the 5´-UTR of pestiviruses have been shown to be palindromic and capable of forming a stable stem-loop structure peculiar to each *Pestivirus* species. Nucleotide substitutions in the stem regions always occur to maintain the palindromic sequence and thereby form a stable stem-loop structure. Thus, this type of mutation was referred to palindromic nucleotide substitutions (PNS). Based on the above mentioned considerations, the observation of nucleotide variations among virus strains at the level of the three specific palindromes in the 5'-UTR has been conceived as method for genotyping (Harasawa and Giangaspero, 1998). The method named palindromic nucleotide substitutions (PNS) appeared to be simple and practical, showing comparable results with other procedures based on the primary structure comparison.

According to palindromic nucleotide substitutions, 534 sequences (Table 1) have been segregated into nine species within the genus *Pestivirus* (Giangaspero and Harasawa, 2011). In addition to the four established species, Harasawa *et al*. (2000) characterized the taxonomic status of a giraffe strain, based on the 5' untranslated region, as a new cluster among *Pestivirus* species. Furthermore, other four tentative species (BVDV-3, BDV-2, Pronghorn and Bungowannah) have been recently proposed (Vilček *et al*., 2005, Kirkland *et al*., 2007, Giangaspero and Harasawa, 2011). Genotypes have been identified in species showing heterogeneity: *Bovine viral diarrhea virus 1* (Giangaspero *et al*., 2001), *Bovine viral diarrhea virus 2* (Giangaspero *et al*., 2008, Giangaspero and Harasawa, 2011), *Border Disease virus* (Giangaspero, 2011) and *Classical Swine Fever virus* (Giangaspero and Harasawa, 2008). The observation made on the nucleotide sequences of the three variable loci at the level of the 5'-UTR genomic region of *Pestivirus* strains allowed to the identification of consensus motifs shared by all the *Pestivirus* species. The characteristic palindromic nucleotide substitutions have been identified at genus, species and genotype level, respectively (Tables 2 and 3). The palindromic loci represented, with about 80 nucleotides, a very limited portion of the virus genome. Within these short sequences, it was sufficient the evaluation of only 21 nucleotides to obtain with certitude the characterization of the genus. Species were characterized through the evaluation of only 6 to 19 nucleotides. Similarly, the genotype was defined with only 6 to 10 nucleotides. These peculiar aspects resumed the high specificity of the PNS method and the reliability of the provided results.

With the aim of improving the PNS method in a full computerization of the procedure and avoid the main limitation due to the manual searching of relevant base pairings and direct observation of the sequence, a specific software was conceived for an easy access of the users and a rapid testing with reliable results applying the patterns for *Pestivirus* identification.

MATERIALS AND METHODS

The program was realized using the "C++" programming language (Ellis and Stroustrup, 1990) and adapted to run under the Windows operative system. The software was constructed in order to evaluate virus nucleotide sequences up to 15,000 bases. In general, RNA virus sequences from 5'-UTR genomic region, deposited in international databases, include approximately 250-350

Table 2. Palindromic nucleotide substitutions (PNS) characteristic to the genus *Pestivirus*. The position of base Pairings is defined by numbering from the bottom of the variable locus

Genus	Locus	Characteristic PNS markers
	V1	Absence in position 22 - size of V1 21 bp (exception U); C C bulge in position 11; A-U in position 10; C-G in position 8 (exceptions U*G, U-A and G G bulge); U-A in position 7 (exception G-C and A A bulge); A in position 6 (exception G); U*G in position 5; U in position 5 right nucleotide; G-C in position 4.
	V2	GGGGU loop (exception GGGGC); C-G in position 8 (exception U*G).
Species	**Locus**	**Characteristic PNS markers**
BVDV-1	V1	U-A in position 15 (exception U*G or C-G);
	V2	G-C in position 5 (exception A-U);
	V3	G-C in position 5; A in position 10 (exceptions A-U, G-C or A C, A A or G A bulges or absence).
BVDV-2	V1	A-U or A C bulge in position 20 (exceptions G*U, C C or A A bulges); A,G or U in position 21 (exception G G);
	V2	U-A or U*G in position 6 (exception C A bulge);
	V3	A-U or A C bulge in position 7 (exception G-C).
BVDV-3 tentative species (HoBi group)	V1	U-A in position 15;
	V3	G-C or G*U in position 3; A-U or G-C in position 7; A in position 10.
BDV	V1	G-C or A-U in position 15 (exceptions C U and A C bulges);
	V3	U C and U U bulges or U*G in position 7 (exceptions A-U, U-A and C C, A C, C U and C A bulges).
BDV-2 tentative species (Italian ovine isolates)	V1	U-A or C A bulge in position 15;
	V3	G*U or G G bulge in position 8.
CSFV	V1	U-A in position 13 (exception U*G);
	V3	U-A in position 2; U or C in position 8 (exception A).
Pronghorn tentative species	V1	G-C in position 2; U-A in position 9; U-A in position 12; U-A in position 15;
	V2	G-C in position 4;
	V3	G A bulge in position 5.
Giraffe tentative species	V1	C-G in position 2; U*G in position 20;
	V2	C-G in position 7;
	V3	C-G in position 4; G*U in position 7.
Bungowannah tentative species	V1	A-U in position 2; G-C in position 7; U-A in position 9; U-A in position 12; G-C in position 13;
	V2	A-U in position 3; G-C in position 4;
	V3	U-A in position 4; G A bulge in position 10; A in position 11.

nucleotides. However, the size of the sequence may change up to a virus genome sequencing performed on the entire RNA. Therefore the program was conceived to allow the analysis of large size sequences. Primary objective was to identify in the test sequence the three variable loci V1, V2 and V3 characteristic for genotyping procedure. The sequences to be tested were prepared in text file type (.txt) as input for the program. No other characters then bases were allowed in the file.

Once verified the compatibility of the input file, trough a first step, the program loaded in memory the sequence. In a second step the three variable sequences were

Table 3. Palindromic nucleotide substitutions (PNS) characteristic to the *Pestivirus* species genotypes. The position of base pairings is defined by numbering from the bottom of the variable locus.

BVDV-1 genotypes	Locus	Characteristic PNS markers
BVDV-1a	V1	U*G or C-G in position 14.
	V2	G*U or G-C in position 7.
	V3	A-U in position 4.
BVDV-1b1	V1	U-A in position 14 (exception G A bulge).
	V2	A-U in position 7 (exception A C bulge).
	V3	G-C in position 4 (exception A-U).
BVDV-1b2	V1	U-A in position 14 (exception A A bulge).
	V2	G*U or G-C in position 7.
	V3	G-C in position 4.
BVDV-1c	V1	C-G in position 14 (exception C A bulge).
	V2	A C bulge in position 7.
	V3	A-U in position 4.
BVDV-1d	V1	C-G in position 14.
	V2	A-U in position 7.
	V3	G-C in position 4.
BVDV-1e	V1	C-G in position 14.
	V2	G-C in position 7.
	V3	A C bulge in position 2; G-C in position 4.
BVDV-1f	V1	C-G in position 14.
	V2	A-U in position 7.
	V3	A-U in position 4.
BVDV-1g	V1	C-G in position 14; U*G in position 17.
	V2	G-C in position 7.
	V3	G*U in position 4.
BVDV-1h	V1	C-G in position 14.
	V2	C U bulge in position 6; G-C in position 7.
	V3	G-C in position 4.
BVDV-1i	V1	C-G in position 14.
	V2	G*U in position 1; G-C in position 7.
	V3	G-C in position 4; A A or G A bulges in position 8.
BVDV-1j	V1	C-G in position 14.
	V2	A-U in position 1; G- C in position 7; A C bulge in position 9.
	V3	G-C in position 4; G-C or G A bulge in position 6.
BVDV-1k	V1	C- G in position 14; C-G or C A bulge in position 16; A A bulge in position 19.
	V2	A-U in position 1; G-C in position 7.
	V3	G-C in position 4.
BVDV-1l	V1	CG in position 14.
	V2	UA in position 1; AC in position 7.
	V3	GC in position 4.
BVDV-1m	V1	C-G in position 14.
	V2	G-C in position 7; A C bulge in position 9.
	V3	G-C in position 4; A A bulge in position 6.
BVDV-1n	V1	C-G in position 14; C-G in position 15.
	V2	G-C in position 7; A-U in position 5.
	V3	G-C in position 4.
BVDV-1o	V1	C-G in position 14.
	V2	G-C in position 7; U-A in position 5.
	V3	G-C in position 4.
BVDV-2 genotypes	**Locus**	**Characteristic PNS markers**
BVDV-2a	V1	C-G in position 16; U*G, C-G or U-A in position 18.
BVDV-2b	V1	G-C in position 12; U-A in position 16; G A, G G or A C bulges in position 17; G G or G A bulges in position 18.
	V3	higher V3 loop, U in position 10.
BVDV-2c	V1	G-C in position 12; C-G in position 14; C-G in position 16; G A, G G or A C bulges in position 17; G A bulge in position 18.
	V3	higher V3 loop, C in position 10;

Table 3. Continues

BVDV-2d	V1	A-U in position 9; A-U in position 12; U-A in position 16; G A, G G or A C bulges in position 17; G A bulge in position 18.
	V3	higher V3 loop, U in position 10.
BVDV-2e	V1	U-A in position 8; C C bulge in position 20.
	V3	G-C in position 7.
BDV genotypes	**Locus**	**Characteristic PNS markers**
BDV-a	V1	A-U or C U bulge in position 9; A A or A G bulges in position 18 (exception G G bulge).
	V2	A-U in position 1.
	V3	A A, G A or A C bulges in position 8.
BDV-b	V1	G-C in position 9; G-C or G G bulge in position 18; G*U or G G bulge in position 20.
	V2	G*U in position 1.
	V3	U-A or C A bulge in position 8.
BDV-c	V1	G-C or U C bulge in position 20; U or U U bulge in position 21.
	V2	A C bulge in position 1.
	V3	C C bulge in position 7.
BDV-d	V1	G-C or A-U in position 9; U*G, G-C, G*U or G G bulge in position 18 (exception A G bulge).
	V2	G*U, G-C in position 1 (exceptions A-U and C U bulge).
	V3	U-A, C-G, U*G, A A or C A bulges in position 8.
BDV-e	V1	U-A, C-G or U*G in position 16.
	V3	C U bulge in position 1; G*U or U U bulge in position 2.
BDV-f	V3	U-A in position 2; U*G in position 7; U or C in position 8.
BDV-g	V1	G-C in position 3; U-A or C-G in position 16.
	V3	G-C in position 4.
BDV-h	V2	G-C in position 5.
	V3	G-C or A C bulge in position 2; C-G in position 7; U U bulge in position 9; U Uor C U bulge in position 10.
CSFV genotypes	**Locus**	**Characteristic PNS markers**
CSFV-1	V1	A C bulge in position 15 (exception G-C).
	V2	U-A in position 5; G:Y in position 7.
	V3	A-U in position 1.
CSFV-2	V1	G-C in position 15; A G bulge in position 19; U-A in position 20.
	V2	A-U in position 5; A C bulge in position 7.
	V3	A-U in position 1; U C bulge in position 6.
CSFV-3	V1	A-U in position 15.
	V2	U-A in position 5; G-C in position 7.
	V3	A_G bulge in position 1.

identified and further the secondary structures were determined according to the Watson-Crick base pairings, with strong bindings as adenine uracil or cytosine guanine, tolerated pairings in secondary structure as guanine uracil and pairings without binding forming bulges as cytosine cytosine. Since timine is equivalent to uracil, the program was adapted to translate as timine in case test sequences were constructed with uracile. In addition, the nucleotides were considered indifferently when expressed in small or capital letters in the sequence file.

PNS class

The most important class of the software was named PNS and, in addition to a series of data structure definitions, it contains two fundamental methods which are respectively: the *SearchPNS,* for loading in memory the test sequence nucleotides and to determine correct V1, V2 and V3 structures; the *CreateHTML*, in charge of V1, V2 and V3 structure analysis by known parameters in order to recognize PNS and create an Html file containing the results of the analysis.

Loading in memory of key data structures

The PNS software uses four data structures to memorize and keep available information during the analytical process.

The memorization of all possible nucleotide pairings and the related bindings was done by a character matrix.

The nucleotides read in the input file were stored in an array (Lodi and Pacini, 1998) for a maximum dimension of 15,000.

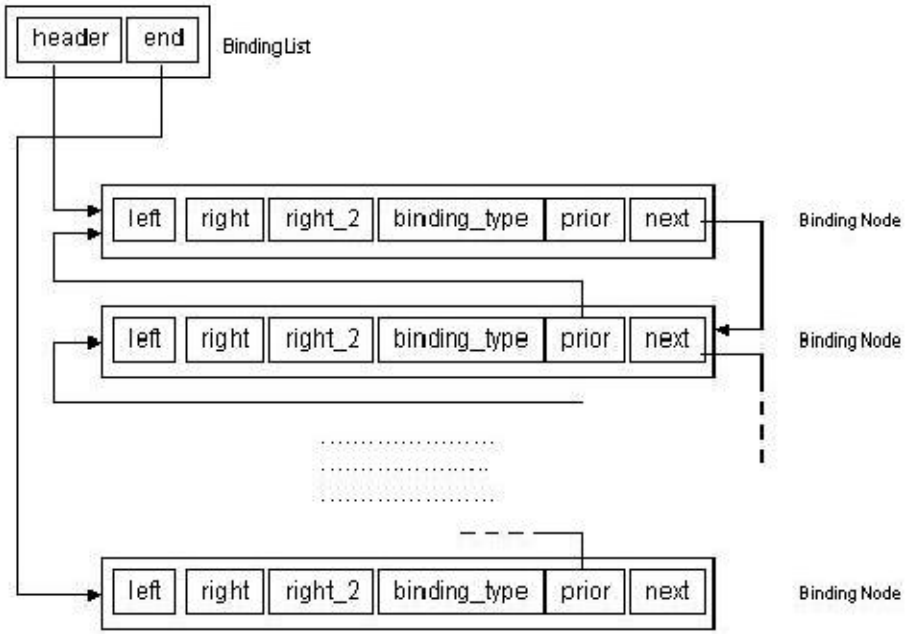

Figure 1. Schematic presentation of the pointer list management in the PNS program memory

The analysis parameters for typing genus, species and genotypes were stored in three arrays of a structure (*struct*) named *Records*, which contains respectively the following fields: *name, structure, position, left, right, binding*.

The V1, V2 and V3 structures were stored using three pointers list (Lodi and Pacini, 1998) defined by two different type of *struct*. The first defined the single node of the list and the second defined the structure of list type. Nodes were linked among them in a sequential order by virtual bindings represented by pointers, addresses of memory, so that a single node contains the pointer at a following node. The list node contains the following fields: *left, right, right_2, binding_type, prior, next*. A schematic presentation of the list management in the program memory is reported in Figure 1.

PSet abstract structure

On the base of the model proposed by Apicella (1999), an abstract parametric class or template, called *PSet,* was realized in order to obtain a dynamic and flexible structure allowing the manage of a double linked pointer list of generic type T objects.

The use of *PSet* was necessary for the program, expecially during the phase of V1, V2 and V3 structure determination where the selection of the correct structure was realized on the base of defined criteria among a large number of possible structures. A *PSet* class opportunely instanced facilitate the entire list recording, the evaluation of the possible structures and the identification of the appropriate one.

Construction of the V1, V2 and V3 palindromic structures

The appropriate structure of type V1, V2 or V3 of a tested sequence was selected among a large number of possible structures identified according to well defined criteria, according to specific program procedures. The three variable loci in the test sequence were identified starting by the V2 locus due to the invariable number of nucleotide composing the palindrome (23 bases) and the characteristic conserved 4 guanine sequence of the loop (5´-GGGG-3´). The calculation of the secondary structure was based on the search of palindromes with the higher number of Watson-Crick base pairings with strong bindings (C-G; G-C; A-U; U-A), which are most probable, with priority for the stability of the structure in confront to tolerated pairings (G*U; U*G) and pairings forming bulges (CC; AA; AC; AG; CA; GG; GA; UU; CU; UC). The construction of the palindrome structures respected the principle of the correct curbing of the sequence at the loop level, thus avoiding formation of base pairing generating strong or weak, tolerated bindings. In other terms, if at loop were present 3 nucleotides n_1, n_2, n_3 the potential binding occurring between n_1 and n_3 , for example G-C, was not considered. Similarly, in case of 4 nucleotides (n_1, n_2, n_3, n_4) the potential binding between

n_1 and n_4 and between n_2 and n_3 were not considered. The first step realized by the software was the verification of the presence of a structure compatible with the V2 locus in a test sequence, giving that the determination of the following structures depended exclusively from this component. In case no V2 structure could be identified, the analyzed sequence had to be considered not corresponding to 5'-UTR of the genus *Pestivirus*, resulting in the termination of the execution of the program. Alternatively, if one or more V2 structures were identifiable in a given sequence, the program selected the appropriate one and successively continued the analysis.

The following palindrome to be identified was the V3 locus, according to the conserved separation by generally 3 nucleotides, and in some cases 2, as for strains VM, CP1874 and Marloie, from the V2 locus end and the starting of the V3 sequence. Consequently, the determination of the V3 was possible only when the V2 locus was identified. Similarly to V2, the construction was based on the search of a secondary structure with the highest number of Watson-Crick base pairings with strong bindings, taking into account the approximate length of the sequence, variable from 15 to 20 nucleotides, with a 6-7 base pairing composed stem and a 3-5 nucleotide loop. The construction of the stable palindrome was more problematic due to the variability of the sequence length.The V3 starting position was the third or the fourth nucleotide located after the end of the V2 locus sequence. In a first step, possible V3 palindromes were identified starting from the fourth nucleotide after the V2 end. From this nucleotide, n possible V3 structures named $v_{i=1,\ldots,n}$, were determined. Per each v_i structure was identified the one with the highest value of strong bindings. In case of two or more v_i had the same maximum value, the selection was done on the lowest index *i* among them, giving that the structure showed the minor number of nucleotides, thus, the more stable one. In a second step, the procedure is repeated starting from the third nucleotide after the V2 end identifying possible V3 structures identified as $v_{j=1\ldots,m}$. At the end of the two steps, the respective resulted structures v_i and v_j were compared in order to determine the correct V3. The structure showing the highest number of strong bindings and the minimum dimension (constructed starting from the fourth nucleotide after V2, irrespective of the number of nucleotides, in case of similarity) was selected as V3 sequence, resulting the output produced by the program.

The V1 was the last identified palindrome with the appropriate construction with a stem, composed by the highest number of strong bindings, interrupted by a lateral palindrome and the characteristic CC bulge and ending with a variable loop. Particular aspects had to be taken into account, as the variable number of nucleotides included in the palindrome, mainly 39 and up to 42 in BVDV 2 and Giraffe, and the variable starting point of the V1 in the genomic sequence, depending from the type of primers used for RT-PCR reaction. The parameters for the determination of the V1 locus were related to the position of V2 locus in the sequence, as per V3.

Starting from the beginning of the V2 locus and proceeding backward on the sequence, for a minimum of ten nucleotides, were identified G or A followed by a nucleotide and then by U and C (the description is following a backward search, in the sequence this appeared as 5'-nnnnCUnG/Ann..nnV2-3'). All the CUnG/A sequences were retained and recorded in a list. In case no CUnG/A sequences could be identified, the program concluded the procedure for the construction of the V1 locus. Furthermore, continuing backward, all the sequences C and A, followed by a nucleotide and a C or A or G (5'-n..nG/A/CnACnn..nnCUnG/An..nV2-3') were identified and recorded in a list. This search was related to the definition of the core of the V1 locus, identifying the two C composing the characteristic bulge located on the stem of the palindrome and considering two nucleotides more, one forward and the other backward (5'-n..n*nG/A/CnACnn..nnCUnG/An*n..nV2-3'). The V1 locus was completed considering backward five nucleotides, of which the first is usually U (5'-n..n*nnnnUnG/A/CnACnn...nnCUnG/An*n..nV2-3') and forward height nucleotides, of which the second, the third and the fourth were characteristic AUG (exception made for BVDV-1b strain Sanders where A is changed with G) (5'-n..n*nnnnUnG/A/CnACnn..nnCUnG/AnnAUGnnnn*n..nnV2-3'). The remaining part of the stem with the lateral palindrome was so constructed. At the level of the lateral palindrome, on the opposite was included a deletion point in order to equilibrate correctly the stem. Per each CUnG/A identified sequences was performed the construction of a palindromic locus according to each defined C/A/GnAC following sequences. All identified possible structure sequences were constructed (30-n+1) and included in a list of candidate V1 in case of the distance between the beginning of the sequence and the first element before a CUnG/A did not exceed the 30 nucleotides. V1 candidate structures were constructed and considered in the list according to the *m* number of C/A/GnAC identified sequences, with variable loop dimension not exceeding 30 nucleotides. The selected sequence showed the highest number of strong bindings in the stem, and in case of similarity, that with the minor number of nucleotides. The absence of the sequence C/A/GnAC indicated the incompleteness of the V1 locus therefore incomplete palindromes were constructed identifying the starting point of the V1, as at the possible level of nAC, AC or C from an incomplete C/A/GnAC sequence, and determining a series of possible V1 sequences, with different dimensions, with a variable loop

starting from the first, second or third nucleotide of the tested sequence, included in the V1 candidate list(the V1 were constructed only with variable loop dimension not exceeding 30 nucleotides).

Log file

The text file *"log.txt"* created during the execution of the program, is important for monitoring the execution of the program. Its content include all the different issues from all necessary steps for the determination of the final result. In case of unclear or incoherent results, the analysis of this file allowed to understand the cause of the occurred problem as malfunctioning of the program or erroneous rationale.

Application of *Pestivirus* identification keys and html file output

After the comparison between the identified structure and all genus, species and genotype parameters, taking also into account all the known exceptions of divergent base pairs in the genotypes, the final result consisted in a .html file, created by the program during the analytical processing and showed at the end of it. This document shows graphically the portions of the sequence reconstructed and identified as the palindromic V1, V2 and V3 variable loci. The corresponding of PNS *Pestivirus* identification keys in the tested sequence was highlighted at the level of genus and species specific PNS parameters indicating the matching with characteristic base pairings. At the genotype parameter level, the highlighting was applied only in relation to the identified species. Per each PNS pattner, the related control result was included in the output file, showing the expected base pairing and the observed one at each position in the structures, in order to evidence any relevant characteristic in the strategic region of the 5'-UTR.

The software was tested considering the nucleotide sequences in the 5'-UTR of five hundred-tirty-four *Pestivirus* strains of the species BVDV-1, BVDV-2, BDV, CSFV, and of the tentative species Giraffe BVDV-3, BDV-2, Pronghorn and Bungowannah. The sequences, with different geographical origin, from different host species or contaminants of biological products, were obtained from the the GenBank DNA database, provided by authors or obtained in our laboratories (Table 1) (detailed list of analysed strains available under request).

RESULTS

The realization of the PNS software resulted in satisfactory prototype, as demonstrated by the successful application of the testing on a large number of virus strains. The sequences were correctly displayed with their palindromes and the application of the keys for *Pestivirus* identification showed clear results presented in the output file. The identification step allowed to three distinct evaluations. The first was the comparison with genus specific PNS, identifying the appurtenance to the *Pestivirus* genus. The further evaluations were applied only in case of matching. The following comparison was performed with species specific PNS for BVDV 1, BVDV 2, BDV, CSFV and the new proposed taxons. The last comparison was performed for genotype determination within a selected species

A file html type was elaborated at the end of the procedure useful for stocking and printing the results, in which were shown the three palindromic fractions of the sequence, V1, V2 and V3, and the related parameters for genus, species and genotype characterization (Figures 2 and 3).

The secondary structure construction procedure showed results with slight differences and more precision at loop level from those obtained by Genetyx-Mac software, based on the algoritm of Zuker and Stiegler (1981) with minimum free energy calculated according Freier *et al*. (1986). For example, the BVDV-2 strain BS-95-II V1 locus presented a loop consisting of 9 nucleotides (5'-AUCAGUUGA-3'). This sequence was reorganized with a loop reduced to four nucleotides (5'-AGUU-3'), followed by two strong binding base pairs G-C and A-U and an unpaired adenine, resulting in an irregular palindromic shape, when calculated by Genetyx-Mac (not shown). However, this discrepancy was occasional and in general the two applications corresponded. In the Giraffe strain, the V1 locus obtained by Genetyx-Mac showed correct palindromic shape and the two potential strong bindings A-U at the level of the loop were not applied, avoiding a similar alteration as in BS-95-II.

The construction of the V2 palindrome did not presented particular difficulty due to the strictly conserved nucleotide sequences composing the locus. The variability of the nucleotide number composing the other two palindromic structures V1 and V3, was, at the contrary an element requiring particular attention. However, the resulted structures were correct and revealed relation with expected genus characteristics. Table 4 shows an example of construction of the V3 palindrome locus with determination of the length of the sequence. Only in four cases, the construction of V3 of the BVDV-1 strains zvr711, 1248/01, G and W was problematic due to specific and uncommon aspects of their sequences.

The V1 was the last identified palindrome. Particular aspects had to be taken into account. The number of nucleotides included in the palindrome was variable, mainly 39 and up to 42-44 in some BVDV 2 and BDV strains, and Giraffe. The starting point of the V1 in the

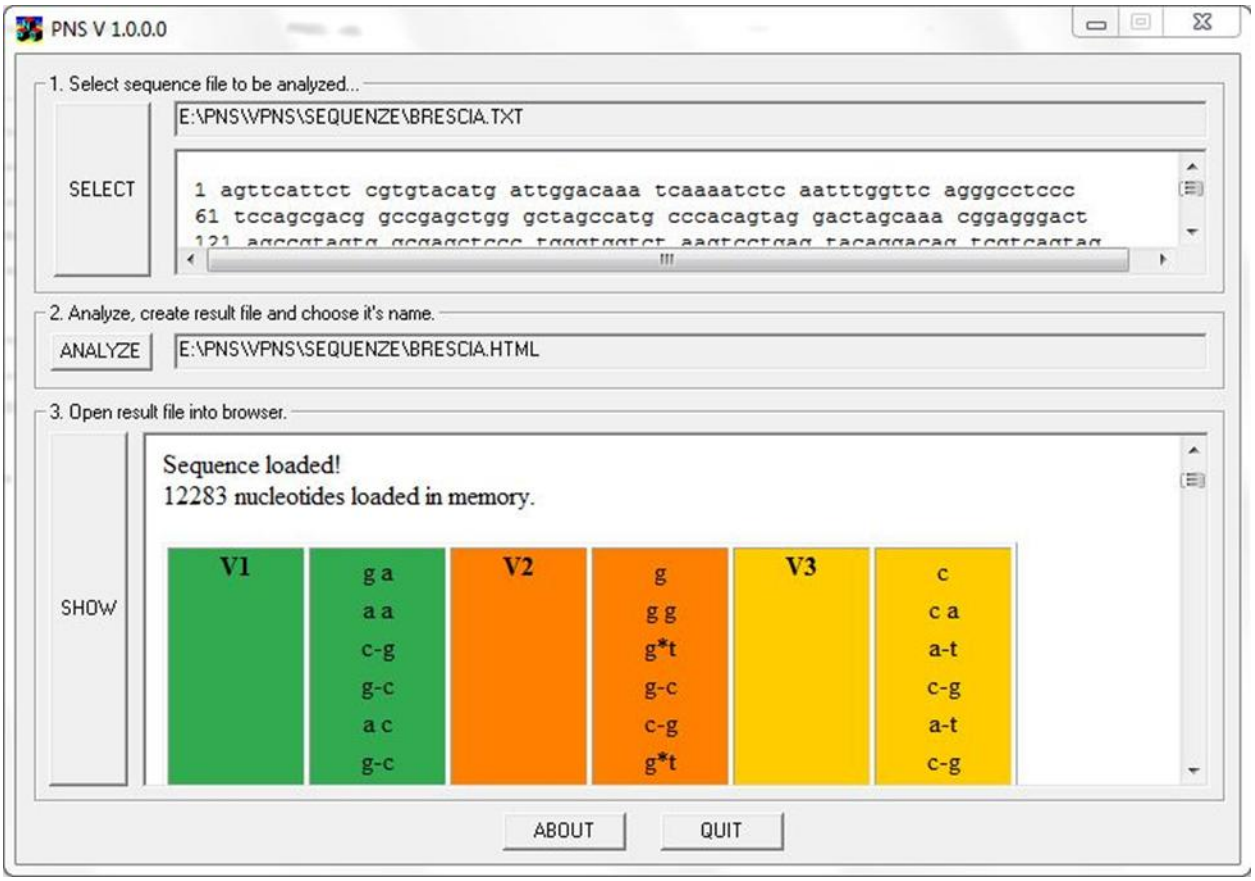

Figure 2. PNS software allowed a simple and intuitive utilization, through selection of sequence to be tested included in a text file format, and display of secondary structure relevant loci with subsequent analytical classification

genomic sequence was variable depending from the type of primers used for RT-PCR reaction. In order to construct the correct V1 palindrome, the specific condition of a minimum number of nucleotides between the C of the characteristic bulge (5'-nG/A/CnACn..nCUnG/An-3'), and between the V1 and V2, had to be applied. This parameter was necessary to avoid the construction of incoherent V1 palindromes, since some strains showed possible combination of G/A/CnAC and CUnG/A consecutive in the sequence. In CSFV strains Alfort and Brescia, starting from the V2, backward for V1 identification, with the first CUnG/A sequence G in position 217 and the following first G/A/CnAC sequence with C in position 210, the program, without correct parameters, constructed a very short V1 (not shown). The correct combination was obtained considering the following G/A/CnAC sequence in position 194. Similar aspects were observed in the BVD-2a strains 713/2, 11mi97 and 552195 and BVDV-1b Influenza2.

In case of incomplete V1 sequences, as for BVDV-1 strains Massimo 1, 2 and 4, the V1 palindrome was constructed according to a predicted structure corresponding to conserved locus in the genus. The program completed the construction of the V2 and V3 palindromes and the related genotyping procedure. The genus and species characteristics were determined. The identified genotype specific PNS were not sufficient to allocate these strains. In case of absence of the V1 sequence, as for BVDV-1 strain M98, BVDV-2 strains 59386 and Scp, and BDV strain L83/l84, the program was orientated to determine species and genotype specific PNS in remaining two loci, V2 and V3. For example, in strain M98 sequence the program could identify the BVDV-1 species specific PNS G-C in position 5 in V2 and in V3.

DISCUSSION

The PNS software represented the transposition of the *Pestivirus* genetic characterization method in a computerized procedure, a relevant improvement with a main advantage represented by the rapidity of the execution of the testing procedure providing data for accurate analyses. The prototype of the program

V1	a a g a g g c a t-a c-g c-g a-t c c a-t g-c c-g t-ag . a t*gt g-c g-c t-a g-c	V2	g g g g*t g-c c-g a c g-c g-c t-a c-g c-g a-t	V3	a g a t t t-a t-a g-c a-t c-g g*t a-t

Genus characteristic PNS: PNS recognized.		
V1:	t in position 22	There is no binding to check.
	t in position 5 right nucleotide	YES (t*gt)
	c c in position 11	YES (c c)
	a t in position 10	YES (a-t)
	c g in position 8	YES (c-g)
	or t a in position 8	NO (c-g)
	or t g in position 8	NO (c-g)
	or g g in position 8	NO (c-g)
	t a in position 7	YES (t-a)
	or a a in position 7	NO (t-a)
	or g c in position 7	NO (t-a)
	. a in position 6	YES (. a)
	or . g in position 6	NO (. a)
	t g in position 5	YES (t*g)
	g c in position 4	YES (g-c)
V2:	g t in position 10	YES (g*t)

Figure 3. Result .hlmt type file of genotyping according to the PNS method of the Europa strain. The 5'-UTR sequence was compared to known *Pestivirus* species BVDV-1, BVDV-2, BDV, CSFV and the new proposed taxons. The three palindromic regions in sequences were identified in the sequence and shown in the first part of the file. Genus *Pestivirus* characteristic PNS and BVDV-1 species characteristic PNS were identified in the palindromes and highlighted in the following section of the file. Genotyping was completed by the evidencing of BVDV-1c specific PNS

successfully demonstrated to be a simple and useful tool for the sequence testing indicating clear results for the allocation of unknown isolates and providing support for research work trough identification of peculiar

Table 4. Palindromic nucleotide substitutions (PNS) genotyping method for genus *Pestivirus*. Construction of the V3 palindrome locus with determination of the length of the sequence. The example was applied on the V3 locus of strain Lees which includes 18 nucleotides. Until now, observations indicated V3 to change form 15 to 20 nucleotides, however, it not possible to exclude new and different data. a) Identification of the linear sequence. The first nucleotide of V3 had a highly conserved position as the fourth nucleotide after the end of V2, and as the third nucleotide only in some strains. Due to the variability of the V3 sequence, the last nucleotide was not determined. b) Determination of the most stable palindromic structure. The sequence was gradually replied onto itself, searching for a maximum number of strong bindings. I. A palindromic structure with 2 strong bindings and 1 weak binding was identified with a 13 nucleotide sequence. II. Sequence with 16 nucleotide showed 3 strong bindings. III. Sequence with 18 nucleotides showed 6 strong bindings and 1 weak binding. IV. A longer sequence showed instable structure. The 18 nucleotide sequence resulted compatible with V3 locus

a)	5’-V2NNNAGCGCCAUUCGUGGCGUUNNNNNNNN-3’ \ last nucleotide of V2	
b)	I. A C U C U G-C C-G G*U 5'-A GGCGUUNNNNNNNNNNN-3'	II. UU A C C-G C U G G C-G G-C 5'-A GUUNNNNNNNNNNN-3'
	III. UC U G A-U C-G C-G G-C C-G G*U 5'-A-U-NNNNNN-3'	IV. C U G U U A G C-G C C G G C U G*U 5'-A-U-3'

characteristics in strategic genomic regions. In addition to recognized PNS, were made available also all structures indicating similarity or divergence, in terms of specific nucleotide base pairings, among virus genomic sequences at the level of the 5’-UTR, possibly expression of evolutionary changes or virus biological activities, such as virulence (Topliff and Kelling, 1998).

The preparation of the software for the PNS method, presented in this study for the first time, named PNS, freely available at www.pns-software.com, with the full computerization of the procedure, eliminated the main limitation due to manual searching of relevant base pairings and direct observation of the sequence, simplified the genotyping procedure for an easy access of the users and a rapid testing with reliable results, allowing the consideration of secondary structures predicted at the three variable regions in the 5'-UTR for the classification of *Pestivirus*. Future improvement will be required to standardize the procedure and increase the performance of the software in order to eliminate any possible incoherence. This aspect could be important also for possible adaptation of the methodology to other positive polarity RNA virus species, as Poliovirus or Hepatitis C virus.

REFERENCES

Apicella C (1999). Una metodologia per la progettazione ad oggetti di applicazioni relazionali: moduli back-end. Thesis of Doctorate, University of Salerno, Italy.

Deng R, Brock KV (1993). 5´ and 3´ untranslated regions of pestivirus genome: primary and secondary structure analyses. Nucleic Acids Res. 21: 1949-1957.

Ellis MA, Stroustrup B (1990). The Annotated C++ Reference Manual. Editors Addison Wesley, London.

Freier SM, Kierzek R, Jaeger JA, *et al*., (1986). Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA 83: 9373-9377.

Giangaspero M (2011). Genetic variation of Border Disease Virus species strains. Vet. Ital. 47: 415-435.

Giangaspero M, Harasawa R (2011). Classification of Pestivirus species genotypes based on palindromic nucleotide substitutions, a genetic marker in the 5' untranslated region of genomic RNA. J. Virol. Meth. 174: 166-172.

Giangaspero M, Harasawa R (2008). Genetic variation of classical swine fever virus based on palindromic nucleotide substitutions, a genetic marker in the 5' untranslated region of RNA. Vet. Ital. 44: 305-318.

Giangaspero M, Harasawa R, Weber EL, Belloli A (2008). Genoepidemiological evaluation of Bovine viral diarrhea virus 2 species based on secondary structures in the 5' genomic untranslated region. J. Vet. Med. Sci. 70: 571-580.

Giangaspero M, Vacirca G, Harasawa R, Büttner M, Panuccio A, De Giuli Morghen C, Zanetti A, Belloli A, Verhulst A (2001). Genotypes of Pestivirus RNA detected in live virus vaccines for human use. J. Vet. Med. Sci. 63: 723-733.

Harasawa R (1994). Comparative analysis of the 5' non-coding region of pestivirus RNA detected from live virus vaccines. J. Vet. Med. Sci. 56: 961-964.

Harasawa R, Giangaspero M (1998). A novel method for pestivirus genotyping based on palindromic nucleotide substitutions in the 5'-untranslated region. J. Virol. Meth. 70: 225-230.

Harasawa R, Giangaspero M, Ibata G, Paton DJ (2000). Giraffe strain of pestivirus. Its taxonomic status based on the 5' untranslated region. Microbiol. Immunol. 44: 915-921.

King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (2012). Virus taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier-Academic Press, Amsterdam.

Kirkland PD, Frost MJ, Finlaison DS, King KR, Ridpath JF, Gu X (2007). Identification of a novel virus in pigs-Bungowannah virus: a possible new species of pestivirus. Virus Res. 129: 26-3

Lodi E, Pacini G (1998). Introduzione alle strutture di dati. Editors Bollati Boringhieri, Bologna, Italy.

Thompson JD, Gibson TJ, Plewniak F, Jaenmougin F, Higgins DG (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4867–4882.

Topliff CL, Kelling CL (1998). Virulence markers in the 5' untranslated region of genotype 2 bovine viral diarrhea virus isolates. Virology 250: 164-172.

Vilček S, Ridpath JF, Van Campen H, Cavender JL, Warg J (2005). Characterization of a Novel pestivirus originating from a Pronghorn antilope. Virus Res. 108:187-193.

Zuker M, Stiegler P (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary. Nucleic Acids Res. 9:133-148.