ASFVdb: an integrative resource of African swine fever for genomics and proteomics analysis

Zhenglin Zhu[1*], Geng Meng[2*]

1. The School of Life Sciences, Chongqing University, Chongqing, China

2. Laboratory of Biomedical research and College of Veterinary Medicine, China Agricultural University, Beijing, China

[*]Corresponding authors

Zhenglin Zhu, The School of Life Sciences, Chongqing University, Chongqing, 400044 China

TEL: (86)23-6512-2686, FAX: (86)23-6512-2689, zhuzl@cqu.edu.cn

Geng Meng, Laboratory of Biomedical research and College of Veterinary Medicine, China Agricultural University, Beijing, 100094 China

TEL: (86)10-6273-3466, FAX: (86)10-6273-3466, mg@cau.edu.cn

Keywords: African swine fever virus, population genomics, proteomics, database

Abstract

The recent outbreaks of African swine fever (ASF) in China and Europe threatened the swine industry globally. With the purpose to control the transmission of the African swine fever virus (ASFV), we developed ASFVdb, the African swine fever virus database, an online data visualization and analysis platform in comparative genomics and proteomics. On the basis of known ASFV genes, ASFVdb re-annotated the genomes of every strain and firstly annotated 4707 possible ORFs. Moreover, ASFVdb did thorough population genetics analysis of the all published ASFV strains' genomes and did function and structural prediction for all genes. For each ASFV gene, visitors can not only get its basic information but also the distribution of the gene in strains, conserved or high mutation region, possible subcellular location and topology. In the genome browser, ASFVdb gives sliding window population genetics analysis results, which facilitate genetics and evolutional analysis in genomic level. The web interface is constructed basing on SWAV 1.0. ASFVdb is freely accessibly at http://asfvdb.popgenetics.net.

KEYWORDS: African swine fever virus; population genomics; database

INTRODUCTION

African swine fever (ASF), caused by the ASF virus (ASFV), is a viral disease in domestic or wild swine (1-3). The mortality rate of ASF is near 100% for infected domestic swine. The outbreaks in China and central Europe in the summer of 2018 caused important economic losses for global swine industry. However, efforts to produce effective vaccines or treatments to ASFV are obstructed by the complexity of the virus. ASF virus, which have an icosahedral capsid with multilayered membrane structure, contains a lot of polypeptides whose function and identities are largely unknown. It was reported that some generated antibodies are not protective but may enhance the disease (4-8), due to virion complexity and intracellular and extracellular localization of infectious particles. Consequently, the identification of functional ASFV polypeptides is important to develop effective vaccine.

On the basis as described above, we developed ASFVdb, the African swine fever database, to assist ASFV researches. Different from previous ASFV relevant database, like Viralzone (9). ASFVdb is specialized for ASFV. It made a comparison of all the published ASFV genomes (10-17), supplemetarily annotated each genome and validated the persistence of each gene in different strains. ASFVdb is developed to facilitate the identification of gene function and identity. Aside from basic annotation, it also gives information on gene subcellular location, gene function, topology prediction, comparative genomic alignments and population genetics analysis results.

MATERIALS AND METHODS

Data collection and processing

We downloaded the genome sequences, CDS, proteins and relevant annotation of 36 sequenced ASFV strains (Table S1) from NCBI genome database. Among the 3 different version (NC_001659.2, U18466.2 and KP055815.1) of Spain_BA71_1971 genome data, we used the original one, KP055815.1. Of the 36 different ASFV genome data, only 21 are fully annotated and their gene ID or names are not unified. To harmonize this inconsistence, we made a unified dataset and used the dataset to re-annotate the 36 genomes. Specifically, we did pairwise alignments for all annotated protein sequences. The sequences were clustered together with the identity > 50% and coverage > 80% according to previous homologues gene identification methods (18,19). Finally, a list containing 308 unified genes was generated. Then we mapped these unified genes to all the 36 genomes. If a unique protein had been mapped with the highest score of identity > 50% and coverage > 80%, we considered the unique protein exists in the strain and the mapped region in the strain genome is marked. To identify the corresponding mapped region of one unique gene that express proteins, we translated the mapped region by NCBI ORFfinder and aligned all possible translations of the unique protein. We took the alignment with the highest score and checked whether its identity and coverage are both higher than 0.5. If yes, we consider the region can express proteins, if no we marked the region as ”Genetic Remains”. The genes those had been annotated in NCBI genome database marked as “Annotation from NCBI”, and the unannotated gene region with high expression possibility marked as “Newly Annotated”. At last, all genes sequenced had been clustered in 253 gene clusters with CD-HIT(20) requiring identify > 50% and coverage > 70%.

To retrieve the annotation of ASFV genes product from Uniprot(21,22), we Blast the protein sequence of ASFV gene against the Uniprot protein database (22) and took hit with the highest score as the best match. If the best match of one ASFV gene is of

E-value <0.05, we extracted the accession of the mapped sequence as the Uniprot ID of the matched ASFV gene. With the Uniprot ID, we got external annotations of the ASFV gene, like corresponding Pubmed ID, EMBL ID, Proteomes ID, Pfam, Interpro, Gene Ontologies, KEGG and function annotations. We also Blast all ASFV protein sequences against the PDB database (23,24) in the same way, and recorded the alignment results. Disappointedly, we got only 14 unique proteins of PDB match with E-value <0.05, which means most of the ASFV expressed proteins have no structural information.

Subcellular localization and topology prediction

The subcellular location and topology annotation from Uniprot only cover several ASFV genes. Thus, we re-do these prediction jobs for all genes. We predicted the subcellular location of ASFV proteins through MSLVP (25) using the parameters of "one-versus-one", "Second-tier Algorithm" and requiring similarity >90%. 308 unique proteins are grouped into 9 subcellular locations. We predicted transmembrane helix within proteins' sequences using TMHMM 2.0 (26). The output images were converted into PNG format to display in website by Magick (www.imagemagick.org).

Comparative genomics and population genetics analysis

To trace the evolutionary history of ASFV, we made whole genome alignment of all 36 ASFV genomes by Mugsy (27), and built the phylogenetic tree by FastTree 2.1 (28), with the parameter '-boot 5000' to test the likelihood of the generated tree. To get a more detailed comparative genomic map of the 36 strains, we used LASTZ (29) to do genome-genome alignments between any two ASFV strains and output the results in AXT format. For one ASFV gene, we retrieved the corresponding sequence of other strains from the genome-genome alignment results, realigned these sequences by MUSCLE (30,31), and got the genomic alignment in CDS region. According to the persistence of each gene in different strains, we also retrieved all protein counterparts of one unique gene and made the multiple alignment by MUSCLE. To

more clearly display the evolution history of each protein, we also drew trees from the multiple alignment like we did for genome.

We slide along the ASFV genome with a window size of 200 bp and a step size of 50 bp. From the genome-genome alignments referred above, we retrieved sequences within the sliding window, with which we calculated Pi, Theta(32) and tajima's D(33) using variscan 2.0 (34,35). We wrote Perl scripts to call Allele Frequencies and used SweapFineder2 (36) to calculate the composite likelihood ratio (CLR)(37) with step size 50. We took the median of population genetics test statistics in the region of each gene as its corresponding value.

Web Interfaces

The web interface of ASFVdb was developed in the structure of Mysql + PHP + CodeIgniter (www.codeigniter.com) + JQuery (jquery.com), and is basing on SWAV (swav.popgenetics.org), which is an open source web application for population genetics visualization and analysis. We used (38) to show the multiple alignments of orthologous CDS or proteins, and phylotree (39) to display of the phylogenetic trees of genomes or proteins. We made changes of the source codes of the referred web application to fit the development of ASFVdb, like add links within the diagram. The search interfaces of sequence alignments are constructed basing on PHP parsing of the results from BLAT (40) and NCBI Blast (41). The workflow of other search blocks, like search by molecular structure or gene names, are mostly SQL query pipelines.

RESULTS

Overview of the genome data in ASFVdb

ASFVdb collected a broad range of published ASFV genome data, analyzed and identified 253 ASFV gene clusters (Table S2) from the genomes of 36 strains (for detail, see Material and Methods). Among the 253 gene clusters, 111 are conserved in ORF in all strains (Table S2). Reannotation of the 36 strains' genomes predicted

4707 genes and 888 genetic remains. In each strain, there are 181 ~ 263 genes in average. These did not contradict previous reports of 151~181 genes per strain (17,42), considering out pipeline included all possible alternative splicing cases (Figure S1) and the discrepancy between annotations for different strains with different strategies at different time (10-17). We did subcellular localization of the ASFV genes to display their roles in the infection process (Figure 2A). We found more than half (60%) are membrane proteins (Figure 2B). 8% (23 cases) are capsid protein. They may provide potential targets for vaccine design. In gene ontology, the highest rate of ASFV genes enriched in membrane and have catalytic activity (Figure 3). Moreover, from population genetics test analysis, we found 302 genes are of significantly low Tajima's D (33) and significantly high composite likelihood ratio (CLR)(37) (Rank Test, P-value<0.05), indicating these genes are possibly under positive selection recently (Table S3). 82 genes are of significantly high tajima's D value (Table S4, Rank Test, P-value<0.05), and they may be in balancing selection. These are informative for future ASFV researches.

ASFVdb offers diverse search way

The main search input of ASFVdb, at the center of the home page, allows users to search a gene by a general gene name, a gene accession or a description of gene's function. We created a gene alias list to make the search result as complete as possible. In the right column of the home page, there are the list of other gene search methods in ASFVdb. First of all, users could BLAT against one ASFV genome like the workflow in UCSC genome browser (43), or BLAST against the CDS sequences or protein sequences. Secondly, users could search for genes in a specific subcellular location. In the result view, we also listed the topology information and function annotation of each gene to convenient users to localize one gene in the host-virus interaction. Thirdly, users can find genes of some specific function according to the gene's GO annotation. "Expression Profiling" lists the persistence of ASFV genes in the 36 strains. Users can view one specific gene by clicking an item in the gene list table. Moreover, to track a list of genes more conveniently, ASFVdb offers to get gene

links by inputting a list of genes' genomic positions or a list of genes' accession names. These operation facilitate users to do analysis in ASFVdb with a personal gene list.

Evolutionary episodes of the ASFV strains in ASFVdb

In the home page, below the main search box is a phylogenetic tree displaying the evolution history of available published 36 ASFV strains. Different from the previous method to make ASFV phylogenetic trees (14,17), we built the tree through doing multiple alignment in genomic scale to better understand the processes and patterns of evolution (44). The tree we made (Figure 4) is in consistent with previous publications (17,42,45). It reflects transmission events in the spread of ASFV. After it was first identified in Kenya in the 1920s (46), this disease was subsequently reported in eastern and southern Africa (47). In the mid of the last century, ASFV spread out from Africa and deposited in Portugal (48) as its first stop in Europe. Then, this disease spread to Malta (1978), Italy (1967, 1980), France (1964, 1967, 1977), Belgium (1985) and Netherlands in 1986 (49). In 2007, ASFV was introduced to Caucasus in 2007(50) and started to spread across Russia and Eastern Europe. In 2018, ASFV came to China and the outbreaks in China and central Europe enhance the seriousness of this disease (51).

Gene Visualization in ASFVdb

Through clicking the taxon name in the tree or selecting an item in the "Gbrwoser" list, users could go to the genome browser page (Figure 5), where gene segments are arranged subsequently along the genome according to genes' genomic positions. Genes annotated from the NCBI GFF file are colored in blood red as "Annotation from NCBI". Genes annotated by mapping (for details, see Material and Methods) are colored in dark red as "Newly annotated". "Genetic Remains" are colored in light gray. Gene Name is above each gene segment and ">" or "<" indicated a gene's transcriptional direction. Three tracks of population genetics analysis, Pi, Theta and Tajima's D, are listed following the gene track. To help to trace selective signatures, we drew a top 5% line for Pi and theta, and two 5% lines in ascending and

descending order for Tajima's D. With the tool bar at the top, users can move or zoom the genome in the browser, set the focus bar of an appointed region, export the drawing of one track with gene segments and export data of one track.

Clicking one gene segment will direct to the gene's information page. A full list of annotations include basic information (strain, gene name, description, location in the genome, Genbank Accession and full name), sequence (CDS and protein), summary (function, Uniprot accession, related Pubmed ID, related EMBL ID, corresponding Proteomes ID, related Pfam ID and correlated Interpro ID), Ontologies (GO and KEGG), subcellular location, topology (transmembrane region prediction), genomic alignment in the CDS region, multiple alignment of orthologues, gene tree of corresponding NCBI annotated or newly annotated proteins, and orthologous in strains. Texts which can link to internal or external sites are hyperlinked to facilitate view and analysis.

Analysis of CD2v(EP402R) in ASFVdb: a case study

CD2v(EP402R) is ASFV gene related to hemadsorption (52). It plays a key role in viral attenuation and may be involved in the generation of the immune protection from the host against ASFV infection (53). Through search "EP402R", we found this gene encoded a single-Pass membrane protein with the transmembrane region from 207AA to 229AA (Figure S2A). In the genome view with population genetics analysis tracks, we found the transmembrane region is of a tajima's D valleys and a CLR peak (Figure S2B), inferring selection signatures. The region from 1AA to 206AA is predicted to be located outside of the viron membrane with peaks of Pi and Theta, inferring its high diversity among strains. From the information of "Genomic alignment in the CDS region" and "Orthologues in Strains", we found this gene only have high conservation among 25 of all compared 36 strains (Figure S2C). These may partly explained the result that pigs vaccinated with baculovirus-expressed EP402R/CD2v did not have fully protection from ASFV infection (54,55). AFSVdb annotated that this protein is of 30% similarity (E-value=0.00079) to the cell adhesion molecule CD2 (56), whose crystal structure had been used to build homology

modelling structure of EP402R (Figure S2D).

CONCLUSION

To assist the veterinary scientist and researchers combating the threat the ASFV outbreak, we developed a database based toolbox ASFVdb. It integrated the data from NCBI, Uniprot(21), VironZone(9) and a broad range of published articles (9). ASFVdb is specialized for ASFV. It made a comparison of all the published ASFV genomes(10-17), summarized data on the proteins of all known ASFV strains, and provided information not only in gene basic information but also in external links, subcellular localization, topology, comparative genomic alignments, gene tree and evolutionary analysis results. These are helpful and convenient in the further accession of proteomics and biochemical data in ASFV research. Hopeful the ASFV database may be useful for further research on ASFV and vaccine design.

DATA AVAILABILITY

All ASFVdb data are publicly and freely accessible at http://asfvdb.popgenetics.net. Feedback on any aspect of the ASFVdb database, and discussion of ASFV gene annotations, are welcome by email to zhuzl@cqu.edu.cn or mg@cau.edu.cn.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (31200941) and the Fundamental Research Funds for the Central Universities (106112016CDJXY290002).

Figure Legends

Figure 1. The number of annotated genes in all ASFV strains. Genes annotated in NCBI, genes newly annotated and genetic remains are marked in dark red, red and gray, separately.

Figure 2. A is a diagram display the infection process of ASFV with MSLVP subcellular locations marked. B displays the percentages of ASFV gene clusters in each subcellular location groups.

Figure 3. Enrichment of gene ontology (GO) of ASFV gene clusters. This figure is drew by WEGO (57).

Figure 4. A, The phylogenetic tree of ASFV strains in ASFVdb with historical events marked. Numbers marked in red are the marginal likelihood of the tree. B. The distribution of strains in ASFVdb in countries. The red arrows with dotted lines pointed out the possible infection paths of ASFV.

Figure 5. The gnome browser view to display ASFV genes and population genetics test statistics.

References

1. Parker, J., Plowright, W. and Pierce, M.A. (1969) The epizootiology of African swine fever in Africa. *Vet Rec*, **85**, 668-674.
2. Thomson, G.R., Gainaru, M.D. and Van Dellen, A.F. (1980) Experimental infection of warthos (Phacochoerus aethiopicus) with African swine fever virus. *Onderstepoort J Vet Res*, **47**, 19-22.
3. Anderson, E.C., Hutchings, G.H., Mukarati, N. and Wilkinson, P.J. (1998) African swine fever virus infection of the bushpig (Potamochoerus porcus) and its significance in the epidemiology of the disease. *Vet Microbiol*, **62**, 1-15.
4. Blome, S., Gabriel, C. and Beer, M. (2014) Modern adjuvants do not enhance the efficacy of an inactivated African swine fever virus vaccine preparation. *Vaccine*, **32**, 3879-3882.
5. Detray, D.E. (1963) African Swine Fever. *Adv Vet Sci*, **8**, 299-333.
6. Mebus, C.A. (1988) African swine fever. *Adv Virus Res*, **35**, 251-269.
7. Stone, S.S. and Hess, W.R. (1967) Antibody response to inactivated preparations of African swine fever virus in pigs. *Am J Vet Res*, **28**, 475-481.
8. Sunwoo, S.Y., Perez-Nunez, D., Morozov, I., Sanchez, E.G., Gaudreault, N.N., Trujillo, J.D., Mur, L., Nogal, M., Madden, D., Urbaniak, K. *et al*. (2019) DNA-Protein Vaccination Strategy Does Not Protect from Challenge with African Swine Fever Virus Armenia 2007 Strain. *Vaccines (Basel)*, **7**.
9. Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I. and Le Mercier, P. (2011) ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res*, **39**, D576-582.
10. Zani, L., Forth, J.H., Forth, L., Nurmoja, I., Leidenberger, S., Henke, J., Carlson, J., Breidenstein, C., Viltrop, A., Hoper, D. *et al*. (2018) Deletion at the 5'-end of Estonian ASFV strains associated with an attenuated phenotype. *Sci Rep*, **8**, 6510.
11. Portugal, R., Coelho, J., Hoper, D., Little, N.S., Smithson, C., Upton, C., Martins, C., Leitao, A. and Keil, G.M. (2015) Related strains of African swine fever virus with different virulence: genome comparison and analysis. *J Gen Virol*, **96**, 408-419.
12. Bishop, R.P., Fleischauer, C., de Villiers, E.P., Okoth, E.A., Arias, M., Gallardo, C. and Upton, C. (2015) Comparative analysis of the complete genome sequences of Kenyan African swine fever virus isolates within p72 genotypes IX and X. *Virus Genes*, **50**, 303-309.
13. Chapman, D.A., Tcherepanov, V., Upton, C. and Dixon, L.K. (2008) Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. *J Gen Virol*, **89**, 397-408.
14. de Villiers, E.P., Gallardo, C., Arias, M., da Silva, M., Upton, C., Martin, R. and Bishop, R.P. (2010) Phylogenomic analysis of 11 complete African swine fever virus genome sequences. *Virology*, **400**, 128-136.
15. Chapman, D.A., Darby, A.C., Da Silva, M., Upton, C., Radford, A.D. and Dixon, L.K. (2011) Genomic analysis of highly virulent Georgia 2007/1 isolate of African swine fever virus. *Emerg Infect Dis*, **17**, 599-605.
16. Olesen, A.S., Lohse, L., Dalgaard, M.D., Wozniakowski, G., Belsham, G.J., Botner, A. and Rasmussen, T.B. (2018) Complete genome sequence of an African swine fever virus (ASFV POL/2015/Podlaskie) determined directly from pig erythrocyte-associated nucleic acid. *J Virol*

Methods, **261**, 14-16.

17. Bao, J., Wang, Q., Lin, P., Liu, C., Li, L., Wu, X., Chi, T., Xu, T., Ge, S., Liu, Y. *et al*. (2019) Genome comparison of African swine fever virus China/2018/AnhuiXCGQ strain and related European p72 Genotype II strains. *Transbound Emerg Dis*.
18. Yue, H., Shu, D., Wang, M., Xing, G., Zhan, H., Du, X., Song, W. and Nie, X. (2018) Genome-Wide Identification and Expression Analysis of the HD-Zip Gene Family in Wheat (Triticum aestivum L.). *Genes (Basel)*, **9**.
19. She, R., Chu, J.S., Wang, K., Pei, J. and Chen, N. (2009) GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res*, **19**, 143-149.
20. Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.
21. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. *et al*. (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*, **32**, D115-119.
22. Patient, S., Wieser, D., Kleen, M., Kretschmann, E., Jesus Martin, M. and Apweiler, R. (2008) UniProtJAPI: a remote API for accessing UniProt data. *Bioinformatics*, **24**, 1321-1322.
23. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**, 235-242.
24. Burley, S.K., Berman, H.M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J.M., Dutta, S. *et al*. (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res*, **47**, D464-D474.
25. Thakur, A., Rajput, A. and Kumar, M. (2016) MSLVP: prediction of multiple subcellular localization of viral proteins using a support vector machine. *Mol Biosyst*, **12**, 2572-2586.
26. Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, **305**, 567-580.
27. Angiuoli, S.V. and Salzberg, S.L. (2011) Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics*, **27**, 334-342.
28. Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.
29. Harris, R.S. (2007) Improved pairwise alignment of genomic DNA. *Ph.D. thesis, Pennsylvania State University*.
30. Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
31. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, **32**, 1792-1797.
32. Watterson, G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*, **7**, 256-276.
33. Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.
34. Hutter, S., Vilella, A.J. and Rozas, J. (2006) Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics*, **7**, 409.
35. Vilella, A.J., Blanco-Garcia, A., Hutter, S. and Rozas, J. (2005) VariScan: Analysis of

evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics*, **21**, 2791-2793.

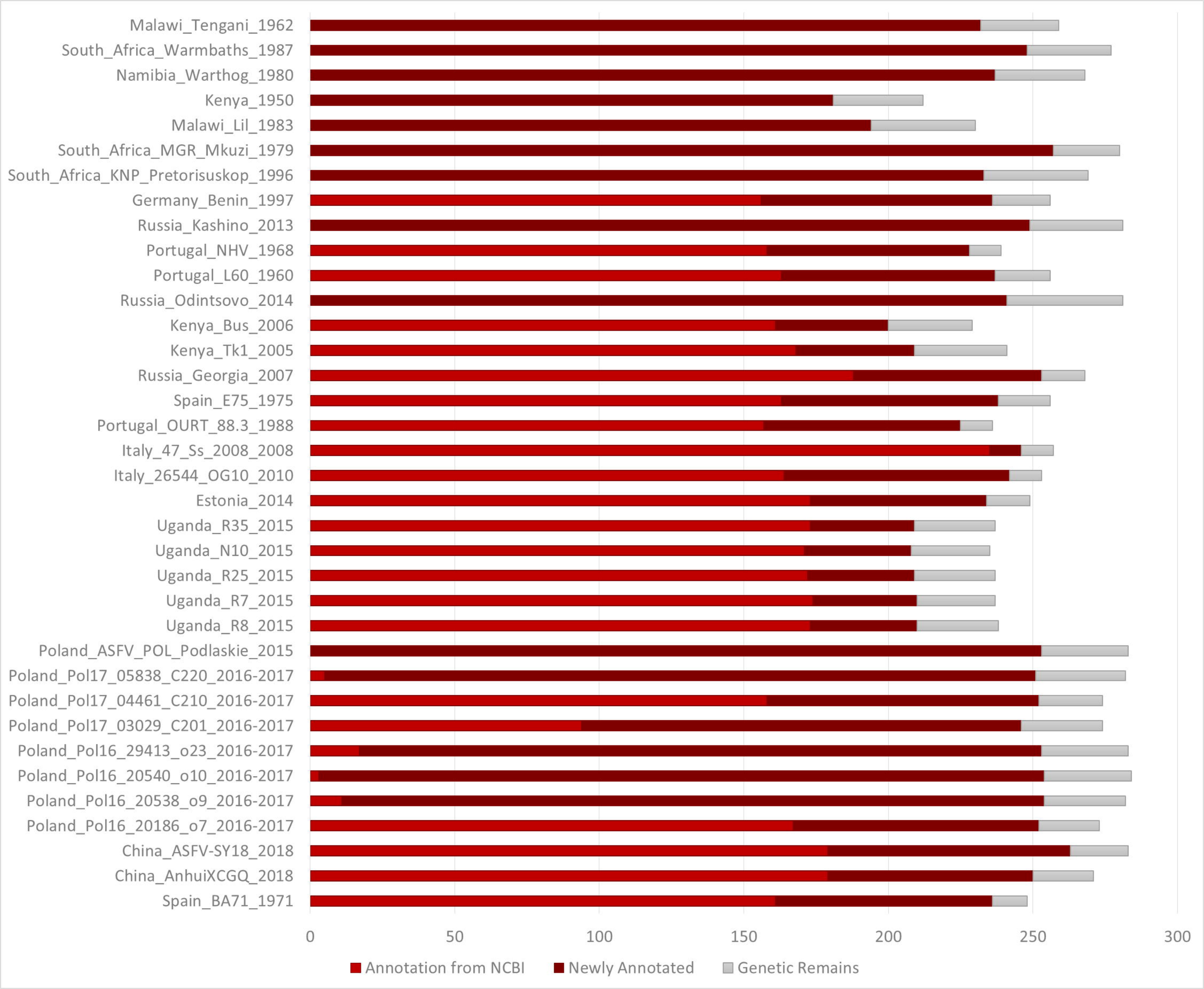
36. DeGiorgio, M., Huber, C.D., Hubisz, M.J., Hellmann, I. and Nielsen, R. (2016) SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*, **32**, 1895-1897.
37. Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G. and Bustamante, C. (2005) Genomic scans for selective sweeps using SNP data. *Genome Res*, **15**, 1566-1575.
38. Yachdav, G., Wilzbach, S., Rauscher, B., Sheridan, R., Sillitoe, I., Procter, J., Lewis, S.E., Rost, B. and Goldberg, T. (2016) MSAViewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics*, **32**, 3501-3503.
39. Shank, S.D., Weaver, S. and Kosakovsky Pond, S.L. (2018) phylotree.js - a JavaScript library for application development and interactive data visualization in phylogenetics. *BMC Bioinformatics*, **19**, 276.
40. Kent, W.J. (2002) BLAT--the BLAST-like alignment tool. *Genome Res*, **12**, 656-664.
41. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res*, **36**, W5-9.
42. Galindo, I. and Alonso, C. (2017) African Swine Fever Virus: A Review. *Viruses*, **9**.
43. Karolchik, D., Hinrichs, A.S. and Kent, W.J. (2007) The UCSC Genome Browser. *Curr Protoc Bioinformatics*, **Chapter 1**, Unit 1 4.
44. Wilgenbusch, J.C., Huang, W. and Gallivan, K.A. (2017) Visualizing phylogenetic tree landscapes. *BMC Bioinformatics*, **18**, 85.
45. Sanchez, E.G., Perez-Nunez, D. and Revilla, Y. (2019) Development of vaccines against African swine fever virus. *Virus Res*.
46. Montgomery, R.E. (1921) On a form of swine fever occurring in british east Africa (Kenya Colony). *J. Comp. Pathol. Ther*, **1921**, 159-191.
47. Thomson G. R., B.A.D.S. (2004) *African swine fever. In Infectious diseases of livestock*. Oxford, UK: Oxford University Press.
48. Sanchez-Botija, A. (1963) Reservorios del virus de la peste porcina Africana. Investigacion del virus la P.P.A en los artropodos mediante la prueba de la hemoadsorcion. Bull. *Bull. Off. Int. Epizoot.*, **60**, 895-899.
49. Costard, S., Wieland, B., de Glanville, W., Jori, F., Rowlands, R., Vosloo, W., Roger, F., Pfeiffer, D.U. and Dixon, L.K. (2009) African swine fever: how can global spread be prevented? *Philos Trans R Soc Lond B Biol Sci*, **364**, 2683-2696.
50. Rowlands, R.J., Michaud, V., Heath, L., Hutchings, G., Oura, C., Vosloo, W., Dwarka, R., Onashvili, T., Albina, E. and Dixon, L.K. (2008) African swine fever virus isolate, Georgia, 2007. *Emerg Infect Dis*, **14**, 1870-1874.
51. OIE. (2019) African Swine Fever (ASF). *Word Animal Health Information and Analysis Department. Report Nº11*, **February 15-March 01**.
52. Rodriguez, J.M., Yanez, R.J., Almazan, F., Vinuela, E. and Rodriguez, J.F. (1993) African swine fever virus encodes a CD2 homolog responsible for the adhesion of erythrocytes to infected cells. *J Virol*, **67**, 5312-5320.
53. Monteagudo, P.L., Lacasta, A., Lopez, E., Bosch, L., Collado, J., Pina-Pedrero, S., Correa-Fiz, F., Accensi, F., Navas, M.J., Vidal, E. *et al.* (2017) BA71DeltaCD2: a New Recombinant Live Attenuated African Swine Fever Virus with Cross-Protective Capabilities. *J Virol*, **91**.
54. Ruiz-Gonzalvo, F., Rodriguez, F. and Escribano, J.M. (1996) Functional and immunological

properties of the baculovirus-expressed hemagglutinin of African swine fever virus. *Virology*, **218**, 285-289.

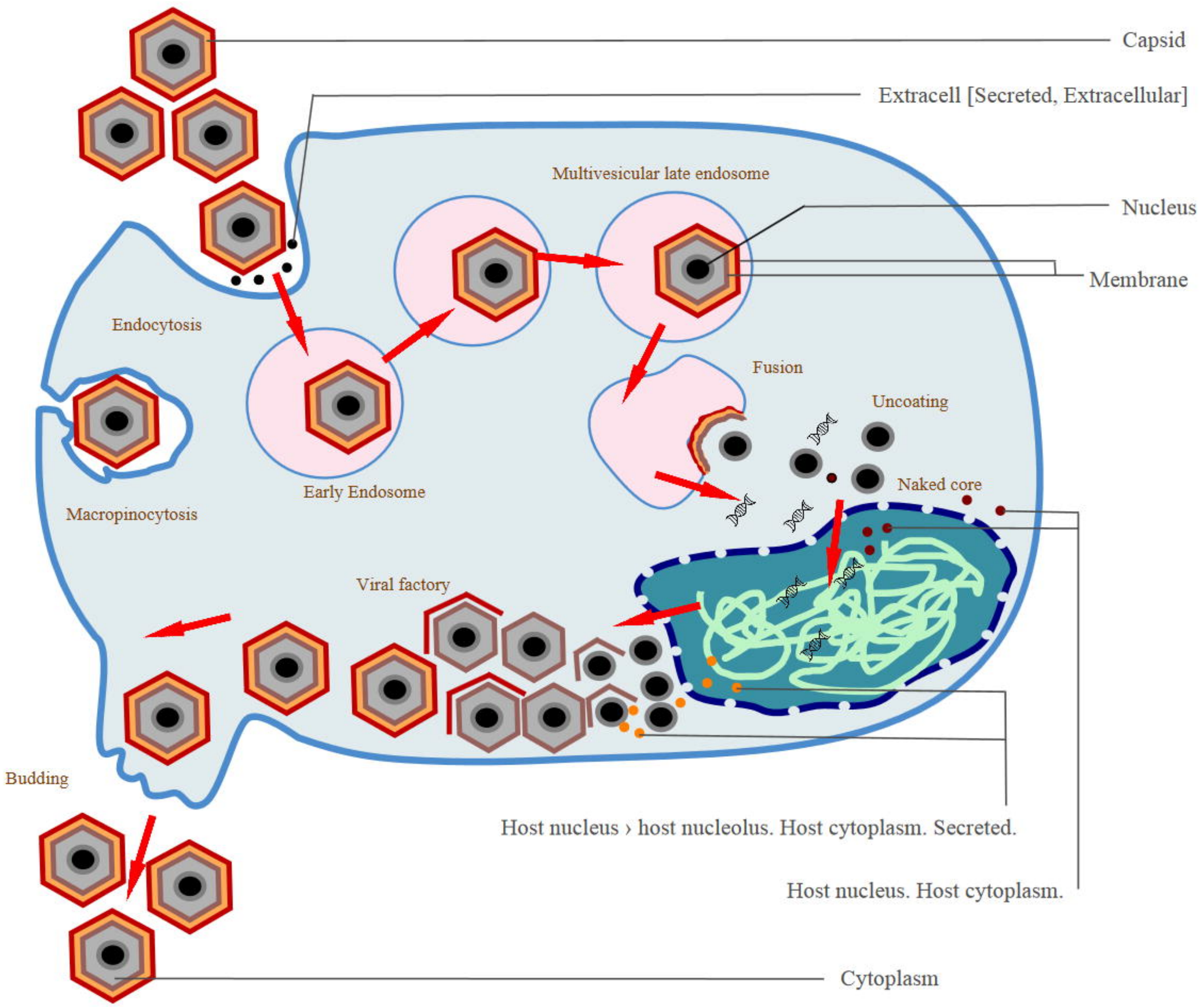
55. Escribano, J.M., Galindo, I. and Alonso, C. (2013) Antibody-mediated neutralization of African swine fever virus: myths and facts. *Virus Res*, **173**, 101-109.

56. Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K. and Stuart, D.I. (1992) Crystal structure at 2.8 A resolution of a soluble form of the cell adhesion molecule CD2. *Nature*, **360**, 232-239.

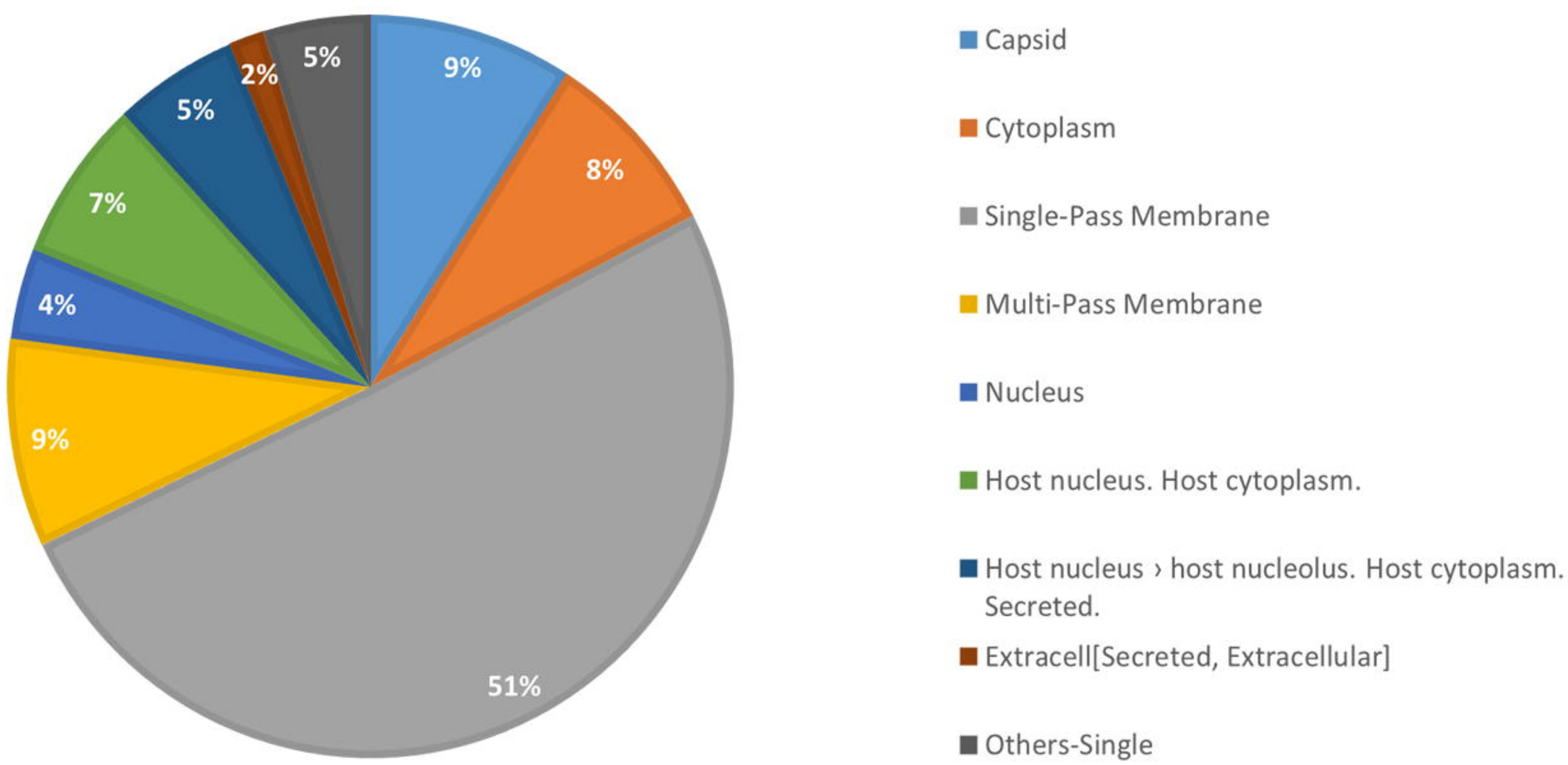
57. Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R. and Bolund, L. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*, **34**, W293-297.

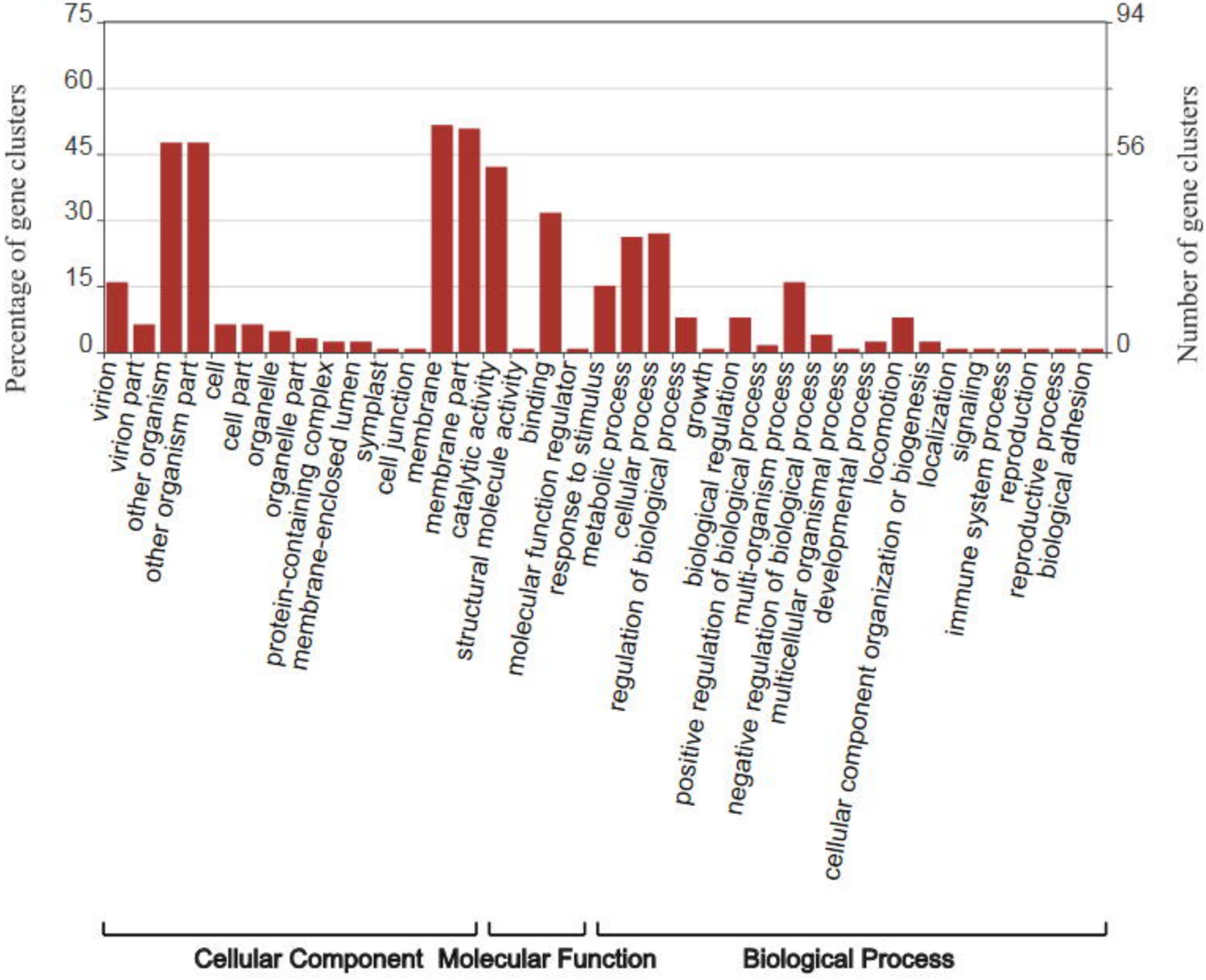

Malawi_Tengani_1962
South_Africa_Warmbaths_1987
Namibia_Warthog_1980
Kenya_1950
Malawi_Lil_1983
South_Africa_MGR_Mkuzi_1979
South_Africa_KNP_Pretorisuskop_1996
Germany_Benin_1997
Russia_Kashino_2013
Portugal_NHV_1968
Portugal_L60_1960
Russia_Odintsovo_2014
Kenya_Bus_2006
Kenya_Tk1_2005
Russia_Georgia_2007
Spain_E75_1975
Portugal_OURT_88.3_1988
Italy_47_Ss_2008_2008
Italy_26544_OG10_2010
Estonia_2014
Uganda_R35_2015
Uganda_N10_2015
Uganda_R25_2015
Uganda_R7_2015
Uganda_R8_2015
Poland_ASFV_POL_Podlaskie_2015
Poland_Pol17_05838_C220_2016-2017
Poland_Pol17_04461_C210_2016-2017
Poland_Pol17_03029_C201_2016-2017
Poland_Pol16_29413_o23_2016-2017
Poland_Pol16_20540_o10_2016-2017
Poland_Pol16_20538_o9_2016-2017
Poland_Pol16_20186_o7_2016-2017
China_ASFV-SY18_2018
China_AnhuiXCGQ_2018
Spain_BA71_1971
0
50
100
150
200
250
300
Annotation from NCBI
Newly Annotated
Genetic Remains

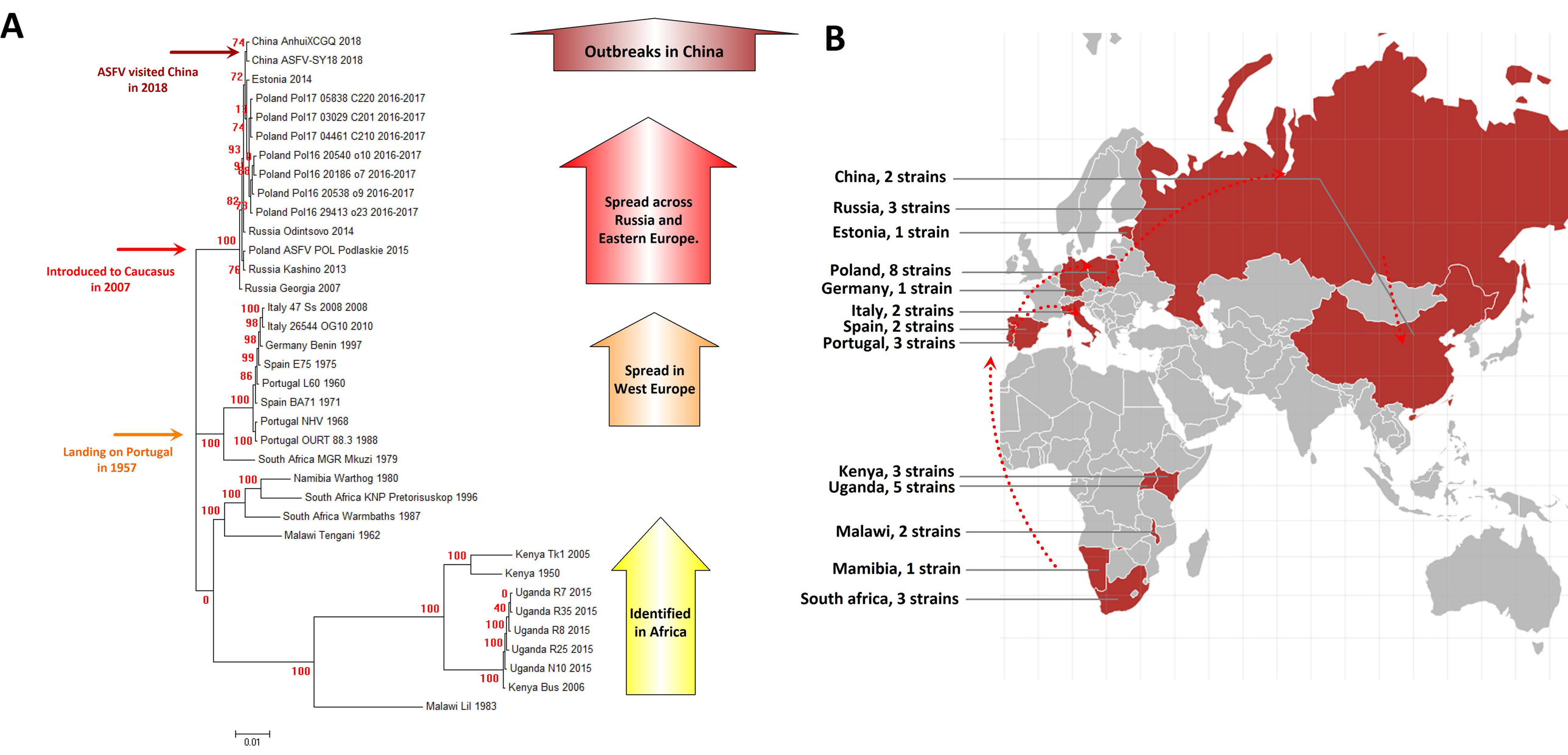
A

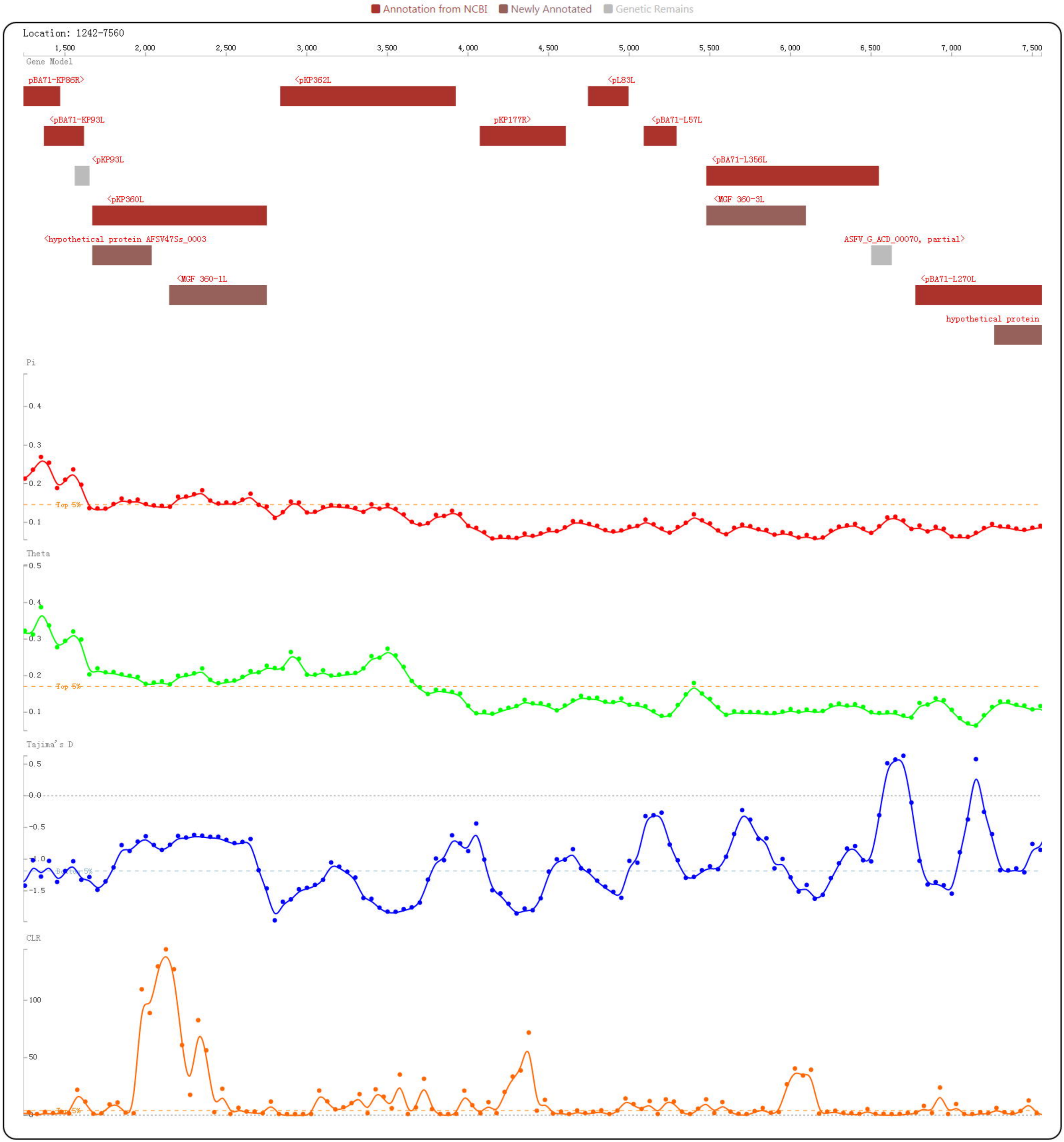

Capsid
Extracell [Secreted, Extracellular]
Multivesicular late endosome
Nucleus
Membrane
Endocytosis
Fusion
Uncoating
Naked core
Early Endosome
Macropinocytosis
Viral factory
Budding
Host nucleus › host nucleolus. Host cytoplasm. Secreted.
Host nucleus. Host cytoplasm.
Cytoplasm

B


9%
8%
51%
9%
4%
7%
5%
2%
5%
Capsid
Cytoplasm
Single-Pass Membrane
Multi-Pass Membrane
Nucleus
Host nucleus. Host cytoplasm.
Host nucleus › host nucleolus. Host cytoplasm. Secreted.
Extracell[Secreted, Extracellular]
Others-Single


Percentage of gene clusters
Number of gene clusters
75
60
45
30
15
0
94
56
0
virion
virion part
other organism
other organism part
cell
cell part
organelle
organelle part
protein-containing complex
membrane-enclosed lumen
symplast
cell junction
membrane
membrane part
catalytic activity
structural molecule activity
binding
molecular function regulator
response to stimulus
metabolic process
cellular process
regulation of biological process
growth
biological regulation
positive regulation of biological process
multi-organism process
negative regulation of biological process
multicellular organismal process
developmental process
locomotion
cellular component organization or biogenesis
localization
signaling
immune system process
reproduction
reproductive process
biological adhesion
Cellular Component
Molecular Function
Biological Process


A
ASFV visited China in 2018
Introduced to Caucasus in 2007
Landing on Portugal in 1957
China AnhuiXCGQ 2018
China ASFV-SY18 2018
Estonia 2014
Poland Pol17 05838 C220 2016-2017
Poland Pol17 03029 C201 2016-2017
Poland Pol17 04461 C210 2016-2017
Poland Pol16 20540 o10 2016-2017
Poland Pol16 20186 o7 2016-2017
Poland Pol16 20538 o9 2016-2017
Poland Pol16 29413 o23 2016-2017
Russia Odintsovo 2014
Poland ASFV POL Podlaskie 2015
Russia Kashino 2013
Russia Georgia 2007
Italy 47 Ss 2008 2008
Italy 26544 OG10 2010
Germany Benin 1997
Spain E75 1975
Portugal L60 1960
Spain BA71 1971
Portugal NHV 1968
Portugal OURT 88.3 1988
South Africa MGR Mkuzi 1979
Namibia Warthog 1980
South Africa KNP Pretorisuskop 1996
South Africa Warmbaths 1987
Malawi Tengani 1962
Kenya Tk1 2005
Kenya 1950
Uganda R7 2015
Uganda R35 2015
Uganda R8 2015
Uganda R25 2015
Uganda N10 2015
Kenya Bus 2006
Malawi Lil 1983
0.01
Outbreaks in China
Spread across Russia and Eastern Europe.
Spread in West Europe
Identified in Africa
B
China, 2 strains
Russia, 3 strains
Estonia, 1 strain
Poland, 8 strains
Germany, 1 strain
Italy, 2 strains
Spain, 2 strains
Portugal, 3 strains
Kenya, 3 strains
Uganda, 5 strains
Malawi, 2 strains
Mamibia, 1 strain
South africa, 3 strains


Annotation from NCBI
Newly Annotated
Genetic Remains
Location: 1242-7560
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
7,000
7,500
Gene Model
pBA71-KP86R>
<pKP362L
<pL83L
<pBA71-KP93L
pKP177R>
<pBA71-L57L
<pKP93L
<pBA71-L356L
<pKP360L
<MGF 360-3L
<hypothetical protein AFSV47Ss_0003
ASFV_G_ACD_00070, partial>
<MGF 360-1L
<pBA71-L270L
hypothetical protein
Pi
0.4
0.3
0.2
Top 5%
0.1
Theta
0.5
0.4
0.3
0.2
Top 5%
0.1
Tajima's D
0.5
0.0
-0.5
-1.0
Bottom 5%
-1.5
CLR
100
50
Top 5%
0