

# Uses of sequencing



## Enric Mateu

Senior lecturer at the Department of Animal Health and Anatomy at the Faculty of Veterinary Medicine/Science (Autonomous University of Barcelona) Researcher at CReSA-IRTA



## Cinta Prieto

Senior lecturer at the Department of Animal Health at the Faculty of Veterinary Medicines/Science (Complutense University of Madrid)

## Highlights

**In most of the cases, the virus genome sequenced area is the ORF5 gene. This is due to the fact that the variability of this gene is high enough to produce a good discrimination between strains.**

**A first use of the sequencing is the distinction of the isolates, mainly in the sense of being able to distinguish if an outbreak has been caused by a new strain or if it is a new outbreak due to an already existing strain.**

**Relatively arbitrarily, it is considered normal that two isolates that share a similarity lower than 97% in the ORF5 gene sequence probably have different origins, although the interpretation of the similarity between two strains should be refined with the data originated from the epidemiologic and clinical situation on the farm.**

**The second aspect for which the sequencing can be very useful is for tracing the 'movements' of the PRRS virus strains in an area or region. This implies the maintenance of updated databases and the development of connection maps between the different farms in a region.**

**The sequencing does not provide us with information on the virulence nor does it allow us to decide which is the vaccine to be used.**

The launching of control programmes, on the farm or at a regional level, has created the need of a tool that allows us to establish the origin of the PRRS viruses that circulate in both places. This knowledge can help us, in the first place, to establish if it is a new outbreak from the external introduction of viruses and, in the second place, it allows us to trace, somehow, the origin of the external introductions or the spreading routes of the virus in a region (connections), provided that we have sufficient database.

In most of the cases, the virus genome sequenced area is the ORF5 gene. This is because the variability of this gene is high enough to produce a good discrimination between strains, keeping segments almost preserved, and allowing the grouping quarantines consistently. Likewise, the ORF5 gene could contain a relevant neutralisation epitope in terms of protection, although its role has been recently questioned. Another segment of the virus genome sometimes used is the ORF7 gene (that encodes the nucleoprotein), although its discrimination power is much lower. We must bear in mind that the ORF5 (or ORF7) gene represents less than 4% of the virus genome, so the conclusions must always be formulated with caution.

Regularly, the sequencing reports usually show the sequence of the strain in question, the similarities with the commercial vaccinal strains, the closest field strain(s), the cluster it belongs to (generally through a phylogenetic tree), and sometimes the protein sequence it encodes and the alterations present in the sequence with regard to a reference strain.

As said before, a first use of the sequencing is the distinction of the isolates, mainly in the sense of being able to distinguish if an outbreak has been caused by a new strain or if it is a new outbreak due to an already existing strain. To this aim, the most practical solution is to compare the sequence of the current isolate with that of the isolate previously detected on the farm. Relatively arbitrarily, but typically considering that two isolates that share a similarity

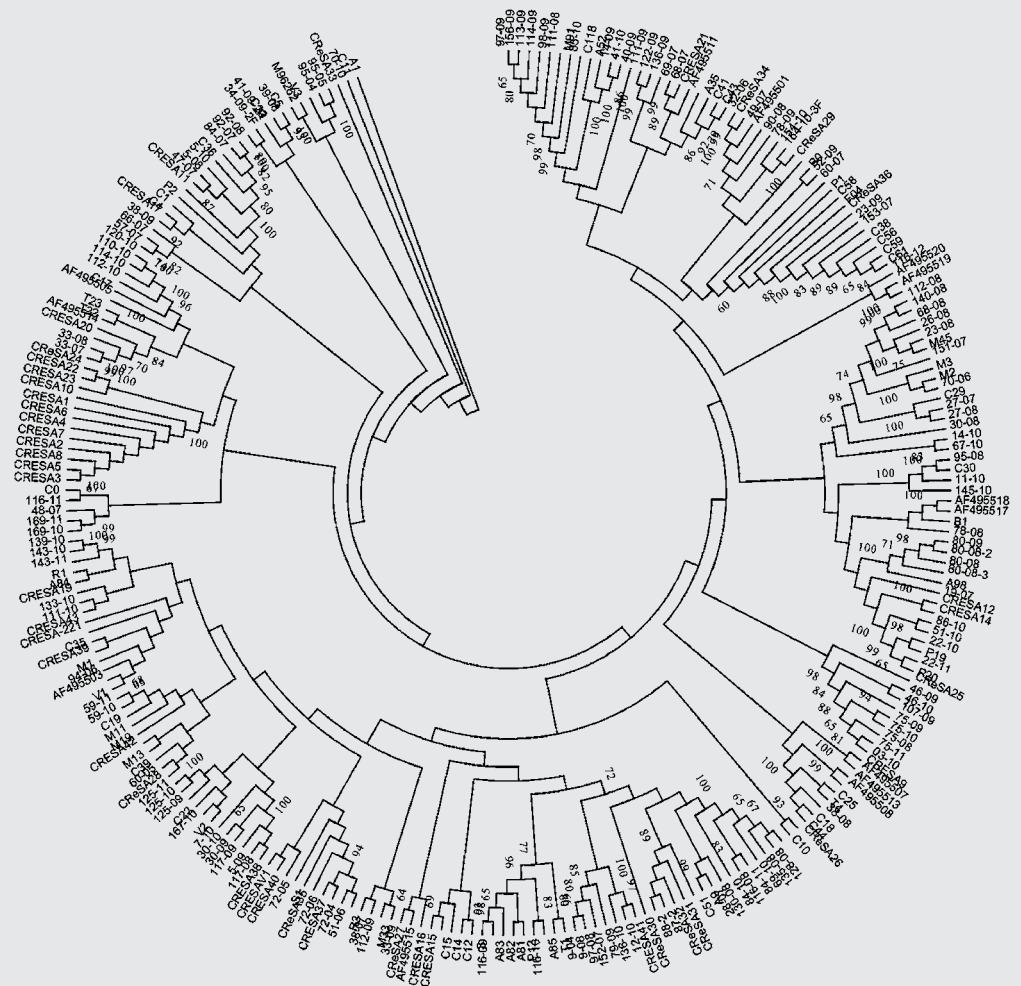
lower than 97% in the ORF5 gene sequence probably have different origins.

As a sign; when this has been studied, it has remained established that the annual ORF5 sequence variation rate is approximately 0.5%, although it can reach 1%, even though in closed herd systems it can be lower. In order to understand this figure and its meaning we must refer to the origin of the genetic variation of the virus. A great part of the variation stems from the random errors produced by the viral RNA-polymerase when replicating the virus genome. The higher the virus transmission and replication on the farm the higher the number of virus particles produced, and therefore the higher the possibility of generating new variants. In addition, the existence of a certain level of immunity on the farm is going to reduce, on one hand, the virus replication, but on the other hand will pick out certain variants. If we understand the complexity of the viral variation phenomenon, we will see that the interpretation of the similarity between two strains should be refined with the data obtained from the epidemiologic and clinical situation of the farm.

The second aspect for which the sequencing can be very useful is for tracing the 'movements' of the PRRS virus strains in an area or region. This implies the maintenance of updated databases and the development of connection maps between the different farms in a region.

In any case, at the present moment, the sequencing does not provide us, in a direct and acceptable way, with information on the virulence of the strain (although the similarity with other extremely virulent strains allows us to make assumptions), and it neither allows us to decide on the vaccine to choose for a specific case, because nowadays the protection cannot be predicted from any sequence. We must underline that the interpretation of the epidemiologic meaning of the sequences obtained improves as we have a higher number of sequences. In other words, it is necessary to constantly sequence to have extra accuracy available when interpreting the results.

**Figure.** Different ways of showing the similarity of a certain PRRS isolate with another specific sequence (A) or with groups of sequences (B). In (A) we can see an alignment between two sequences in which we can check the specific differences in the nucleotides between both as well as their global similarity percentage. In (B) we show two kinds of phylogenetic trees with the same data. The trees show groups of strains related phylogenetically; this is, groups that have a common evolutionary origin. The numbers in the branches tell the probability that each branch really reflects a statistically different group (in terms of percentage).

**A****B**

## References

- Diaz *et al.* 2006. Virology, 351: 249-259.
- Murtaugh M. 2012. Proc. Allen D. Leman Swine Conference, 49-55 pp.
- Prieto *et al.* 2009. Vet J, 180: 363-370.
- Prieto *et al.* 2008. Vet J, 175: 356-363.

