Recombination is known to be an important driving force in plant virus evolution (García-Arenal et al., 2001; Nagy, 2008). Some plant viruses have attracted particular interest in this connection, the family Geminiviridae, comprising viruses with small DNA genomes, being one of the most studied. Within the genus Begomovirus, several species have been described that clearly result from recombination events (Fauquet et al., 2005; García-Andrés et al., 2007b; Monci et al., 2002; Padidam et al., 1999; Zhou et al., 1997).

Tomato yellow leaf curl virus (TYLCV) and Tomato yellow leaf curl Sardinia virus (TYLCSV) are two begomoviruses that, throughout the past two decades, have spread to most Mediterranean countries, and cause serious problems particularly to tomato crops: in Italy, Spain and Portugal they are frequently detected, with some tomato plants co-infected by both viruses. The presence of both viruses in the same nuclei, shown by Morilla et al. (2004), constitutes a favourable environment for recombination.

In Italy, TYLCSV and TYLCV are known to have co-existed since 2002 (Accotto et al., 2003), when TYLCV was detected in Sicily, where TYLCSV had been present at least since 1989. TYLCV quickly invaded the area colonized by TYLCSV, and its incidence, often in mixed infections with TYLCSV, became important (Davino et al., 2006). Intriguingly, when some years later the same samples were reanalysed, recombinant molecules were detected in field tomatoes collected in 2002 (Davino et al., 2008), indicating that recombination actually occurred almost immediately after the second virus arrived. To clarify this question, we here examine recombination events, in natural field conditions and in experimental infections, and compare distribution and location of recombination breakpoints along the viral genomes.

Over 4 years (2006–2009) 100 samples per year were collected in Ragusa Province (Sicily) from tomato-protected crops showing yellow leaf curl symptoms. Total DNA was extracted from 0.2 g of tissue as described elsewhere (Noris et al., 1994) and resuspended in 500 μl of TE buffer. Multiplex PCR (Davino et al., 2008) was used for amplification, followed by digestion with Psp1406I (AclI) restriction enzyme (Fermentas) and electrophoresis on
2.5 % agarose gels. Control reactions used artificial mixtures of DNA extracted from plants infected with TYLCV and TYLCV (not shown) to rule out the possibility that recombinants detected in field samples could be artefacts generated during PCR.

The pattern obtained (Fig. 1) indicated no evidence of displacement of one viral species by the other, and showed that TYLCSV-, TYLCV- and recombinant-type molecules were all detectable even in 2009, 7 years after the change in virus population due to the spread of TYLCV. It is noteworthy that, throughout the period considered, the most common pattern was co-infection with molecules of the three kinds, and that recombinant molecules were never found alone in any plant. Furthermore, each parent virus maintained its own share in single infections: the overall pattern appeared stable in a sort of equilibrium, neither virus species nor recombinants showing selective advantage. In our survey all recombinants had TYLCSV sequences in the 3’ half of the intergenic region (IR), or in ORF V2, and TYLCV sequences in the 5’ half of the IR, as reported in other studies (Davino et al., 2008, 2009; García-Andrés et al., 2007a).

Twenty samples were selected randomly from those in which recombination events had been detected (five from each year) and used to specifically amplify a 2260 nt long genomic segment from the recombinant molecules, representing more than 80 % of the viral genome. Amplifications were run using a primer designed on the TYLCSV-Sar-IT:Sar:88 genomic segment from the recombinant molecules, representing more than 80 % of the viral genome. Amplifications were run using a primer designed on the TYLCSV-Sar-IT:Sar:88 (GenBank accession no. X61153, Kheyr-Pour et al., 1991) and TYLCSV-IL-IT:Sic:04 (GenBank accession no. DQ144621, Davino et al., 2006). Beginning 10 days post-inoculation (p.i.), plants were checked for systemic infection by the two viruses and by possible recombinants, using the multiplex PCR approach described above. At 24 days p.i., TYLCSV and TYLCV were detected in young leaves of all plants, and at 45 days p.i., recombination events were detected in three of ten plants. At 60 days p.i., when TYLCSV, TYLCV and recombinants were detected in all plants, young tissue was used to extract DNA and amplify recombinant fragments, as described above, with primers TY224 + and TY2482 –; 50 clones of the amplified DNAs (five from each plant) were sequenced. The distribution of breakpoints showed a recombination pattern having similarities and differences with field samples (Fig. 2a). No recombination was detected in the CP (V1) ORF, while a hotspot was mapped at the 3’ terminus of ORF C3, with a peak at position 1157, followed by a region without breakpoints until position 1518 overlapping between C2/C3 ORFs. This cold-spot was not detected in field samples. Following this region, numerous (44 events) and diverse breakpoints were detected until position 1992. The final portion examined (1992–2482) displayed few breakpoints, located near position 2390. The finding that most breakpoints map at antisense ORFs supports the speculation that, as a consequence of the rolling-circle replication mechanism and the existence of sense and antisense ORFs, clashes between replication and transcription complexes may occur (Lefevre et al., 2009).

The local degree of sequence similarity is one of the factors that impacts on the efficiency of homologous recombination (Baird et al., 2006). We analysed this feature in our samples, and found that in field conditions (Fig. 2b) 75 % of breakpoints were located next to short common stretches (3–8 nt) and only 5.5 % next to long stretches (19 nt or more). In laboratory conditions, 54.2 % of breakpoints were located next to short common stretches (3–8 nt) and 16.7 % next to long stretches (19 or more). We are unable to offer any reasonable explanation for these results, but a preference for short common sequences in breakpoints was also detected.
by Martin et al. (2011b) in artificial co-infections; in their case the most frequent length was 5–12 nt, which compares relatively well with our laboratory data. Whether the relative abundance of short common stretches in field versus laboratory experiments is due to selection deserves future study.

Very few laboratory experiments on induction and analysis of recombination have been reported in monopartite begomoviruses. Co-inoculation in tomato of TYLCSV-ES[ES:Mur1:92] and TYLCV-Mld[ES:72:97], two different strains of the same virus species used in this study, lead to the detection of recombinants after 130 days (with recombination breakpoints being analysed at 400 days p.i.) (García-Andrés et al., 2007b), while in our study recombinants were first detected at 45 days p.i., and analysis was at 60 days p.i. García-Andrés et al. (2007b) detected breakpoints only in the central portion of the genome, approximately 700 nt between the 3’-end of ORF V1 and the 3’-proximal portion of ORF C1. The hotspot found in the region where ORFs V1 and C3 co-terminate, corresponding to the termini of the two transcription units of the genome, was also present in our field and laboratory samples, and is probably the most frequent recombination hotspot in begomoviruses, apart from that in the stem–loop containing the origin of replication (Lefeuvre et al., 2009). Another analysis of breakpoints resulting from artificial co-infections with tomato begomoviruses was recently reported (Martin et al., 2011b); the two viruses co-inoculated in tomato plants were TYLCV-Mld[RE:02] and Tomato leaf curl Comoros virus – Mayotte (ToLCKMV-YT[YT:Dem:03]) sharing about 82 % of sequence similarity. Plants were analysed 4 months after inoculation and, apart from the usual hotspot around the origin of replication, only one other hotspot was found, situated within ORF C4; this hotspot was not present in our case. Curiously, the

Fig. 2. (a) Hotspot profiles of recombinant TYLCSV/TYLCV sequences obtained in laboratory and field conditions, with genomic organization represented below. Each rod represents a recombination breakpoint; multiple breakpoints at the same genomic position are piled. The grey rectangle delimits the genomic region studied. The coordinates of breakpoints on the genome of TYLCSV-Sar[IT:Sar:88] are in Table S1. (b) Correlation between the length of common nucleotide stretches between TYLCSV and TYLCV, and number of independent breakpoints detected (vertical axis, per cent).
hotspot found by García-Andrés et al. (2007b) and ourselves in the region where the two ORFs V1 and C3 co-terminate, was considered a cold-spot in Martin’s study. To our knowledge, this is the first attempt to compare recombination profiles in natural and artificial co-infections of two geminiviruses, and this deserves some consideration. If new recombinants arise at sites where the two parents share a short stretch of nucleotides, as short as 3 nt in our case, then one would expect a large number of different breakpoints in laboratory specimens, where selection had little time to act. In parallel, when studying field samples, one would expect to detect breakpoints mostly in molecules that have maintained (or improved) their fitness. However, although we detected hotspots in both conditions, some genomic regions showed no breakpoints in the laboratory samples, while some were found in the field samples in ORF CP and in the C2/C3 overlapping region (Fig. 2a). One should then conclude that the number of recombination events that can be generated is extremely high, provided time is sufficient (see field data, from samples collected during 4 years), and that most of the resulting variants may continue to exist in an extremely diverse population of viral molecules, whether viable or defective. In the field, whitefly vectors can spread virus variants generated in a single plant to many others, with the result that one plant will eventually host virus variants generated elsewhere. Conversely, in laboratory experiments, where time (60 days) and plants (10) are limited, and vectors lacking, there is less chance of finding diversity. In this view, selection would presumably not play a major role in determining the breakpoints found. Finally, it must be remembered that what are detected are simply recombination breakpoints, not full, infectious, transmissible and fully viable viruses.

Recombination in RNA and DNA viruses, in addition to genetic mutations, and to genome reassortment in the case of multicomponent viruses, contributes greatly to virus genome variability (Martin et al., 2011a; Nagy, 2008). In the case of the two virus species examined, only four TYLCSV/TYLCV recombinant viruses have been extensively characterized from field samples thus far, two from Italy (Davino et al., 2009) and two from Spain (García-Andrés et al., 2006; Monci et al., 2002). Their ability to infect plants systemically, induce disease, and be transmitted by the natural whitefly vector, has been demonstrated using infectious clones. It should be stressed that, where the viability of TYLCV/TYLCV recombinants has been demonstrated, mapping of

---

**Fig. 3.** Phylogenetic network, generated by SplitsTree 4.12.6, of complete genomes of virus isolates causing the tomato yellow leaf curl disease constructed by split-decomposition analysis to visualize reticulated evolutionary relationships produced by recombination. Alignments were obtained using CLUSTAL W. Distance transformation was calculated with the neighbour-net algorithm. GenBank accession numbers of full-length genomes are reported in the figure. Isolates found in Italy are in bold.
breakpoints simply gives indications on genomic recombination sites, but does not predict which recombinant viruses will be viable and eventually spread in field conditions.

Despite fears of new epidemics involving the four well-characterized TYLCV/TYLCSV recombinants, their relevance appears limited thus far: only for Tomato yellow leaf curl Malaga virus (TYLCMalV) has an epidemic been reported, in the bean, in Almería (Spain) in 2000 (Monci et al., 2002). No recent data of field surveys highlighting the importance of TYLCV/TYLCSV recombinants in causing epidemics are available. This can be taken as an indication that the fitness of viable recombinant virus variants, generated continuously in mixed infections between similar begomoviruses, is not comparable to that of the parent viruses, at least not in the host plant where they are generated (Sánchez-Campos et al., 2002).

Our results also show that co-infection with two begomoviruses sharing a certain degree of similarity (above 70% in our case) results in the rapid emergence of a huge number of recombinant variants. Even were it possible to study most of them in depth, and remove defective ones from the list, the final results would still be a multitude of fully viable ‘entities’. How should they be classified and named? Are the criteria used thus far for species demarcation in begomoviruses (89% sequence similarity) still valid? Application of this criterion has resulted in the creation of new species (TYLCMalV in Monci et al., 2002; TYLCaxV in García-Andrés et al., 2006) for ‘entities’ that are obviously recombinants between two well-defined viruses. Also, application of the 89% threshold to the case of a multitude of recombinant viral entities deriving from two parents, A and B, would lead to some variants being named strains (or isolates, if very close) of parent species A, others of parent species B, and others as belonging to a different virus species. The resulting taxonomic chaos, with different names for very close viral entities, would not help studies on evolution and phylogeny and, moreover, such a classification of recombinant variants would not indicate their true origins. Since conventional phylogenetic trees do not provide a reliable picture of the evolutionary relationships between viruses in which recombination has played a major role, a more convenient ‘split tree’ has been used (Huson & Bryant, 2006). This representation highlights the origins of recombinant viruses, as observed for the two isolates from Italy: the two recombinant viruses described in Sicily (Davino et al., 2009), TYLCaxV-Sic1-[IT:Sic2/2:04] and TYLCaxV-Sic2-[IT:Sic2/5:04], are correctly positioned between parent viruses isolated in the same country (Fig. 3).

Since genetic variation occurs discontinuously along the genome, classification of recombinants is particularly challenging. The difficulty in demarcating begomovirus species when recombination creates new viral entities has been recognized (Fauquet et al., 2005), but unfortunately the International Committee on Taxonomy of Viruses has not yet found valid alternatives.

Acknowledgements

The authors wish to thank Dr Ricardo Flores (IBMCP-CSIC, Valencia, Spain) for critical reading of the manuscript. S. P. was supported by the Italian Ministry of Education (MIUR) with a PhD grant.

References


