Plant virus genome studies using novel databases and bioinformatics tools for text compression and entropy

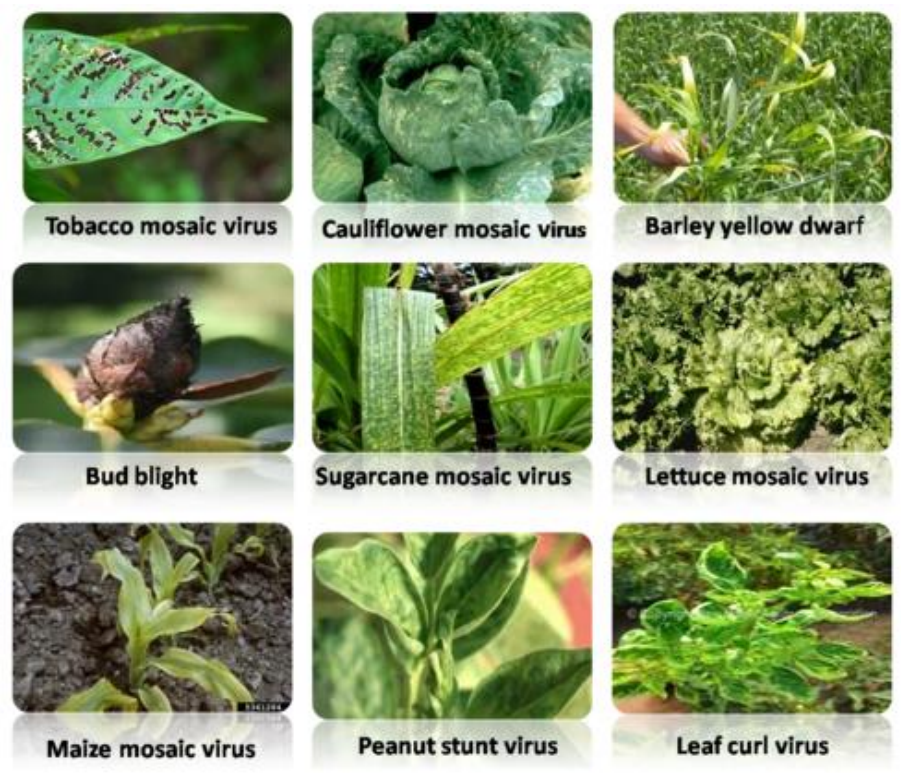
Ignatov A.N., Orlov Y.L., Luzin A.N., Pakina E.N., Dobrovolskaya O.B.

Peoples' Friendship University of Russia (RUDN University), Moscow, Russia
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia

Systems biology approach for analysis of the relationship of plants, phytopathogens and other components of the pathogenesis is a challenging task. Currently, several plant viruses and bacteria are model objects in the study of various types of plant responses aimed at creating resistant plant genotypes, because the resistance of varieties is the most reliable way to reduce the harm caused by these pathogens to crop production.

Novosibirsk, 2021

DNA sequence analysis of plant virus genomes

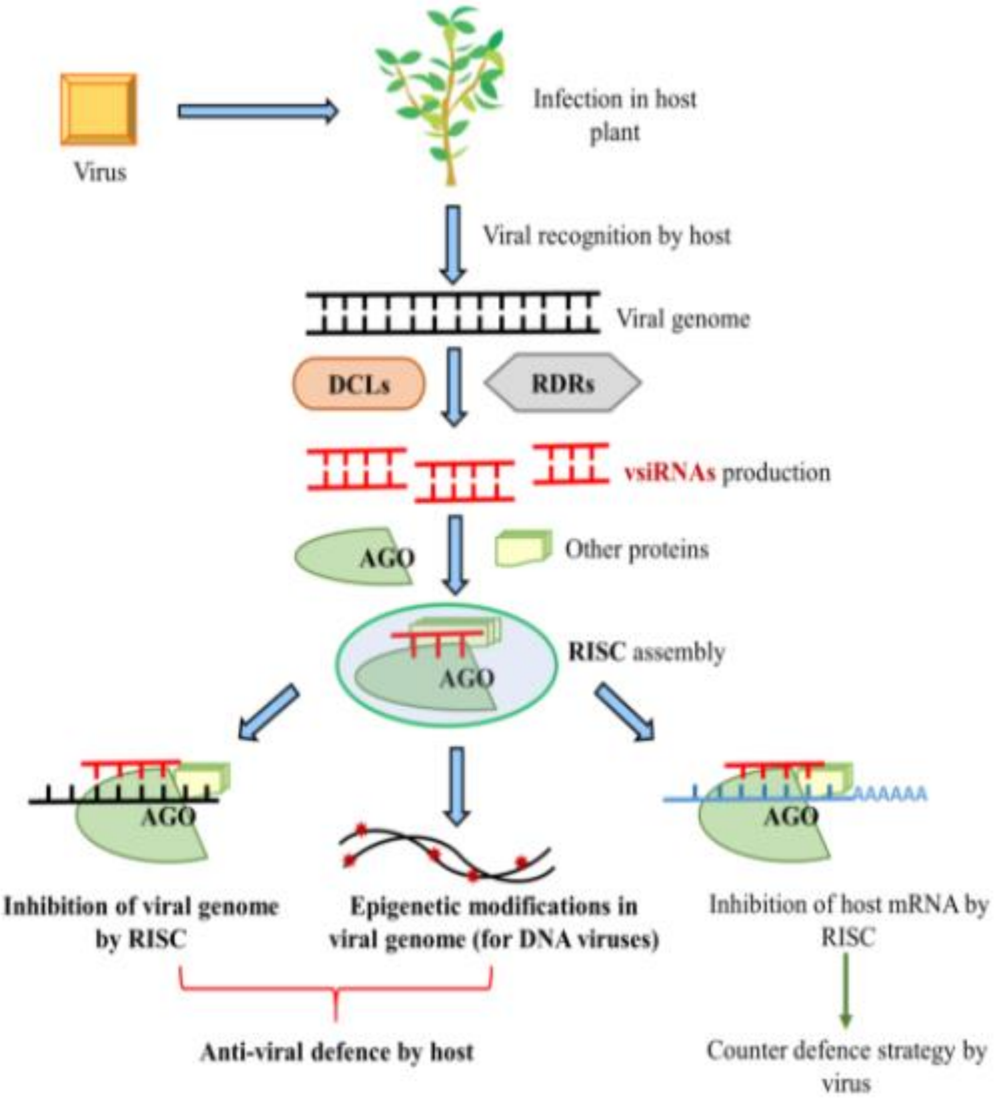


https://www.bighaat.com/blogs/kb/viral-diseases-in-plants-and-its-control

Examples of plant virus phenotypes

RNA viruses face fluctuating environments and are incredibly effective at adaptation when a selective pressure is applied. Central to this adaptive capacity is the enormous genetic diversity that characterizes RNA virus populations, which is mainly due to the distinctive low fidelity of the RNA-dependent RNA polymerases of these viruses and large population sizes. Before the advent of the new generation sequencing, detection of new viruses and plant pathogenic bacteria that show latent infection was an unsolvable task.

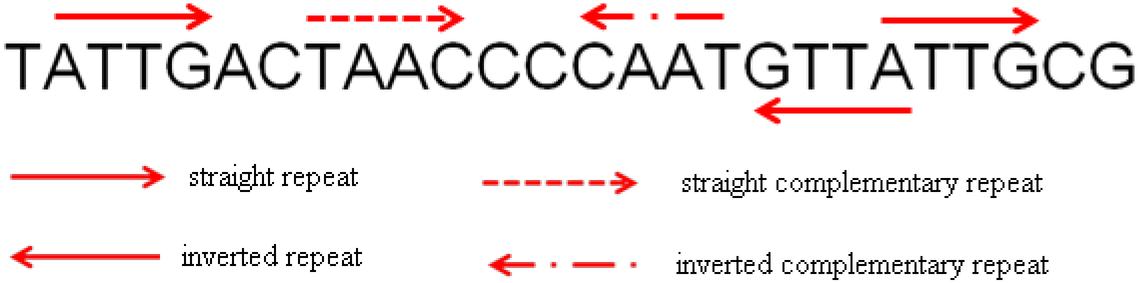
.


Virus
Infection in host plant
Viral recognition by host
Viral genome
DCLs
RDRs
vsiRNAs production
AGO
Other proteins
RISC assembly
AGO
AGO
AAAAAA
AGO
Inhibition of viral genome by RISC
Epigenetic modifications in viral genome (for DNA viruses)
Inhibition of host mRNA by RISC
Anti-viral defence by host
Counter defence strategy by virus

Important bioinformatics problems of plant virus genome organization could be solved using set of novel databases on plant virus data such as PVsiRNAdb (Plant Virus-derived small interfering RNAs database), plant RefSeq.

http://14.139.61.8/PVsiRNAdb/index.php

PVsiRNAdb, is the plant exclusive database dedicated to virus-derived small interfering RNAs (vsiRNA)



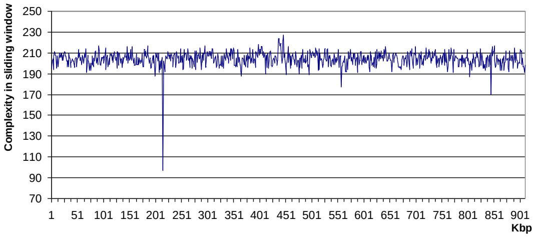
TATTGACTAACCCCAATGTTATTGCG

straight repeat

straight complementary repeat

inverted repeat

inverted complementary repeat



- Lempel-Ziv's complexity

$$H(S) = S[1; i_1]S[i_1 + 1;\ i_2] \dots S[i_{m-1} + 1; N]$$

$S[i_k + 1; N]$ – the part of the given string which is generated on the step k

Now, it is possible to determine all possible variants of RNA transmitted inside the plant cell, which makes it possible to detect viruses that do not have clearly defined symptoms. We used bioinformatics tool Complexity for estimation of plant virus genome complexity, search for genome repeats and rearrangement sites. We estimate nucleotide and dinucleotide entropy in several classes of plant viruses in relation to adaptation. The complexity of symbolic sequences is the important feature used in different scientific directions.

We used Lempel-Ziv complexity measure for DNA sequences (Orlov, Potapov, 2004) to estimate low complexity regions in viral genomes.

http://wwwmgs.bionet.nsc.ru/mgs/programs/complexity