The Evolution and Emergence of Plant Viruses; Past, Present and Future

Anita Kumari* and Sumit Shekhar

Department of Plant Pathology, Bihar Agricultural University, Sabour – 813210, India

*Corresponding author

A B S T R A C T

Over the years, agriculture across the world has been compromised by a succession of devastating epidemics caused by evolving viruses that spilled over from reservoir species or by new variants of classic viruses that acquired new virulence factors or changed their epidemiological patterns. Population genetics can be used as a powerful tool for identification of disease dynamics over population across large-scale geographic regions. Knowledge of life-history and origin of pathogen can greatly benefit from emergence and expansion of spatial genetics. This branch of genetics uses information of pathogen divergence at the spatial level to gain insights into a pathogen niche and evolution and to characterize pathogen dispersal within and between host populations. The assessment of pathogen transmission across different geographical region, and specifically the evaluation at long-distance dispersal events, has major significance for disease management strategies. To focus on these problems, pathogen tracing relies on indirect approaches that derive epidemiological information from the spatiotemporal structure of pathogen genetic diversity. Viruses are particularly compliant to such studies because of their evolutionary and epidemiological dynamics exists for very short timescales. Moreover, the high number of polymorphisms in their small genomes can be accessed relatively easily and increasingly in real time, during epidemics; such viruses are “measurably evolving” pathogens.

Keywords
Virus, Evolutionary driver, Index of association, Fitness tradeoff, Spatial genetic

Introduction

Evolution of virus is very closely associated with domestication which gives rise to many disease attributes of an agricultural origin. The unrivaled human population densities, domesticated animals and plants in which efficient transmission rates were possible provided new pools for viral disease. The combination of proximity of species as domesticates came into contact with others, each other and indigenous wild species in new environments facilitated the transfer of diseases between species, often with an associated increased virulence in the adopted host. While this process has long been appreciated as an origin of many plant diseases, more recently it has become apparent that the origins of many domesticated plant diseases are recent, and can be categorized into three principal time periods of origin (Jones, 2009; Gibbs et al., 2010).

Firstly, many plant diseases may be associated with the agricultural origin as increased plant densities and intensified agricultural practices caused both the
augmentation of existing diseases from the wild ancestor as well as transmission from other wild species of the centre of crop origin for crops such as wheat, maize and rice (Jones, 2009; Munkacsi et al., 2007). Secondly, subsequent to domestication, the spread to new environments with agricultural magnification caused domesticated plants to come into contact with new indigenous wild populations resulting in host transfer to crops in the past few years (Nguyen et al., 2013; Brunner et al., 2007). Thirdly, more recent global shift of plants and disease vectors in the past few hundred years have also caused the emanation of significant pathogens from wild hosts from quite disparate geographies.

Emergence of infectious plant diseases are recognized as a growing threat to global food security, and among them viruses account for almost half (Anderson et al., 2004). Therefore a better understanding of the origins of viral plant diseases is of significantly important for global food resources management. Ancient DNA and RNA of viruses obtained directly from herbaria and long-term field sampling have manifests that heterochronous sampling serves to improve phylogenetic based estimates by retaliating recent calibration bias and often resulting in a greater time depth for the estimate of viral origins (Gibbs et al., 2010; Fargette et al., 2008; Fraile et al., 1997; Simmons et al., 2008). However, the oldest specimens have been around 100–150 years which have been used to date in age (Fraile et al., 1997; Malmstrom et al., 2007). Therefore, it is possible that further improvements on the estimate of age of virus origins could be obtained from viruses recovered from older, archaeological material.

Origin

The tempo and time scale of plant virus evolution, molecular sequence analyses may also probe spatial population structure and shed light on the transmission dynamics that gave rise to the current spatial distribution of plant viral lineages. It is therefore not surprising that the field of plant virus epidemiology has started to adopt recent statistical inference methodology that integrates temporal and spatial dynamics in a phylogenetic context (Lemey et al., 2009, 2010; Drummond et al., 2012).

As an example of this, the ongoing global spread of tomato yellow leaf curl virus (TYLCV) has attracted significant interest as a potential threat to tomato production in all temperate parts of the world. Viruses are likely to originate in the Middle East during the first half of the 20th century; this area remained epidemiologically relatively isolated. Instead, many global movements of TYLCV appear to have been seeded from the Mediterranean basin.

As another example of a tropical plant virus that poses a threat to food security, maize streak virus (MSV) has caused severe epidemics throughout the maize growing regions. As the etiological agent of the most damaging plant virus disease in the world, cassava mosaic-like virus (CMV) has caused devastating crop losses across world. This epidemic was estimated to have originated in the late 1930s in mainland Africa with subsequent introductions to the southwest Indian ocean islands between 1988 and 2009 (De Bruyn et al., 2012).

Among the fast evolving plant viruses, RYMV is also of particular interest because it circulates in most rice growing countries (Bakker et al., 1974; Abubakar et al., 2003), impacting the lives of millions of impoverished population that rely on rice agriculture for subsistence (Abo et al., 1998). The virus is transmitted by chrysomelid beetles (Bakker et al., 1974), by mammals (Sarra and Peters, 2003), and by contact
during cultural practices (Traore et al., 2006), but no evidence of seed transmission has been found (Konate et al., 2001). The known natural host range of RYMV is limited to the two species of cultivated rice Oryza sativa L. and Oryza glaberrima, and a few related wild grasses (Bakker et al., 1974).

Early spatial genetic analyses have suggested a fairly regular pattern of spread with a correlation between genetic and geographic distances and no evidence of long-range dissemination. Based on comparisons of genetic diversity, these analyses have also implicated East Africa as the area of early diversification (Abubakar et al., 2003). Specifically, A relatively long history of co-existence of RYMV strains in conditions that support habitat fragmentation indeed point at this region as a putative origin for the virus (Fargette et al., 2004).

**Evolution and adaptation to the plant host**

Virus adaptation to novel hosts is an example of the more general evolutionary phenomenon of invasion of and adaptation to a new niche. The new host may generate challenges at the level of entry of virus into the cells, replication of virus and its transmission from the host.

Only a small minority of the initial pool of viral genotypes may survive these hindrances, but if a population is established in the new host, subsequent adaptation will be likely to lead to improved adaptation into the virus. The highest mutation rates among all living entities present in RNA virus because of the lack of proofreading activity associated with RNA-dependent RNA polymerase (RdRp), with extremely high genetic variability being generated rapidly within virus populations. RNA virus populations are typically unruffled with assortment of sequence variants. Often, the variation within an RNA virus population is being depicted by synonymous with a quasispecies (Eigen and Biebricher, 1988; Eigen, 1996; Domingo et al., 2008). The quasispecies model requires those populations which have a high mutation rate and is extremely large in number, with all population members being in direct competition with one another.

As a consequence, a quasispecies population structure is driven entirely by selection. Being deterministic, the quasispecies model does not allow for stochastic changes in population structure, such as those due to genetic drift. Evidence for reduced effective population sizes and genetic drift in plant viruses has been amply documented by several researchers (Garcia-Arenal et al., 2001; Schneider and Roossinck, 2001; French and Stenger, 2003; Hughes, 2009).

A rigorous definition of an emerging virus may be described as “the causal agent of an infectious disease of viral aetiology whose incidence is increasing following its first introduction into a new host population or whose incidence is increasing in an preexisting host population as a result of long-term changes in its underlying epidemiology” (Woolhouse and Dye, 2001), and could also be added as “often accompanied by a significant increase in symptom severity” (Cleaveland et al., 2007).

Accordingly, the epidemic spread 20 years ago of necrogenic strains of Cucumber mosaic virus (CMV) on tomato crops in eastern Spain (Escriu et al., 2000) or the worldwide ongoing epidemic of Pepino mosaic virus in tomatoes should both be considered as paradigms of emerging viral infection. Emerging viruses come from host species in which they are already established and which play the role of a reservoir host during emergence. Species jumps, or spillovers, have given rise to devastating
epidemics, as exemplified above, but there are numerous examples of species jumps that have had far less dramatic consequences. There are even many viruses that have a long history of routinely jumping between species without triggering major epidemics (Thresh, 2006).

The study of viral emergence could be split into three phases. The first phase accounts for the mechanisms and limitations involved in jumping the species barrier. The second phase includes the subsequent evolutionary dynamics that lead to a virus well adapted to its new host. The third phase comprises the epidemiological spread of this well-adapted virus in the new host population.

A detailed description of these three phases is beyond the scope (and length) of this review. Therefore, we will only concentrate on the evolutionary genetic principles underlying first and second phases. Nevertheless, this division in phases is somewhat capricious, since, some of the mechanisms operate during more than one phase.

**Genetics of virus**

The first process in emergence of viral disease is the vulnerability of the new host species to the virus. The rate of exposure will be a function of the ecology of the two hosts and of the transmission biology of the virus, including any relevant vectors.

The pivotal step in emergence of virus is infection of individual of the new host species initially. However, most viruses transferred to new hosts replicate poorly and are inefficiently transmitted. Therefore, the preexistence of host-range mutants within the standing genetic variation in the reservoir host increases the probabilities of a successful jump to new host. The amount of standing genetic variation would depends mainly on i) the rates of mutation and recombination, ii) the distribution of mutational effects on viral fitness, and iii) the strength of genetic drift and gene flow among subpopulations. In addition, it is important to note that host interference with replication allegiance can consequence mutation rates (Pita et al., 2007). Recombination potentially increases fitness by creating advantageous genotypes and removing deleterious mutations, suggesting that will strengthen the process of emergence.

However, this possibility is still controversial. While some studies have proclaimed that recombination may assist the process of cross-species transmission (Chare and Holmes, 2006; Codoner and Elena, 2008), others have pointed out that the association between recombination and emergence is circumstantial (Holmes, 2008). The vast majority of references illustrating examples of recombinant genotypes among plant viruses are based on the analyses of epidemiological sequence data (Awadalla, 2003).

Phylogenetic data have at least one major drawback; they do not represent an unbiased sample of all recombination events but only epitomize successful recombinant genotypes sorted out by natural selection or those genotypes that generally induce new pathologies. Recombination rates are controlled by two factors, the ability of the viral replicases to undergo template switching and the multiplicity of infection (MOI) during infection.

The first factor clearly varies among viruses as a function of their biology and, for example, negative-strand RNA viruses are expected to be less recombinogenic because their RNA is never naked (Chare et al., 2003). The second factor depends on the peculiarities of each virus-host pair and has started receiving attention only very recently.
A fundamental challenge for host-switching viruses is that different hosts impose different selective requirements for viruses; so acquiring the ability to replicate in a new one may impose a fitness burden in the original. These fitness tradeoffs can be generated by different mechanisms, antagonistic pleiotropy (AP) being the simplest and most intuitive one. AP means that mutations that are beneficial in one host may be deleterious in an alternative one.

A second mechanism that promotes tradeoffs results from mutation accumulation by genetic drift. Accumulated mutations may be neutral in the current host but may be essential in a future one (Kawecki1994). Although both mechanisms involve differences in mutational effects across hosts, it is necessary to stress that they are by no means equivalent phenomena. While natural selection is the only reason for the tradeoff in the former, genetic drift is important in the latter.

Most of the accumulated evidence suggests that AP is the principal but not the only reason for fitness tradeoffs (Elena et al., 2009). AP may be an unavoidable consequence of the small size of viral genomes, which in many instances contain overlapping genes and encode multifunctional proteins, making it extremely difficult to optimize one function without jeopardizing another.

Fitness tradeoffs across alternative hosts have been reported for several plant viruses. For instance, Jenner and associates (2002) found that Turnip mosaic virus (TuMV) capable of infecting two different genotypes of turnips paid a fitness penalty compared with the ancestral virus, which was only capable of infecting a given genotype. Similarly, Wallis and associates (2007) have shown that, following serial passages in peas, Plum pox virus increased infectivity, viral load, and virulence in the new host with a concomitant reduction in transmission efficiency in the original host peach trees. Some pieces of evidence also suggest that the fitness of a virus simultaneously facing multiple hosts is either constrained by the most restrictive one or is not subject to a tradeoff at all. In this respect, theory predicts that the extent to which multi host viruses evolve depends on the frequency at which viruses transmit among heterologous hosts (Wilke et al., 2006).

When transmission among heterologous hosts represents an infrequent event, the viral population essentially adapts to the current one. However, if heterologous transmissions are frequent, the viral population behaves as if the fitness landscape did not change at all but was the average of the changing landscapes (Wilke et al., 2006).

The distribution encompasses all possible mutations and can be divided into fractions, beneficial, neutral, deleterious, or lethal. Given the compactness of viral genomes for a well-adapted virus, most mutations are expected to fall into the last two categories. However, the distribution of fitness effects on a given genotype is rarely constant across hosts, and the contribution of each category to the overall fitness will vary depending on the overlap between the alternative hosts (Martin and Lenormand, 2006).

A compelling suggestion is that the more closely related the reservoir and the new host are, the greater the chances for a successful spillover (DeFilippis and Villareal, 2000). There is a good mechanistic reason to believe that a relationship exists between hosts’ phylogenetic distance and the likelihood of viral emergence. If the ability to recognize and infect a host cell is important for cross-species transmission, then related species are
more likely to share related vectors, cell receptors, and defense pathways. However, others state that there are no rules to predict the susceptibility of a new host; spillovers have occurred between hosts independently of their relatedness (Holmes and Drummond 2007). Moreover, viral host switches between closely related species (e.g., species within the same genera) may be limited by cross-immunity to related pathogens (Parrish et al., 2008).

In a very recent study, Cronin and associates (2010) evaluated the relative importance that the following four variables had in key epidemiological parameters that determine potential of different species to serve as reservoirs for Barley yellow dwarf virus species (BYDV) and promote spillovers: i) phylogenetic relatedness between host species, ii) differences in physiological phenotype (rapidly growing short-lived leaves and high metabolic rates vs. slow-growing long-lived leaves and low metabolic rates), iii) provenance (exotic vs. naïve), and iv) host lifespan.

Host physiological phenotype and not the degree of phylogenetic relatedness was the variable better explaining variation among species in their potential as BYDV reservoirs. Indeed, differences among host species in the probability of transmission of BYDV from an infected host to an uninfected feeding vector were only explained by this variable. Additional beneficial mutations or new genetic combinations would be needed to further ensure adaptation to the new host. The evolutionary fate of a population in a constant environment depends on the distribution of mutational effects on fitness. This host-dependence of the distribution of mutational effects may impact the likelihood of adaptation after host switching. For instance, if the host provides new opportunities for the virus, the fraction of beneficial mutations may be increased either by moving the average of the distribution towards more positive values while keeping the shape constant or, alternatively, by increasing the variance without affecting the mean.

**Spatial genetic**

Plant architecture creates a spatially structured environment for plant viruses. This means that the viral population replicating within an infected plant must be considered as a collection of subpopulations, each replicating in different parts, from the arrangement of different tissues within a leaf to individual leaves and, finally, branches. Spatial structure imposes strong conditions on the spread of beneficial mutations that may improve the fitness of an emerging virus on its new host.

Spatial structure and mutual exclusion also reduces the opportunity for recombination and, thus, of generation of genetic variation. In recent years, different groups have evaluated the strength of population bottlenecks during the colonization of distal tissues. Sacristán and associates (2003) used a similar coinoculation approach and estimated that, during systemic colonization by TMV.

Characterizing the distribution of mutational effects across a panel of possible alternative hosts varying in genetic relatedness to the natural one is a pending task. Given the high mutation rate of RNA viruses, mutations may not appear as single events, but genomes may contain multiple hits (Malpica et al., 2002; Tromas and Elena, 2010). The way in which mutations interact in determining viral fitness, a concept known as epistasis, conditions whether certain evolutionary pathways are more likely than others (Weinreich et al., 2005). If mutational effects are multiplicative, the shape of the landscape will be smooth, with a single peak emerging from a flat
surface. By contrast, the stronger the deviation from multiplicatively, the more fitness peaks of different heights may exist in a landscape.

**Evolutionary drivers**

Evolution of virus populations depends on several forces including mutation, recombination, genetic drift, selection and migration, acting concomitantly but exerting pressures that vary widely in direction and intensity. It makes therefore difficult to predict viral emergences or the durability of control strategies. The relative intensity of these forces will determine whether evolution follows predominantly stochastic or deterministic patterns.

Amongst the many known plant pathogens, viruses are responsible for the majority of the emerging diseases that threaten food production worldwide (Anderson et al., 2004). However, viruses in their native environments rarely cause damaging diseases (Jones, 2009). Within the undisturbed ecological contexts of such environments, the numerous interactions that viruses encounter with their natural host and transmission vector species are generally both evolutionarily ancient and relatively stable (Malmstrom et al., 2011).

The rise of modern agriculture has been accompanied by the dissemination of large numbers of exotic plant species, transmission vectors and viruses into foreign environments, which has precipitated multitudes of novel evolutionarily recent virus-host-vector-environment interactions (Fig 1). It is possible that the instability of some of these “unnatural” interactions, has in many cases triggered the emergence of devastating new viral diseases (Jones, 2009). The key to understanding the emergence/re-emergence of novel viruses is to know the intricate “host pathogen-environment” relationship in the evolution of pathogens. While the emergence of infectious diseases in naive regions is caused primarily by the movement of pathogens via trade and travel, local emergence is driven by a combination of environmental and social change.

The molecular evolutionary changes that accompany changes in the host ranges of animal and plant viruses have been studied using susceptible hosts; viral populations have been transferred serially in a single or in different host(s), as reported for some viruses (Kurath and Palukaitis, 1989; Schneider and Roossinck, 2000, 2001; Hall et al., 2001; Liang et al., 2002; Novella, 2004; French and Stenger, 2005; Carrillo et al., 2007; Elena and Sanjuan, 2007; Iglesias and Elena, 2007; Wallis et al., 2007).

Similar studies of serially transferred bacteriophages (Wichman et al., 1999; Bull et al., 1997) found two sorts of convergent change in the genomic sequences of adapted variants: some sites in independently passaged isolates had identical mutations, whereas others had different mutations, and they distinguished these as resulting from ‘parallel evolution’ and ‘directional evolution’. However, Sacristan et al., (2005) did not find evidence of convergent evolution in cucumber mosaic virus strains passage into different host plant species. There are three major ways of vertical transmission of plant viruses via the contamination of true seeds.

In only a few examples, particularly stable viruses such as tobamoviruses can be retained in the seed coat and then transmitted to the seedling after germination (Broadbent, 1965). In that case, there is no contamination of the embryo and the process of seedling infection resembles horizontal transmission through contact with an infected plant. The two other ways of contamination correspond to invasion
of the embryo by the virus, either from infected maternal tissues or, more rarely, via infected pollen. Although seed embryos are usually protected against invasion by viruses that affect the mother plant, many viruses have the capacity to circumvent this barrier. Even low rates of seed transmission can be epidemiologically important because secondary spread of viruses can begin as soon as the germination stage (Coutts et al., 2009) and virus seed transmission can be economically significant for at least 18% of plant viruses (Johansen et al., 1994).

![Fig.1](image)

**Fig.1** A cartoon depiction of important emerging/re-emerging viral infections and their possible origins, evolutionary drivers, and risk factors

Most of the material we brought together for this review explores the role of viral evolution in the early stages of emergence. We would like to argue here that the viral genetic variability contained in the reservoir population is the most important genetic determinant of viral emergence. We know viruses of wild plant species that probably work as a large reservoir generating spillovers on cultivated plants or between wild species, so there is a whole evolutionary space that we totally ignore, making it more difficult to predict and prevent emerging plant viral diseases. Natural selection will operate upon this genetic variability to optimize viral fitness. After reading the presentation we made above, one may consider that successful emergence, characterized by sustained host-to-host transmission, may be a far more difficult process than expected given the remarkable evolutionary plasticity of RNA viruses. Fitness tradeoff is a strong bottleneck at different levels of emergence, an excess of deleterious mutations, spatial constraints, and fragmented host populations will limit the chances for new viruses to emerge.

**References**


Nguyen, H. D. et al., Turnip mosaic potyvirus probably first spread to Eurasian Brassica crops from wild orchids about 1000 years ago. Plos One 8, e55336; DOI: 10.1371/journal.pone.0055336 (2013).


How to cite this article:
doi: https://doi.org/10.20546/ijcmas.2020.905.306