

Comment on "Large Bottleneck Size in Cauliflower Mosaic Virus Populations during Host Plant Colonization" by Monsion et al. (2008)

Gael Thébaud, Yannis Michalakis

▶ To cite this version:

Gael Thébaud, Yannis Michalakis. Comment on "Large Bottleneck Size in Cauliflower Mosaic Virus Populations during Host Plant Colonization" by Monsion et al. (2008). PLoS Pathogens, 2016, 12 (4), pp.e1005512. 10.1371/journal.ppat.1005512 . hal-02003715

HAL Id: hal-02003715 https://hal.umontpellier.fr/hal-02003715v1

Submitted on 1 Feb 2019

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.



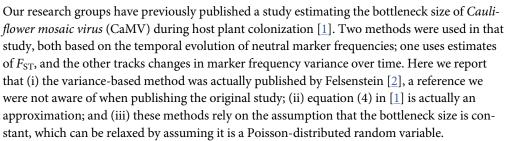
FORMAL COMMENT

Comment on "Large Bottleneck Size in Cauliflower Mosaic Virus Populations during Host Plant Colonization" by Monsion et al. (2008)

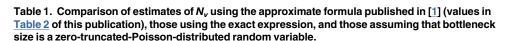
Gaël Thébaud¹*, Yannis Michalakis²

1 INRA, UMR 385 BGPI, Montpellier, France, 2 MIVEGEC, UMR 5290 CNRS IRD Université de Montpellier, Montpellier, France

^{*} thebaud@supagro.inra.fr



Equation (4) in [1] reads $N_v = p(1-p)/[\operatorname{Var}(p') - \operatorname{Var}(p)]$, where N_v is the estimated bottleneck size, and p and p' represent the frequency of a neutral marker before and after the bottleneck, respectively. The approximation in this equation concerns the numerator. Its exact formulation should be E(p(1-p)) and not E(p)(1-E(p)), which is the expression implied in [1] and used in the numerical application on CaMV (E denotes the expected value). Based on the definition of the variance, it can be shown that E(p(1-p)) = E(p)(1-E(p)) - Var(p), which is equivalent to the numerator of equation (14) in [2]. Thus, the approximation will give relatively accurate results whenever Var(p) is negligible relative to E(p)(1-E(p)). This was indeed the case in the study by Monsion et al. [1], as can be seen in Table 1, which compares the N_v values published in [1] to those obtained when applying the exact expression. The results of that paper are thus not qualitatively affected and, consequently, its conclusions remain unchanged. This method was also used in another study [3], the results of which are also marginally affected quantitatively, but not at all qualitatively (Table 2). To illustrate the application of these methods in a situation in which the estimated bottleneck size is lower, we used data from Tromas et al. (Table 3) [4].



Marker	N_{ν} approximate	N_{ν} exact	n _{vP}
VIT1	423.00	421.16	422.16
VIT3	298.45	296.31	297.31
VIT4	484.61	481.42	482.42

doi:10.1371/journal.ppat.1005512.t001





Citation: Thébaud G, Michalakis Y (2016) Comment on "Large Bottleneck Size in Cauliflower Mosaic Virus Populations during Host Plant Colonization" by Monsion et al. (2008). PLoS Pathog 12(4): e1005512. doi:10.1371/journal.ppat.1005512

Editor: Raul Andino, University of California San Francisco, UNITED STATES

Received: November 12, 2015
Accepted: February 29, 2016

Published: April 14, 2016

Copyright: © 2016 Thébaud, Michalakis. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors received no specific funding for this work

Competing Interests: The authors have declared that no competing interests exist.

Table 2. Comparison of estimates of N_v published in [3] using the approximate formula (values in table S1 of [3]), those using the exact expression, and those assuming that bottleneck size is a zero-truncated-Poisson-distributed random variable. Note that due to copying errors, the values reported in table S1 of [3] for leaves 5 and 21 were slightly wrong—we report here the correct values.

Leaf level	N _ν approximate	N _ν exact	n _{vP}
# 5	8.78	8.63	9.79
# 16	127.70	116.98	117.99
# 21	15.45	13.90	14.99

doi:10.1371/journal.ppat.1005512.t002

Both Felsenstein's formula [2] and its approximation assume that the bottleneck size N is constant (i.e., identical in all the sampled plants). However, we can make the more realistic assumption that N follows a zero-truncated Poisson (ZTP) distribution (i.e., N can vary among experimental replicates according to a Poisson distribution, but p' cannot be measured when N = 0 and the plant is discarded, hence the zero-truncation):

$$N \sim \text{ZTP}(n_{vP}), \text{ so } \forall k \in N^*, P(N=k) = \frac{n_{vP}^k}{k!(e^{n_{vP}}-1)}.$$

Among the N genomes that go through the bottleneck, the number X bearing the neutral marker is drawn according to its pre-bottleneck frequency p from the binomial distribution: $X \sim B(N,p)$. Thus, the marker frequency after the bottleneck is p' = X/N. The variance of this marker frequency can be expressed as: Var(p') = Var[E(p'|N,p)] + E[Var(p'|N,p)].

Because E(p'|N,p) = E(X|N,p)/N and $Var(p'|N,p) = Var(X|N,p)/N^2$, and because N and p are independent, we then get: $Var(p') = Var(p) + E[p(1-p)] \times E(1/N)$.

Using the properties
$$E(1/N) = \sum_{k=1}^{\infty} \left(\frac{1}{k}P(N=k)\right)$$
 and $\sum_{k=1}^{\infty} \frac{n_{vP}^k}{k!k} = \text{Ei}(n_{vP}) - \ln(n_{vP}) - \gamma$,

where Ei = $\int_{-\infty}^{x} \frac{e^{t}}{t} dt$ is the exponential integral function and γ the Euler-Mascheroni constant, one can estimate the bottleneck size $n_{\nu P}$ by numerically solving the following equation:

$$\frac{E(p)(1-E(p))-Var(p)}{Var(p')-Var(p)} = \frac{e^{n_{vp}}-1}{\operatorname{Ei}(n_{vp})-\ln(n_{vp})-\gamma}.$$

We recommend using this estimation procedure when N_{ν} values are below 20, i.e., when the relative estimation error is higher than 5%. The following lines of R code may be used to estimate n_{ν} :

```
library(gsl)
EulerMaschCst<-(-digamma(1))
p<-c(...,...) # vector of initial frequencies
p_prime<-c(...,....) # vector of final frequencies
funestim<-function(n_vP) (mean(p)*(1-mean(p))-var(p))/(var(p_prime)
-var(p)) -
(exp(n_vP)-1)/(expint_Ei(n_vP)-log(n_vP)-EulerMaschCst)
uniroot(funestim,interval = c(1e-6,600))</pre>
```

Table 3. Estimates of N_{v} **using data published in [4] for leaf level 5.** The authors of that paper had estimated a bottleneck size of 6 (see their figure 5B) using the F_{ST} method. To obtain the reported estimates, Leaf 3 and Leaf 5 were considered as the levels before and after the bottleneck, respectively. From the histogram 4A in [4], we calculated a mean frequency of 0.622 for the marker in Leaf 3, and we used the variances in marker frequency reported for Leaf 3 and Leaf 5.

Leaf level	N_{ν} approximate	N_{ν} exact	n _{vP}
# 5	6.91	6.83	8.02

doi:10.1371/journal.ppat.1005512.t003



References

- Monsion B, Froissart R, Michalakis Y, Blanc S. Large bottleneck size in Cauliflower mosaic virus populations during host plant colonization. PLoS Pathog. 2008; 4: e1000174. doi: 10.1371/journal.ppat. 1000174 PMID: 18846207
- Felsenstein J. Inbreeding and variance effective numbers in populations with overlapping generations. Genetics. 1971; 68: 581–597. PMID: 5166069
- 3. Gutiérrez S, Yvon M, Pirolles E, Garzo E, Fereres A, Michalakis Y, et al. Circulating virus load determines the size of bottlenecks in viral populations progressing within a host. PLoS Pathog. 2012; 8: e1003009. doi: 10.1371/journal.ppat.1003009 PMID: 23133389
- Tromas N, Zwart MP, Lafforgue G, Elena SF. Within-host spatiotemporal dynamics of plant virus infection at the cellular level. PLoS Genet. 2014; 10: e1004186. doi: 10.1371/journal.pgen.1004186 PMID: 24586207