PREreview of bioRxiv article "Genome-wide increased copy number is associated with emergence of super-fit clones of the Irish potato famine pathogen *Phytophthora infestans*"

Sophien Kamoun[1]

[1]The Sainsbury Laboratory

April 28, 2020

Abstract

This is a review of Knaus et al. bioRxiv doi: https://doi.org/10.1101/633701 posted on May 16, 2019. In this paper, the authors studied variations in ploidy in a wide range of isolates of the potato blight pathogen *Phytophthora infestans*.

Summary

This study examines variations in ploidy in a wide range of isolates of the potato blight pathogen Phytophthora infestans. The main addition is a sample of isolates from Mexico, the presumed center of origin of this pathogen. The main finding is the preponderance of triploid strains among clonal isolates and elevated levels of copy number variation (CNV) and gene deletions. Similar studies have been reported before for P. infestans populations and clonal lineages although with more limited samples. They may want to further emphasize the novelty of the study and the findings.

Major comments

* The model of clonal lineage emergence proposed by the authors does not consider the possibility that sexual reproduction might also be occurring outside Central Mexico. This is a serious issue throughout the paper as there is an assumption that sexual reproduction is restricted to Mexico despite a rich literature that indicates the contrary ever since the work of Drenth and other in the 1990s (see also (Li et al., 2012)). Note also this statement in (Fry et al., 2015): "In northern Europe, there is now convincing evidence that there are residential sexual populations of *P. infestans*, particularly in the Nordic countries (Yuen and Andersson 2013)."

* In the absence of experimental data, it is inappropriate to refer to the examined clonal lineages as "super-fit" clones. A more appropriate term would be "invasive" or "dominant" clones. Note that the reasons behind the increase in population frequency of particular clones may have nothing to do with their fitness per se but rather due to other factors, such as spread through nurseries or introduction to new geographical regions.

* Fitness is a relative term. A clone that is fit in one year may not be fit in another. Thus I'm disappointed by the loose use of the term here. See for example (Cooke et al., 2012) experiments on "fitness" which showed that aggressiveness of *P. infestans* 13_A2 is more evident at lower temperatures.

* The ploidy analyses which are central to this paper fail to include any statistical tests. Although these reviewers appreciate the modified methods introduced here, we can't tell how good they are. There was no comparison to prior methods developed to detect ploidy in Phytophthora. Specifically, the authors should apply the method of (Weiß et al., 2018) to rigorously test their conclusions about ploidy using a validated and benchmarked statistical test. They should also compare their method to that of (Weiß et al., 2018).

* Genes showing copy number variations (deletions and duplications) in P. infestans 13_A2 strain 06_-3928A isolate are more frequently located in gene sparse regions (GSRs) (Fig. S17 of (Cooke et al., 2012)). Here, they conclude that CNV did not adhere to this "two-speed genome" hypothesis. Given that their findings apparently contradict at least one previous observation, they should make sure that their methods are sensitive enough to detect the signal, and thus apply similar methods as (Cooke et al., 2012) Fig. S17. This would be the positive control for their analyses.

* Note that the two-speed genome hypothesis was primarily developed based on inter-species comparative analyses first for *P. infestans*, *P. ramorum* and *P. sojae* (Haas et al., 2009), and then with *P. infestans* and its clade 1c sister species (Raffaele et al., 2010). The patterns could be explained by increased rates of mutation or genetic instability in the GSRs, which is plausible for deletions and CNVs even for intraspecific comparisons. To these reviewers, the points made in the current paper on this topic appear somewhat superficial and fail to integrate previous knowledge and analyses (see (Raffaele & Kamoun, 2012) (Dong et al., 2015) and several other papers on this topic).

* Gene loss analysis seems to indicate that the RXLR effector genes tend to be at higher frequency in the repeat-rich regions. However, authors did not mention or discuss this pattern in this section.

* In contrast, the authors reported that CNV did not support the two-speed genome hypothesis because genes with copy number of three are enriched in the gene-dense regions (Figure 6). This might be a mis-interpretation if the core-orthologous genes were enriched in the gene dense regions. A plot of the positions of core-orthologous genes over the heatmap of gene abundance as in Figure 4 should be included to support the authors' claim.

* The method used for determining CNV may have missed recently expanded genes/genome segments where there might be no or very few heterozygous sites in those regions. Thus, CNV calculations for individual genes should be described as an approximation.

* It would be reassuring if select regions of the genome with CNV or deletions are displayed as alignments as in (Cooke et al., 2012) Fig. S19 and S20. If anything, this is a sanity check they should go through.

Other comments

Line 31. "and did not to adhere to" should be "and did not adhere to"

Line 74. It really upsets people to say that P. infestans "caused" the Great Famine. M<any blame the English. A more sensitive term would be "triggered".

Line 93-94. The statement "dramatic changes to the gene-sparse, transposon and effector rich portion of the genome are responsible for most of the adaptation in clonal lineages." Is incorrect. As stated above the two-speed genome hypothesis was primarily developed as a macroevolutionary concept based on inter-specific comparisons (Raffaele & Kamoun, 2012). It was never stated that model would explain adaptation in clonal lineages.

Line 114-115. It is incorrect to state that previous studies "have not included a representative sample from sexual populations." Many of the European isolates are probably from sexual populations.

Figure 2A. Why are the "America" and "Europe" isolates that are not assigned to a clonal lineage assumed to be clonal? These isolates in the orange and blue histogram bars could definitely include sexual isolates. This assumption is not supported by any data.

Line 152. Why 12X? Is this based on any benchmarking as per (missing citation).

Figure 3. No statistical analysis to support the ploidy assignments.

Line 184-194. This section on the historical samples is odd. First, it reads more like a discussion than a results paragraph. Second, why is it surprising that historical isolates have CNV. CNV is present in other Phytophthora spp. And has certainly been documented in the clade 1c sister species of P. infestans.

Line 204-206. These analyses lack a control in the form of core orthologous genes (COGs).

Line 259. "(Figure 6)" should be "(Figure 7)"

Figure 4. Here and elsewhere the use of pathogenicity factors to refer to effectors do not conform with the literature. For example, elicitins are not viewed as "pathogenicity factors".

Figure 5. The results were not clearly described and discussed. There is clearly variation between classes but how do we know whether this deviates from random effects or not. Also, is it possible that pairwise comparisons would be more sensitive and would reveal differences between gene classes.

Figure 8 is plain wrong given that sexual reproduction was reported outside Mexico.

Reviewers

Joe Win and Sophien Kamoun. The Sainsbury Laboratory, Norwich Research Park, University of East Anglia, Norwich, UK.

References

Population dynamics of Phytophthora infestans in the Netherlands reveals expansion and spread of dominant clonal lineages and virulence in sexual offspring.. (2012). *G3 (Bethesda)*, *2*, 1529–1540.

Five Reasons to Consider Phytophthora infestans a Reemerging Pathogen.. (2015). *Phytopathology*, *105*, 966–981.

Genome analyses of an aggressive and invasive lineage of the Irish potato famine pathogen.. (2012). *PLoS Pathog*, *8*, e1002940.

nQuire: a statistical framework for ploidy estimation using next generation sequencing.. (2018). *BMC Bioinformatics*, *19*, 122.

Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans.. (2009). *Nature*, *461*, 393–398.

Genome evolution following host jumps in the Irish potato famine pathogen lineage.. (2010). *Science*, *330*, 1540–1543.

Genome evolution in filamentous plant pathogens: why bigger can be better.. (2012). *Nat Rev Microbiol*, *10*, 417–430.

The two-speed genomes of filamentous pathogens: waltz with plants.. (2015). *Curr Opin Genet Dev*, *35*, 57–65.