




# Peer Community In Evolutionary Biology

## The scandalous pest

**Nicolas Galtier**  based on peer reviews by 2 anonymous reviewers

Georgios D Koutsovoulos, Eder Marques, Marie-Jeanne Arguel, Laurent Duret, Andressa CZ Machado, Regina MDG Carneiro, Djampa K Kozlowski, Marc Bailly-Bechet, Philippe Castagnone-Sereno, Erika VS Albuquerque, Etienne GJ Danchin (2019) Population genomics supports clonal reproduction and multiple gains and losses of parasitic abilities in the most devastating nematode plant pest. bioRxiv, ver. 1, peer-reviewed and recommended by Peer Community in Evolutionary Biology. <https://doi.org/10.1101/362129>

Submitted: 24 August 2018, Recommended: 10 July 2019

### Cite this recommendation as:

Galtier, N. (2019) The scandalous pest. *Peer Community in Evolutionary Biology*, 100077. [10.24072/pci.evolbiol.100077](https://doi.org/10.24072/pci.evolbiol.100077)

Published: 10 July 2019

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

---

Koutsovoulos et al. [1] have generated and analysed the first population genomic dataset in root-knot nematode *Meloidogyne incognita*. Why is this interesting? For two major reasons. First, *M. incognita* has been documented to be apomictic, i.e., to lack any form of sex. This is a trait of major evolutionary importance, with implications on species adaptive potential. The study of genome evolution in asexuals is fascinating and has the potential to inform on the forces governing the evolution of sex and recombination. Even small amounts of sex, however, are sufficient to restore most of the population genetic properties of true sexuals [2]. Because rare events of sex can remain undetected in the field, to confirm asexuality in *M. incognita* using genomic data is an important step. The second reason why *M. incognita* is of interest is that this nematode is one of the most harmful pests currently living on earth. *M. incognita* feeds on the roots of many cultivated plants, including tomato, bean, and cotton, and has been of major agricultural importance for decades. A number of races were defined based on host specificity. These have played a key role in attempts to control the dynamic of *M. incognita* populations via crop rotations. Races and management strategies so far lack any genetic basis, hence the second major interest of this study. The authors newly sequenced the full genome of eleven strains from Brazil and added nine already available samples from Africa and North-America. They report that, in all likelihood, *M. incognita* is indeed a purely asexual species. This is supported by (i) the confirmation that the genome is in its major part haploid, and (ii) a spectacularly high level of linkage disequilibrium, which does not decline with genetic distance between loci at a 100kb scale. The absence of sex and recombination is associated in *M. incognita* with a remarkably low amount of genetic diversity - one order of magnitude less than in typical sexual nematodes - and an heavy load of deleterious mutations, as measured by the ratio of non-synonymous (=amino-acid changing) to synonymous (=amino-acid conservative) diversity in coding sequences. The other important result of this study is that the population substructure in *M. incognita* is in no way related to host races or geography. The tree genetic clusters that are identified include strains from

several continents and feeding on a diversity of host plants. The implications of this work are numerous. First, the results suggest that *M. incognita* is an ancient asexual. Asexuality, which was here demonstrated *via* linkage disequilibrium analysis, must be ancient enough for diploidy (or, in this case, maybe triploidy) to have been lost - i.e., formerly homologous chromosomes have accumulated enough mutations to be assembled as distinct entities. So we are not talking about a highly successful clone having recently spread the world - rather a long-term obligate parthenogen. Asexual organisms are deprived of the source of genetic variation offered by recombination, which is why asexuality is thought to be an evolutionary dead-end. Long-term asexuals are uncommon and even the most famous ones, bdelloid rotifers, are suspected to experience between-individual genetic transfers [3]. *M. incognita* is apparently a true 'evolutionary scandal', and as such deserves particular attention from molecular evolutionary geneticists. The lack of any host race effect on the genetic diversity of *M. incognita* is another important finding. So-called 'races' have largely contributed to shape researchers' view of the structure of the species so far. This study demonstrates that a mental effort is now needed to forget about races, and consider host-specificity for what it is - a phenotypic trait. This result implies that many host shifts must have independently occurred in the three *M. incognita* genetic lineages, suggesting an arms race between plants and nematodes, which in the absence of sex and recombination must be entirely mutation-driven on the nematode side. Genes functionally involved in the arms race might therefore be expected to have experienced convergent evolution, if distinct *M. incognita* lineages have adopted the same solutions to overcome plant defenses. The present study paves the way for such a genome scan. The authors rightly discuss that the strong adaptive potential of *M. incognita*, at least in terms of host shift, despite no sex and tiny amounts of genetic diversity, is a paradox that would deserve to be further investigated.

### **References:**

- [1] Koutsovoulos, G. D., Marques, E., Arguel, M. J., Duret, L., Machado, A. C. Z., Carneiro, R. M. D. G., Kozłowski, D. K., Bailly-Bechet, M., Castagnone-Sereno, P., Albuquerque, E. V., & Danchin, E. G. J. (2019). Population genomics supports clonal reproduction and multiple gains and losses of parasitic abilities in the most devastating nematode plant pest. bioRxiv, 362129, ver. 5, peer-reviewed and recommended by Peer Community in Evolutionary Biology. doi: [10.1101/362129](<https://dx.doi.org/10.1101/362129>)
- [2] Hartfield, M. (2016). Evolutionary genetic consequences of facultative sex and outcrossing. *Journal of evolutionary biology*, 29(1), 5-22. doi: [10.1111/jeb.12770](<https://dx.doi.org/10.1111/jeb.12770>)
- [3] Debortoli, N., Li, X., Eyres, I., Fontaneto, D., Hespels, B., Tang, C. Q., Flot, J. F. & Van Doninck, K. (2016). Genetic exchange among bdelloid rotifers is more likely due to horizontal gene transfer than to meiotic sex. *Current Biology*, 26(6), 723-732. doi: [10.1016/j.cub.2016.01.031](<https://dx.doi.org/10.1016/j.cub.2016.01.031>)

## **Reviews**

### **Evaluation round #2**

DOI or URL of the preprint: <https://doi.org/10.1101/362129>

Version of the preprint: 3

### **Authors' reply, 06 July 2019**

[Download author's reply](#)

[Download tracked changes file](#)

## Decision by [Nicolas Galtier](#) , posted 25 June 2019

### Decision on: "Parallel adaptations...", by Koutsovoulos et al.

I concur with the reviewer that the manuscript has been substantially improved. The scope of the study has broadened, and I find the overall message clear and compelling. The analysis of coverage, the distinction between so-called "heterozygous" and "homozygous" variation and the linkage disequilibrium analysis are important, informative additions. The title was appropriately amended and reflects, I think, the more ambitious nature of the study.

The reviewer has a couple of comments, which deserve to be considered.

First, the reviewer suggests analysing the variation between homeologous regions within a sample (major comments 1 to 4), when the text currently focuses on the between-samples ("haploid") variation. This would be a really nice addition, if possible. I am not sure, however, that separating true variants (between homeologs) from spurious variants (due to assembly/duplication issues) based on coverage is easy to do in this case - figure S1 suggests that the coverage distributions of the two categories of variants overlap quite a bit. Please let us know what you think is doable here.

The other important comment made by the reviewer (major comments 5-6) is that the population genetic analyses have been done in an unusual way, i.e., by comparing each sample to the reference. This has an unclear meaning, which depends on how the reference was generated (single individual? pool of individuals? from which origin?). The reviewer rather suggests analysing multiple alignments across the newly sequenced strains, which indeed should provide more reliable estimates of, particularly,  $\pi_N$ ,  $\pi_S$  and their ratio. This is clearly a sensible recommendation, which I think should be followed.

I have a related, minor comment: for the same reason, I find the "homozygous SNP" vs "heterozygous SNP" terminology quite misleading. A SNP is normally a position in a genome at which between-individual variation has been detected - i.e., a variable colon in a within-species alignment, or a vector of genotypes. Such a vector should not be qualified as homozygous or heterozygous. Furthermore, because the *M. incognita* genome is haploid, one would not expect to find any heterozygous genotype at all. Yet, because the authors have applied a variant calling method that assumes diploidy, and because of assembly/duplication errors or partial di/triploidy, a large number of apparently heterozygous variants were called. I would suggest refraining from calling a SNP "homozygous" or "heterozygous", and using "variant" rather than "SNP" when referring to a genotype predicted by the variant caller. The word "SNP" should be restricted to the new analysis suggested by the reviewer, when sequences from the distinct strains have been multiply aligned.

I would suggest following these very last suggestions, which I think should help improve this excellent manuscript even further.

## Reviewed by anonymous reviewer 1, 13 June 2019

In this revised version the authors have done several additional analyses and have extensively rewritten the manuscript. The new results strengthen the manuscript and broaden its interest. However, some problems remain about data analysis, especially for polymorphism analysis, which were not performed properly if my understanding is correct. They should be easily corrected but some results may change.

### Major comments

- The title of the first paragraph of the results is "...the genome is mostly haploid...". If the species is triploid due to hybridization it means that two sets of chromosomes should pair whereas the third one should be alone. If I understood correctly, 80% of SNPs were heterozygotes. Does it mean that they correspond to the diploid pair and to the Meselson effect between the two chromosomes?

- The fact that most heterozygotes could be duplicates because of the doubling of the coverage is convincing. But in that case, the two parts of the genome should be split based on coverage. This would allow to analyse separately the haploid and the diploid genome. This would clearly help to better understand the reproductive system of this species.
- The choice of excluding heterozygote sites can be justified so it indeed prevents to compute  $F_{is}$ . However, it would be important to know how behaves the diploid genome. Is there an excess of heterozygotes or not. If not, or it varies along the genome it could be informative of the kind of asexuality. Modification of meiosis could also be a possibility instead of mitotic reproduction (for example see (Lenormand et al., 2016, Engelstadter, 2017)). We could imagine a form of automixis with one set of haploid chromosome transmitted as a block without segregation.
- Related to this important question, and still if my understanding is correct, it's important to note that the lack of recombination is shown (nicely, see below) for the haploid genome (which is expected) but not for the diploid genome. Absence of recombination for the diploid genome ( $+ F_{is} < 0$ ) would be a strong argument for mitotic recombination (but still some for of automixis can lead to a similar pattern)..
- If I understood correctly, SNPs are defined as variant compared to the reference. Then polymorphism is computed for each strain as the % of SNP. But this is not a measure of polymorphism of a strain but of the genetic distance between the strain and the reference. If only homozygote variants are kept,  $\pi_S$  and  $\pi_N$  cannot be computed for a strain. Here what should be done is to compute  $\pi_S$  and  $\pi_N$  for the whole species and also interestingly for the three clusters detected by the PCA. And when computing these statistics the reference genome should not be considered, except if the reference strain is added as a data point.
- If the order of magnitude of  $\pi_S$  and  $\pi_N/\pi_S$  is still valid after correction, it is interesting to note that, although  $\pi_N/\pi_S$  is three times higher than in outcrossing nematodes, it is still lower than many other species, including human (around 0.2).
- The PCA and phylogenetic tree do not support clustering by host. This could be used as an opportunity to try to identify the few SNPs (if any) that could be associated with hosts. It is not possible for all host but for those that can be found in different clusters such as soybean, cotton and tobacco.

Minor comments - The test of absence of recombination is a nice addition and the idea of comparing to a recombining species is a nice control. This part could be move earlier (second or third part of results for example). - It was not one of my previous comments but I'm not really convinced by the argument stating that an ancient polyphageous strain is not very likely. Here the number of strains is too low for ancestral state reconstruction. I don't mean that this idea is wrong but that the alternative proposed by the other reviewer could also be discussed and that additional data would be needed to test it properly.

References: - Engelstadter, J. 2017. Asexual but Not Clonal: Evolutionary Processes in Automictic Populations. Genetics. - Lenormand, T., Engelstadter, J., Johnston, S. E., Wijnker, E. & Haag, C. R. 2016. Evolutionary mysteries in meiosis. Philos Trans R Soc Lond B Biol Sci 371.

## Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/362129>

Version of the preprint: 2

### Authors' reply, 10 May 2019

[Download author's reply](#)

[Download tracked changes file](#)

## Decision by [Nicolas Galtier](#) , posted 19 September 2018

### **Decision on: "Parallel adaptations...", by Koutsovoulos et al.**

The two reviewers have expressed relevant and important comments on various aspects of the study. I concur that the current manuscript should be extensively revised in order to reach, and convince, a wider evolutionary biology audience, as expected for a manuscript recommended by PCI.

Reviewer 2 recapitulates the main results of the study and identifies a number of issues requiring clarification, rewriting, and/or re-analysis. This reviewer also suggests that the current title does not optimally reflect the content of the study - I agree.

In addition, Reviewer 1 mentions a number of analyses that could be made in order to better characterize the population genomics and molecular evolution of *M. incognita*, with a focus on its supposed asexuality. I agree that this is a missed opportunity, especially knowing that previous publications on the subject, some by authors of this manuscript, have opened very interesting questions (eg Castagnone-Sereno & Danchin 2014 JEB).

The two reviews are highly complementary and provide a number of clearly expressed recommendations, which I think should greatly help improve the manuscript.

### **Additional requirements of the managing board**

We ask you to carefully verify that your manuscript complies with the following requirements (indicated in the 'How does it work?' section and in the code of conduct) and to modify your manuscript accordingly:

-Data must be available to readers after recommendation, either in the text or through an open data repository such as Zenodo, Dryad or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.

-Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) must be available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused.

-Details on experimental procedures must be available to readers in the text or as appendices.

## Reviewed by anonymous reviewer 2, 25 August 2018

In this manuscript, the authors collected 11 'isolates' of the parasitic root knot nematode *Meloidogyne incognita* from 6 different crops. These isolates were then assigned to four 'host races' traditionally recognized in the species based on the ability of nematodes to infect particular reference crop species (these reference species differ from the crops the nematodes were collected from). The authors then re-sequenced the genomes of the isolates and inferred SNPs relative to an available reference genome of the species. Two independent clustering approaches based on the SNP data (PCA and a phylogenetic network based on the subset of SNPs in coding regions) indicate that the 11 isolates form 3 diverged clusters. From these clusters the authors draw the following 3 conclusions:

1. The clusters do not correspond to 'host races' and the use of the term 'host race' should be abandoned for *M. incognita*. I have no problem with this conclusion - in more 'standard' terminology, host races are polyphyletic and correspond to a phenotype rather than a lineage. However, the authors also state that these data indicate multiple independent adaptations to different host ranges. In the absence of information of ancestral host ranges, this is clearly an over-interpretation. For example, one could imagine a highly polyphagous ancestral lineage (i.e., with extreme phenotypic plasticity for host plant use), but the ability to infect specific hosts was lost independently in different lineages. The title "Parallel adaptations to different host plants despite clonal reproduction in the world most devastating nematode pest" should therefore be abandoned as it does not reflect the findings of the paper.
2. The authors state that there is no correlation between the diverged clusters and the geographic origin

of the samples. However, according to Fig 5 this is not entirely true as there is some clustering of geographically close samples. Instead of eyeballing whether or not the genetic clusters correspond some grouping of samples from a given country, I suggest the author conduct standard IBD analyses and calculate the % of genetic variance explained by geographic distance (i.e., using pairwise geographic distances between isolates). This % will be small, but represents a more objective evaluation of the effect of geography.

3. Similar to point 2, the authors also state that there is no correlation between the clusters and the crop species where the isolates were collected. This is difficult to evaluate given the small number of isolates (n=11) relative to the number of crops (n=6). Nevertheless, 3 of the 4 isolates from cotton are members of the same cluster, as are the 2 out of 2 isolates from tobacco, so there appears to be some correlation. Again, I suggest that the authors quantify the amount divergence of isolates of the same vs different crop species to show that the amount of variance between crops is not larger than the variance within.

Finally, there are parts in the methods that should be clarified. For example, I believe the term 'isolate' usually refers to a strain derived from a single female, but here apparently it is a pool of individuals collected at a given location (or even from multiple locations for isolate R3-4, see line 117). How did the authors deal with population variation – was only the major allele considered at each position? Along the same lines: *M. incognita* is a hybrid species and highly heterozygous. I believe the reference assembly is largely haploid (the two parental genomes are assembled separately), meaning there will be no heterozygous positions for a genome based on a single genotype. These points do not affect the results but should be clarified in the methods.

Minor comments. - In the discussion of divergence from the reference genome for nuclear and mitochondrial sequences (L241 and following), I am surprised the authors are not mentioning lack of recombination – the lack of recombination in mt genomes contributes to their relatively high substitution rates in comparison to the nuclear genome in sexual species. This difference no longer exists in *M. incognita*. -L134, 135: do you mean divergence from the Morelos reference strain? (are these uncorrected p-distances?) -The map in Figure 1 has 13 isolates listed for Brasil, not 11. -L108: I dont understand the sentence "Considering the clonal reproduction of this species, the route of adaptation to different hosts is evolutionary important since it is unknown whether it happens independently or a consequence of four ancestral states." -'Host preference' usually refers to a neurological mechanism of preferring one host over another, not to being able to infect one host but not another. I would chose a different term.

## **Reviewed by anonymous reviewer 1, 25 August 2018**

This manuscript presents a survey of genetic diversity at the whole genome level in the root-knot nematode *Meloidogyne incognita*, which is a clonal worldwide plant pathogen with different host races. The main question was to determine whether these races corresponded to distinct genetic clusters. Three main genetic clusters were found but they are neither associated with host races nor geographical origins. This result is interesting with important practical implications and the ms clearly presents this main result. However, I think that the dataset (whole genome sequence of 11 strains + additional published genomes) is clearly under-analysed and that a better knowledge of the system could be reach, also with potential practical implications. This version should be sufficient for a specific audience specifically interested in the biological model. But to reach a broader audience, additional analyses should be done

### **MAJOR COMMENTS**

A major assumption is that the species is clonal. However this is not discussed while the data could help to evaluate more precisely the breeding system of the species. Is the species really completely clonal? Is it recent or not? For example, Fig 4 represents a network and not a fully resolved tree. In a purely clonal species, a perfect tree should be expected. Potential signature of genetic exchanges could be assessed with the current dataset: - Under pure clonality all genomic regions should give the same history. So first did you get the same

nuclear and mitochondrial tree? Then, if you split the nuclear dataset into blocks (chromosomes or shorter blocks) do you obtain the same tree for all blocks? This is a simple way to test whether genetic exchanges have occurred or not? - There are also different methods to test the occurrence of recombination (or gene conversion) that should be applied here (ex: decrease of linkage-disequilibrium with distance, four-gamete test, and more elaborated methods) - Individual heterozygosity is not given. Under pure clonality, excess of heterozygosity ( $F_{is} < 0$ ) is also expected I think this point is crucial to correctly interpret the results.

Genetic diversity is only briefly analysed and not very precisely. P6 I133: What is exactly "level of variation"? Does it mean Tajima's  $\pi$  or another statistics? To allow comparison with other species it would be interesting to compute  $\pi$  synonymous (or 4-fold): is it of the same order of other clonal or selfing species? The comparison with selfing and outcrossing nematodes should be particularly relevant. More generally, recent surveys of genetic diversity can be used for comparison, for ex: - Romiguier et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261-263. - Chen, J. et al. 2017. Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Mol Biol Evol* 34:1417-1428. To better interpret  $\pi_S$  in term of effective population size ( $N_e$ ), an idea of the mutation rate should be important also. In addition to  $\pi_S$ , the  $\pi_N/\pi_S$  ratio should also be computed and given. It gives an idea of the efficiency of selection and is usually rather well correlated with  $N_e$ . This could be compared also with previous studies.

The analyses suggested above could be done at the whole species level but also for the three different genetic clusters separately. In addition,  $F_{st}$  should be given to get a more quantitative idea of population structure than just the PCA.

It would also be interesting to present the distribution of genetic diversity along chromosomes. This could also bring information about potential rare or past recombination events, for example if there is more genetic diversity in telomeric than in centromeric regions.

"M. incognita is particularly versatile and adaptive despite its clonal mode of reproduction": this interpretation implicitly assumed that adaptation to the host is highly multigenic and complex. However, if only a few key genes determine host compatibility, the problem of being clonal is much less important and it maybe not necessary to invoke CNV or epigenetic mechanisms. It would also explain why there is no association between genetic cluster and host race.

Because there is no association between genetic cluster and host races it is potentially a good situation to identify genes potentially involved in adaptation to the different hosts. The sample size may be too small but more elaborate genomic scan could be done rather than simply searching for specific SNPs. In particular, the genomic location should be taken into account to increase the power of detection.

#### MINOR COMMENTS

P4 I73: "no clear genetic differences underlying the phenotypic plasticity" this is not well formulated because strictly speaking phenotypic plasticity does not require genetic variation, otherwise we would rather use "local adaptation"

P5 I96: "in this analysis; no relation" -> ";" should be replaced by ","

Fig 2: Number of variants per isolate: to what are variants defined? The reference strain?

Fig4: the scale on the figure is too small and hardly readable. Is it 0.1?