Phylogenetic relationships and percent sequence identity in combination with RFLP to interpret virus relationships

D. Marthaler, M. Murtaugh

Key Points:

- Using a standard nucleotide percent identity cut-off value (97% or 98%) or RFLP cut-patterns can be misleading.
- Phylogenetic tree (dendrogram) is necessary to understand the relationship of PRRSV strains with similar nucleotide percent identity and the same RFLP cut pattern.

A reader inquired about the lineages used in last week's report. Leung and colleagues (Shi et al. 2010. Journal of Virology 84:8700-8711) described a PRRSV classification system with groupings of related viruses, they referred to lineages and sub-lineages, based on 8,624 ORF5 globally distributed type 2 PRRSV sequences.

Figure 1 shows a dendrogram of lineage 1 sequences from last week's phylogenetic tree. The 1-8-4 RFLP family first identified in 2000 in MN is in this lineage. Sub-lineages are presented by different colors. The light blue cluster (clade or sub-lineage) of sequences at the top of the dendrogram contains RFLP cut patterns 1-4-2, 1-4-4, and 1-7-4. Within the cluster, a single sequence (C) and two other groups of sequences (A and B) were labeled to illustrate how phylogenetic and nucleotide percent identity provides useful information that is missing in an RFLP.

These sequences all have the RFLP cut pattern of 1-7-4. The group A sequences shared a 99.0-99.5% nucleotide identity, and the group B sequences share a 98.8-99.2% nucleotide identity. Pairwise identity between group A and group B sequences ranges between 96.7-97.4%.

Strain C's based identity is 98.8-99.2% with the sequences in group A and has 97.7-98.3% with the sequences in group B. These similarities look about the same. How can a person tell where strain C came from? The answer is in the phylogenetic tree. The phylogenetic tree shows genetic relationships based on the minimum number of base changes that would link two sequences. The closest neighbor most similar to strain C is in group B, indicated by the asterisk in Figure 1. Even though the viruses in group A have about the same percent similarity, the exact base differences are not the same as in the group B and strain C viruses.

The closer genetic relationship of strain C to group B was made evident by construction of the phylogenetic tree. We see here how the pairwise comparisons give more information than a simple RFLP type. We also see that viruses that are 98% identical can be essentially the same while a 97%, 99%, or whatever cut-off that is being used can be misleading. Viruses can mutate in different ways to create similar but distinct sub-groups. The small differences can be significant when you need to know where a virus came from, and can be discerned in a phylogenetic tree.

In conclusion, the phylogenetic tree is the most important part of PRRSV sequence analysis since both nucleotide percent identity and RFLP cut-pattern can mislead the interpretation. We always start with the phylogenetic tree and use the nucleotide percent identity to support the analysis. RFLP cut-pattern may be convenient but used by itself can be misleading since it does not provide useful information when investigating the relationship among PRRSV sequences.