RESEARCH REPORT
SWINE HEALTH

Title: Identification of genetic signatures for African swine fever virus serologic group specificity – **NPB #12-106** revised

Investigator: Daniel L. Rock

Institution: University of Illinois, Urbana-Champaign and the National Research Institute for Veterinary Virology and Microbiology, Pokrov, Russian Federation

Co-Investigator: Denis Kolbasov

Date Submitted: July 1, 2013

Industry Summary:

African swine fever (ASF) is an acute viral hemorrhagic disease in domestic swine with mortality rates approaching 100%. Devastating ASF disease outbreaks and the continuing disease epidemic in the Caucasus region and Russia (2007 – to date) highlight the significance of this disease as arguably the most significant emerging disease threat for the swine industry worldwide. There is no vaccine for ASF available; however, it is clear that vaccination is possible since protection against homologous reinfection has been definitively demonstrated. Vaccine progress is hindered by lack of knowledge concerning the extent of ASFV strain variation and the viral antigens responsible for protective immunity. To date, eight ASFV serogroups have been identified although more likely exist. Notably and of great significance with respect to vaccine design and development, viruses within serogroups provide cross-protection from challenge with viruses within the serogroup. Unlike other viruses currently challenging the swine industry, such as PRRS, ASFV is a DNA virus and thus much less variable over time. Once the viral diversity currently circulating in natural reservoirs has been determined, a set of vaccines could be developed to be effective against relevant ASFV strains.

Here, we have used a collection of serologically-grouped ASFV isolates and a large and diverse collection of ASF viruses to identify genetic signature(s) for ASFV serologic group specificity and to further define ASFV strain variability. We have demonstrated through gene sequencing and comparative analysis of ASFV strains a correlation between the genotype of the ASFV CD2v gene and virus grouping based on serospecificity. Overall, the concordance between CD2v region phylogenetic data and serogroup-specific typing provides predictive value of CD2v locus genotyping in predicting serologic, and potentially cross protective, virus groups.

Results outlined here will have broad impact on vaccine-orientated approaches for ASF disease control thus reducing the threat posed by this high consequence viral disease. Knowledge of ASFV strain diversity and the breath of strain variation in nature as well as rapid genotyping methods to serotype viruses and to predict efficacy of a given vaccine to provide cross protection for a newly identified field isolate will facilitate vaccine design, development and emergency use. (For additional information contact D. L. Rock - dlrock@illinois.edu)

Keywords:

African swine fever**,** ASF**,** strain variability, genotyping, viral CD2v, cross protective immunity

Scientific Abstract:

African swine fever (ASF) is arguably the most significant emerging disease threat for the swine industry worldwide. There is no vaccine for ASF available; however, it is clear that vaccination is possible since protection against homologous reinfection has been definitively demonstrated. Vaccine progress is hindered by lack of knowledge concerning the extent of ASFV strain variation and the viral antigens responsible for protective immunity. Eight ASFV serogroups have been identified and notably, and of great significance with respect to vaccine design and development, viruses within serogroups provide cross-protection from challenge with viruses within the serogroup. Here, we have used a collection of serologically-grouped ASFV isolates and a large and diverse collection of ASF viruses to identify genetic signature(s) for ASFV serologic group specificity and to further define ASFV strain variability. Overall, the concordance between CD2v region phylogenetic data and serogroup-specific typing provides predictive value of CD2v locus genotyping in predicting serologic, and potentially cross protective, virus groups.

Introduction:

ASF is an acute viral hemorrhagic disease in domestic swine with mortality rates approaching 100%. Devastating ASF disease outbreaks and the continuing disease epidemic in the Caucasus region and Russia (2007 – to date) highlight the significance of this disease threat for domestic swine populations worldwide. ASF is arguably the most significant emerging disease threat for the swine industry worldwide. There is no vaccine for ASF available; however, it is clear that vaccination is possible since protection against homologous reinfection has been definitively demonstrated. Vaccine progress is hindered by lack of knowledge concerning the extent of ASFV strain variation and the viral antigens responsible for protective immunity (Costard et al. 2009; Tulman et al., 2001 and 2009). Unlike other viruses currently challenging the swine industry, such as PRRS, ASFV is a DNA virus and thus much less variable over time. Once the viral diversity currently circulating in natural reservoirs has been determined, a set of vaccines could be developed to be effective against relevant ASFV strains.

Scientists at the National Research Institute for Veterinary Virology and Microbiology, Pokrov, Russian Federation (VNIIVViM) have developed and refined a serologic typing assay for ASFV based on hemadsorption inhibition (HAI). To date, eight serogroups have been identified although more likely exist. Notably and of great significance with respect to vaccine design and development, viruses within serogroups provide cross-protection from challenge with viruses within the serogroup (Baluishev et al., 1995; Baluishev et al., 2010; Vishnyakov et al., 1995, and VNIIVViM unpublished data). VNIIVViM holds a very large and diverse ASF virus collection. The ASFV strains represented are from geographically diverse regions of the African continent, Europe and the Caribbean. Notably, the collection is particularly well characterized with respect to viral virulence and host range in animals.

The VNIIVViM ASFV HAI - based serologic typing assay, a collection of serologically-grouped ASFV isolates and a large and diverse collection of ASF viruses provides a truly unique resource for defining ASFV strain variability and establishing its relationship to cross-protective immunity.

Objective:

Identify genetic signature(s) for ASFV serologic group specificity and determine serological group specificity for currently untyped ASFV isolates in VNIIVViM strain collection

Materials & Methods:

VNIIVViM ASFV isolate collection. Our experiments were based on the unique ASFV collection maintained at VNIIVViM. Viral isolates were collected during the former Soviet ASFV program at VNIIVViM and represent significant geographic and temporally diversity. These include isolates from disease outbreaks in Africa, Europe, the Caribbean, and more recently from the Russian/Trans-Caucasian epizootic. Also included are high passage derivatives phenotypically distinct (eg. attenuated) from outbreak strains. Critical for our objective was the data previously generated at VNIIVViM in characterizing serological group specificity for specific ASFV isolates by hemagglutination inhibition (HAI) assay. Serotyping at VNIIVViM has identified eight discernable serogroups (SG), although more are thought to exist. Many characterized VNIIVViM isolates fall within four serogroups (SG1-SG4). Other characterized isolates form an additional four serogroups (SG5-8). Not all isolates in the VNIIVViM collection have been serogrouped, leaving isolates of unknown genetic diversity and serotype.

These serogroups formed the phenotypic basis for subsequent genotypic association. This study assessed CD2v region sequences across the spectrum of serogoups established for ASFV isolates collected at VNIIVViM. Viruses selected for sequencing represented all eight characterized serogroups, including the reference strains for each serogroup and multiple isolates from most serogroups. Also selected were isolates not typed or considered untypable using SG1-SG8 reference antisera. Selected strains represented the best documented and most extensively characterized strains available from the VNIIViM collection as well as viruses that did not fit into the current typing scheme, and isolates for which serotyping data was inconclusive. Viral DNAs used in this study were extracted from either archived, frozen samples of infected swine tissue or from blood of viremic animals using the DNeasy Blood and Tissue kit (Qiagen, Inc). Samples were resuspended in deioinized water and kept at -20ºC until further analysis.

PCR and Direct Amplicon Sequencing. CD2v and adjacent genomic regions were PCR amplified and directly sequenced from most ASFV isolates of the VNIIVViM collection characterized for HAI phenotype. Primers specific for CD2v regions were designed using primou (http://www.genome.ou.edu/ informatics/primou.html) and compared to available nucleotide sequences to select primers conserved across known isolates. Primers were selected to cover an approximately 1300 base pair CD2v gene region with multiple overlapping amplicons (Table 1). PCR was conducted using high fidelity PCR kits and the following general PCR conditions (0.75U Taq, 10 pmol each primer, 0.5mM dNTP mix, 0.5mM MgCl2, in 25ul reaction, 95ºC 5 min. denaturation, 30 amplification cycles (95ºC 40 sec, 52-56ºC 45 sec, 72ºC 1-1.5 min), and 72ºC 7 min. final extension, in a PalmCycler thermocycler (Corbett Research). PCR products were resolved electrophoretically on 1-2% agarose gels, purified using the Qiaquick Gel Extraction Kit (Qiagen) and quantified (OD260) for use as templates in Sanger sequencing reactions. Ten microliter sequencing reactions were performed per manufacturers protocol (BigDye ® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems), purified using the BigDye® XTerminator Purification Kit (Applied Biosystems), and run on an Applied Biosystems 3130 Genetic Analyzer.

Sequence assembly and analysis. Chromatograms were basecalled and sequences assembled and edited with phred, phrap, and consed, respectively (www.phred.org). Assembled sequences were screened with Consed and evaluated manually for sequence quality and consistency in determining consensus sequences. Regions/reads yielding insufficient data were reamplified and/or resequenced. Consenus sequences were manipulated with the EMBOSS package (Rice et al., 2000). Edited assembly consensus sequences were conceptually translated, and relevant ORF sequences aligned using MAFFT (Katoh et al., 2002). ORFs containing functional lesions (i.e., frameshifts or in-frame stops) were manually corrected for purposes of phylogenetic comparisons and signature evaluation. SEAVIEW4 (Gouy et al., 2010) was used for manual review, refinement of alignments using CLUSTAL and MUSCLE (Edgar, 2004), and for selection of subregions. Automated removal of gapped and ambiguously aligned regions was done using gblocks (Castresana, 2000). Neighbor joining analysis was used as implemented in SEAVIEW4. Maximum parsimony and maximum likelihood analyses were used as implemented in PROTPARS (Felsenstein, 1989), and PHyML (Guindon and Gascuel, 2003).

Results:

CD2v Protein Domain Architecture. Initial sequencing involved full CD2v-locus coverage, and included complete CD2v gene sequences for a subset of VNIIVViM isolates comprising reference strains for all eight serogroups. Observed genes and predicted products demonstrated the overall structural elements described previously (J.Virol. 1993 67(9): 5312-20; Virology 1994 199:463-8), including amino-terminal signal sequences followed by: repetitive amino-terminal domain, conserved extracellular immunoglobulin domain, transmembrane domain, conserved predicted intracellular domain, proline-rich repetitive domain, and conserved carboxyl-terminal domain. Copy-number variants were observed in repetitive domains. For certain isolates (x2, x6, x7, x14, x20) functional lesions in coding sequences were observed and, where possible (x2, x6, x7, x14), were logically corrected to maximize data for protein sequence phylogenies.

Phylogenetic clustering of complete ASFV CD2v protein sequences correlated with HAI serogrouping. Phylogenetic analyses in search of CD2v gene sequence groupings consistent with grouping based on HAI were performed. Notably, simple distance-based phylogenetic reconstructions utilizing whole-CD2v protein sequence alignments yielded results indicating that CD2v-based groupings correlated with serogrouping (Figure 2). Sequences of similar serotype grouped reliably with each other, and they formed closely related clusters within the tree relative to isolates of other serotypes. Observed were significantly greater genetic distances between serogroup clusters than within them. Certain instances of exceptional intra-serogroup distance could be explained by lesions (i.e. deletion) relative to the wild-type sequence and which increased distance (i.e. x20).

Similar results and reliable genotypic grouping of serotypic groups were obtained using alternate alignment (Muscle, data not shown) or phylogenetic algorithms (Fig 3), supporting serotype/CD2v genotype clustering. Apparent between-serogoup relationships were observed in whole CD2v phylogenies; however, support for these relationships were lower than intra-serogroup clusters in any given tree. Also, consistency and support for inter-serogroup relationships were diminished comparing between phylogenetic methods. Overall, the correlation between whole CD2v protein typing and serotyping is robust when comparing reference strains.

CD2v genotyping indicated that additional serogroups likely exist, and that genetic diversity may vary within different serogroups. Notable in these data are three untyped strains (x19, s23, x24). These were distinct from other serogroups, having genetic distances to other serogrouped isolates greater than most within-serogoup distances and less reliably grouping with other isolates (Figs 2-3). Such outlier isolates potentially represent novel, and yet to be established, serogroups – indeed at least one of these (x19) could not be typed using existing reference antisera. Also notable was distance observed between isolates characterized as belonging to SG8 (Figs 2-3). While still reliably grouping together, taxa within the SG8 subgroup showed above average within-serogroup distances, suggesting greater diversity existing within the serogroup.

CD2v genotyping resolves serogrouping where P72 genotyping does not. Importantly, use of data from CD2v locus sequencing better matched serotypic grouping than does grouping based on P72 capsid protein genotyping, which has become the de facto standard for typing ASFV. Whereas P72 genotyping places multiple ASFV isolates into single groups, CD2v genotyping more accurately places viruses into groups that reflect phenotypic serological reactivity (Table 2). In fact, all known serogroups and a non-reactive type (x23) were indicated as separate groups based on CD2v genotype, where P72 typing failed to resolve multiple serogroups (Table 2). In addition, CD2v typing was able to group distinct isolates within the diverse SG8 cluster, where P72 typing has placed them into different groups. Overall, data here suggest that CD2v genotypic data provide greater resolution then P72 genotype and more accurately allow prediction of serogroup.

Use of CD2v protein data to predict serotypes of untyped strains. The consistent CD2v genotypic grouping of serotypically similar viruses suggested that ASFV sequences falling within or very near to a CD2v-based cluster may be predicted to also have a similar serotype. Such is the case for VNIIVViM isolate x15, which lacks serotype designation but using CD2v data here clearly falls within serogroup 2 (Figs 2-3). Given establishment of the CD2v/serotype correlation based on reference and well-typed VNIIVViM isolates, it was of interest to assess these data with previously collected CD2v data from unserotyped ASFV isolates. Inclusion of previous CD2v data indicates that many unserotyped isolates fall within or very close to the serotype-specific CD2v-based clusters defined by the VNIIViM collection, and might be predicted to be of the same serotype (Figure 4). It also indicates that many unserotyped ASFV are very distinct by CD2v genotype, falling outside of VNIIViM CD2v clusters and potentially representing completely novel serogroups. Finally, it indicates that many previously sequenced ASFV isolates are located in the CD2v tree at positions intermediate to known VNIIVViM serogroups, suggesting potential for serologically cross reactive, and potentially cross-protective, types relative to SG1-8. Thus, CD2v/serotype correlation gives genotypic reference to ASFV serospecificity and predictive power in defining serotypes of novel ASFV strains.

CD2v signature subregion groups the VNIIVViM ASFV collection. Given the structure of the CD2v gene, with repeat-rich and variable regions in the amino and carboxyl terminal regions of the predicted protein, much of the informative data for phylogenetic analyses was contributed by more conserved subregions of the protein. Assessment of these conserved regions allowed selection of a protein subregion as a candidate signature for serotypic reactivity. This signature was subsequently used to screen additional, less-well characterized isolates of the VNIIVViM collection to probe for additional genotypic diversity and to continue associating CD2v genotype with serotype. Use of this signature sequence maintained whole CD2v-based serotypic clustering (Figure 5). Apparent with increasing sequence representation is the diversity represented within the SG8 cluster, with additional, closely related isolates (x48) affecting discrete SG8 grouping, particularly in relation to novel virus x24. Refinement of the sequence alignment and method of analysis still allowed discrimination of SG8 as a group within the CD2v tree (Fig 5). In addition, several less-well characterized VNIIVViM isolates (x49, x66) appeared discrepant within the tree; however, their true serotype status is currently unclear, and data here provide a screen with which to target such isolates for serotypic reevaluation.

Overall, the concordance between CD2v region phylogenetic data and serogroup-specific typing suggests predictive value of CD2v locus genotyping in predicting serologic, and potentially cross protective, types.

Discussion:

There is no vaccine for ASF available; however, it is clear that vaccination is possible since protection against homologous reinfection has been definitively demonstrated. Vaccine progress is hindered by lack of knowledge concerning the extent of ASFV strain variation and the viral antigens responsible for protective immunity. To date, eight ASFV serogroups have been identified although more likely exist. Notably and of great significance with respect to vaccine design and development, viruses within serogroups provide cross-protection from challenge with viruses within the serogroup.

Here, we have used a collection of serologically-grouped ASFV isolates and a large and diverse collection of ASF viruses to identify genetic signature(s) for ASFV serologic group specificity and to further define ASFV strain variability. Identification of genetic signatures for ASFV strain specificity may also identify relevant ASFV antigens associated with serotype- specific protective immunity.

The ASFV CD2v protein was selected as an initial target for identifying genetic signatures of serospecificity. The CD2v gene is a genetic locus previously identified as being highly variable in the ASFV genome. While all examined strains contained the CD2v locus, it exhibited an above average level of genetic diversity, despite is location in an otherwise conserved region of the genome. This variability, the known role of CD2v in hemadsorption by ASFV-infected cells, and the fact that the HAI-based serologic phenotype correlated with cross-protection, all suggested that CD2v was associated with ASFV serospecificity and possibly crossprotective immunity.

Data presented here indeed indicated that the CD2v locus contains genetic diversity and specificity to allow resolution of viruses into groups, recapitulating groups based on HAI serotyping. Use of phylogenetic models applied to whole protein sequence alignments discerned serotypic reference isolates from each other, and placed other well-characterized viruses of the same serotype clearly into the same sequence-based clusters. Serotypic clusters were reproducible among different alignments and phylogenetic models, indicating that this method in general captures the correlation of CD2v grouping with serotype. In addition, subregions of the CD2v gene retained information sufficient to maintain a genotype/serotype correlation, indicating that ASFV typing is possible with subregion signatures.

The ability to correlate serotype to CD2v genotype suggests that typing of CD2v has predictive value in assessing cross-protective potential of uncharacterized ASFV strains. Among VNIIVViM isolates, most clearly clustered within the eight characterized serogroups with relatively little within-cluster diversity, including isolates for which serotyping had not been conducted or was not clear. Also falling within or very near to serotype-specific clusters were previously sequenced CD2v genes from uncharacterized ASFV isolates. Given relationships between known serotypes, these isolates of uncharacterized serotype might naturally be predicted to be of the serotype specific to that CD2v cluster. Also observed were clusters of viruses characterized as having the same serotype and falling within a single CD2v-based cluster, but whose intra-cluster genetic diversity was greater than that seen within other serotypic clusters. Notable among these were viruses of serotype 8, which demonstrated multiple relatively diverse isolates within the serotype.

Addition of a larger number CD2v sequences from unserotyped ASFV isolates, or isolates failing to react with reference antisera, indicated increasing diversity outside the CD2v clusters specific to the eight VNIIVViM serotypes. These were represented both as discrete, divergent clusters that might be predicted to represent completely novel serogroups, and as clusters more closely related to or falling between serogroups. These observations suggest that 1) additional ASFV diversity exists and that additional serogroups likely remain to be characterized, and 2) there may be a continuum of ASFV diversity and that cross-reactive types remain to be qualitatively and quantitatively characterized. Ultimately, both potentially affect seroreactivity and cross-protection between viral subpopulations.

Finally, though CD2v genotyping here reflects serogrouping, it remains to be seen if CD2v is responsible for the cross-protective immunity observed for viruses within a serogroup. This important question will need to be addressed directly in vaccination /challenge experiments in animals. Although is seems highly likely that ASFV CD2v protein is an antigen involved in protective immunity, it may not be the only protective antigenic determinant(s). CD2v genotyping may ultimately reflect overall phylogeny. If so, other antigenic determinants may confer cross protection based on the same phylogenetic relationship, making CD2v genotypic data an indirect but highly significant signature for cross-protection.

Table 1: CD2v Region Primers

Primer	Sequence (5'>3')	Primer	Sequence (5'>3')
ga2453f	GGAATGTGGCATGGAGATC	ga3616r	TTAAACTACAATTATTTG
ga2576f	ATGGAAACCAAAATACCTTGTAC	ga3880r	GTATTATTAATCATACATGT
ga2629f	ACCTATTTTTGAATTTCATGT	ga4124f	CTGAATCTAATGAAGAAGA
ga2736r	CCTTATACAATACAGCATAAATAACGTGA	ga4220r	AGGGACGCATGTAGTAAATAG
ga3010f	TGTCCTAAAGATTGGGTTGG	ga4540f	GATCATCGAGAACATCATGTGATTC
ga3039r	ATTATTATATCCAACCCAAT	ga4638f	GCAGCTTTTTACCGGAGATTATCT
ga3278r	TATTTACTACAAATAAATAATA	ga4698r	AAGTCTTTGTAGGTTTTTCGTTCA
ga3611f	TATAATATAACAAATAATTGTAG	ga4808r	AATGCCGGGCCTTCTACAA

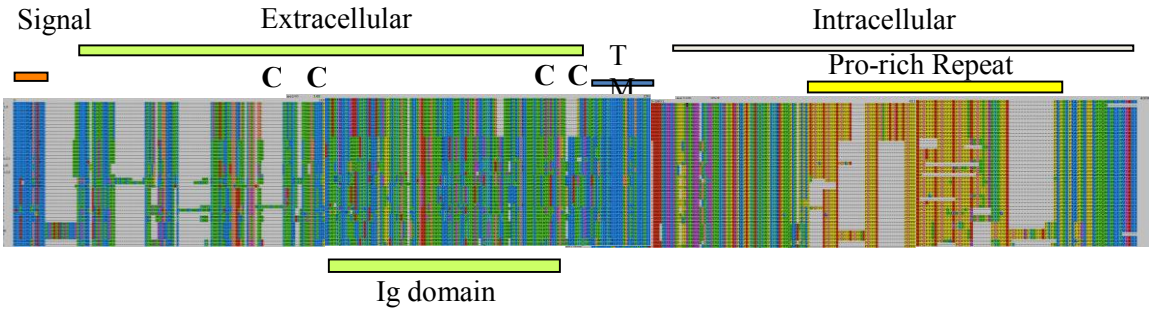

Figure 1: Predicted structure of CD2v protein. Predicted are N-terminal signal sequence, extracellular domain containing immunoglobulin (Ig) fold domain, transmembrane (TM) domain, and carboxyl-terminal intracellular domain containing proline-rich repeats bounded by conserved domains.

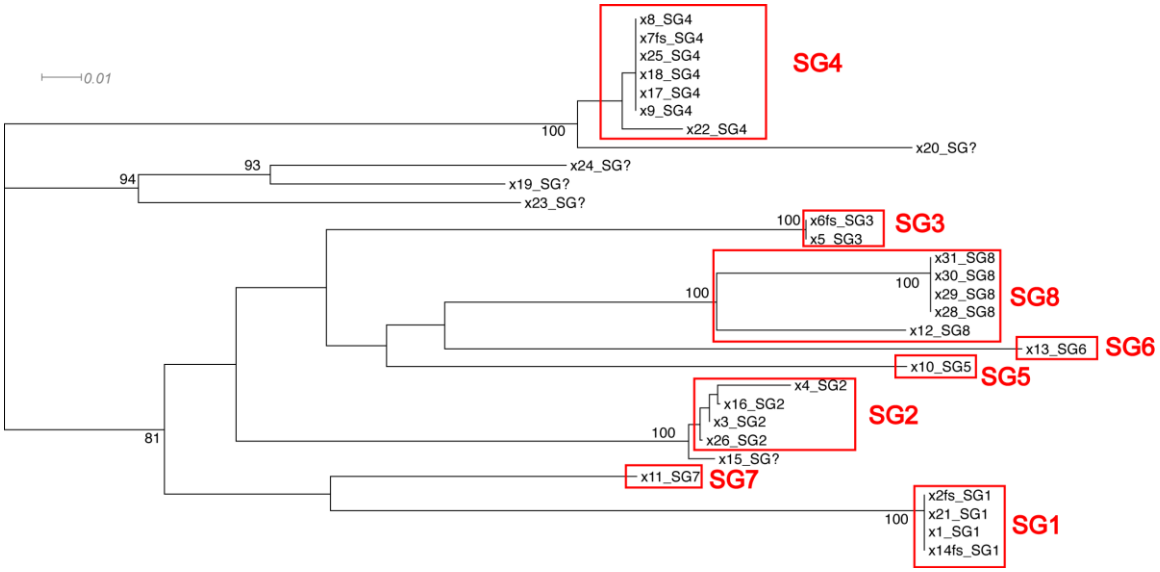

Figure 2. Neighbor joining phylograms of VIIMMViM full-length CD2v protein sequences. BioNJ algorithm using Kimura correction applied to 277 sites as aligned with Mafft. Bootstrap replicate support greater than 70% is indicated at appropriate nodes.

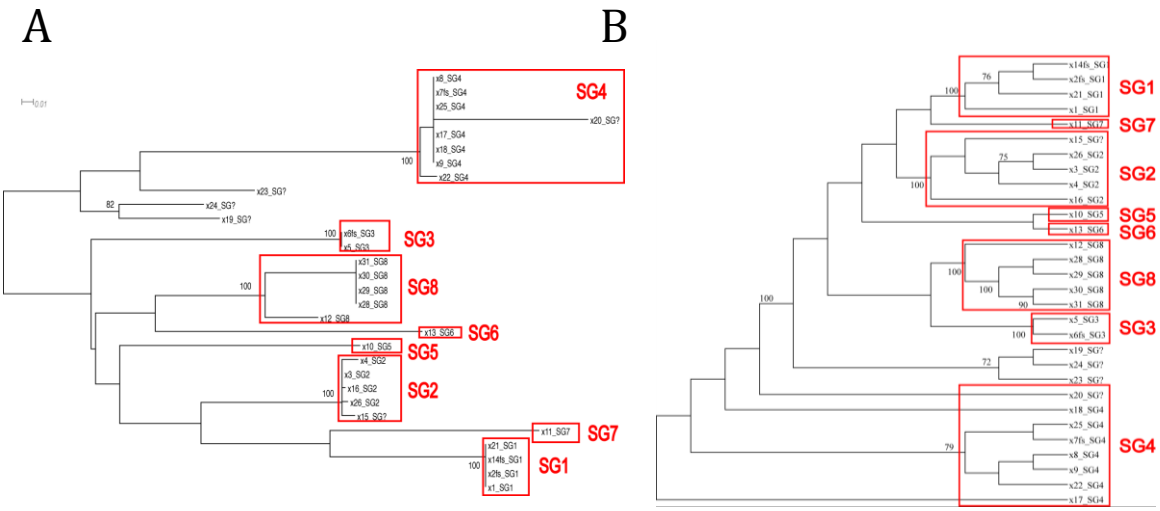

Figure 3: Maximum likelihood (A) and maximum parsimony (B) analyses of VIIMMViM full-length CD2v protein sequences. A) Phylogram of PhyML analysis using LG model. B) Cladogram of ProtPars analysis (328 sites, 159 informative). Bootstrap supports greater than 70% are indicated on branch leading to appropriate nodes.

Table 2: CD2v vs P72 Genotypes Relative to HAI Serotype

Strain	Virulence	HAI Serotype	P72 Genotype	CD2v Genotype
x1	virulent	I	I	I
x21	avirulent			
x2	avirulent			
x14	virulent			
x14/105	avirulent			
x3	virulent	II		II
x26	avirulent			
KK-203	avirulent			
x16	virulent			
x4	avirulent			
x9	virulent	IV		IV
x22	virulent			
x8	virulent			
x17	virulent			
FK-32/135	avirulent			
x7	avirulent			
x5	virulent	III	V	III
x6	avirulent			
x10	virulent	V	X	V
x11	virulent	VII		VII
x23	virulent	N/t		Unique
x28	virulent	VIII	II	VIII
x12	virulent		VIII	

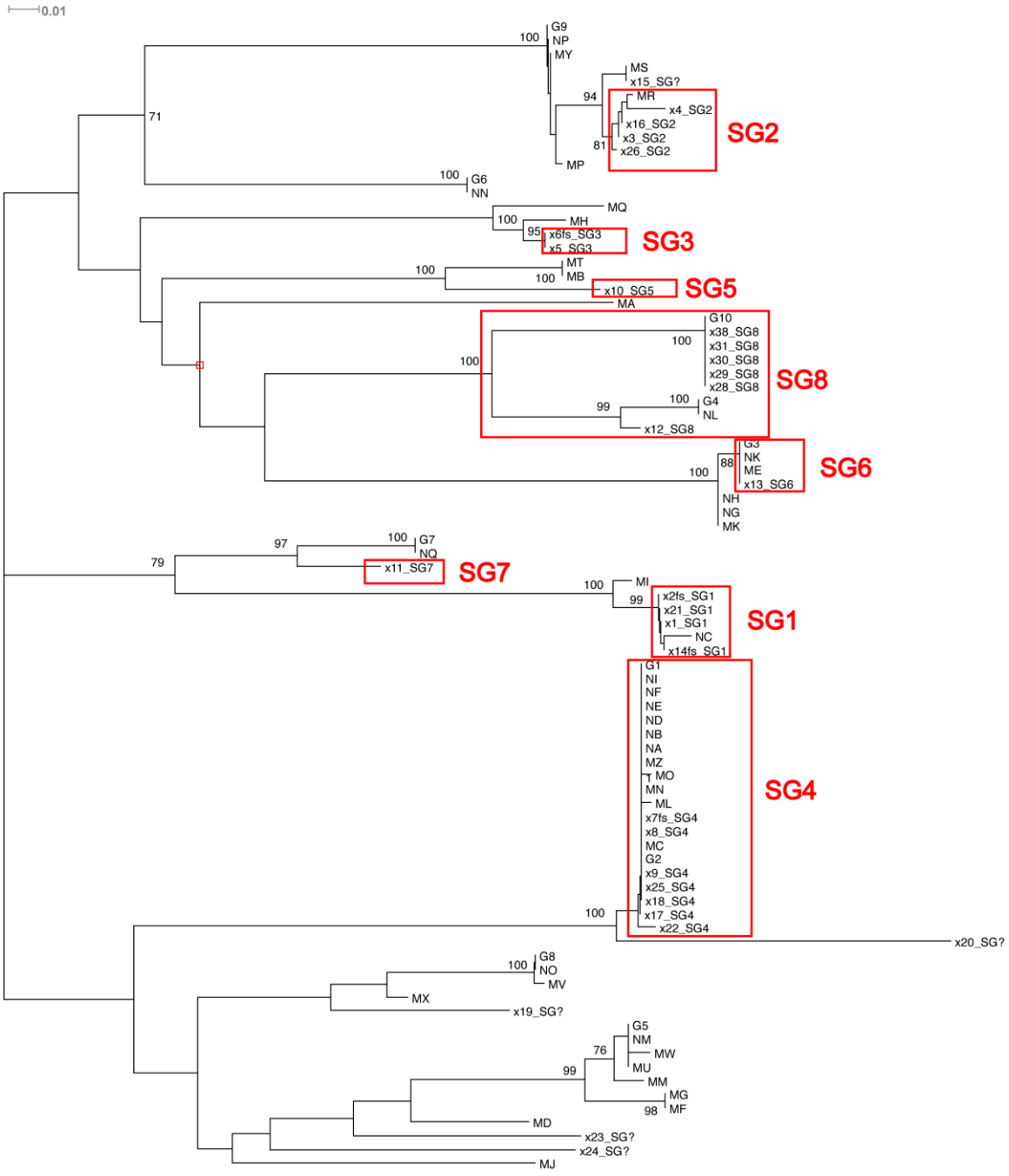

Figure 4: Neighbor joining analyses of complete CD2v protein sequences from reference VNIIVViM isolates and from previously sequenced, unserotyped isolates. Phylogram of BioNJ analysis on 284 sites aligned with Mafft and using Kimura correction. Bootstrap branch supports greater than 70% are indicated on branch leading to appropriate nodes.

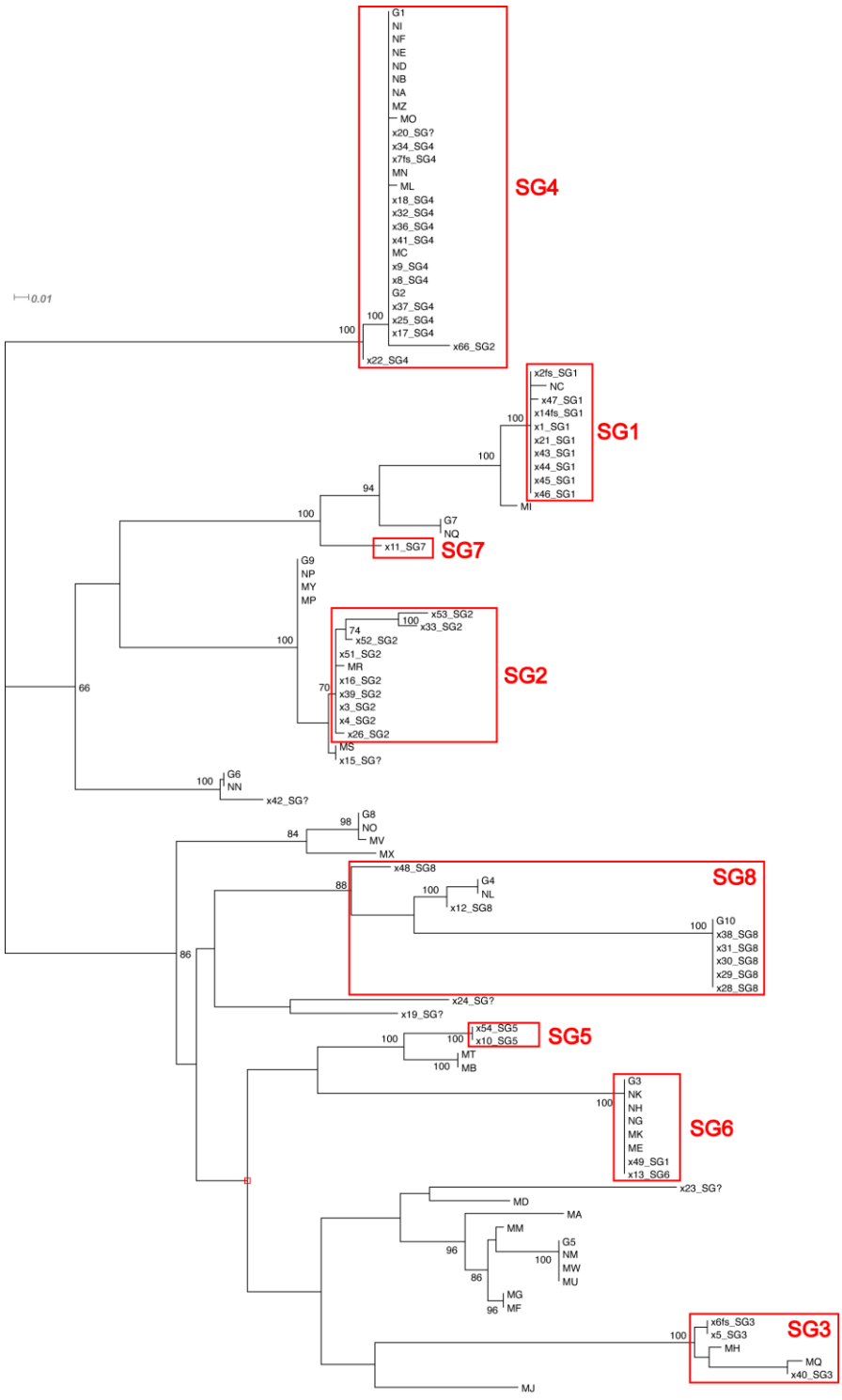

Figure 5: Maximum likelihood analyses of CD2v signature region protein sequences. Phylogram of PhyML analysis on 198 sites aligned with Mafft and refined with Muscle, using LG model. Bootstrap support greater than 70% are indicated at appropriate nodes. Similar SG8 grouping results were observed using parsimony analysis with gap positions as novel character states.