



**HAL**  
open science

## Barriers to gene flow between emerging populations of Watermelon mosaic virus in southeastern France

Benoit Joannon, Claire Lavigne, Hervé Lecoq, Cecile Desbiez

► **To cite this version:**

Benoit Joannon, Claire Lavigne, Hervé Lecoq, Cecile Desbiez. Barriers to gene flow between emerging populations of Watermelon mosaic virus in southeastern France. *Phytopathology*, 2010, 100 (12), pp.1373-1379. 10.1094/PHYTO-04-10-0118 . hal-02662396

**HAL Id: hal-02662396**

**<https://hal.inrae.fr/hal-02662396v1>**

Submitted on 31 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Barriers to Gene Flow Between Emerging Populations of *Watermelon mosaic virus* in Southeastern France

B. Joannon, C. Lavigne, H. Lecoq, and C. Desbiez

First, third, and fourth authors: INRA, Unité de Pathologie Végétale UR407, F-84140 Montfavet, France; and second author: INRA, Unité de Recherches Plantes et Systèmes de culture Horticoles UR1115, F-84000 Avignon, France.  
Accepted for publication 17 August 2010.

## ABSTRACT

Joannon, B., Lavigne, C., Lecoq, H., and Desbiez, C. 2010. Barriers to gene flow between emerging populations of *Watermelon mosaic virus* in southeastern France. *Phytopathology* 100:1373-1379.

Since 1999, “emerging” (EM) strains of *Watermelon mosaic virus* (WMV) have been detected in cucurbit crops of southeastern France, probably as a result of recent introductions. Population genetic approaches were used to study the structure of EM isolates in southeastern France and to identify factors involved in their spatial distribution. A population clustering method (SAMOVA) and a maximum-difference algorithm (Monmonier’s algorithm) were combined to visualize and quantify barriers to gene flow between populations. Both methods yielded similar results

and two main barriers were identified. A North/South oriented barrier may be related to physical obstacles to gene flow (Rhône River, presence of an area with few cucurbit crops). Although the barrier was very strong, some “crossing” events were detected. A second barrier, oriented Northwest to Southeast, was not correlated with obvious geographical features. The two methods used here are complementary and confirm the limited spread of WMV-EM isolates. This approach can be useful in epidemiology studies to characterize the structure of viral populations and identify barriers to gene flow.

*Additional keywords:* epidemiology, potyvirus, spatial structure.

*Watermelon mosaic virus* (WMV, a member of the genus *Potyvirus*) is the most common virus infecting cucurbit crops in France (12,31). WMV is transmitted by numerous aphid species in a nonpersistent manner, i.e., with short acquisition and retention times (30). WMV epidemics can spread very rapidly at the field level (29), but the distance of migration at a larger scale is still poorly known. Since its first description in France in 1974, WMV has been associated with severe symptoms and important yield losses in melon crops but usually with mild symptoms without economic impact in zucchini squash crops (29). All WMV isolates collected in France before 1999 belong to the same phylogenetic group 1 (10) and will be thereafter referred to as classic (CL) isolates. In early 2000, new WMV strains that caused severe symptoms in zucchini squash were isolated in southeastern France (10,11). Molecular studies showed that these new strains, referred to as emergent (EM) isolates, are distributed in four molecular subgroups, EM1 to EM4, belonging to phylogenetic group 3 (11). These new WMV isolates are closely related to strains from Asia, but their introduction routes have not yet been identified (11). A survey conducted from 2004 to 2007 in the major growing areas in France revealed that EM strains increased in frequency in southeastern France, rapidly replacing CL strains, but were not detected in the other cucurbit production areas (southwestern and central-western France) (11).

In this study we applied population genetic approaches in order to study the structure of EM isolates in southeastern France, and to identify factors which could be involved in their particular distribution. Attempts based on phylogeography for the estimation of plant virus migration, as for *Rice yellow mottle virus* (RYMV, a member of the genus *Sobemovirus*) in Africa (1,47) or *Turnip mosaic virus* (TuMV, a member of the genus *Potyvirus*) in Asia

(46), usually deal with large time scales (several decades), but no spatial studies were performed shortly after the introduction of new strains in a given environment. In the case of WMV, results from epidemiological surveys suggested that long-distance dissemination of EM strains was limited in the few years following their introduction (11). This can be related to an intrinsically low efficiency of spread beyond a few hecto- or kilometers, and/or to the presence of barriers to gene flow. Studying the spatial pattern of EM strains aims at understanding better the host–vector–pathogen interactions in an agricultural landscape (40). A better knowledge of ecological or genetic factors that can reduce the spread of nonpersistently transmitted plant viruses (7,8) and the barriers limiting virus spread may help to characterize more accurately the mechanisms of viral emergence and define ways to limit the risks associated with such emergences.

The concept of barrier to gene flow is now widely used in population genetics (14,16,32). The number of gene flow and population structure studies has greatly increased in recent years in the ecology and conservation biology fields (41). Genetic barriers are in some cases associated with geographic elements (mountains, rivers...) (13,26,48) or climatic constraints (dominant winds...) (15). Several methods, commonly used or newly developed, allow studying population structure of microorganisms (25,37). These new approaches begin to be used for studying epidemiology of infectious diseases (2,9). However, the studies were usually conducted for diploid and sexually reproducing organisms. These methods, particularly those using Bayesian analysis, frequently require multiple independent loci with neutral evolution, which is not available for a monopartite virus such as WMV whose 10-kb genome is mostly submitted to strong negative selection (12). They also usually rely on strong assumptions of Hardy-Weinberg and linkage equilibrium, even if they are sometimes considered as robust enough to be extended to populations where these assumptions are violated (e.g., use of STRUCTURE [39] for Hepatitis B virus [43]). At least two methods, however, can be applied to haplotypic data and can do

Corresponding author: C. Desbiez; E-mail address: cecile.desbiez@avignon.inra.fr

doi:10.1094/PHYTO-04-10-0118

© 2010 The American Phytopathological Society

without these a priori assumptions. Besides, they rely on different concepts and are complementary. Monmonier's algorithm (35) aims at finding genetic barriers between populations, whereas spatial analysis of molecular variance (SAMOVA) (14) identifies groups of populations with a maximal intergroup differentiation. Both methods are supposed to be efficient when gene flow between groups is low (14). The association of Monmonier's algorithm and a clustering method is increasingly used in population genetics, whether with SAMOVA (3) or with a model-based Bayesian clustering (45).

This study is to our knowledge the first one analyzing barriers to gene flow using these approaches for a plant RNA virus at these time and geographical scales. We used two complementary methods, Monmonier's maximum-difference algorithm (35) and SAMOVA (14) to analyze population structure and look for barriers to gene flow for EM WMV populations from 2004 to 2008 in southeastern France.

## MATERIALS AND METHODS

**Sampling and characterization of WMV populations.** The strains used for the spatial analysis of WMV were collected between 2004 and 2008 during a nationwide epidemiological survey. Samples of melon, zucchini, squash, and cucumber were collected each year during the major cucurbit growing season, i.e., between June and September. In addition to the 1702 (including 1349 WMV-positive) samples collected in southeastern France from 2004 to 2007 and already characterized molecularly (11), 335 samples were obtained in 2008, mostly from farmers, farm advisers, and seed companies. From 1 to 25 samples were obtained from each sampling location. Information associated with each sample included sampling date, geographic origin at the town level, crop species, and cultivar.

All samples were tested by double-antibody sandwich enzyme-linked immunosorbent assay with a WMV polyclonal antiserum. Positive samples were characterized by one-step reverse transcription-polymerase chain reaction (RT-PCR) and direct sequencing of a 408-bp fragment encompassing the variable N-terminal part of the coat protein coding region (for more details see literature citation 11). Sequence fragments were aligned with ClustalW included in DAMBE (51) and BioEdit (22). Strains were clustered into molecular subgroups by computing genetic distances between isolates and building neighbor joining phylogenetic trees with MEGA 3.1 software (28). Sequences from samples with CL/EM mixed infections were checked by eye or by RT-PCR with subgroup-specific primers designed in the P1 coding region to determine the EM subgroup(s) present in the mixed infections: W-EM1-930-5' (TACATTGGAACAACACTGCTGC), W-EM2-540-5' (TGGTATTGCCAATGGCACAC), W-EM3-190-5' (CTGAAA

CACAGCACAGTAACA), W-EM4-550-5' (GGTTCTACCAGTG GTGGGT), and the nonspecific reverse primer WMV-1090-3' (TATTTGCCTTCTTCATGTGG). These samples were included to the analyses.

For spatial structure analysis, only EM samples from three contiguous "départements" (French administrative units) (Gard, Vaucluse, and Bouches-du-Rhône, in southeastern France) were used. In France, EM strains were very likely first introduced in these areas (11) since they were not detected in the other production areas (mostly southwestern and central-western France) until 2008 (11; this study). These three départements correspond to important cucurbit production areas (circa 50% of the French production, for melon and zucchini squash) and represent nearly 70% (2037/2994) of all samples in the national epidemiological survey between 2004 and 2008. Only sampling sites with at least three WMV-EM-infected samples were taken into consideration for spatial and statistical analysis. Between 21 and 33 localities per year were thus included in the analyses (Table 1) with a total of 1,060 WMV-EM samples—ranging from 166 samples (2004) to 299 samples (2006) per year. In the absence of GPS coordinates for the sampling sites, each site was represented by the geographic coordinates of the nearest town or village as indicated by the INSEE (<http://www.insee.fr>).

**$F_{ST}$  calculation.** To analyze the structure of WMV-EM populations,  $F_{ST}$  values were calculated according to Weir (49) with the software GENETIX (4). To test isolation by distance,  $F_{ST}$  values were calculated between all population pairs (one population is a set of samples collected in the same place during the same growing season), using raw sequences of all isolates, without correction for multiple substitutions (default options). For barrier analyses, since intra-EM-subgroup variability was very limited (see results), all isolates from each EM subgroup were considered as identical in sequence to the major haplotype of the EM-subgroup, i.e., FMF00-LL1 (GenBank accession number EU660581) for subgroup EM1, FMF03-141 (EU660583) for EM2, FBR04-37 (EU660586) for EM3, and C05-270 (EU660585) for EM4, and  $F_{ST}$  values were calculated accordingly.

**Isolation by distance.** Isolation by distance is defined by the decrease in genetic similarity among populations as geographic distance between them increase (50). If dispersal is spatially limited, individuals in populations that are close may interact easily and genetic exchange is facilitated between them. Allele frequencies in populations that are closer are more similar than frequencies from more distant populations. In the case of viruses or clonal organisms, isolation by distance would rather correspond to a progressive accumulation of mutations during the spread from the point of introduction. Relationships between genetic and geographic distances were tested by a Mantel test for spatial autocorrelation (34). For WMV, the test was first per-

TABLE 1. Spatial analysis of molecular variance (SAMOVA) and Monmonier's algorithm to identify barriers to gene flow on *Watermelon mosaic virus* (WMV) emergent (EM) populations from southeastern France from 2004 to 2008

| Year                                           | 2004              | 2005              | 2006              | 2007              | 2008              |
|------------------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Number of populations                          | 21                | 24                | 33                | 21                | 21                |
| Number of samples from epidemiological surveys | 166               | 214               | 299               | 212               | 169               |
| Mean number of samples per site                | 7.9 (±5.6)        | 8.9 (±10.4)       | 9.1 (±5.9)        | 10.1 (±5.2)       | 8 (±5.7)          |
| <b>SAMOVA</b>                                  |                   |                   |                   |                   |                   |
| Number of clusters (K)                         | 4                 | 4                 | 7                 | 4                 | 3                 |
| $F_{CT}$                                       | 0.846             | 0.743             | 0.767             | 0.699             | 0.897             |
| $F_{SC}$                                       | 0.052             | 0.047             | 0.014             | 0.097             | 0.107             |
| $F_{ST}$                                       | 0.854             | 0.756             | 0.770             | 0.728             | 0.908             |
| P value ( $F_{CT}$ )                           | <10 <sup>-5</sup> | <10 <sup>-5</sup> | <10 <sup>-5</sup> | <10 <sup>-5</sup> | <10 <sup>-5</sup> |
| <b>Monmonier's algorithm</b>                   |                   |                   |                   |                   |                   |
| Number of groups                               | 5                 | 6                 | 9                 | 5                 | 5                 |
| $F_{CT}^a$                                     | 0.834             | 0.742             | 0.770             | 0.733             | 0.908             |
| $F_{SC}^a$                                     | 0.003             | 0.024             | -0.042            | -0.060            | -0.077            |
| $F_{ST}^a$                                     | 0.834             | 0.749             | 0.760             | 0.717             | 0.901             |

<sup>a</sup> Values calculated with AMOVA on the groups defined with Monmonier's algorithm.

formed on all EM isolates taken together. The analysis was also performed on each EM subgroup independently. Mantel tests were performed with the web software IBDWS v3.16 (24).

**Monmonier's algorithm: Search for barriers to gene flow.** Monmonier's algorithm (35) aims at identifying zones of sharp genetic changes in a geographical area, thus drawing "barriers" between populations that display a maximal genetic distance. Monmonier's algorithm was applied to study WMV-EM distribution (each year from 2004 to 2008 was analyzed separately) using BARRIER 2.2 software (33).

The first step of the analysis consists in connecting sampling localities using Delaunay network (5), fastest method for connecting a set of point on a plane. Pairwise  $F_{ST}$  values were calculated between each pair of directly connected localities in the Delaunay network. The second step in Monmonier's algorithm consists in computing barriers between localities. A first barrier is drawn across the edge with the largest genetic distance. The line is then extended across adjacent edges with the next largest genetic distances, until it reaches the border of the network, meets another barrier or closes a circle around one or several localities.

An option in BARRIER 2.2 allows subreplicate analyses with resampled bootstrap matrices to test the robustness of barriers. Bootstrapped matrices ( $n = 100$  bootstraps) were obtained with the "R" software (44) at each sampling locality, conserving the average frequency of the different EM subgroups in each sampling point. For each iteration, new  $F_{ST}$  matrices were calculated using GENETIX (4) and were then analyzed with BARRIER 2.2 (33). The barriers obtained are displayed with different thickness depending on their robustness, i.e., the number of appearances of the barrier in the 100 bootstrapped populations. The number of barriers displayed is limited to four to limit interference and artificial separation between populations.

**SAMOVA.** SAMOVA allows defining partitions of local population into clusters that are maximally differentiated from each other and geographically contiguous. As a by-product, groups are separated from each other by barriers to gene flow. Contrary to Bayesian methods such as STRUCTURE (39), this method makes no assumptions about Hardy-Weinberg equilibrium within population and linkage equilibrium between loci (14). The analysis was performed using the software SAMOVA 1.0 (14).

For each year, the  $n$  sampled populations are initially distributed into  $K$  clusters ( $K$  is defined a priori). Then a simulated annealing procedure is used to find the composition of the  $K$  clusters. This simulation aims at maximizing the fixation index  $F_{CT}$  which is the proportion of total variance due to differences between groups of populations.

For each simulation, the population indexes  $F_{CT}$ ,  $F_{ST}$ , and  $F_{SC}$ , quantifying variation between groups, between populations, and within populations, respectively, were calculated. The observed values were compared with those expected in a completely non-structured population and the statistical significance of the population clustering is estimated. For each year, values of  $K$  ranging between 2 and 10 were tested, and  $F_{CT}$ ,  $F_{SC}$ , and  $F_{ST}$  values were plotted.  $K$  value was considered as optimal when the  $F_{CT}$  was maximal or only marginally increasing with  $K$ , with a minimal but positive  $F_{SC}$  value. This low  $F_{SC}$  value means that clusters are genetically as homogenous as possible. The optimal  $K$  also aimed at limiting the number of clusters containing only one population. To avoid an effect of a given initial configuration on the final configuration of the  $n$  populations in the  $K$  clusters, the simulated annealing process is repeated 100 times with a different initial configuration for each iteration. The configuration with the largest  $F_{CT}$  value after the 100 independent simulated processes is considered as the best grouping of populations.

In order to compare clusters obtained with SAMOVA and Monmonier's analysis, the grouping obtained with Monmonier was tested by analysis of molecular variance (AMOVA) (18) with

the software Arlequin 3.1 (17). Using AMOVA,  $F_{CT}$  values were calculated for the groups, i.e., populations separated by barriers, obtained with Monmonier's algorithm, using the same settings as in the SAMOVA procedure (pairwise difference, 1,000 permutations). Similar population clustering and  $F_{CT}$  values obtained with both methods indicate that the results of both analyses are consistent.

## RESULTS

**Variability of EM isolates.** In 2008, 196 WMV-positive samples were analyzed in Vaucluse (48.5%), Bouches-du-Rhône (11.2%), and Gard (40.3%). CL strains represented only 10% (19/196) of all samples, mostly in Gard, confirming the fast local replacement of CL strains by EM strains (11). As a consequence of the decrease of CL strains, mixed infections with CL+EM isolates represented less than 4% (8/196) of the samples in 2008 (Fig. 1). Contrary to previous years, EM1 populations represented 10% (8/80) of the EM isolates detected in Gard (Fig. 1), and were observed not only close to the Rhône River that constitutes the limit between this département and Vaucluse/Bouches-du-Rhône, but also in more distant sites apparently unconnected with other EM1 populations.

For all samples collected between 2004 and 2007, 60 haplotypes (i.e., different sequences) had been obtained in the four subgroups of EM strains in southeastern France (11). Eighteen haplotypes belonged to EM1 subgroup, 25 to EM2, 6 to EM3, and 11 to EM4 (11). In 2008, 26 distinct haplotypes were found in the three départements. Among them, 14 new haplotypes were identified and deposited into GenBank (accession numbers HMO44202 to HMO44215): 7 belonged to subgroup EM1, 2 to EM2, 1 to EM3, and 4 to EM4. For the 5 years, the divergence (i.e., maximal diversity) between EM subgroups was 6 to 8%, while the divergence within subgroups did not exceed 1.2%. The mean diversity within subgroups did not exceed 1%, whereas the mean diversity between subgroups was 4.2%. For each subgroup, one major haplotype, corresponding to the type strain of the subgroup, represented more than 70% of the sequences. Since intra-subgroup diversity was very low, only the subgroup level was used to perform Monmonier's algorithm and SAMOVA of the different data sets. As a control, all sequence data including intra-subgroup variants were used for year 2006, and the results were not modified (data not shown).

**Number of samples and subgroups per site.** The mean number of samples per locality ranged between 7.9 and 10 (Table 1). The difference between years was not significant (Kruskal-Wallis test,  $P = 0.22$ ). The mean number of subgroups identified per site each year ranged between  $1.2 \pm 0.5$  (2004) and  $1.5 \pm 0.6$  (2006). No significant difference between years was observed ( $P = 0.21$ ), but the difference between départements was significant ( $P = 0.005$ ): except in one site, only EM1 was present in Bouches du Rhône (average number of subgroups per site  $N_{site} = 1.06$ ), whereas at least two EM subgroups were repeatedly present in Gard and Vaucluse ( $N_{site} = 1.47$  and  $N_{site} = 1.48$ , respectively).

**Isolation by distance.** No pattern of isolation by distance was observed with the genetic marker used in this study (e.g.,  $r = 0.0854$ ,  $P = 0.1990$ , Mantel test on EM1 strains for year 2006; not shown in the other cases). This can be related to the high genetic homogeneity within each EM subgroup: for EM1 to EM4, one major haplotype represented 70 to 90% of the isolates and the genetic differentiation within EM subgroups was very low regardless of the geographic distance.

**Barriers to gene flow.** Monmonier's algorithm allowed defining geographic barriers each year among WMV-EM populations (Fig. 2). An important North-South barrier separating Gard from Vaucluse and Bouches-du-Rhône was consistently observed between 2004 and 2008. This barrier to genetic flow was the most

significant each year (score between 80 and 100). It seemed to follow roughly the Rhône River separating Gard from Vaucluse and Bouches-du-Rhône and distinguished mostly strains EM3/EM4 from strains EM1/EM2. Despite its high significance, this barrier seemed to be crossed repeatedly by EM1 strains, detected in Gard generally at low frequency almost each year, and by EM3 or EM4 strains occasionally present in Bouches-du-Rhône and Vaucluse. Another barrier, roughly oriented North-west-Southeast, was observed each year but was more fluctuating geographically (Fig. 2). This barrier was also highly significant, with scores ranging from 70 to 100. It distinguished populations from northeastern Vaucluse consisting mostly of EM2 strains from those of southern Vaucluse and northern Bouches-du-Rhône where EM1 strains were the most frequent. It did not follow the Durance River which separates Vaucluse from Bouches-du-Rhône, but it was located within Vaucluse.

Fluctuating barriers with lower bootstrap values (30 to 50) were also detected in 2005 and 2006. They usually separated populations displaying different proportions of the same strain subgroups. In particular, a zone in central and/or eastern Vaucluse containing EM1 and EM2 populations at an approximate 50/50 ratio was detected repeatedly between 2005 and 2007 (Fig. 2B, C, and D).

**Clustering of populations (SAMOVA).** Using the same set of populations as for Monmonier's analysis, SAMOVA analysis allowed defining homogeneous clusters among WMV-EM populations. The number of clusters (*K* value) ranged between 3 (2008) and 7 (2006). Very high genetic differentiation between clusters was detected, with *F<sub>CT</sub>* values ranging from 0.699 (2007) to 0.897 (2008). *F<sub>CT</sub>* values were always statistically significant ( $P < 10^{-5}$ ) (Table 1).

Every year, three main clusters were observed (Fig. 2). The first one includes a majority of the sampling points from Gard, and contains mostly EM4 strains. At least one minor cluster was also observed each year in Gard, corresponding to sampling points where EM3 or EM1 strains were present at a frequency higher than 50% (Fig. 2).

The second cluster contains mostly isolates from EM1 subgroup and corresponds to sampling points from southern Vaucluse

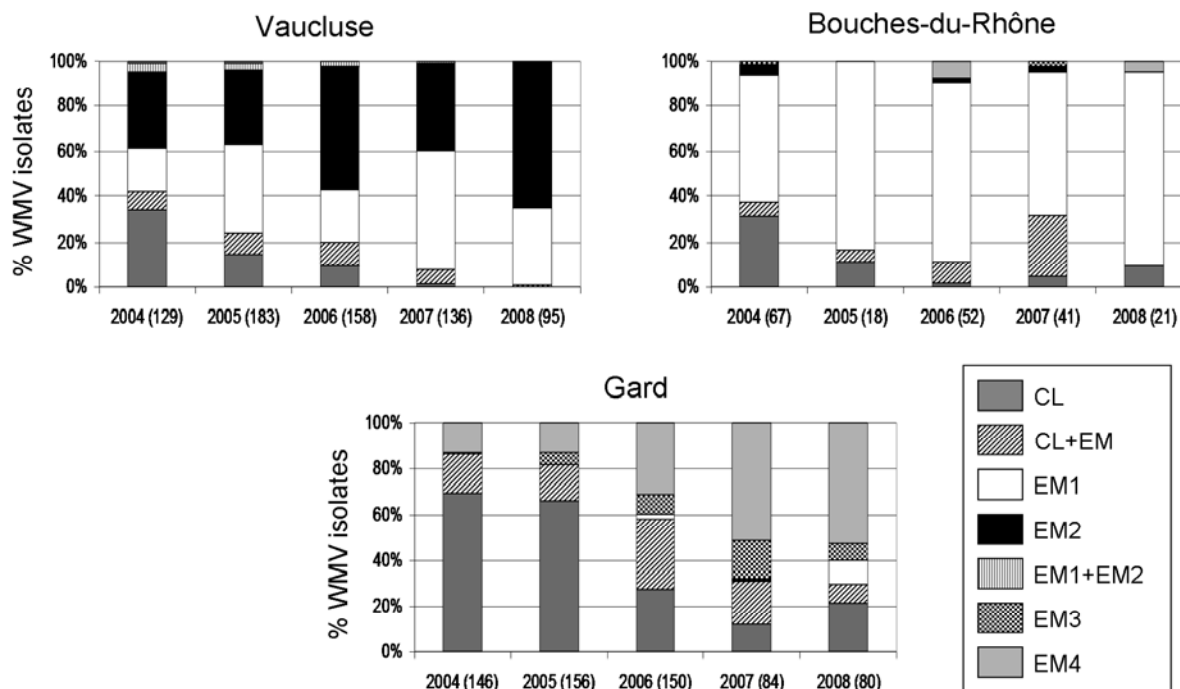
and from Bouches-du-Rhône. The third cluster assembles sampling points from Northern Vaucluse where EM2 subgroup was found almost exclusively. Clusters obtained with SAMOVA were genetically homogeneous and mainly not fragmented. Mixed populations, corresponding to localities where several EM subgroups were found, were generally included in one of the three major clusters. However in 2006, a distinct cluster characterized by mixed infections of EM1 and EM2 subgroups at an approximate 50/50 ratio was defined in eastern Vaucluse (Fig. 2C).

## DISCUSSION

This work aimed at characterizing the structure of WMV-EM subgroups in southeastern France during the first years after the introductions of EM strains, and determining if such structure could be explained at least partially by the presence of barriers to gene flow.

Many methods allow testing the geographical distribution in a landscape and/or presence of barriers to gene flow (42). An important step for studying the effect of landscape is to design a sampling scheme adapted to specific questions, but a problem is that the optimal sampling depends on the true pattern which is not known in advance (20). In our study, the set of data available was from samples collected from different sources which could differ in number, timing and location from year to year. Only three localities were sampled every year during the 5 years, and only 12 sites have been sampled for 4 years. This could introduce biases locally but the general pattern of barriers remained stable in the study. Changes in cucurbit varieties within the 5 years of the surveys could have introduced biases in our studies. However, there was no obvious change in the cucurbit species or cultivars grown during the 5 years of the survey (data not shown).

A regular distribution of sampling is frequently considered as the ideal case (27) but a lot of studies do not satisfy this point. In the case of Monmonier's algorithm, a regular distribution of sampling points is required for a maximal accuracy of the method. Barriers tend to separate the most distant populations (33). So, it is difficult to distinguish a sharp change in genetic distribution from a nonsampled population. For instance, for year



**Fig. 1.** Distribution (in percentage) of *Watermelon mosaic virus* (WMV) populations in three “départements” of southeastern France: Vaucluse, Bouches-du-Rhône, and Gard, from 2004 to 2008. Number of samples per year is indicated in brackets.

2007 there are only a few sampling points in Gard. They are distant from each other and there is a lack of information between the north and the south of this département (Fig. 2D). The connectivity and the interaction between these sampling points are not obvious.

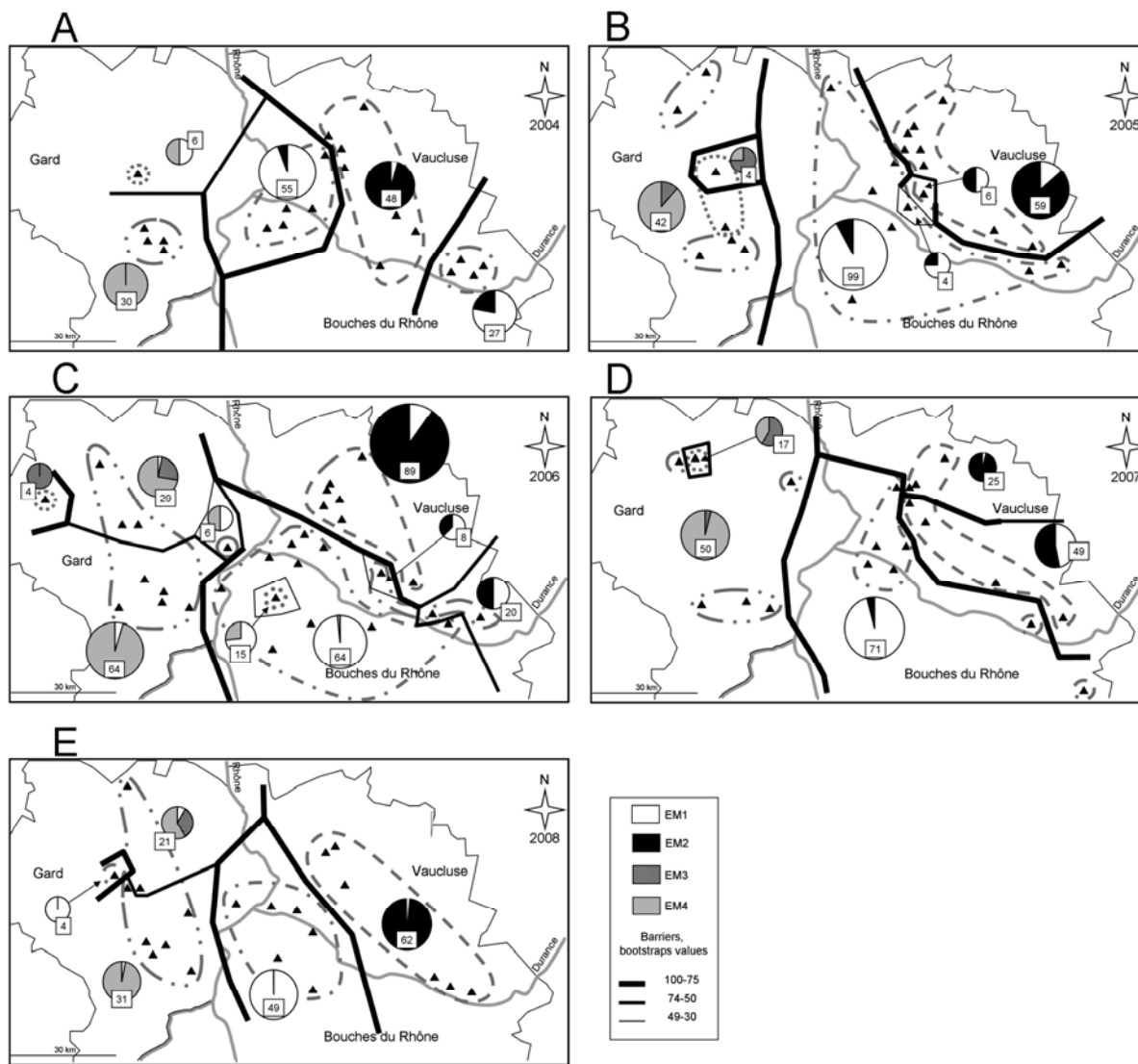
The sampling scheme used in this study provided us with a variable number of individuals per population (Table 1). Tests on the significance of genetic barriers usually do not address the topic of populations containing few individuals (36). Sampling bias can be very high in populations containing only two samples which can lead to artefactual barriers, and statistical analyses using bootstrap resamplings may put an excessive weight on these barriers. In our study, populations containing less than three samples were not included in the analysis, in order to limit the risk of such artefactual barriers devoid of any biological significance, even though this may induce some loss of information on WMV-EM distribution.

As observed between 2004 and 2007 (11), the genetic variability within WMV-EM subgroups remained low and their geographic distribution was rather stable in southeastern France in 2008. However, for the first time, a relatively high proportion (10%) of samples from Gard contained EM1 strains. EM1 isolates from Gard belonged to the most frequent EM1 haplotype

representing 80% of the isolates characterized so far in Vaucluse and Bouches-du-Rhône. Few EM1 isolates were previously detected in Gard between 2004 and 2007, usually in mixed infection with CL or EM4 isolates (Fig. 1). The year 2008 was relatively low in virus incidence in France (Lecoq, Wipf-Scheibel, and Desbiez, *unpublished data*), and infections occurred later than in previous years. EM1 strains introduced or already present in reservoirs in Gard may have met a higher proportion of uninfected plants than in the previous years, thus favoring their spread.

No isolation by distance was observed among EM isolates, whether using all sequences together or analyzing the sequences of each subgroup independently. The lack of correlation between geographic and genetic distances is consistent with the hypothesis that WMV-EM isolates did not arise from genetic differentiation after a single introduction, but from several introduction events in different locations.

Intra-subgroup variability was very low. For each EM subgroup, one major haplotype represented 70 to 90% of the isolates throughout the sampled area between 2004 and 2008, and intra-subgroup divergence did not exceed 1.2%. Divergent haplotypes—differing from the type strain by one or two mutations—were usually detected at only one place and 1 year during the surveys, and may represent mutants that appeared locally and



**Fig. 2.** Results of Monmonier's algorithm and spatial analysis of molecular variance (SAMOVA) analysis on *Watermelon mosaic virus* (WMV) samples from southeastern France for **A**, 2004, **B**, 2005, **C**, 2006, **D**, 2007, and **E**, 2008. Sampled points are spotted with triangles, the weight of Monmonier's barriers is represented by the thickness of black lines. SAMOVA clusters are represented by gray dotted lines. Circles are proportional to sample numbers (indicated in rectangles) for each area defined by Monmonier's algorithm.

were subsequently eliminated by genetic drift and/or negative selection. This high intra-subgroup homogeneity did not allow characterizing the genetic exchanges and transmission pathways between different locations within each subgroup. Increasing the genetic information available by using whole genome sequences probably would help to identify the origins of introduction, pathways of dissemination (6) and molecular evolution (19) of emerging WMV isolates.

Isolation by distance could have introduced biases in the spatial analyses using Monmonier's algorithm and SAMOVA (14,33), since it is very difficult to discriminate its effects from the presence of actual genetic breaks (36). Since no isolation by distance was detected, the reliability of the spatial analyses of barriers to gene flow was enhanced.

Among mechanisms limiting gene flow in populations, the presence of physical barriers is usually considered as a very important factor: Rieseberg et al. (41) reported that half of the studies published in *Molecular Ecology* related population genetic structure to the presence of a physical barrier. The present study associates Monmonier's algorithm and SAMOVA in order to detect barriers to gene flow in a plant RNA virus. These methods have a wide range of use, since Monmonier's algorithm does not necessarily require genetic data.

Using both approaches, robust barriers to gene flow were detected in WMV-EM populations. Similar results were obtained with Monmonier's algorithm and SAMOVA. A North-South barrier was detected each year with Monmonier's algorithm, roughly separating populations of Gard from those of Vaucluse/Bouches-du-Rhône, and SAMOVA also attributed to different clusters populations from Gard and Vaucluse/Bouches-du-Rhône. A second barrier observed each year was oriented from northwest to southeast within Vaucluse. SAMOVA distinguished two clusters in the Vaucluse/Bouches-du-Rhône populations, following the barrier found with Monmonier's algorithm. Furthermore, the  $F_{CT}$  values obtained with groups defined with Monmonier's algorithm were very similar to those obtained with SAMOVA (Table 1).

However, some differences were detected between the two methods. Monmonier's algorithm often separates populations containing different ratios of isolates from two EM-subgroups by a barrier, like in Gard and Vaucluse in 2005 (Fig. 2B) and in Vaucluse in 2007 (Fig. 2D). But with our criteria, SAMOVA made this distinction only in 2006 in the east Vaucluse (Fig. 2C). Contrary to Monmonier's algorithm, SAMOVA allows the detection of fragmented clusters of populations. Monmonier's algorithm split adjacent populations; if populations belonging to the same cluster are not connected, only SAMOVA analysis will cluster these populations together (Fig. 2A, C, and E).

There is a distinct zone in eastern and central Vaucluse where EM1 and EM2 strains were observed in the same fields and occasionally in the same plants. Monmonier's algorithm and SAMOVA handled this zone differently. SAMOVA included this zone in one of the major clusters in 2005 and 2007 (Fig. 2B and D), whereas Monmonier's algorithm separated these populations from others with a barrier (Fig. 2B and D). The distinction of this zone may result partially from sampling bias, but can also correspond to zones where the major barriers are more "permeable" and where exchanges between populations can take place. The coexistence of strains EM1 and EM2 for several years in the same area seems to indicate that they do not differ very much in relative fitness. In contrast during the same period, competition between EM and CL isolates has led to the quasi-disappearance of CL isolates in areas where both types of strains were present in southeastern France (11).

The stability in the spatial distribution of WMV-EM strains could be explained at least partially by landscape barriers to gene flow between populations. The North-South barrier approximately follows the Rhône River that separates Gard from Vaucluse/Bouches-du-Rhône. This barrier also follows the

dominant wind in the area (Mistral, a strong North wind blowing in the Rhône valley). Besides, in the east of Gard close to the Rhône River, there is an area circa 20 km wide covered mostly by grapevine or fruit trees and garrigue, with very few cucurbit or vegetable crops. The second barrier was not correlated with any obvious geographical or environmental pattern: no mountains or rivers are present in the area, the dominant wind is oriented North-South and the density of susceptible hosts is rather high during the growing seasons in all this area.

The distribution of WMV, as for other plant viruses, is mainly determined by hosts (and vectors) number and density (23). Host abundance is a major factor in disease epidemics. It has an influence on inoculum sources, pathogen refuge and virus life cycle (38). The low density of susceptible hosts in the East of Gard may have limited the dissemination of WMV. However, even though several landscape elements contribute to the North-South barrier, it must have been crossed occasionally since EM1 isolates were detected repeatedly in Gard and EM3 or EM4 isolates were very occasionally present in Vaucluse and Bouches-du-Rhône at least in 2006 and 2007. Whether the crossing occurred through flights of viruliferous aphids or plant/seedlings transfers has not yet been established. Although EM1 isolates were detected in Gard since 2004, their frequency remained very low until 2008, suggestive of strong competition with the well-established EM4 strains.

Besides geographical barriers, molecular and biological mechanisms may play a role in limiting gene flow between WMV populations. Cross-protection, i.e., the "protection" of a plant infected by one virus strain from overinfection with another strain of the same virus, is a mechanism enhancing genetic isolation between different closely related viral strains (21). In the samples collected, some mixed infections of one plant by two different EM subgroups were detected. This seems to indicate that cross-protection is not complete between EM subgroups, as also suggested by experiments in laboratory conditions (data not shown). This incomplete cross-protection can limit but does not completely preclude the dissemination of an EM isolate in crops where another EM subgroup is already present. However this mechanism may contribute to, and partially explain, the observed structure of WMV-EM populations.

This study is the first one combining a clustering and a tessellation method based on geographical and epidemiological data to identify the spatiotemporal dynamic of a RNA plant virus population structure. The results show that obvious geographical elements can not explain completely the geographic structure of WMV-EM populations. Only one stable barrier to gene flow was correlated with a geographic pattern (Rhône River), and even this barrier was not completely "virus-proof" since evidence for crossing(s) on both ways were observed. The other barriers, although detected consistently during the 5 years, were not correlated with obvious landscape characteristics but molecular and biological mechanisms can contribute together to make barriers more robust. Although barriers to gene flow were occasionally crossed by viruses, the populations did not tend to homogenize over the 5-year survey. This suggests that the average distance of virus dissemination is low—below a few kilometers. Other mechanisms, probably involving competition between minor variants and preestablished populations—either for infecting crops, overwintering in weeds, transmission by aphid vectors...—may impose genetic barriers to gene flow and may also be involved in the maintenance of highly structured populations of WMV in France.

## ACKNOWLEDGMENTS

This work was partly funded by the Région Provence-Alpes-Côte-d'Azur and the Institut National de la Recherche Agronomique as part of a co-financed Ph.D. thesis. We thank C. Wipf-Scheibel and C. Chandeysson for their excellent technical assistance.

## LITERATURE CITED

- Abubakar, Z., Ali, F., Pinel, A., Traore, O., N'Guessan, P., Notteghem, J. L., Kimmins, F., Konate, G., and Fargette, D. 2003. Phylogeography of Rice yellow mottle virus in Africa. *J. Gen. Virol.* 84:733-743.
- Archie, E. A., Luikart, G., and Ezenwa, V. O. 2009. Infecting epidemiology with genetics: A new frontier in disease ecology. *Trends Ecol. Evol.* 24:21-30.
- Bagnoli, F., Vendramin, G. G., Buonamici, A., Doulis, A. G., Gonzalez-Martinez, S. C., La Porta, N., Magri, D., Raddi, P., Sebastiani, F., and Fineschi, S. 2009. Is *Cupressus sempervirens* native in Italy? An answer from genetic and palaeobotanical data. *Mol. Ecol.* 18:2276-2286.
- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., and Bonhomme, F. 1996-2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France.
- Brassel, K. E., and Reif, D. 1979. Procedure to generate Thiessen polygons. *Geogr. Anal.* 11:289-303.
- Cottam, E. M., Thebaud, G., Wadsworth, J., Gloster, J., Mansley, L., Paton, D. J., King, D. P., and Haydon, D. T. 2008. Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proc. R. Soc. B-Biol. Sci.* 275:887-895.
- Dallot, S., Gottwald, T., Labonne, G., and Quiot, J. B. 2003. Spatial pattern analysis of sharka disease (*Plum pox virus* strain M) in peach orchards of southern France. *Phytopathology* 93:1543-1552.
- Dallot, S., Gottwald, T., Labonne, G., and Quiot, J. B. 2004. Factors affecting the spread of *Plum pox virus* strain M in peach orchards subjected to roguing in France. *Phytopathology* 94:1390-1398.
- de Meeûs, T., McCoy, K. D., Prugnolle, F., Chevillon, C., Durand, P., Hurtrez-Boussès, S., and Renaud, F. 2007. Population genetics and molecular epidemiology or how to "débusquer la bête". *Inf. Gen. Evol.* 7:308-332.
- Desbiez, C., Costa, C., Wipf-Scheibel, C., Girard, M., and Lecoq, H. 2007. Serological and molecular variability of watermelon mosaic virus (genus *Potyvirus*). *Arch. Virol.* 152:775-781.
- Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C., and Lecoq, H. 2009. Emergence of new strains of Watermelon mosaic virus in South-eastern France: Evidence for limited spread but rapid local population shift. *Virus Res.* 141:201-208.
- Desbiez, C., and Lecoq, H. 2008. Evidence for multiple intraspecific recombinants in natural populations of *Watermelon mosaic virus* (WMV, *Potyvirus*). *Arch. Virol.* 153:1749-1754.
- Dixo, M., Metzger, J. P., Morgante, J. S., and Zamudio, K. R. 2009. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. *Biol. Conserv.* 142:1560-1569.
- Dupanloup, I., Schneider, S., and Excoffier, L. 2002. A simulated annealing approach to define the genetic structure of populations. *Mol. Ecol.* 11:2571-2581.
- Dzialuk, A., Muchewicz, E., Boratynski, A., Montserrat, J. M., Boratynska, K., and Burczyk, J. 2009. Genetic variation of *Pinus uncinata* (Pinaceae) in the Pyrenees determined with cpSSR markers. *Plant Syst. Evol.* 277:197-205.
- Excoffier, L., and Heckel, G. 2006. Computer programs for population genetics data analysis: A survival guide. *Nat. Rev. Genet.* 7:745-758.
- Excoffier, L., Laval, G., and Schneider, S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.
- Excoffier, L., Smouse, P. E., and Quattro, J. M. 1992. Analysis of molecular variance Inferred from metric distances among DNA haplotypes—Application to human mitochondrial-DNA restriction data. *Genetics* 131:479-491.
- Garcia-Andres, S., Accotto, G. P., Navas-Castillo, J., and Moriones, E. 2007. Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean Basin. *Virology* 359:302-312.
- Guillot, G., Leblois, R., Coulon, A., and Frantz, A. C. 2009. Statistical methods in spatial genetics. *Mol. Ecol.* 18:4734-4756.
- Hall, J. S., French, R., Hein, G. L., Morris, T. J., and Stenger, D. C. 2001. Three distinct mechanisms facilitate genetic isolation of sympatric *Wheat streak mosaic virus* lineages. *Virology* 282:230-236.
- Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95-98.
- Holmes, E. C. 2004. The phylogeography of human viruses. *Mol. Ecol.* 13:745-756.
- Jensen, J. L., Bohonak, A. J., and Kelley, S. T. 2005. Isolation by distance, web service. *BMC Genetics* 6:13.
- Jombart, T., Pontier, D., and Dufour, A. B. 2009. Genetic markers in the playground of multivariate analysis. *Heredity* 102:330-341.
- Karlsson, S., Saillant, E., and Gold, J. R. 2009. Population structure and genetic variation of lane snapper (*Lutjanus synagris*) in the northern Gulf of Mexico. *Marine Biol.* 156:1841-1855.
- Kayser, M., Lao, O., Anslinger, K., Augustin, C., Bargel, G., Edelmann, J., Elias, S., Heinrich, M., Henke, J., Henke, L., Hohoff, C., Illing, A., Jonkisz, A., Kuzniar, P., Lebioda, A., Lessig, R., Lewicki, S., Maciejewska, A., Monies, D., Pawlowski, R., Poetsch, M., Schmid, D., Schmidt, U., Schneider, P., Stradmann-Bellinghausen, B., Szibor, R., Wegener, R., Wozniak, M., Zoledziewska, M., Roewer, L., Dobosz, T., and Ploski, R. 2005. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum. Genet.* 117:428-443.
- Kumar, S., Dudley, J., Nei, M., and Tamura, K. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* 9.
- Lecoq, H. 1992. Les virus des cultures de melons et de courgettes de plein champ (II). *PHM Rev. Hortic.* 324:15-25.
- Lecoq, H., and Desbiez, C. 2008. Watermelon mosaic virus and *Zucchini yellow mosaic virus*. Pages 433-440 in: *Encyclopedia of Virology*, 3rd. ed. Vol. 5. B. W. J. Mahy and M. H. V. Van Regenmortel, eds. Elsevier, Oxford.
- Lecoq, H., Wipf-Scheibel, C., Chandeysson, C., and Desbiez, C. 2008. Results of a survey for cucurbit viruses in France and new insights on *Watermelon mosaic virus* and *Zucchini yellow mosaic virus* epidemiology. Pages 24 in: *Proceedings of the 3rd Conference of the International Working Group on Legume and Vegetable Viruses*. Ljubjana, Slovenia.
- Manel, S., Schwartz, M. K., Luikart, G., and Taberlet, P. 2003. Landscape genetics: Combining landscape ecology and population genetics. *Trend Ecol. Evol.* 18:189-197.
- Manni, F., Guérard, E., and Heyer, E. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Hum. Biol.* 76:173-190.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209-220.
- Monmonier, M. S. 1973. Maximum-difference barriers: An alternative numerical regionalization method. *Geogr. Anal.* 5:245-261.
- Paquette, S. R., and Lapointe, F.-J. 2009. A statistical procedure to assess the significance level of barriers to gene flow. *J. Genet. Genomics* 36:685-693.
- Perez-Losada, M., Porter, M. L., Tazi, L., and Crandall, K. A. 2007. New methods for inferring population dynamics from microbial sequences. *Inf. Genet. Evol.* 7:24-43.
- Plantegenest, M., Le May, C., and Fabre, F. 2007. Landscape epidemiology of plant diseases. *J. R. Soc. Interface* 4:963-972.
- Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Real, L. A., and Biek, R. 2007. Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *J. R. Soc. Interface* 4:935-948.
- Rieseberg, L., Vines, T., and Kane, N. 2009. Molecular ecology. Editorial and retrospective 2008. *Mol. Ecol.* 18:1-13.
- Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Dezzani, R., Delmelle, E., Vierling, L., and Waits, L. P. 2007. Putting the "landscape" in landscape genetics. *Heredity* 98:128-142.
- Szmaragd, C., and Balloux, F. 2007. The population genomics of hepatitis B virus. *Mol. Ecol.* 16:4747-4758.
- Team R Development Core. 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Computing, ed., Vienna, Austria.
- Thiel-Egenter, C., Holderegger, R., Brodbeck, S., and Gugerli, F. 2009. Concordant genetic breaks, identified by combining clustering and tessellation methods, in two co-distributed alpine plant species. *Mol. Ecol.* 18:4495-4507.
- Tomitaka, Y., and Ohshima, K. 2006. A phylogeographical study of the *Turnip mosaic virus* population in East Asia reveals an 'emergent' lineage in Japan. *Mol. Ecol.* 15:4437-4457.
- Traore, O., Pinel-Galzi, A., Sorho, F., Sarra, S., Rakotomalala, M., Sangu, E., Kanyeka, Z., Sere, Y., Konate, G., and Fargette, D. 2009. A reassessment of the epidemiology of *Rice yellow mottle virus* following recent advances in field and molecular studies. *Virus Res.* 141:258-267.
- Vonlanthen, P., Excoffier, L., Bittner, D., Persat, H., Neuenschwander, S., and Largiadèr, C. R. 2007. Genetic analysis of potential postglacial watershed crossings in Central Europe by the bullhead (*Cottus gobio* L.). *Mol. Ecol.* 16:4572-4584.
- Weir, B. S. 1990. *Genetic Data Analysis: Methods for Discrete Population Analysis*. Sinauer Associates, Sunderland, MA.
- Wright, S. 1943. Isolation by distance. *Genetics* 28:139-156.
- Xia, X., and Xie, Z. 2001. DAMBE: Data analysis in molecular biology and evolution. *J. Hered.* 92:371-373.