Marilyn J. Roossinck
Editor

Plant Virus
Evolution

Springer

Marilyn J. Roossinck

Editor

Plant Virus Evolution

Springer

Plant Virus Evolution

Marilyn J. Roossinck

Editor

Plant Virus Evolution

Springer

Dr. Marilyn J. Roossinck
The Samuel Roberts Noble Foundation
Plant Biology Division
P.O. Box 2180
Ardmore, OK 73402
USA

Preface

The evolution of viruses has been a topic of intense investigation and theoretical development over the past several decades. Numerous workshops, review articles, and books have been devoted to the subject. Medical practitioners have recognized the importance of viral evolution when treating patients with viral diseases. Farmers have recognized the importance of understanding virus evolution in combating emerging viral diseases in their crop plants. As with any field where knowledge is rapidly expanding, many controversies have also arisen about the nature of virus evolution, how to describe virus populations, how to analyze sequence data and estimate phylogenies, etc. Differing points of view will also be found in the various chapters of this book, and I leave it to the readers to decide for themselves which side they find most helpful. In some cases it seems to me that all sides are correct. In other cases, future historians will decide.

This book focuses on the evolution of plant viruses, although some chapters also draw on the more extensive knowledge of animal viruses. It covers topics on evolutionary mechanisms, viral ecology and emergence, appropriate methods for analysis, and the role of evolution in taxonomy. It includes RNA viruses, DNA viruses, integrated viruses and viroids. I hope that this book will provide a much needed reference for all virologists, teachers, plant pathologists, and evolutionists, and that it will inspire young investigators to explore the topic of plant virus evolution in their research. In many cases plant viruses make excellent models for understanding basic principles of evolution, ecology, and animal/human viral evolution. Plant viruses provide experimental systems that cannot be established for animal viruses, such as the generation of unlimited numbers of genetically identical hosts and the inexpensive cultivation and infection of these hosts. Plant viruses were the first viruses discovered, and they have been studied for more than 100 years. With this book plant virology has finally come of age.

October 2007   Marilyn J. Roossinck

Contents

Contributors

Akhtar Ali
Department of Biological Science, University of Tulsa, 600 South College Avenue, Tulsa, OK 74104-3189, USA

Keith A. Crandall
Genoma LLC, Woodland Hills, UT 84653, USA, and Department of Biology, Brigham Young University, Provo, UT 84602, USA

Aurora Fraile
Centro de Biotecnología y Genómica de Plantas and Departamento de Biotecnología, Universidad Politécnica de Madrid, E.T.S.I. Agrónomos, 28040 Madrid, Spain

Fernando García-Arenal
Centro de Biotecnología y Genómica de Plantas and Departamento de Biotecnología, Universidad Politécnica de Madrid, E.T.S.I. Agrónomos, 28040 Madrid, Spain

Robert L. Gilbertson
Department of Plant Pathology, University of California, Davis CA 95616, USA

Anne-Lise Haenni
Institut Jacques Monod, CNRS et Universités Paris 6 et 7, 2 Place Jussieu – Tour 43, 75251 Paris Cedex 5, France, and Grupo Inmunovirologia, Facultad de Medicina, Universidad de Antioquia, Medellin, Colombia

Glyn Harper
John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

Thomas Hohn
Botanical Institute, University of Basel, Schönbeinstrasse 6, 4056 Basel, Switzerland, and Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

Roger Hull
John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

Marie-Line Iskra-Caruana
CIRAD/UMR BGPI 83, 34398 Montpellier Cedex 5, France

Peter D. Nagy
Department of Plant Pathology, 201F Plant Science Building, University of Kentucky, Lexington, KY 40546, USA

Robert A. Owens
Molecular Plant Pathology Laboratory, Beltsville Agricultural Research Center, (USDA/ARS), 10300 Baltimore Blvd., Beltsville, MD 20705, USA

Marcos Pérez-Losada
Genoma LLC, Woodland Hills, UT 84653, USA

Justin S. Pita
Plant Biology Division, The Samuel Roberts Noble Foundation Inc., P.O. Box 2180, Ardmore, OK 73402, USA

Megan Porter
Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

Alison G. Power
Department of Ecology & Evolutionary Biology, Corson Hall, Cornell University, Ithaca, NY 14853, USA

Katja R. Richert-Pöggeler
Federal Biological Research Centre for Agriculture and Forestry (BBA), Messeweg 11–12, 38104 Braunschweig, Germany

Maria R. Rojas
Department of Plant Pathology, University of California, Davis CA 95616, USA

Marilyn J. Roossinck
Plant Biology Division, The Samuel Roberts Noble Foundation Inc., P.O. Box 2180, Ardmore, OK 73402, USA

Trude Schwarzacher
University of Leicester, University Road, Leicester LE1 7RH, UK

Christina Staginnus
Gregor Mendel Institute of Molecular Plant Biology GmbH, Dr. Bohr-Gasse 3 1030 Vienna, Austria, and GenXPro GmbH, Altenhöferallee 3, 60438 Frankfurt, Germany

Chee How Teo
University of Leicester, University Road, Leicester LE1 7RH, UK

Pierre-Yves Teycheney
CIRAD-UPR75, Station de Neufchâteau, Sainte-Marie, 97130 Capesterre Belle-Eau, Guadeloupe

Chapter 1
Questions and Concepts in Plant Virus Evolution: a Historical Perspective

Fernando García-Arenal(✉) and Aurora Fraile

Abstract The interest in plant virus evolution can be dated to the late 1920s, when it was shown that plant virus populations were genetically heterogeneous, and that their composition changed according to the experimental conditions. Many important ideas were generated prior to the era of molecular virology, such as the role of host- and vector-associated selection in virus evolution, and also that small populations, gene coadaptation and evolutionary trade-offs could limit the efficiency of selection. The analysis of viral genomes in the 1980s and 1990s established the quasispecies-like structure of their populations and allowed extensive analyses of the relationships among virus strains and species. The concept that virus populations had huge sizes and high rates of adaptive mutations became prevalent in this period, with selection mostly invoked as explaining observed patterns of population structure and evolution. In recent times virus evolution has been coming into line with evolutionary biology, and a more complex scenario has emerged. Population bottlenecks during host colonization, during host-to-host transmission or during host population fluctuations may result in smaller population sizes, and genetic drift has been recognized as an important evolutionary factor. Also, particularities of viral genomes such as low levels of neutrality, multifunctionality of coding and encoded sequences or strong epistasis could constrain the plasticity of viral genomes and hinder their response to selection. Exploring the complexities of plant virus evolution will continue to be a challenge for the future, particularly as it affects host, vector and ecosystem dynamics.

Fernando García-Arenal
Centro de Biotecnología y Genómica de Plantas and Departamento de Biotecnología, E.T.S.I. Agronomos, Ciudad Universitaria, Madrid 28040, Spain
fernando.garciaarenal@upm.es

1.1   Introduction

As is the case with all living entities, reproduction of plant viruses may result in the generation of individuals that differ genetically from their parents, which are called mutants or, more vaguely, variants. Hence, populations of plant viruses are genetically heterogeneous, and the frequency distribution of genetic variants in the population (i.e., the genetic structure of the population) may change with time. This process is called evolution. A major area in the study of evolution aims at understanding the mechanisms of evolution and how they shape the genetic structure of populations. Another area aims at understanding the evolutionary history of organisms and the resulting taxonomic relationships among them. Both aspects of evolutionary studies have a long history in plant virology and have attracted much interest in the last few decades, particularly since the availability of molecular analytical techniques, such as those allowing the rapid determination of nucleotide sequences.

   In this chapter we will review how the analysis of plant virus evolution has itself evolved. We do not pretend to make an exhaustive review, but we hope rather to put emphasis on the concepts that have driven the development of the field, illustrated with references to the publications that introduced those concepts or that, in our opinion, best developed them.

1.2   The Early Period

By this, we refer to the period from the origins of plant virology until the widespread use of molecular techniques for nucleic acid analyses. The heterogeneous nature of plant virus populations was evident as early as 1926, by the isolation of symptom variants from areas with atypical symptoms in systemically infected plants (Kunkel 1947) or after biological cloning through single-lesion passage, once necrotic local lesion hosts (i.e., hypersensitive hosts) had been discovered (Holmes 1929). It was also soon perceived that the major components of virus preparations could vary according to the conditions in which the virus was multiplied and passaged. Numerous reports of serial passage experiments including host shifts showed host-associated changes in viral properties, what led to the concept of host adaptation (Yarwood 1979). These observations were interpreted as due to selection in the new conditions. A major concern was whether selection acted on variants present in the original population, or on variants generated under the new conditions. This conceptual dispute was related to a second one about the possibility of obtaining genetically homogeneous virus preparations by single-lesion cloning. Some virologists, particularly Milton Zaitlin, claimed that the frequent appearance of mutants in virus stocks, known from earlier research with *Tobacco mosaic virus* (TMV; Gierer and Mundry 1958), prevented population homogeneity. The reversibility of host adaptation and the first molecular characterization of the phenomenon (Donnis-Keller et al. 1981) supported the

hypothesis of host-associated selection of pre-existing variants. Early molecular analyses also showed that continuous generation of mutants prevented genetic homogeneity in single-lesion-derived stocks (García-Arenal et al. 1984). Hence, the confrontation of the two hypotheses was irrelevant, but it promoted research that showed the relevance of selection as an evolutionary process in plant viruses and the intrinsic heterogeneity of plant virus populations.

Evidence that selection could operate rapidly in viral populations also came from natural populations, particularly in relation to the overcoming of resistance factors in crops. The analysis of the selection of pathotype P1 of *Tomato mosaic virus*, which overcomes *Tm-1* gene resistance in tomato, continues to be a classic (Pelham et al. 1970). However, it was also noticed that selection would not always be so effective, as evidenced by the durability of some resistance factors to viruses in crops. Bryan D. Harrison was responsible for three seminal concepts in this respect. He proposed that the evolutionary relevant size of virus populations could approach the number of infected plants or of viruliferous vectors, being thus much smaller than suggested by the high number of virus particles accumulating in the infected plant. Relatively small population sizes could hinder the efficiency of selection in virus populations (Harrison 1981). In addition, his work on *Raspberry ringspot virus* showed two phenomena also limiting the efficiency of selection: selection for mutual compatibility between RNAs 1 and 2 of the virus, and the existence of evolutionary trade-offs, two concepts that became very important in pathogen evolution theory (Hanada and Harrison 1977).

Interest in the evolution of viruses as taxonomic entities (the concept of virus species was slow to be accepted by plant virologists) also originated in this period. Analyses of relatedness among viruses or strains were initially based on biological assays, such as the extent of cross-protection. Later, serological differentiation indices or the amino acid composition of the coat proteins allowed development of quantitative analyses (Van Regenmortel 1975). The work of Adrian Gibbs pioneered the establishment of phylogenetic relationships among plant viruses, and he was also a pioneer in the development of analytical tools, as exemplified by his work on the relationships among the species of tobamoviruses (Gibbs 1986).

Thus, many of the ideas and conceptual approaches relevant to understanding virus evolution, to be developed later on, were generated in this early period on the bases of sound biological experiments or observations, in spite of limited experimental tools.

1.3 The Analysis of Viral Genomes and Its Impact on Virus Evolution Research: Quasispecies and Phylogenetics

The development in the 1970s of methods for the analyses of nucleic acids had a deep impact on the study of virus evolution. These methods allowed the comparison of virus isolates on the basis of genomic regions or viral proteins other than the structural ones, and eventually allowed the comparison of complete genomes.

Comparison of viral variants made much use of ribonuclease T1 fingerprinting, restriction fragment length polymorphisms (RFLPs), ribonuclease protection assay (RPA) of a labeled complementary RNA probe or single-stranded conformation polymorphisms (SSCPs), in addition to nucleotide sequence determination of genomes or parts of genomes. Data from fingerprints, RFLPs and, of course, nucleotide sequences can be used to directly estimate genetic distances between genotypes, while data from RPA and SSCPs cannot, as they depend on sequence context. These methodological limitations often were overlooked because initial analyses of virus variability focused just on the detection of variants, but later handicapped the development of quantitative analyses of population structure.

The availability of methods allowing the differentiation of closely related genotypes, and the availability of biologically active complementary DNA (cDNA) clones of RNA genomes, definitively determined that virus populations are intrinsically heterogeneous owing to errors during replication. Following the trend with animal- and bacteria-infecting viruses, research focused on RNA viruses, and heterogeneity of cDNA-derived populations was initially shown for *Cucumber mosaic virus* (CMV) satellite RNA and for TMV (Aldahoud et al. 1989; Kurath and Palukaitis 1989). It was shown also, initially for *Tobacco mild green mosaic virus* (TMGMV; Rodríguez-Cerezo and García-Arenal 1989), that the frequency distribution of genotypes in virus populations was gamma, with a major genotype plus a set of minor variants newly generated by mutation or kept at a low level by selection. It was shown later on that the shape of this distribution depended on both the virus and the host plant (Schneider and Roossinck 2000, 2001). This genetic structure had been previously reported for RNA viruses infecting bacteria or animals and had been named a quasispecies (Domingo and Holland 1997), as it corresponded to that predicted by Eigen's quasispecies theory, proposed to describe the evolution of an infinite population of asexual replicators at high mutation rate (Eigen and Schuster 1977). The quasispecies concept has been used often in virology as a mere description for genetically heterogeneous virus populations ("swarms" of mutants), with no concern or awareness for further implications, or for the specific conditions required for the quasispecies concept to materialize, as pointed out by Eigen (1996) himself and developed in the next section. Regardless of the limited appreciation of its implications, the quasispecies concept was crucial in making virologists in the 1980s aware of the intrinsic heterogeneity of virus populations, an early discovery that had been overlooked in an era focused on the molecular analyses of viral genomes.

The quasispecies concept assumed high mutation rates for RNA viruses. It was indeed shown with bacteriophages and with lytic viruses infecting mammalian cells that RNA-dependent RNA polymerases lacked a proofreading activity, and had error rates several orders of magnitude higher than DNA-dependent DNA polymerases of large DNA phages or of cellular organisms (on the order of 10^{-4}–10^{-6} per position and replication round; Drake et al. 1998). Because of high mutation rates of RNA viruses and high accumulation levels in host cells, it was concluded that RNA viruses had large and highly diverse populations. As a consequence, viral populations would easily respond to changing selection pressure, and the evolution

of high mutation rates would have an adaptive value, allowing the virus to survive in changing environments. This concept became the "dogma" that has presided over analyses of RNA virus evolution for more than two decades since the early 1980s. Challenges to this dogma, coming initially from the plant virus field, will be described in the next section.

Nucleotide sequence determination, and the development of methods for the comparison of distantly related sequences, led to phylogenetic analyses of proteins with a similar function in viruses belonging to different genera. These analyses, first done with RNA-dependent RNA polymerases (Kamer and Argos 1984), allowed the classification of viruses in large groups or "superfamilies" (Koonin and Dolja 1993; Goldbach and de Haan 1994) although the validity of the higher-order comparisons was later seriously questioned (Zanotto et al. 1996). Availability of nucleotide sequences of complete viral genomes showed that phylogenies of different gene families were not congruent and that gene organization within the genomes could vary between viral taxonomic groups that were otherwise related. This could be explained by "reassortment of functional modules of coding and regulatory sequences" (Haseloff et al. 1984) according to the concept of "modular evolution," first proposed for bacteriophages (Botstein 1980). Also, availability of whole genome sequences showed that virus genes were often contained totally or partially within another gene, in a different reading frame. This observation led Adrian Gibbs to propose the very novel concept of de novo generation of genes by "overprinting," and methods to analyze which of the two overlapping genes was the novel one (Keese and Gibbs 1992).

The ease of comparing viral genomes also prompted analyses of the genetic structure of natural populations of plant viruses. Phylogenetic approaches were generally preferred to population genetics ones. Both approaches showed from the early 1990s that virus populations could be structured according to different factors, such as geographic or host origin, and different selection pressures were invoked to explain the observed population structures. Again, Gibbs's work on tymoviruses infecting wild plants (Skotnicki et al. 1993, 1996) was pioneering in this field. Major selection pressures acting on virus genomes were identified in this period. Selection was associated with the need to maintain a functional structure, for instance, in the capsid protein of tobamoviruses (Altschuh et al. 1987) or in noncoding subviral pathogenic nucleic acids such as satellites or viroids (Fraile and García-Arenal 1991; Elena et al. 1991). Host-associated selection, already known from passage experiments, was also invoked to explain population structure, for instance, in *Kennedya yellow mosaic virus* (KYMV; Skotnicki et al. 1996), *Hop stunt viroid* (Kofalvi et al. 1997) or *Barley yellow dwarf virus* (BYDV; Mastari et al. 1998). Evidence of vector-associated selection initially derived from loss of transmissibility upon mechanical passage or vegetative propagation of the virus host (Reddy and Black 1977). Population structure in relation to vector transmission has been analyzed in few instances, mostly with begomoviruses (Harrison and Robinson 1999; Simón et al. 2003) supporting vector-associated selection.

Because most analyses of virus population structure followed a phylogenetic approach and because analytical methods were able to differentiate between closely