



SPECIES DEMARCATION IN PLANT VIRUS TAXONOMY: PRINCIPLES AND RATIONALE

Boopathi N, Jothishri C, Balasubramanian V, Selvarajan R and Nagendran K*

*Department to Molecular Virology Laboratory, ICAR-National Research Centre for
Banana, Tiruchirappalli, Tamil Nadu*

**Corresponding Author Mail ID: krishnagendra@gmail.com*

Introduction

Species demarcation constitutes a fundamental component of virus taxonomy, providing a structured framework for the classification and nomenclature of plant viruses. Given the high mutation rates, recombination potential, extensive genetic diversity exhibited by plant viruses and complete dependence on biological properties such as host range or symptomatology is insufficient for accurate classification. The International Committee on Taxonomy of Viruses (ICTV) therefore employs molecular criteria to define species boundaries, ensuring consistency, reproducibility and global standardization. Such demarcation is essential not only for taxonomic clarity but also for advancing studies on virus evolution, epidemiology, diagnostics and disease management.

Concept and application of sequence identity thresholds

Species demarcation in plant viruses is primarily based on pairwise sequence comparison; wherein defined percentage identity thresholds (cut-off values) are used to distinguish species. These thresholds delineate the upper limits of intra-species variability and the lower limits of inter-species divergence. When sequence identity between two isolates exceeds the prescribed threshold, they are considered members of the same species. On contrary, values below the threshold indicate distinct species. Importantly, these cut-off values are not arbitrary but are derived from comprehensive analyses of sequence datasets encompassing multiple isolates across diverse geographic and host ranges.

Selection of genomic regions for taxonomic comparison

The choice of genomic regions for species demarcation is a critical aspect of ICTV taxonomy. Although complete genome comparisons provide the most comprehensive assessment of genetic relatedness, they are not always practical or necessary. Instead, specific genomic regions are selected based on their evolutionary stability, functional conservation and representation across all members of a given taxon.

Role of conserved and variable genomic domains

Viral genomes comprise regions that differ markedly in their rates of evolution. Highly conserved domains, such as those encoding replication-associated proteins (e.g., RNA-dependent RNA polymerase or replicase), evolve slowly and are suitable for resolving higher-order taxonomic relationships. In contrast, moderately variable regions, particularly those encoding structural proteins such as the coat protein, provide sufficient resolution for species-level discrimination.

Functional constraints and gene selection

The genomic regions employed for species demarcation typically correspond to genes of critical functional importance, including replication, encapsidation, and movement. Examples include the RNA-dependent RNA polymerase (RdRp), coat protein (CP), heat shock protein 70 homolog (HSP70h) in certain virus families, and polyprotein coding regions in monopartite RNA viruses (Table 1).

Table 1. ICTV Species Demarcation Criteria for Plant Virus Families

Family	Genus (Representative)	Cut-off Limit	Target Region for Comparison
Geminiviridae	<i>Begomovirus</i>	< 91% nt identity	Complete genome (DNA-A component)
Nanoviridae	<i>Nanovirus</i>	~75-85% nt identity	Complete genome components
Caulimoviridae	<i>Badnavirus</i>	~80% nt identity	RT/RNase H region (Pol gene)
Potyviridae	<i>Potyvirus</i>	< 76-78% nt identity	Complete polyprotein ORF (CP, NIb, P1 regions)
Secoviridae	<i>Nepovirus</i>	< 73% aa (CP); < 80% aa (Pro-Pol)	Coat protein (CP), Protease–Polymerase
Bromoviridae	<i>Cucumovirus</i>	< 90% aa identity (CP or replicase)	Coat protein (CP), RdRp
Closteroviridae	<i>Closterovirus</i>	< 75% aa identity	HSP70h, RdRp, CP
Tombusviridae	<i>Tombusvirus</i>	< 75% nt identity	Complete genome or RdRp/CP
Virgaviridae	<i>Tobamovirus</i>	< 90% nt identity	Complete genome
Virgaviridae	<i>Furovirus</i>	< 80% nt identity	RNA1 and RNA2 segments
Alphaflexiviridae	<i>Potexvirus</i>	< 72% nt or 80% aa identity	CP or RdRp
Betaflexiviridae	<i>Carlavirus</i>	< 72% nt or 80% aa identity	CP or RdRp
Gammaflexiviridae	<i>Mycoflexivirus</i>	< 70-75% nt identity	RdRp
Partitiviridae	<i>Alphapartitivirus</i>	≤ 90% aa (RdRp); ≤ 80% aa (CP)	RdRp and CP
Endornaviridae	<i>Endornavirus</i>	~75% nt identity	Complete genome
Reoviridae (plant-infecting genera)	<i>Phytoreovirus</i>	~75-80% nt identity	Genome segments (RdRp, CP)
Rhabdoviridae (plant viruses)	<i>Nucleorhabdovirus</i>	~75% nt identity	Complete genome or L gene (RdRp)
Tospoviridae (now within Bunyavirales)	<i>Orthospovirus</i>	< 90% aa identity (N protein)	Nucleocapsid (N) protein gene
Kitaviridae	<i>Cilevirus</i>	~75% nt identity	RNA1 (replicase)
Aspiviridae	<i>Ophiovirus</i>	~85% aa identity	Coat protein (CP)
Deltaflexiviridae (plant-associated)	<i>Deltaflexivirus</i>	~70-75% nt identity	RdRp
Umbraviridae	<i>Umbravirus</i>	~70-75% nt identity	RdRp
Solemoviridae	<i>Sobemovirus</i>	< 75% nt identity	Whole genome or CP
Luteoviridae (now split into Solemoviridae etc.)	<i>Luteovirus</i>	~ <90% aa (CP)	Coat protein
Tymoviridae	<i>Tymovirus</i>	< 80% nt identity	Whole genome or CP
Marafiviridae	<i>Marafivirus</i>	~70-80% nt identity	Polyprotein
Closteroviridae	<i>Crinivirus</i>	< 75% aa identity	HSP70h, CP

Closteroviridae	<i>Ampelovirus</i>	< 75% aa identity	RdRp, HSP70h
Ben yviridae	<i>Benyvirus</i>	~75% nt identity	RNA1 (replicase)
Ourmiaviridae	<i>Ourmiavirus</i>	~75% nt identity	RdRp
Virgaviridae	<i>Hordeivirus</i>	~72–75% nt identity	Genome segments

The functional indispensability of these genes imposes evolutionary constraints that limit random sequence variation, thereby ensuring that observed divergence reflects genuine evolutionary separation rather than stochastic mutations. This functional conservation enhances the robustness and biological relevance of taxonomic criteria.

Empirical determination of demarcation thresholds

ICTV species demarcation thresholds are established through empirical analyses of large-scale sequence datasets. Pairwise sequence identities are calculated among multiple isolates within and between proposed species, and the resulting distributions are examined to identify discontinuities or “barcode gaps.” These gaps represent natural separations between intra-species and inter-species variation. The placement of cut-off values at these boundaries ensures that species demarcation is grounded in observed genetic structure rather than arbitrary assignment. This data-driven approach underpins the reliability and acceptance of ICTV taxonomy.

Variation in demarcation criteria among virus families

Species demarcation criteria are inherently taxon-specific, reflecting differences in genome type, replication strategy, mutation rate, and recombination frequency among virus families. DNA viruses, which generally exhibit lower mutation rates, tend to have higher sequence identity thresholds for species separation. In contrast, RNA viruses, characterized by error-prone replication mechanisms, display greater genetic variability and consequently lower thresholds. Additionally, genome organization (monopartite vs multipartite) and evolutionary dynamics further influence the selection of target regions and cut-

off values. As a result, ICTV adopts flexible, family- or genus-specific criteria rather than imposing a universal standard.

Relevance to plant virology research and diagnostics

Accurate species demarcation has significant implications for plant virology. It underpins the development of reliable diagnostic tools, including serological assays and nucleic acid-based detection methods, which often target conserved genomic regions. Furthermore, precise taxonomic classification enables effective monitoring of virus emergence, evolution, and geographic spread. In applied contexts, such as crop protection and quarantine regulation, clear species boundaries are essential for risk assessment and management strategies. For researchers and students, an understanding of these principles is critical for interpreting molecular data and contributing to taxonomic refinement.

Conclusion

Species demarcation in plant virus taxonomy is a rigorously defined, data-driven process that integrates sequence identity thresholds with biologically meaningful genomic regions. The selection of target regions is guided by considerations of evolutionary conservation, functional significance, and empirical sequence variation, while cut-off values are established based on observed genetic discontinuities. The resulting framework, as standardized by ICTV, provides a robust and adaptable system for classifying the vast diversity of plant viruses, thereby facilitating both fundamental research and practical applications in plant health management.