PNAS

Supplementary Information for
Non-concomitant host-to-host transmission of multipartite virus genome segments may lead to complete genome reconstitution.

Jérémy Di Mattia, Babil Torralba, Michel Yvon, Jean-Louis Zeddam, Stéphane Blanc and Yannis Michalakis

Yannis Michalakis
Email: yannis.michalakis@ird.fr

This PDF file includes:

Supplementary Information Text

Defining qPCR Ct thresholds beyond which segments are considered absent through tolerance limits.

To reliably claim that full genome reconstitution has occurred, we must be able to ascertain when any specific segment can be considered absent from or present in an infected plant. To do this, for each segment, we performed qPCR using the focal segment's primers on infected plants lacking this segment, that is plants inoculated with all segments but the focal segment. We thus recorded the Ct's obtained for a focal segment when it is itself lacking from an infected plant but when all other segments were present. We then established a Ct distribution for each of the three "missing segments" of interest, C, N and U4 from a number of infected plants with two technical replicates for each sample (sampling and qPCR conditions were performed as described in the main text).

Based on these Ct distributions we calculated for each segment the (0.95, 0.95)-one-sided lower tolerance interval, i.e. the Ct value above which would lie 95% of future comparable samples with 95% confidence, given the observed distribution of Ct values. Based on this method, in the genome reconstitution experiments we considered a segment to be absent from a sample if its Ct value lied above the corresponding threshold.

In order to calculate these thresholds we used the method described by Francq et al. 2019 (1) for the one random factor design (in our case the random factor is the sampled plant, accounting for the fact that the two technical replicates originate from the same sample). We adapted the R script provided by the authors in the appendix of their paper by adjusting the quantiles of the Normal, t and X2 distributions to our desired 95% coverage and 95% confidence proportions, and the sample-size related parameters to our sample sizes for each segment.

The sample sizes to obtain these tolerance limit values were 36 infected plants (2 technical replicates each) for missing-segment N and 56 infected plants (2 technical replicates each) for missing-segment U4. In the case of U4, we did not consider one of the infected plants because its two technical replicates yielded Ct values around 24, and were identified as outliers (below the lower quartile – 1.5 the interquartile range), indicating contamination at some stage (the next Ct values were >26 and these were included in our estimation). Incorporating this sample would have yielded a threshold Ct value of 27.66, instead of 28.29. Including this discarded individual plant in establishing the threshold value changes the status, from containing to non-containing segment U4, of only three plants: one in the U4-||C- parallel transmission, one in the C-||U4- sequential transmission with one day of time spacing, and one in the U4-||N- sequential transmission with one day of time spacing treatments. Hence considering or not this individual does not affect our results and inferences.

As mentioned in the main text, it was very difficult to obtain infected plants without segment C. We only had Ct values from 7 such plants (with two replicates each). Because we did not want to base distributional properties on such a small sample size we also used Ct values from 15 uninfected plants (with two replicates each). We first ran a mixed model with infection or not as a fixed factor and sampled plant as random factor. This model showed that the infection status did not significantly affect the Ct values of the missing C segment ($p=0.68$), with very similar means (infected/non-infected: 35.13/34.75). Subsequently, we calculated the threshold Ct using the distribution from these 22 plants and performed all our analyses with this threshold value.

Fig. S1. Localization of FBNSV segments in aphid AMG cells. Viral DNA is labeled by FISH in the AMG of viruliferous aphids and observed by confocal microscopy. The green probe targets segment U4 and the red probes target either C or N in the respective panels. Each panel corresponds to the control experiment of C/U4 (light blue part) and N/U4 (light orange part) of figure 4. The “without C”, “without N” or “without U4” panels show midguts of aphids fed on plant lacking the corresponding segments and thus control for probe specificity. The “All segments” panels show the localization of C/U4 or N/U4 in midguts from aphid fed on plant containing all eight segments (concomitant acquisition of all segments). The accumulation of FBNSV DNA was similarly revealed in all observed cells (>10 cells per midgut) from 18, 10, 23, 10, 13 and 26 viruliferous aphids, respectively from top to bottom rows; A representative image of each case is shown to illustrate the results. All images correspond to single optical sections. Cell nuclei are stained with DAPI (4,6-diamidino-2-phenylindole; blue). The scale bar represents 25 µm.

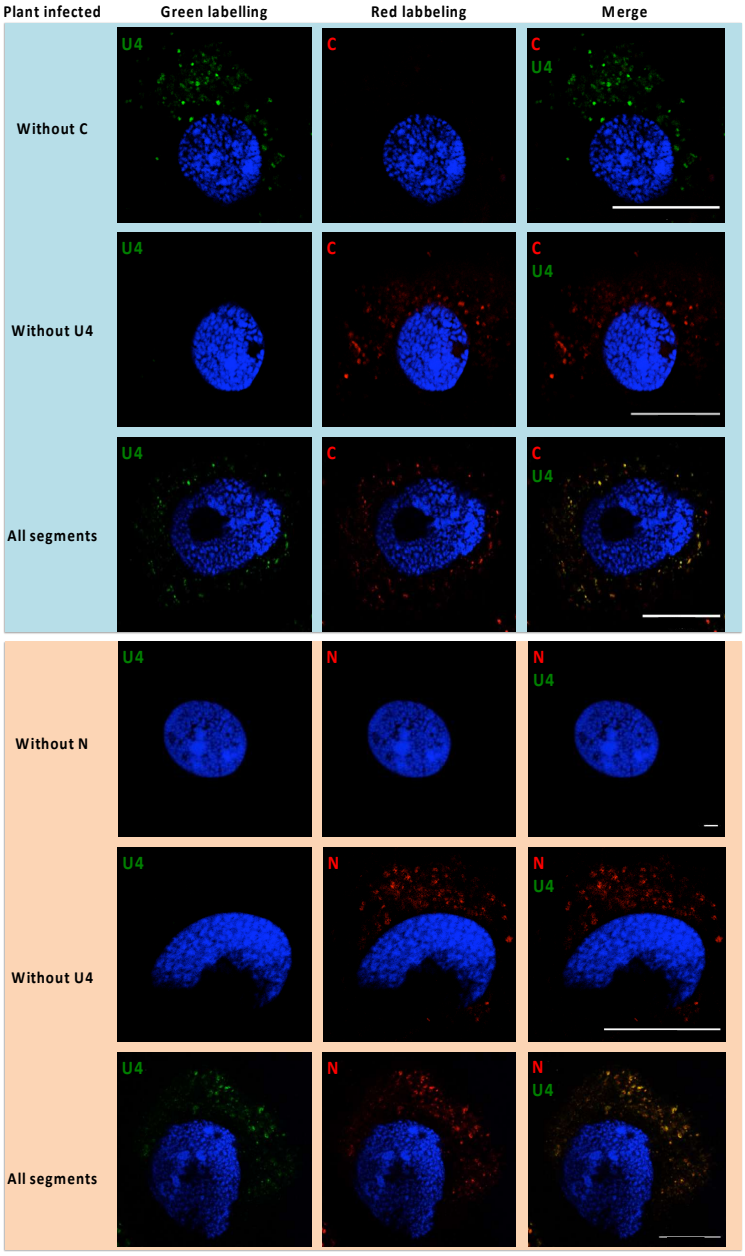

Table S1. Missing segments per spacing time. Number of inoculated (N) and infected (N_inf) plants, number of plants containing each segment (N_segment name) and percentage of plants containing each segment among infected plants (% segment name) as a function of spacing time (in days).

Spacing	N	N_inf	N_C	% C	N_N	% N	N_U4	% U4
0	170	65	51	78	62	95	49	75
1	172	51	37	73	50	98	26	51
2	174	28	24	86	27	96	10	36
3	172	13	6	46	13	100	5	38

Table S2. Missing segments per condition, replicate and spacing time. Number of inoculated (N) and infected (N_inf) plants, number of plants containing each segment (N_segment name) and percentage of plants containing each segment among infected plants (% segment name) as a function of spacing time (in days) per condition and replicate (rep).

spacing	condition	rep	N	N_inf	N_C	% C	N_N	% N	N_U4	% U4
0	U4-||C-	1	21	13	11	85	12	92	11	85
0	U4-||C-	2	20	8	7	88	8	100	4	50
0	C-||U4-	1	22	6	5	83	5	83	4	67
0	C-||U4-	2	21	10	7	70	10	100	7	70
0	U4-||N-	1	21	7	7	100	7	100	7	100
0	U4-||N-	2	23	12	6	50	12	100	8	67
0	N-||U4-	1	21	4	3	75	4	100	4	100
0	N-||U4-	2	21	5	5	100	4	80	4	80
1	U4-||C-	1	22	8	6	75	8	100	6	75
1	U4-||C-	2	20	7	6	86	6	86	3	43
1	C-||U4-	1	20	4	4	100	4	100	3	75
1	C-||U4-	2	24	10	6	60	10	100	7	70
1	U4-||N-	1	21	5	4	80	5	100	3	60
1	U4-||N-	2	24	14	8	57	14	100	3	21
1	N-||U4-	1	21	0						
1	N-||U4-	2	20	3	3	100	3	100	1	33
2	U4-||C-	1	23	6	5	83	6	100	2	33
2	U4-||C-	2	22	4	4	100	4	100	0	0
2	C-||U4-	1	23	1	1	100	1	100	0	0
2	C-||U4-	2	21	4	3	75	4	100	4	100
2	U4-||N-	1	20	1	1	100	0	0	0	0
2	U4-||N-	2	21	5	4	80	5	100	0	0
2	N-||U4-	1	22	6	5	83	6	100	3	50
2	N-||U4-	2	22	1	1	100	1	100	1	100
3	U4-||C-	1	21	0						
3	U4-||C-	2	24	5	2	40	5	100	3	60
3	C-||U4-	1	17	0						
3	C-||U4-	2	20	2	0	0	2	100	2	100
3	U4-||N-	1	23	0						
3	U4-||N-	2	22	3	3	100	3	100	0	0
3	N-||U4-	1	23	2	1	50	2	100	0	0
3	N-||U4-	2	22	1	0	0	1	100	0	0

SI References

1. B. G. Francq, D. Lin, W. Hoyer, Confidence, prediction, and tolerance in linear mixed models. *Statistics in Medicine* **38**, 5603–5622 (2019).