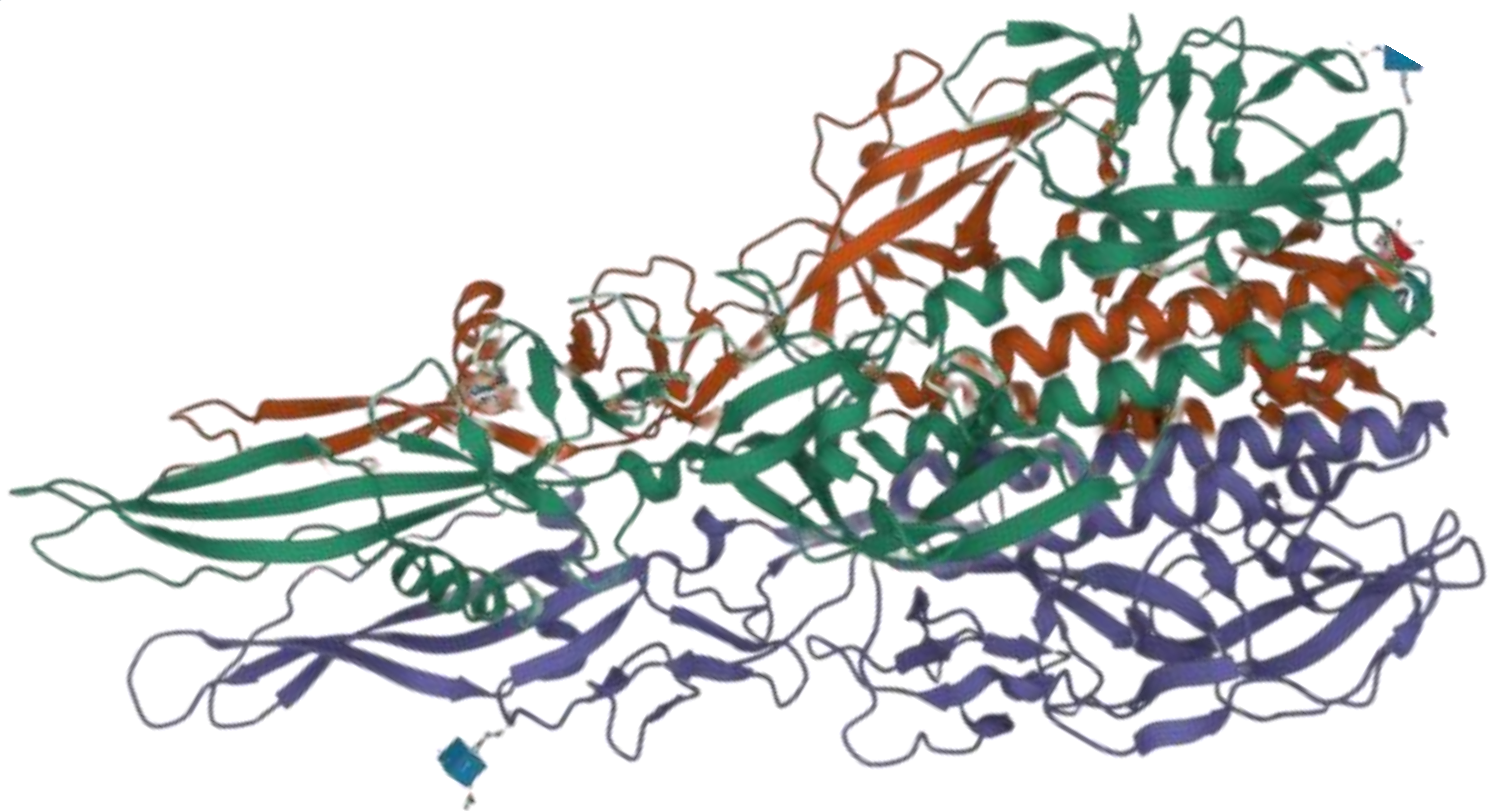


Biophysical Features and Local Early Conformational Propensities in Intrinsically Disordered Regions of Rhabdoviral Glycoproteins

Nikolay Simankov ^{1,2}, H  l  ne Soyeurt ², S  bastien Massart ¹, Rachid Tahzima ^{1,3*}

¹Laboratory of Plant Pathology – TERRA – Gembloux Agro-BioTech – University of Li  ge (ULg) - 5030 Gembloux, Belgium. ²Statistics, Computer Science and Modeling applied to bioengineering (SIMa) – TERRA – University of Li  ge (ULg) 5030 Gembloux, Belgium. ³Flanders Research Institute for Agriculture, Plant Sciences, Fisheries and Food (ILVO), 9820 Merelbeke, Ghent, Belgium. *Contact: rachid.tahzima@uliege.be

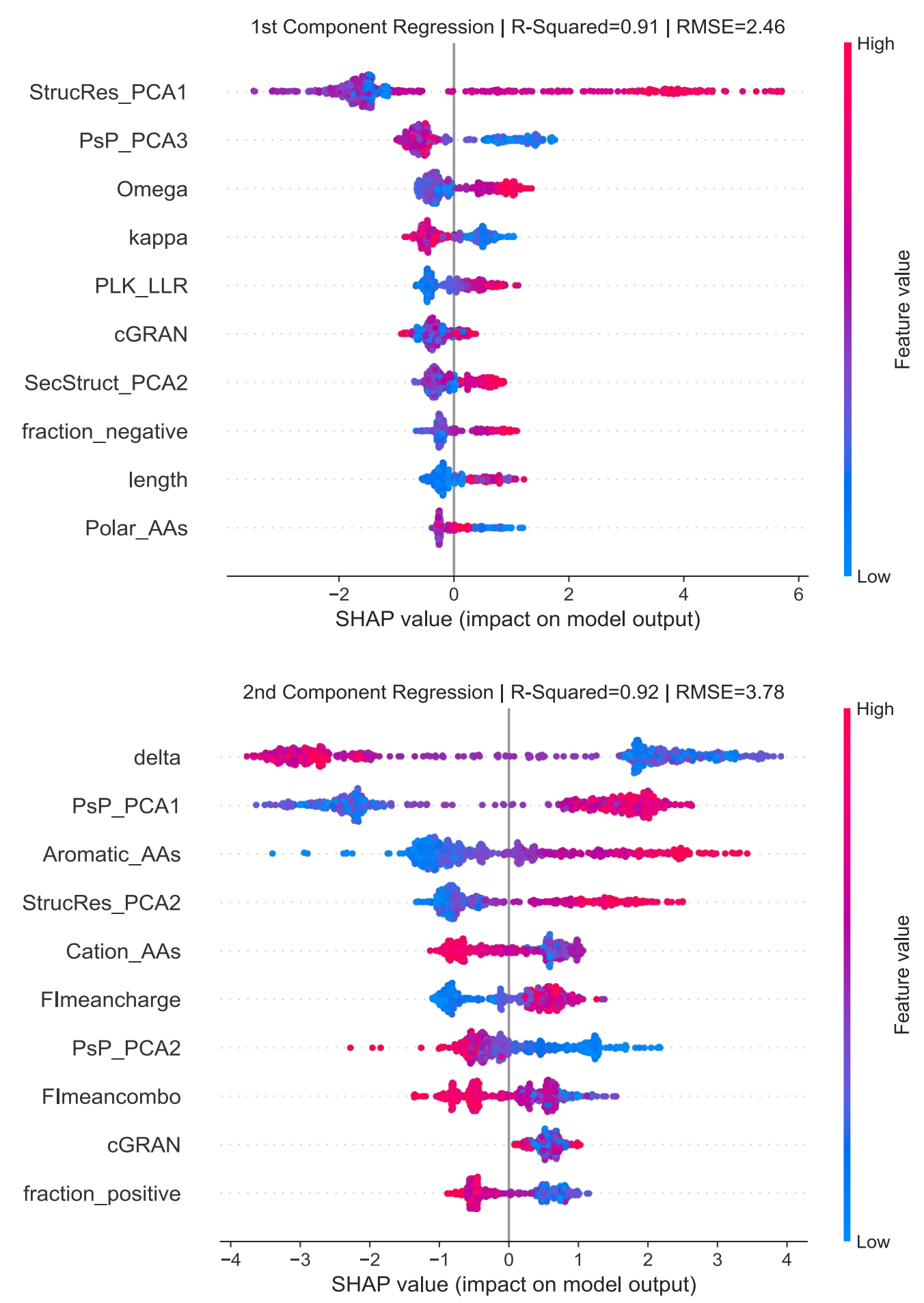
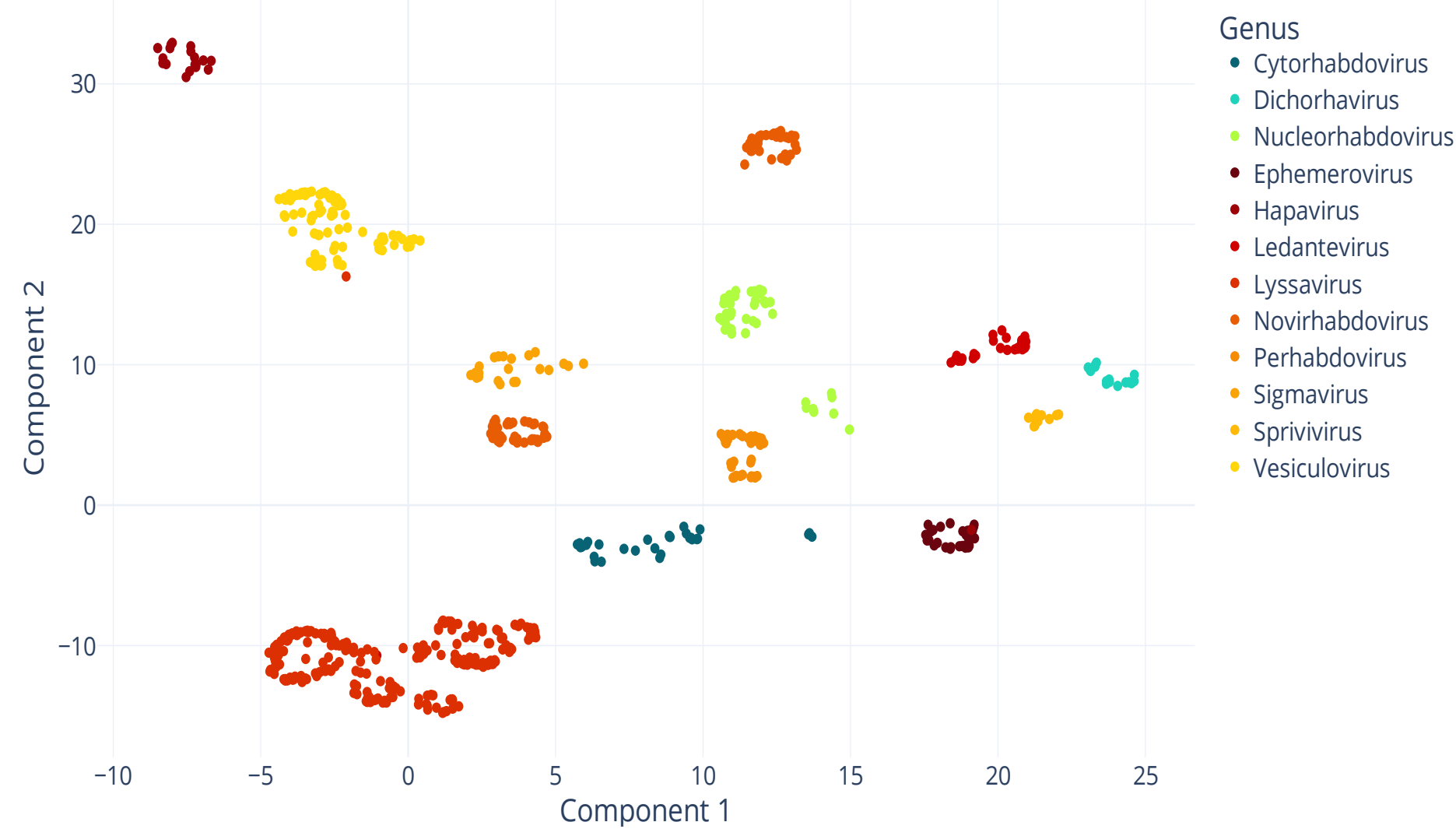


Rhabdoviral Glycoprotein (Gp)

Most harmful rhabdoviruses cause disease that is invariably lethal to humans, animals and plants. Throughout viral infection, protein refolding is a complex process critical to both receptor recognition and membrane-interacting fusion domains mediated by transmembrane rhabdoviral Gp. Little is known about the early stages of context-sensitive structural transitions of the rhabdoviral Gp. To solve this issue, we combined state-of-the-art machine learning-driven biophysical and intrinsic disorder-based metapredictors.

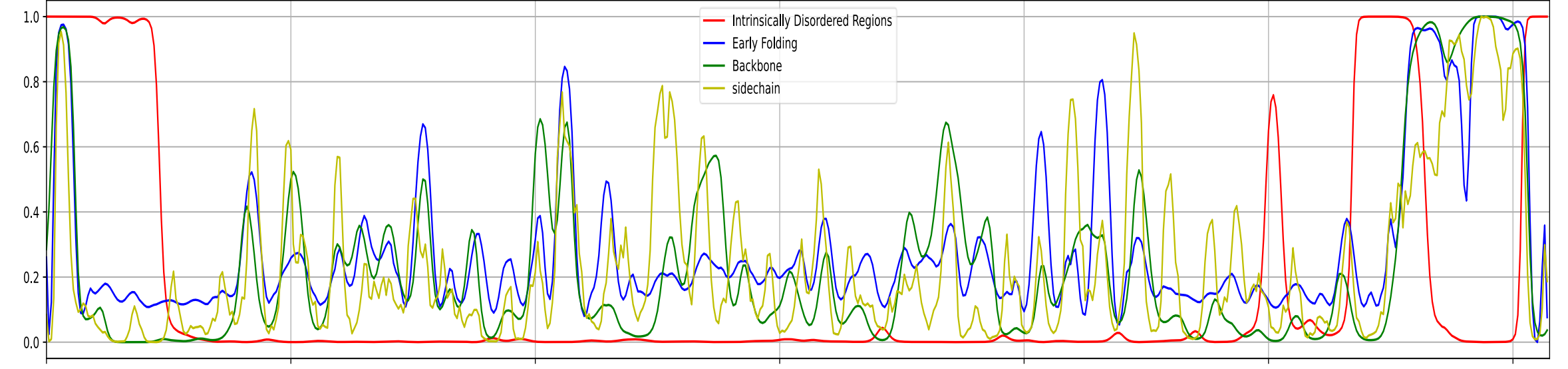
Across the Diverse Rhabdoviral Gps

Through meticulous examination of the Gps from the most well-known genera of Rhabdoviridae, a high level of conservation in biophysical and structural properties is apparent among genera. Furthermore, various gradients of secondary structures, phase separation determinants, amino acid composition and properties display strong intra-genus impact. In contrast, Rabies viruses show weak sequence-based biophysical variations within a wide range of hosts, suggesting importantly this group adopts "biophysical hotspot" allowing efficient multi-host functioning.



Local Conformational Propensities in Intrinsically Disordered Regions (IDRs)

Gp are essential for virion attachment and penetration of host cells prior to transmission. Here, we investigate the host-specific biophysical IDRs-determinants encoded in the primary amino acid sequence of rhabdoviral Gps, which are predicted to modulate early conformational events. This can be related to host-specific biophysical features of the local backbone flexibility and secondary structure propensities mostly close to or in the C-terminal IDRs.

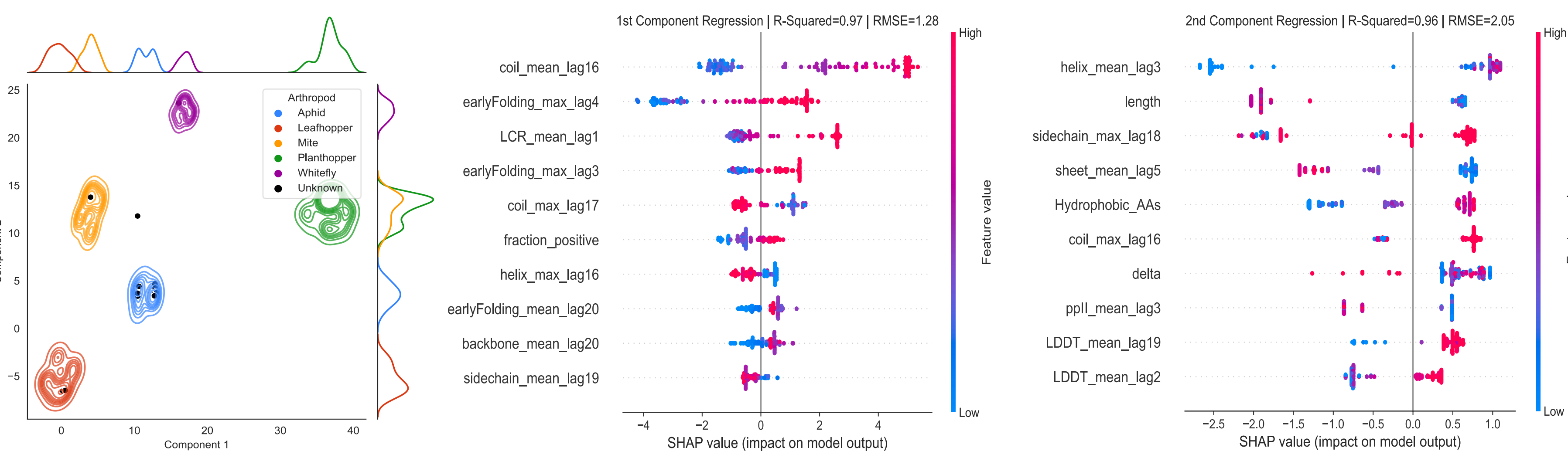
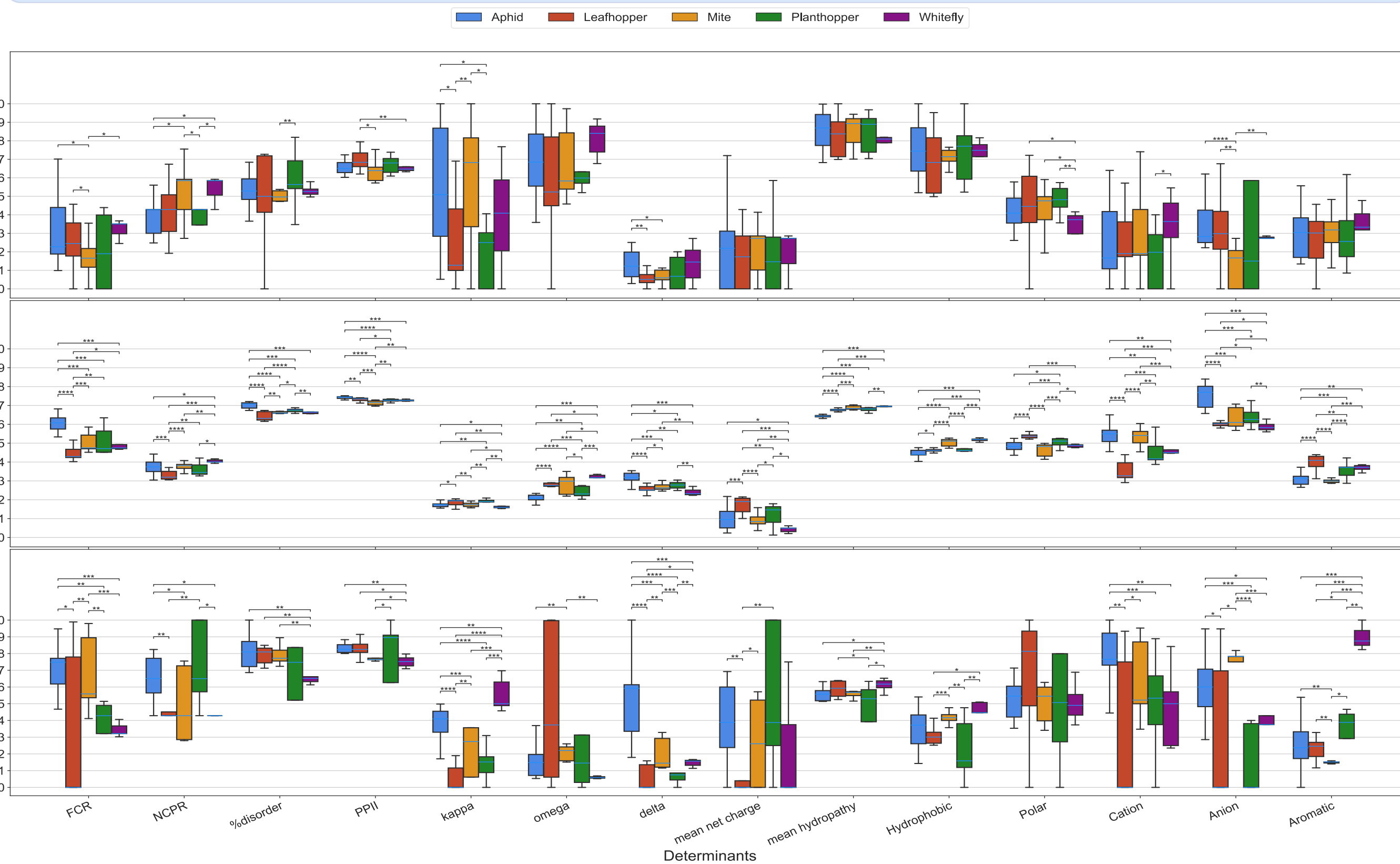


High Conservation and Specificity in Biophysical Determinants

Importantly, the essential plant rhabdoviral Gp is still PFAM-unassigned and poorly characterized. While showing striking differences between IDRs of N- and C-termini and the core, our results give direct insights into the biophysical folding signals located in these variable regions and are in congruence with independent epidemiological observations.

On a quantitative scale, these termini-residues contain biophysical features and sequence segregation compositional biases, such as hydrophobicity, charges and aromaticity, involved mostly in binding interactions and display evolutionary covariation, reflecting a general tendency toward host-specificity.

Hence, an in-depth comparative structural study of the conserved core protein and the C-terminal IDRs patterns would most likely allow the robust identification and assignment of a new PFAM domain candidate. This analysis highlights these regions' functional significance and potential roles in the viral life cycle, host interactions, and viral adaptation across different hosts.



Linking Biophysical Features to the Biological Context

Using supervised uniform manifold approximation and projection (UMAP) together with robust gradient boosting regressors (XGBoost) helped unravel the most relevant features.

Therefore, characterizing the impact of secondary structures, sidechains and conformational properties in IDRs could help elucidate vector-specific transmission differences.

Conclusion and Perspectives

In conclusion, the accurate and statistically significant connection of structural features with both IDRs conformational propensities and context-sensitive folding data suggests their critical role in local biophysics with lasting effects on subsequent conformational changes during virus-host interactions and disease-related pathogenicity outcomes.

Our results also highlight the importance of a comprehensive understanding of the combined effect of the conserved structural patterns in the core protein domain and the biophysical features of the genus-specific C-terminal IDRs among rhabdoviral Gps, such as, for instance, in novel plant-virus-vectors associations.

Hence, more is known about the early stages of context-sensitive structural transitions of the rhabdoviral Gp.

Future studies building upon these findings could potentially lead to the development of fine-tuned epidemiological predictions, novel antiviral strategies and deepen our knowledge of complex virological processes.

Entry	Genus	Species	Prediction
A0A2D2PYL4	Cytorhabdovirus	Cabbage cytorhabdovirus 1	Aphid
A0A482PGH7	Cytorhabdovirus	Raspberry vein chlorosis virus	Mite(?)
A0A482PGU2	Cytorhabdovirus	Raspberry vein chlorosis virus	Aphid
A0A6F9EYP3	Cytorhabdovirus	Trichosanthes associated rhabdovirus 1	Aphid
A0A7G3W8J0	Cytorhabdovirus	Medicago cytorhabdovirus A	Aphid
A0A7T7FQX1	Cytorhabdovirus	Kenyan potato cytorhabdovirus	Aphid
A0A7T7FQY6	Cytorhabdovirus	Kenyan potato cytorhabdovirus	Aphid
A0A7T7JPK0	Cytorhabdovirus	Kenyan potato cytorhabdovirus	Aphid
A0A8E6YK00	Cytorhabdovirus	Actinidia cytorhabdovirus JS27	Aphid
A0A8F4M7Q7	Cytorhabdovirus	Passionfruit-associated rhabdovirus YN	Whitefly
A0A1D8FVH5	Nucleorhabdovirus	Physostegia chlorotic mottle virus	Leafhopper
A0A1Y0JW81	Nucleorhabdovirus	Physostegia chlorotic mottle virus	Leafhopper
A0A4P2UVB8	Nucleorhabdovirus	Green Sichuan pepper nucleorhabdovirus	Aphid
A0A4Y6GLZ0	Nucleorhabdovirus	Physostegia chlorotic mottle virus	Leafhopper
A0A6M6R7W4	Nucleorhabdovirus	Cardamom vein clearing nucleorhabdovirus 1	Aphid
A0A8K1X5H2	Nucleorhabdovirus	Cnidium virus 1	Aphid
A0A0H4IR51	Unclassified	Citrus leprosis virus	Mite
A0A2U8J9E2	Unclassified	New discovered	Aphid
A0A2U8J9E8	Unclassified	New discovered	Aphid