ARTICLE (Discoveries)

Population genomics supports clonal reproduction and multiple gains and losses of parasitic abilities in the most devastating nematode plant pest

Georgios D. Koutsovoulos[1*], Eder Marques[2], Marie-Jeanne Arguel[3], Laurent Duret[4], Andressa C.Z. Machado[5], Regina M.D.G. Carneiro[2], Djampa K. Kozlowski[1], Marc Bailly-Bechet[1], Philippe Castagnone-Sereno[1], , Erika V.S. Albuquerque[2*], Etienne G.J. Danchin[1*], ~~Erika V.S. Albuquerque[2*]~~

[1]INRA, CNRS, Université Côte d'Azur, ISA, France

[2]Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brasil

[3]CNRS, Université Côte d'Azur, IPMC, France

[4]Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR 5558, F-69100, Villeurbanne, France

[5]Área de Proteção de Plantas, Instituto Agronômico do Paraná, Londrina, Paraná, Brasil

Corresponding author: Etienne GJ Danchin, etienne.danchin@inra.fr

* equal contributions

Abstract

The most devastating nematodes to worldwide agriculture are the root-knot nematodes with *Meloidogyne incognita* being the most widely distributed and damaging species. This parasitic and ecological success seem surprising given its supposed obligatory clonal reproduction. Clonal reproduction has been suspected based on cytological observations but, so far, never confirmed by population genomics data. At the species level, *M. incognita* is highly polyphagous with thousands of host plants. However, the host range varies among different *M. incognita* isolates that may present distinct and more restricted host compatibilities. Historically, four 'host races' had been defined as a function of ranges of compatible and incompatible plants. We sequenced the genomes of 11 isolates across Brazil, covering these four distinct races to assess (i) how clonal reproduction is and (ii) how the level of genome variability associates with biological traits such as the host races, affected agronomic culture, and geographical distribution. By aligning the genomic reads of the isolates to the *M. incognita* reference genome assembly, we identified ~~SNPs~~ SNV and small-scale insertions/deletions. Analysis of linkage disequilibrium and 4-gametes test, showed no sign of recombination, confirming the clonal mode of reproduction of *M. incognita.* We showed that there are relatively few point variations between the different isolates, and these variations show no significant association with either the host races, the geographical origin of the samples or the host plant on which they have been collected. Due to the lack of phylogenetic signal underlying their existence, we recommend the discontinuation of the terminology "race". Overall, these results suggest that multiple gains and losses of parasitic abilities and adaptations to different environmental conditions account for the broad host spectrum and wide geographic distribution of *M. incognita*. Hence, this ~~metazoan~~ nematode constitutes a model species to study adaptability without sexual recombination and overall low genomic variations in animals.

Introduction

Nematodes cause severe damages to the world agricultural production every year with the root-knot nematodes (RKN, genus *Meloidogyne*) being the most economically harmful in all temperate and tropical producing areas (Moens et al. 2009; Jones et al. 2013). Curiously, the most polyphagous RKN species, able to parasitize the vast majority of flowering plants on Earth (Trudgill and Blok 2001), are described as mitotic parthenogenetic, based on cytogenetics comparisons with outcrossing relatives (Triantaphyllou 1981; Triantaphyllou 1985). This suggests absence of meiosis and obligatory asexual reproduction. Among these mitotic parthenogenetic RKN, *M. incognita* is the most widespread and is present, at least, in all the countries where the lowest temperature exceeds 3°C. Greenhouses over the world also extend its geographic distribution (Sasser et al. 1983). *M. incognita* is so widely distributed that it is not even included on the list of regulated pests (Singh et al. 2013). Due to its worldwide distribution and extremely large range of hosts, *M. incognita* has been deemed the most damaging species of crop pest worldwide (Trudgill and Blok 2001).

However, it has become more and more evident that the ~~full~~ whole broad host range of *M. incognita* as well as the other major RKN species is not present in all the individuals within the species but that different 'populations' or 'isolates' have different and overlapping ranges of compatible hosts (Moens et al. 2009). Variations regarding host range within one given species gave rise to the concept of 'host race' as soon as 1952 (Sasser 1952). Although RKN species can be differentiated based on morphological descriptions (Eisenback and Hunt 2009), isozyme phenotypes (Esbenshade and Triantaphyllou 1985; Carneiro et al. 2000), and molecular analysis (Blok and Powers 2009), this is not the case of host races within a species (Triantaphyllou 1985). Consequently, the pattern of compatibility/incompatibility of the nematode interaction with a set of different host plants was standardised into the North Carolina Differential Host Test (NCDHT, (Hartman and Sasser 1985)) to differentiate races within Meloidogyne spp. In *M. incognita*, all the populations originally tested reproduced on tomato, watermelon, and pepper and none infected peanut, but they differed in their response to tobacco and cotton defining four distinct host races (Hartman and Sasser 1985) (Table S1). Whether some genetic features are associated with the *M. incognita* races remains unknown. Indeed, diversity studies using RAPD and ISSR markers across eight isolates and the four races of *M. incognita* found no correlation between phylogeny and host races (Cenis 1993; Baum 1994; Santos et al. 2012). Although in one of these studies (Santos et al. 2012), two out of three esterase phenotypes were monophyletic in the phylogenetic tree of the *M. incognita* isolates, they did not segregate according to the host races. A different molecular approach to try to differentiate host races was also proposed based on repeated sequence sets in the mitochondrial ~~sequence~~ genome (Okimoto et al. 1991). Although the pattern of repeats allowed differentiating one isolate of race 1, one of race 2 and one of race 4; the study

encompassed only one isolate per race, and thus the segregation could be due to differences between isolates unrelated to the host race status itself.

Hence, no clear genetic determinant underlying the phenotypic diversity of *M. incognita* isolates in terms of host compatibility patterns have been identified so far (Castagnone-Sereno 2006). This lack of phylogenetic signal underlying the host races is surprising because it would suggest multiple independent gains and losses of host compatibly patterns despite clonal reproduction. Theoretically, animal clones have poorer adaptability because the efficiency of selection is impaired, advantageous alleles from different individuals cannot be combined and deleterious mutations are predicted to progressively accumulate in an irreversible ratchet-like mode (Muller 1964; Hill and Robertson 1966; Kondrashov 1988; Glémin and Galtier 2012).

For these reasons, the parasitic success of *M. incognita* has long been described as a surprising evolutionary paradox (Castagnone-Sereno and Danchin 2014). However, this apparent paradox holds true only if this species does actually reproduce without sex and meiosis while presenting substantial adaptability. So far, no whole genome level ~~conclusive~~ study ~~conducted at the whole genome level,~~ conclusively support these tenets.

A first version of the genome of *M. incognita* was initially published in 2008 (Abad et al. 2008) and re-sequenced at higher resolution in 2017, providing the most complete *M. incognita* reference genome available to date (Blanc-Mathieu et al. 2017). This study showed that the genome is triploid with high average divergence between the three genome copies most likely because of hybridization events. Due to the high divergence between the homoeologous genome copies, and the supposed lack of meiosis, it was assumed that the genome was effectively haploid. The genome structure itself showed synteny breakpoints between the homoeologous regions and some of them formed tandem repeats and palindromes. These same structures were also described in the genome of the bdelloid rotifer *Adineta vaga* and considered as incompatible with meiosis (Flot et al. 2013; Blanc-Mathieu et al. 2017). However, whether these structures represent a biological reality or artefacts of genome assembly remains to be clarified. Indeed, both genomes have been assembled using the same techniques and no independent biological validation for these structures has been performed. Hence, so far no strong evidence supporting the absence of meiosis was available at the genome level.

Furthermore, because the reference genome was obtained from the offspring of one single female (originally from Morelos, Mexico), no information about the genomic variability between different populations or isolates was available. More recently, a comparative genomics analysis, including different strains of *M. incognita*, showed little variation at the protein-coding genome level between strains collected across different geographical locations (Szitenberg et al. 2017), confirming previous observations with RAPD and ISSR markers (Cenis 1993; Baum 1994; Santos et al. 2012). However, no attempt was made to associate these few variations with biological traits

such as the host-race or geographic origin. Moreover, the variability between isolates at the non-coding level, which represents the ~~vast~~ majority of the genome, was not described in this initial analysis.

In the present study, we used population genomics analyses to investigate, (i) whether the supposed absence of meiosis is supported by the properties of genome-wide ~~SNP~~ SNV markers between ~~different~~ isolates, (ii) the level of variation between ~~different~~ isolates at the whole genome level, and (iii) whether these variations follow a phylogenetic signal underlying ~~several~~ life history traits such as the host compatibility patterns, the geographic distribution or the current host crop plant.

To address these questions, we have sequenced the genomes of 11 isolates representing the four *M. incognita* host-races from populations parasitizing six crops ~~from~~ across geographically different sites ~~across~~ in Brazil (Figure 1). We used isozymes profiles, SCAR markers, and the NCDHT to characterize the biological materials before the DNA extraction and high coverage genome sequencing. We identified ~~SNPs~~ short-scale variations at the whole genome level by comparing the *M. incognita* isolates to the reference genome (Morelos strain from Mexico). We conducted several ~~SNP~~SNV-based genetic tests to investigate the evidences (or lack thereof) of recombination. Using two different approaches, we classified the *M. incognita* isolates according to their ~~SNP~~ SNV patterns and investigated whether the classification was associated to the following biological traits: host compatibility, geographical localisation and current host plant.

Our population genomics analysis allowed addressing key evolutionary questions such as how asexual is reproduction in this animal species. We were also able to clarify the adaptive potential of this devastating plant pest in relation to its mode of reproduction. In particular, we determined whether there is a phylogenetic signal underlying variations in biological traits of agro-economic importance such as the patterns of host compatibility (host races) ~~in *M. incognita*~~. While association between phylogenetic signal and patterns of host compatibilities would tend to show stable inheritance from ancestral states, the non-association would support multiple gains and losses of parasitic abilities and substantial adaptability.

This resolution has important agricultural and economic implications since crop rotation and other control strategies should take into account the adaptive potential of this nematode pest.

Results

The *M. incognita* genome is mostly haploid and shows few short-scale variations

We collected 11 *M. incognita* populations from six different states across Brazil and from six different crops (soybean, cotton, coffee, cucumber, tobacco, watermelon) (Figure 1). Each isolate was reared by multiplication of the egg mass of one single female on tomato plants (methods). After having confirmed that the 11 isolates we collected showed the characteristic isozyme profiles and molecular signatures of *M. incognita*, we characterized their host race status using the NCDHT (methods, Table 1). We characterised three isolates as race 1, two as race 2, three as race 3, and three as race 4.

We generated ~~high-quality~~ paired-end genome reads (~76 million per isolate) ~~for each *M. incognita* isolate. These reads~~ which covered the ~184Mb *M. incognita* genome assembly (Blanc-Mathieu et al. 2017) at a depth > 100X (Table S2) for each isolate. Variant calling, performed in regions with at least 10x coverage per sample, identified 338,960 polymorphic positions (~0.19% of the total number of non-ambiguous nucleotides). Around 20% of these positions corresponded to 1/1 ~~SNPs~~SNV, ~~homozygous~~ fixed within each isolate~~s~~ but variable between isolates and the reference genome. We examined the distribution of base coverage of ~~homozygous~~ SNV fixed within all isolates (1/1 fixed SNV) and ~~heterozygous~~ SNV that presented variations within at least one isolate (0/1 SNV) ~~SNPS~~. We observed that the ~~heterozygous~~ 0/1 ~~SNPs~~ SNV, which were variable within isolates, showed a peak of distribution at ~twice the coverage of the peak for ~~homozygous~~ fixed 1/1 ~~SNPs~~ SNV in the 11 isolates (Fig. S1). This parallels the distribution of base coverage in the *M. incognita* reference genome scaffolds which shows a major peak at ~65X and a second minor peak at ~130X (twice the coverage; Fig. S2). These genome regions at double coverage were considered as representing highly similar pairs of homoeologous genome copies that were collapsed during the assembly (Blanc-Mathieu et al. 2017). Although these regions are minority in the genome assembly, they seem to be responsible for many ~~heterozygous (~~0/1~~)~~ ~~SNPs~~SNV (presenting within isolate variations). The ~~SNPs~~ SNV in these minority regions of double coverage probably result from genome reads of two homoeologous regions mapped to a single collapsed region in the reference assembly. Hence, most of these ~~heterozygous SNPs~~0/1 SNV might not represent variations ~~between isolates or~~ between individuals within an isolate but between the few collapsed homoeologous regions. ~~Since~~ Because ~~most of~~ the reference genome is mostly assembled in haploid status (Fig. S2), and the ~~status~~ nature of ~~heterozygous~~ 0/1 ~~SNPs~~ SNV is unsure, we will utilise only ~~homozygous~~ 1/1 ~~SNPs~~ SNV fixed within isolates for~~or~~ all downstream analyses. Although this precludes analyses of variations between individuals within isolates this allows a comparison of variations between isolates based on >66,000 solid fixed markers.

No evidence for meiotic recombination in *M. incognita*

Based on cytogenetics observation, *M. incognita* and other tropical root-knot nematodes have been described as mitotic parthenogenetic species (Triantaphyllou 1981; Triantaphyllou 1985). However, this evolutionary important claim has never been confirmed by genome-wide analyses so far. Using the SNVP fixed markers at the whole-genome scale, we conducted linkage disequilibrium (LD) analysis as well as 4-gametes test to search for evidence for recombination (or lack thereof). In an outcrossing species, although physically close markers should be in high LD, this LD should substantially decrease with distance between the markers, because of recombination, and eventually reach absence of LD as for markers present on different chromosomes. In clonal species, however, in the absence of recombination, the LD between markers should remain high and not rapidly decrease with distance between markers. By conducting an analysis of LD, we did not find any trend for a decrease of LD between markers as a function of their physical distance (Figure 52). In contrast, the LD values remained high regardless the distance and oscillated between 0.85 – 0.94. Hence, we did not observe the expected characteristic signatures of meiosis. An inversely contrasted situation between outcrossing and clonal genomes should be observed for the 4-gametes test. Taking fixed SNPV markers that exist in two states among the 11 isolates, the proportion of pairs of markers that pass the 4-gametes test (i.e. that represent the 4 products of meiosis) should rapidly increase with distance between the markers, in case of recombination. In contrast, in the absence of recombination, no trend for an increase of the proportion of pairs or markers passing the 4-gametes test with distance between markers should be observed. By conducting an analysis of 2-states markers, we observed no trend for a change in the proportion of markers passing the test with distance. In contrast, the distribution remained flat and close to a value of 0.0. Again, this trend does not correspond to the expected characteristic of meiotic recombination.

To assess the sensitivity of our methods in finding evidence for recombination, we conducted the same analyses (LD and 4-gametes tests) in the outcrossing diploid meiotic plant-parasitic nematode *Globodera rostochiensis* (Eves-van den Akker et al. 2016). Because the *G. rostochiensis* genome assembly mostly consists of merged paternal and maternal haplotypes, we had to phase the SNVPs before conducting LD and 4-gametes tests. The results were totally contrasted between *M. incognita* and *G. rostochiensis* (Figure 63). In *G. rostochiensis*, the LD and 4-gametes curves started at relatively lower (<0.7) and higher (>0.15) values, respectively (Figure 6). Furthermore, we observed a rapid exponential decrease of r^2 in the first kb for LD. At an inter-marker distance of 3kb the r^2 value was <0.37. In parallel, we observed a concomitant rapid and exponential increase in the proportion of markers passing the 4-gametes test, which was >0.38 at the same inter-marker distance. Hence, while *G. rostochiensis* appears to display all the expected characteristics of meiotic recombination, this was not the case for *M. incognita*. This

confirms at a whole genome-scale the lack of evidence for meiosis previously observed at the cytological level in *M. incognita*.

The levels of variations between isolates are low and not specific to races

Each isolate showed a different level of divergence from the reference genome with R1-2 having the highest number of ~~homozygous variable~~ fixed SNV~~positions~~ (41,518) and R1-6 having the least (17,194) variants (Figure ~~2~~4). The R3-4 isolate originated from a pool of four populations. However, the low number of ~~SNPs~~ SNV compared to the reference indicates either that the genomes of these four populations were very close or that a specific population displaced the other three (Figure ~~2~~4). Thus, the R3-4 isolate was analysed exactly as the other isolates. Overall, the percentages of ~~homozygous variable positions~~ fixed SNV on the nuclear genomes of the eleven isolates, compared to the Morelos reference strain, ranged between 0.01 % and 0.02 %. In comparison, the percentages of ~~variable positions~~ SNV in the mitochondrial genome ranged between 0.04% and 0.18%.

Interestingly, race-specific variants exist only for race 2, which exhibited 30 race-specific variations. This is possibly due to the fact that race 2 is represented by only two isolates (vs. 3 for the rest of the isolates). The vast majority (~78%) of ~~SNPs~~ SNV were outside of coding regions; only 14,704 variable positions fell in coding regions and covered 7,259 out of 43,718 predicted protein-coding genes. In these coding regions, 8,179 were synonymous substitutions, 3,854 ~~SNPs~~ SNV yielded non-synonymous substitution, 93 nonsense mutations and the rest other disruptive mutations.

From the ~~SNPs~~ SNV falling in coding regions, we constructed a multiple alignment ~~also~~ and measured nucleotide diversity at synonymous (π_s) and non-synonymous (π_n) sites for the 11 isolates as well as the π_n/π_s ratio as a measure of the efficiency of selection. Consistent with the overall low number of ~~SNPs~~SNV, the π_s value~~s were quite low for~~ across the 11 isolates was low ($1.29\ 10^{-03}$)~~(ranging from $2.82\ 10^{-04}$ (R2-1) to $5.60\ 10^{-04}$ (R1-2) and median value = $3.53\ 10^{-04}$, Table S3)~~. This is ~~two~~ one order~~s~~ of magnitude lower than the values measured for two outcrosser nematodes from the *Caenorhabditis* genus (Romiguier et al. 2014), *C. doughertyi* (formerly sp. 10: $4.93\ 10^{-02}$) and *C. brenneri* ($3.22\ 10^{-02}$). A similar difference of ~~two~~ one order~~s~~ of magnitude was also observed for the diversity at non-synonymous sites with a π_n value~~s ranging from~~ of $1.66\ 10^{-04}$ ~~$4.63\ to\ 8.56\ 10^{-05}$~~ for *M. incognita* and values reaching $2.53\ 10^{-03}$ and $1.28\ 10^{-03}$ for *C. doughetryi* and *C. brenneri*, respectively. However the π_n/π_s ratio was substantially higher for *M. incognita* (0.129~~median ratio = 0.156 and range 0.149-0.164~~) than for the two outcrossing Caenorhabditis (0.051 and 0.040 for *C. doughetryi* and *C. brenneri*, respectively). ~~This~~ These results ~~would~~ suggest a lower efficacy of selection in the obligate parthenogenetic *M. incognita* than in these two outcrossing *Caenorhabditis* nematodes.

There is no significant association between the short-scale variants and biological traits

Using principal component analysis on the whole set of ~~homozygous~~ fixed ~~SNPs~~ SNV~~(PCA, methods)~~, we showed that the eleven *M. incognita* isolates formed three distinct clusters, which we named A, B and C (Figure ~~3~~5). Cluster A is represented by isolate R1-2 alone, which has the highest number of variants. Cluster B is constituted by R3-2 and R4-4. The rest of the isolates fall in a single dense cluster C of overall low variation. There was no significant association between the clusters and the host race status (Fisher's exact test p-value=1, Sup. Text, Table ~~S4~~S3). This implies that isolates of the same host race are not more similar to each other than isolates of different host races. There was also no significant association between the ~~SNP~~SNV-based clusters and the original host plant from which the nematodes have been collected (Fisher's exact test p-value=0.69, Sup. Text, Table ~~S4~~S3). Interestingly, the four different host races are all represented in one single cluster (C). Within this cluster, the total number of variable positions was 29,597. Meaning that the whole range of host-race diversity is present in a cluster that represents only 44% of the total existing genomic variation. We also conducted an isolation by distance (IBD) analysis, which showed no correlation between the genetic distance and the geographical distance (Fig. S3).

To assess the levels of separation vs. past genetic exchanges between these clusters, we calculated fixation index values (F_{ST}). Weighted F_{ST} values between clusters were all >0.83, suggesting a lack of genetic connections between the clusters (Table ~~S5~~S4). Using the mean F_{ST} values, in contrast, while we observed a mean F_{ST} >0.98 between clusters A and B, indicating a lack of genetic connection between R1-2 and cluster B, the F_{ST} values were much lower between A and C (0.35) and between B and C (0.52). This would suggest isolates from clusters A and B both result from a past bottlenecked dispersal and propagation from some isolates in cluster C. We also conducted the same π_n/π_s analysis than the one performed at the whole species level for each cluster of the PCA containing at least 2 isolates. These cluster-specific statistics yielded similar π_n/π_s ratio than the one observed at whole species level (Cluster C: π_s 3.8 10^{-04}, π_n 5.36 10^{-05}, π_n/π_s 0.141; Cluster B: π_s 2.08 10^{-05}, π_n 2.64 10^{-06}, π_n/π_s 0.127).

Phylogenetic networks confirm the lack of association of ~~SNPs~~ SNV with biological traits and support clonal evolution

Using a phylogenetic network analysis based on ~~SNP~~ SNV present in coding regions, we could confirm the same three clusters (Figure ~~4~~6). This further supports the absence of

phylogenetic signal underlying the host races (patterns of host compatibilities). Interestingly, this network analysis based on ~~homozygous~~ fixed ~~SNPs~~ SNV yielded a bifurcating tree and not a network. This result suggests a lack of genetic exchanges between the isolates, as expected from a clonal species. To confirm this result, we conducted separate phylogenetic analyses for each of the 14 longest ~~14~~ scaffolds with sufficient number of phylogenetically informative variable positions and the mitochondrial genome. All these analyses showed a clear separation between the three clusters (A, B, and C) with some minor polytomies within cluster C (Fig. S5 and Fig; S6).

According to the two classification methods (PCA and phylogenetic network), isolate R1-2 seemed to be the most divergent from the rest of isolates, which is consistent with its higher total number of ~~SNPs~~ SNV and number of isolate-specific ~~SNPs~~SNV. Then, a small cluster was composed of isolates R3-2 and R4-3 (equivalent to cluster B of the PCA). Finally, a cluster (equivalent to PCA cluster C) grouped the rest of the eight isolates and covered all the defined host races as well as 5 of the 6 different host plants

Consistent with the PCA and phylogenetic network analysis, we also did not observe significant association between the number of repeats in the two repeat regions in the mitochondrial genome (63R and 102R) and races, geographical origin or host plant of origin (Table 2).

~~No evidence for meiotic recombination in *M. incognita*~~

~~Based on cytogenetics observation, *M. incognita* and other tropical root-knot nematodes have been described as mitotic parthenogenetic species (Triantaphyllou 1981; Triantaphyllou 1985). However, this evolutionary important claim has never been confirmed by genome-wide analyses so far. Using the SNP markers at the whole-genome scale, we conducted linkage disequilibrium (LD) analysis as well as 4-gametes test to search for evidence for recombination (or lack thereof). In an outerossing species, although physically close markers should be in high LD, this LD should substantially decrease with distance between the markers, because of recombination, and eventually reach absence of LD as for markers present on different chromosomes. In clonal species, however, in the absence of recombination, the LD between markers should remain high and not rapidly decrease with distance between markers. By conducting an analysis of LD, we did not find any trend for a decrease of LD between markers as a function of their physical distance (Figure 5). In contrast, the LD values remained high regardless the distance and oscillated between 0.85 – 0.94. Hence, we did not observe the expected characteristic signatures of meiosis. An inversely contrasted situation between outerossing and clonal genomes should be observed for the 4-gametes test. Taking SNP markers that exist in two states among the 11 isolates, the proportion of pairs of markers that pass the 4-gametes test (i.e. that represent the 4 products of meiosis) should rapidly increase with distance between the~~

~~markers, in case of recombination. In contrast, in the absence of recombination, no trend for an increase of the proportion of pairs or markers passing the 4-gametes test with distance between markers should be observed. By conducting an analysis of 2-states markers, we observed no trend for a change in the proportion of markers passing the test with distance. In contrast, the distribution remained flat and close to a value of 0.0. Again, this trend does not correspond to the expected characteristic of meiotic recombination.~~

~~To assess the sensitivity of our methods in finding evidence for recombination, we conducted the same analyses (LD and 4-gametes tests) in the outcrossing diploid meiotic plant-parasitic nematode *Globodera rostochiensis* (Eves-van den Akker et al. 2016). Because the *G. rostochiensis* genome assembly mostly consists of merged paternal and maternal haplotypes, we had to phase the SNPs before conducting LD and 4-gametes tests. The results were totally contrasted between *M. incognita* and *G. rostochiensis* (Figure 6). In *G. rostochiensis*, the LD and 4-gametes curves started at relatively lower (<0.7) and higher (>0.15) values, respectively (Figure 6). Furthermore, we observed a rapid exponential decrease of r^2 in the first kb for LD. At an inter-marker distance of 3kb the r^2 value was <0.37. In parallel, we observed a concomitant rapid and exponential increase in the proportion of markers passing the 4-gametes test, which was >0.38 at the same inter-marker distance. Hence, while *G. rostochiensis* appears to display all the expected characteristics of meiotic recombination, this was not the case for *M. incognita*. This confirms at a whole-genome scale the lack of evidence for meiosis previously observed at the cytological level in *M. incognita*.~~

Addition of further geographical isolates does not increase the genomic diversity and confirms the lack of association between genetic distance and biological traits

To investigate more widely the diversity of *M. incognita* isolates in relation to their mode of reproduction and other biological traits, we included whole-genome sequencing data for additional geographical isolates ~~recently published~~ (Szitenberg et al. 2017). These genome data included one isolate from Ivory Coast, one from Libya, one from Guadeloupe (French West Indies) and five from the USA ~~(Szitenberg et al. 2017)~~ (Figure 1). We pooled these eight new isolates with the eleven Brazilian isolates produced as part of this analysis as well as the *M. incognita* Morelos strain (reference genome) and performed a new PCA with the same methodology. Astonishingly, adding these new isolates recovered the same separation in three distinct clusters (A, B and C) (Figure 7). All the new isolates from additional and more diverse geographical origins fell in just two of the previous Brazilian clusters (A and C). Cluster A that previously contained R2-1 alone, now encompasses the Ivory Coast, Libyan and Guadeloupe isolates. Cluster C that previously contained eight of the Brazilian isolates and covered all the

host races now includes the five US isolates as well as the Mexican isolate (Morelos, reference genome). Cluster B remains so far Brazilian-specific with only R3-2 and R4-4 in this cluster. Addition of these new geographical isolates (Szitenberg et al. 2017) did not substantially increase the number of detected variable positions in the genome. Analyses ran with this whole set of available *M. incognita* isolates also further supported the lack of association of ~~SNP~~SNV-based clusters with biological traits such as host races, nature of the host of origin and geographical distribution (Text S1, Fig. S7).

Discussion

Is the parasitic success of *M. incognita* an evolutionary paradox? This proposition would be true only if *M. incognita* is adaptive despite having a fully parthenogenetic reproduction. Our results ~~seem to~~ support these two aspects.

The lack of sexual reproduction in *M. incognita* was so far only assumed based upon initial cytogenetic observations (Triantaphyllou 1981; Triantaphyllou 1985) but never further supported at whole-genome scale. Here, the different analyses we performed at the population genomics level converge in supporting the lack of recombination and genetic exchanges in *M. incognita*. The phylogenetic network analysis based on ~~homozygous~~ fixed SNV~~P~~s returned a bifurcating tree that separated the different isolates and not a network. This suggests a lack of genetic exchange between the isolates. In sexual 'recombining' species, the mitochondrial genome accumulates mutations much faster than the nuclear genome. This is also true in the model nematode *C. elegans* where the mitochondrial mutation rate is at least two orders of magnitude higher than the nuclear mutation rate (Denver et al. 2004; Denver et al. 2009). The higher mitochondrial accumulation of mutations is supposed to be the combined result of extremely rare or total lack of recombination, the low effective population size and the effectively haploid inheritance in mitochondria (Neiman Maurine and Taylor Douglas R 2009). In *M. incognita*, as opposed to *C. elegans*, we found that the percentage of variable positions in the mitochondrial genome is only one order of magnitude higher than in the nuclear genome. This suggests that the nuclear genome evolves at a comparable rate to the mitochondrial genome and reinforces the idea that the nuclear genome is mostly effectively haploid and non-recombining. Theoretically, the efficacy of selection should be lower in non-recombining species than recombining ones. We showed that the ratio of diversity at non-synonymous sites / diversity at synonymous sites (π_n/π_s) was indeed one order of magnitude ~~~3 times~~ higher in *M. incognita* than in two outcrossing Caenorhabditis species. Finally, the proportion of markers passing the 4-gametes test and linkage disequilibrium did not show the exponential ~~decrease~~increase, respectively ~~increase~~decrease, with physical distance as expected under recombination. In contrast, a rapid exponential decrease of linkage

disequilibrium was recently observed and considered as an evidence for recombination in the bdelloid rotifer *Adineta vaga* (Vakhrusheva et al. 2018). Collectively, these results strongly suggest absence (or extremely rare) recombination and support the mitotic parthenogenetic reproduction of *M. incognita*.

Despite its clonal reproduction, ~~I~~it was already evident that *M. incognita* has an adaptive potential. Indeed, experimental evolution assays have shown the ability of *M. incognita* to overcome resistance conferred by the Mi gene in tomato in a few generations (Castagnone-Sereno et al. 1994; Castagnone-Sereno 2006). Naturally virulent *M. incognita* populations (i.e. not controlled by the resistance gene) have also been observed in the fields and probably emerged from originally avirulent populations (Verdejo-Lucas et al. 2012; Tzortzakakis et al. 2014; Barbary et al. 2015), although it is unknown if this resistance breaking is as rapid as under controlled lab conditions. However, adapting from a compatible host plant to another very different incompatible plant is certainly more challenging ~~that~~ than breaking down a resistance gene in a same plant. Here, we showed that the different host races defined in *M. incognita* as a function of patterns of (in)compatibilities with different plants do not follow a phylogenetic signal. This ~~could~~ would ~~represent~~ imply multiple independent gains and losses of parasitic abilities to arrive at the current phylogenetic distribution of host compatibility patterns (i.e. host races). Whether these multiple gains and losses occurred from a hyper-polyphagous common ancestor or an ancestor with a more restricted host range remains to be clarified. ~~adaptations or gradual loss of compatibilities from a hyper-polyphagous ancestor.~~ To address this question, we have reconstructed host compatibilities at each ancestral node based on the SNP-based phylogenetic classification of the *M. incognita* isolates (Fig. S8). This reconstruction showed that the two hypotheses concerning the host range status of the last common ancestor were equally likely. Addition of other isolates characterized for their host race might allow to favour one or the other hypothesis in the future. ~~This maximum-likelihood reconstruction shows that the hypothesis of a hyper-polyphagous ancestor that then progressively lost ability to either parasitize Cotton or Tobacco is not realistic. The ancestral reconstruction rather suggests multiple independent gains and losses of the ability to parasitize these plants.~~ Consistent with ~~this suggestion~~multiple gains and losses of parasitic abilities, host race switching within an isolate over time has already been observed. Isolates of *M. incognita* race 2 and 3, which parasitize tobacco and cotton plants respectively, switched to behaviour similar to race 3 and 2 after staying for eight months on coffee plants (Rui Gomes Carneiro, personal communication). Together with the previously reported ability to break down resistance gene in plants, the ability of *M. incognita* to loose and gain ability to infect different plants highlights its adaptive potential.

~~The lack of sexual reproduction in *M. incognita* was so far only assumed based upon initial cytogenetic observations (Triantaphyllou 1981; Triantaphyllou 1985) but never further supported at whole-genome scale. Here, the different analyses we performed at the population genomics level converge in supporting the lack of recombination and genetic exchanges in *M. incognita*. The phylogenetic network analysis based on homozygous SNPs returned a bifurcating tree that separated the different isolates and not a network. This suggests a lack of genetic exchange between the isolates. In sexual 'recombining' species, the mitochondrial genome accumulates mutations much faster than the nuclear genome. This is also true in the model nematode *C. elegans* where the mitochondrial mutation rate is at least two orders of magnitude higher than the nuclear mutation rate (Denver et al. 2004; Denver et al. 2009). The higher mitochondrial accumulation of mutations is supposed to be the combined result of extremely rare or total lack of recombination, the low effective population size and the effectively haploid inheritance in mitochondria (Neiman Maurine and Taylor Douglas R 2009). In *M. incognita*, as opposed to *C. elegans*, we found that the percentage of variable positions in the mitochondrial genome is only one order of magnitude higher than in the nuclear genome. This suggests that the nuclear genome evolves at a comparable rate to the mitochondrial genome and reinforces the idea that the nuclear genome is mostly effectively haploid and non-recombining. Theoretically, the efficacy of selection should be lower in non-recombining species than recombining ones. We showed that the ratio of diversity at non-synonymous sites / diversity at synonymous sites (π_n/π_s) was indeed ~3 times higher in *M. incognita* than in two outcrossing Caenorhabditis species. Finally, the proportion of markers passing the 4-gametes test and linkage disequilibrium did not show the exponential decrease, respectively increase, with physical distance as expected under recombination. In contrast, a rapid exponential decrease of linkage disequilibrium was recently observed and considered as an evidence for recombination in the bdelloid rotifer *Adineta vaga* (Vakhrusheva et al. 2018). Collectively, these results strongly suggest absence (or extremely rare) recombination and support the mitotic parthenogenetic reproduction of *M. incognita*.~~

Overall, we provided here additional evidence for adaptability and the first whole-genome level assessment for the lack of recombination in *M. incognita*, consolidating this species as a main model to study the paradox of adaptability and parasitic success in the absence of sexual reproduction.

The adaptability of *M. incognita* despite its obligatory asexual reproduction and the lack of phylogenetic signal underlying the host races have important practical implications at the agricultural level. Characterizing populations that differ in their ability to infest a particular host (that carries specific resistance genes) is of crucial importance for growers and agronomists. Indeed, the main *Meloidogyne* spp. control strategies consist in deploying resistant cultivars and appropriate crop rotation against a specific given race. If the identity of a population is unknown,

the crop selected for use in a management scheme may cause dramatic increases in nematode populations (Hartman and Sasser 1985). However, the adaptability of *M. incognita* casts serious doubts on the durability of such strategies and must be taken into account in rotation schemes. Furthermore, the biological reality of host races themselves is challenged by the lack of underlying genetic signal. Actually, the initial host race concept, has never been universally accepted, in part because it covered only a small portion of the whole potential variation in parasitic ability (Moens et al. 2009). Although *M. incognita* was already known to parasitize hundreds of host plants, only six different host standards were used to characterise four races. New host races might be defined in the future when including additional hosts in the differential test. Furthermore, using the same six initial host plant species, two additional *M. incognita* races that did not fit into the previously published race scheme have already been described (Robertson et al. 2009). Although the terminology 'races' of *Meloidogyne* spp. has been recommended not to be used since 2009 (Moens et al. 2009), several papers related to *M. incognita* diversity of host compatibility or selection of resistant cultivars are still using this term; including on coffee (Lima et al. 2015; Peres et al. 2017); cotton (Mota et al. 2013; da Silva et al. 2014) or soybean (Fourie et al. 2006). This reflects the practical importance to differentiate *M. incognita* populations according to their different ranges of host compatibilities. However, because these variations in host ranges are not monophyletic and thus do not follow shared common genetic ancestry, we recommend ~~to abandon~~abandoning the term 'race'.

Another main question relates to the level of intra-specific genome polymorphism required to cover the different ranges of host compatibilities in *M. incognita* and their ability to survive in different environments, despite their clonal reproduction. In this study, we found that the cumulative ~~homozygous~~ fixed divergence across the eleven isolates from Brazil and the reference genome (sampled initially from Mexico) reached ~0.02% of the nucleotides. Addition of isolates from Africa, the West Indies and the USA did not increase the maximal divergence. This relatively low divergence is rather surprising, considering the variability in terms of distinct compatible host spectra (host races). Host-specific ~~SNPs~~ SNV were found only for Race 2 and no functional consequence for these SNPs could be found, as they did not fall in predicted coding or evident regulatory regions. Furthermore, the existence of race-specific ~~SNPs~~ SNV themselves is even questionable as addition of other isolates might disqualify the few Race 2-specific SNPs in the future. Similarly, when grouped by host plant species there were no disruptive variations identified in the coding regions, we found no SNV associated to cotton, only one synonymous variant for soybean, and only one synonymous variant for tobacco.

Collectively, our observations indicate that *M. incognita* is versatile and adaptive despite its clonal mode of reproduction. The relatively low divergence at the ~~SNP~~ SNV level suggests acquisition of point and short scale mutations followed by selection of the fittest haplotype is

probably not the main or at least not the sole player in the adaptation of this species to different host plants and environments. Other mechanisms such as epigenetics, copy number variations, movement of transposable elements or large-scale structural variations could be at play in the surprising adaptability of this clonal species. Consistent with this idea, convergent gene copy number variations (CNV), have recently been shown to be associated with adaptation to a resistance gene-bearing plant in *M. incognita* (Castagnone-Sereno et al. 2019). Interestingly, the parthenogenetic marbled crayfish has multiplied by more than 100 its original area of repartition across Madagascar, adapting to different environments despite showing a surprisingly low number of nucleotide variation (only ~400 ~~SNPs~~ SNV on a ~3Gb genome representing a proportion of variable positions of 1.3 10^{-7} only). This also led the authors to suggest that mechanisms other than acquisition of point mutations and selection of the fittest haplotype must be involved (Gutekunst et al. 2018).

Previously, we have shown that the genome structure of *M. incognita* itself, could participate in its versatility. Indeed, being allopolyploid, *M. incognita* has >90% of its genes in multiple copies. The majority of these gene copies show diverged expression patterns one from the other and signs of positive selection between the gene copies have been identified (Blanc-Mathieu et al. 2017). How the expression patterns of these gene copies vary across different geographical isolates with different host compatibilities would be interesting to explore in the future.

Material and Methods

Purification, species identification of *M. incognita* isolates and determination of host races.

The *M. incognita* isolates involved in this study (Table 1) originate from populations collected from different crops and geographically distant origins in Brazil (Figure 1). For each isolate, one single female and its associated egg mass were retrieved as explained in (Carneiro & Almeida 2001). To determine the species (here *M. incognita*), we used esterase isozyme patterns on the female (Carneiro et al. 2000). The corresponding single egg mass was used for tomato plant infection and multiplication. –We reproduced egg-mass isolates on tomato plants (*Solanum lycopersicum* L. cv. Santa Clara) under greenhouse conditions at a temperature of 25-28 ° C. After three months, we confirmed the *M. incognita* species using esterase phenotypes (Carneiro and Almeida 2001). Once enough nematodes were multiplied, a pool was collected and we performed the North Carolina Differential Host Test (NCDHT) (Hartman and Sasser 1985) with the following plants: cotton cv. Deltapine 61, tobacco cv. NC95, pepper cv. Early California Wonder, watermelon cv. Charleston Gray, peanut cv. Florunner and tomato cv. Rutgers to determine the

host race status. We inoculated these plants with 5,000 eggs and J2 of *M. incognita* and maintained them under glasshouse conditions at 25-28° C for three months, with watering and fertilisation as needed. Two months after inoculation, the root system was rinsed with tap water, and egg masses were stained with Phloxine B (Hartman and Sasser 1985) to count the number of galls and eggs masses separated for each root system. We assigned a rating index number according to the scale: 0 = no galls or egg masses; 1 = 1-2 galls or egg masses; 2 = 3-10 galls or egg masses; 3 = 11-30 galls or egg masses; 4 = 31-100 galls or egg masses; and 5 > 100 galls or egg masses per root system (Table 1). Host–plants types that have an average gall and egg mass index of 2 or less are designated non-host (-). The other plants (index ≥ 3) are designated hosts (+). We categorised *M. incognita* host races based on their ability to parasitize tobacco and cotton (Table 1). Classically, the index for Rutgers tomato (susceptible control) is higher than 4 (+) (Hartman & Sasser, 1985). The rest of the population was kept for multiplication on tomato plants to produce enough nematodes for sequencing (typically >1 million individuals pooled together).

DNA preparation and SCAR test

For each characterized nematode isolate, we extracted and purified the genomic DNA from pooled eggs with the supplement protocol for nematodes of the QIAGEN Gentra® Puregene® Tissue Kit with the following modifications: incubation at 65 ° C in the cell lysis buffer for 30 min and incubation at 55 ° C with proteinase K for 4h. We verified DNA integrity on 0.8% agarose gel and the DNA quantification on Nanodrop. We confirmed isolate species purity by SCAR-PCR (Zijlstra et al. 2000; Randig et al. 2002) using the SCAR primers specified in Table ~~S6~~ S5 for the RKN *M. javanica, M. paranaensis, M. incognita,* and *M. exigua*.

Sequencing library preparation

We assessed input gDNA quantity using Qubit and normalised the samples to 20ng/ul as described in TruSeq®DNA PCR-Free Library Prep Reference Guide (#FC-121-3001, Illumina) prior fragmentation to 350bp with Covaris S2. We assessed the quality of fragments after size selection and size control of the final libraries using High Sensitivity DNA Labchip kit on an Agilent 2100 Bioanalyzer.

Whole genome sequencing

We quantified isolated sample libraries with KAPA library quantification kit (#7960298001, Roche) twice with two independent dilutions at 1:10,000 and 1:20,000. We calculated the average concentration of undiluted libraries and normalised them to 2nM each then pooled them for sequencing step.

We generated high-coverage genomic data for the 11 *M. incognita* isolates by 2x150 bp paired-end Illumina NextSeq 500 sequencing with High Output Flow Cell Cartridge V2 (#15065973, Illumina) and High Output Reagent Cartridge V2 300 cycles (#15057929, Illumina) on the UCA Genomix sequencing platform, in Sophia-Antipolis, France. We performed two runs to balance the read's representation among the isolates and obtain homogeneity of coverage for the different samples (Table S2).

Variant Detection

We trimmed and cleaned the reads from each library with cutadapt tool (Martin 2011) to remove adapter sequences and bases of a quality inferior to 20. We ~~aligned~~ mapped the clean reads to the *M. incognita* reference genome (Blanc-Mathieu et al. 2017), using the BWA-MEM software package (Li 2013). This reference genome is described as triploid with three equally highly diverged A, B and C genome copies as a result of hybridization events. Most of the ~~duplicated~~ triplicated regions have been correctly separated during genome assembly, according to genome assembly size (183.53 Mb) that is in the range of the estimated total DNA content in cells via flow cytometry (189±15Mb) (Blanc-Mathieu et al. 2017)). Hence, the genome was considered in this analysis as mostly haploid. However, the distribution of per-base coverage on the genome assembly presented a 2-peaks distribution with a second minor peak at ~twice the coverage of the main peak (Fig. S2). Genome regions of double coverage most likely represent cases where two of the three ~~collapsed~~ homoeologous loci have been collapsed during the assembly, probably due ~~of~~ to lower divergence. ~~and~~ Such regions will systematically be responsible for 'artefactual' ~~heterozygous~~ 0/1 ~~SNPs~~ SNV (presenting variations within isolates) as the reads from the two homoeologous copies will map a single collapsed region in the reference genomes. To avoid confusion between ~~SNPs~~ SNV representing true variations between individuals within isolates from those being artefacts due to collapsed homoeologous regions, ~~heterozygous (~~0/1~~)~~ ~~SNPs~~ SNV were discarded from the analysis and only ~~homozygous (~~1/1~~)~~ ~~SNPs~~ SNV fixed within isolates were considered.

We used SAMtools (Li et al. 2009) to filter alignments with MAPQ lower than 20, sort the alignment file by reference position, and remove multi-mapped alignments.

We used the FreeBayes variant detection tool (Garrison and Marth 2012) to call ~~SNPs~~ SNV and small-scale insertions/deletions, incorporating all the library alignment files simultaneously and produced a variant call file (VCF). We filtered the resulting VCF file with the vcffilter function of vcflib (Anon 2018) , retaining the positions that had more than 20 Phred-scaled probability (QUAL) and a coverage depth (DP) > 10. To conduct the same analyses on the genome of the meiotic diploid nematode *Globodera pallida*, we first phased the ~~SNPs~~ SNV to

haplotypes using WhatsHap (Martin et al. 2016) because the genome assembly mainly consist of collapsed paternal and maternal haplotypes.

Genetic tests for detection of recombination

We used custom made scripts (cdf. Data Availability section) to calculate the proportion of fixed markers passing the 4-gametes test and Linkage Disequilibrium (LD) r^2 values as a function of inter-marker distance along the *M. incognita* genome scaffolds.

Genetic diversity between isolates, clusters and efficacy of purifying selection

We used ~~SNPGenic (Nelson et al. 2015)~~ bpppopstats from the Bio++ libraries (Guéguen et al. 2013) to estimate the nucleotide variability at non-synonymous and synonymous sites as well as efficacy of purifying selection (π_N , π_S and π_N/π_S) ~~for each isolate using as input the reference genome, the gene annotation file, and the variant call file~~using a multiple alignment of the coding regions. We calculated fixation index (F_{ST}) for the three clusters using vcftools (Danecek et al. 2011).

~~Genetic tests for detection of recombination~~

~~We used custom made scripts (cd. Data Availability section) to calculate the proportion of markers passing the 4-gametes test and Linkage Disequilibrium (LD) r^2 values as a function of inter-marker distance along the *M. incognita* genome scaffolds.~~

Principal component analysis

We performed a principal component analysis (PCA) to classify the isolates according to their ~~SNP~~ SNV patterns and mapped the race characteristics, geographic location, or current host plants on this classification. We used the filtered VCF file as input in the statistical package SNPRelate (Zheng et al. 2012) to perform the PCA with default parameters.

Phylogenetic analysis

Based on the VCF file and the *M. incognita* gene predictions (Blanc-Mathieu et al. 2017), we selected 85,413 positions that contained synonymous or non-synonymous mutations (i.e. in coding regions). We aligned these positions and then used them as an input in SplitsTree4 with default parameters. The resulting network produced a bifurcating tree that was identical to the one obtained with RAxML-NG using GTR+G+ASC_LEWIS model. The bifurcating tree was used as input to PastML (Ishikawa et al. 2018) for reconstruction of the ancestral states of ability to parasitize tobacco and cotton (Fig. S8). Phylogenetic inferences for the largest scaffolds

containing at least 20 ~~SNPs~~ SNV and the mitochondrial genome were conducted with RAxML-NG (Kozlov et al. 2019) using the GTR+G substitution model (except for scaffolds 10 and 20 for which the K80+G model was used because not enough phylogenetically informative positions were available).

Test for association between biological traits and genetic clusters

We used a Fisher's exact test in R to assess whether there was a significant association between the ~~SNP~~SNV-based clusters and the host races or the crop species from which the isolates were originally collected. We also conducted an Isolation By Distance (IBD) analysis using the adegenet R package (Jombart and Ahmed 2011) to check how well the genetic distances correlate with geographic distances between the sampling points of the isolates. Geographic distances were calculated from exact sampling locations, when available, or centre points if the region was known but not the exact sampling location. Sample R3-4 was excluded from this analysis since it was a mix of samples pooled together from different geographical locations. L27 was also excluded since the sampling location was unknown.

Mitochondrial genome analysis

We sub-sampled genomic clean reads to 1% of the total library for each *M. incognita* isolate. Then, we assembled them independently using the plasmidSPAdes assembly pipeline (Antipov et al. 2016). We extracted the mitochondrial contigs based on similarity to the *M. incognita* reference mitochondrial genome sequence (NCBI ID: NC_024097). In all cases, the mitochondrion was assembled in one single contig. We identified the two repeated regions (63 bp repeat and 102 bp repeat), described in (Okimoto et al. 1991) and we calculated the number of each repeat present in these regions.

Acknowledgements

This work was supported by the Agence Nationale de la Recherche program (grant number ANR-13-JSV7-0006—ASEXEVOL), the Insitut National de la Recherche Agronomique program (grant number Accueil de Chercheur Etranger SPE 2016). GK has received the support of the EU in the framework of the Marie-Curie FP7 COFUND People Programme, through the award of an AgreenSkills+ fellowship (under grant number 609398), EM received a post-doc fellowship from the Consórcio Pesquisa Café, EVSA received funding from Embrapa for sequencing at UCA Genomix platform. We thank the SPIBOC Bioinformatics platform for the data hosting.

Figure Legends

Figure 1. World map showing geographical origins for all samples used in the study.

Expanded map of Brazil showing the states where the 11 isolates ~~that were~~ sequenced in this study were collected. Each state is highlighted with a different colour. The countries listed in the literature for other sequenced genomes are completely coloured. The cultures from which the samples were isolated are illustrated by photographs, which are pointed by arrows coming from the name of the respective isolate. Only 3 of the isolates described previously in the literature have their culture of origin reported. The names of the Brazilian isolates are in 4 different colour sources for each race (race 1 in green, 2 in red, 3 in grey and 4 in blue). The names of the isolates of the literature are written in white or black.

Figure ~~5~~2. Linkage Disequilibrium (red) and 4-gametes test of *M. incognita* isolates.

Based on SN~~P~~Vs ~~that were homozygous~~ fixed within isolates: proportion of pair of markers that pass the 4-gametes test (blue) and linkage disequilibrium measured as r2 between markers (red), both as a function of the inter-markers distance on *M. incognita* scaffolds.

Figure ~~6~~3. Linkage Disequilibrium (red) and 4 gametes test (blue) for phased SN~~P~~Vs in *Globodera rostochiensis* isolates.

The r^2 correlation between markers, indicating linkage disequilibrium (LD) is given in the red (upper) plot, as a function of the physical distance between the markers. The proportion of pairs of two-state markers that pass the 4-gamete test is given in the blue (lower) plot as a function of the distance between the markers.

Figure ~~2~~4. Distribution of the number of variants per race and isolate.

Number of variants per isolate and isolate-specific variants for the 11 Brazilian isolates.

Figure ~~3~~5. PCA analysis of the different *M. incognita* isolates groups them into three clusters (A, B, and C).

The geographic origins are associated to coloured shapes*:* black circle: Paraná, orange diamond: Santa Catarina, green square: São Paulo, red triangle: Mato Grosso, blue star: pool. Host plant representative pictures are displayed next to the isolates: soybean pod (R1-2 and R3-2); cotton

flower (R3-1, R3-4, R4-4, and R4-1); coffee grain (R2-6); cucumber vegetable (R1-3); tobacco leaves (R1-6 and R2-1); and watermelon fruit slice (R4-3).

Figure 46. Phylogenetic network for *M. incognita* isolates based on ~~SNPs~~ SNV present in coding sequences.

The phylogenetic network based only on changes in coding sequences shows the same grouping into 3 distinct groups.

~~Figure 5. Linkage Disequilibrium (red) and 4-gametes test of *M. incognita* isolates.~~

~~Based on SNPs that were homozygous within isolates: proportion of pair of markers that pass the 4-gametes test (blue) and linkage disequilibrium measured as r2 between markers (red), both as a function of the inter-markers distance on *M. incognita* scaffolds.~~

~~Figure 6. Linkage Disequilibrium (red) and 4-gametes test (blue) for phased SNPs in *Globodera rostochiensis* isolates.~~

~~The r^2 correlation between markers, indicating linkage disequilibrium (LD) is given in the red (upper) plot, as a function of the physical distance between the markers. The proportion of pairs of two-state markers that pass the 4-gamete test is given in the blue (lower) plot as a function of the distance between the markers.~~

Figure 7. PCA analysis of all known *M. incognita* genomes.

The isolates were regrouped based on ~~SNP~~ SNV patterns confirming the same three clusters. Origin countries are indicated by flags (Brazil for R1-2, R1-3, R1-6, R2-1, R2-6, R3-1, R3-2, R3-4, R4-1, R4-3, R4-4; USA for L27, 557R, HarC, W1, VW6; Mexico for Morelos; Libya for A14; Ivory Coast for L9; Guadeloupe for L19).

Data Availability

All the sequence data generated during this study have been deposited at the NCBI under GEO accession GSE116847 and available at this URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116847

The different scripts and R codes used to process the data are available on GitHub at the following URL:https://github.com/GDKO/gdk_scripts/tree/master/popgenvcf

Author contributions

EGJD, PC-S, GDK, EVSA and LD contributed to the design of the project; RMDGC, ACZM, EM contributed to collection of samples. Classification of samples into races was done by RMDGC, ACZM, EM; library preparations were performed by M-JA; GDK, EGJD, LD, EVSA, MBB and DKK contributed to the analysis and interpretation of data. EGJD and GDK wrote the manuscript with contributions of EVSA, RMDGC, PC-S, M-JA and LD. All authors approved the final manuscript.

Tables & Figures

Table 1. Host race characterization of the 11 *Meloidogyne incognita* isolates used in this study

Race	ID[a]	Host Crop[b]	PID[c]	Geographic origin	Esterase phenotype	NCDHT note[d]						Reference
						tb	tm	wm	pt	pr	ct	
1	R1-2	soybean	INC15	Londrina – PR	I2	0	5	4	0	5	0	(Mattos et al. 2016)
	R1-3	cucumber	ND	Piracicaba – SP	I2	0	5	5	0	4	0	
	R1-6	tobacco	LGM39	Mercedes – PR	I2 / I1	0	5	5	0	4	0	(Filho et al. 2016)
2	R2-1	tobacco	LGM09	Sombrio – SC	I2	5	5	4	0	4	0	(Filho et al. 2016)
	R2-6	coffee	22B	São Jorge do Patrocínio – PR	I1	4	5	5	0	4	1	
3	R3-1	cotton	PR-3	Umuarama – PR	I2	0	5	5	0	5	5	(da Silva et al. 2014)
	R3-2	soybean	ND	Londrina – PR	I2	0	5	5	0	5	5	
	R3-4	cotton (pool)	PR-3 Umu, PR-3 Lon, MTS-R3, BA-R3	Umuarama - PR, Londrina - PR, Dourados - MS, L.E.Magalhães – BA	I2	0	5	5	0	5	5	(da Silva et al. 2014)
4	R4-1	cotton	MT-4	Campo Verde – MT	I2	5	5	5	0	5	5	(da Silva et al. 2014)
	R4-3	watermelon	NG	Londrina – PR	I2	5	5	5	0	5	5	
	R4-4	cotton	GEN 306	Vargem Grande do Sul – SP	I1	5	5	5	0	5	5	

a, isolate identification code

b, host plant crop (soybean, *Glycine max*; cucumber, *Cucumis sativus*; tobacco, *Nicotiana tabacum*; coffee, *Coffea arabica*; cotton, *Gossypium hirsutum*.; watermelon, *Citrullus vugaris*

c, population id

d, host range results for the North Carolina Differential Host Test (NCDHT), numbers represent gall index with 0 = no galls; 1 = 1 to 2; 2 = 3 to 10; 3 = 11 to 30; 4 = 31 to 100; and 5 = more than 100 galls. (tb=tobacco 'NC95', tm=tomato 'Rutgers', wm=watermelon 'Charleston Gray', pt=peanut 'Florunner', pr=pepper 'Early California Wonder', ct=cotton 'Deltapine 61'

Table 2. Number of repeats per region (63nt and 102nt) in the mitochondrial DNA of each isolate, decimals indicate truncated repeats

ID	63nt Region	102nt Region	Location	Host plant
R1-2	7.3	5.5	Londrina - PR	soybean
R1-3	7	13	Piracicaba - SP	cucumber
R1-6	1.2	7	Mercedes - PR	tobacco
R2-1	7	15.4	Sombrio - SC	tobacco
R2-6	7	9	São Jorge do Patrocínio - PR	coffee
R3-1	7	13	Umuarama - PR	cotton
R3-2	14	8.3	Londrina - PR	soybean
R3-4	6	13	Umuarama – PR Londrina – PR Dourados – MS L.E.Magalhães - BA	cotton
R4-1	6	14.7	Campo Verde - MT	cotton
R4-3	3	9	Londrina - PR	watermelon
R4-4	14	8.3	Vargem Grande do Sul - SP	cotton

759 Figure 1. World map showing geographical origins for all samples used in the study.

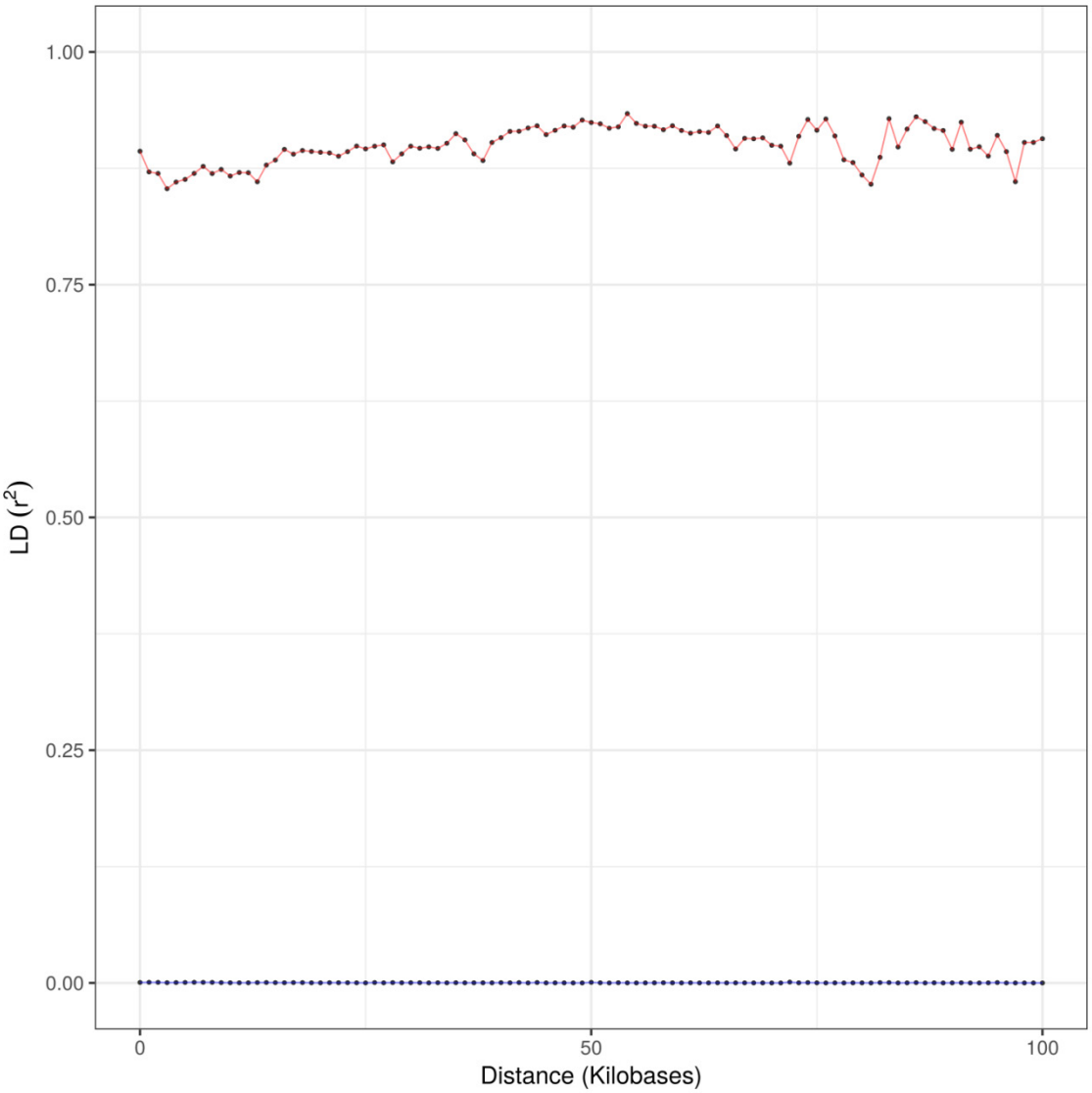


Figure 2. Linkage Disequilibrium (red) and 4-gametes test (blue) of *M. incognita* isolates.

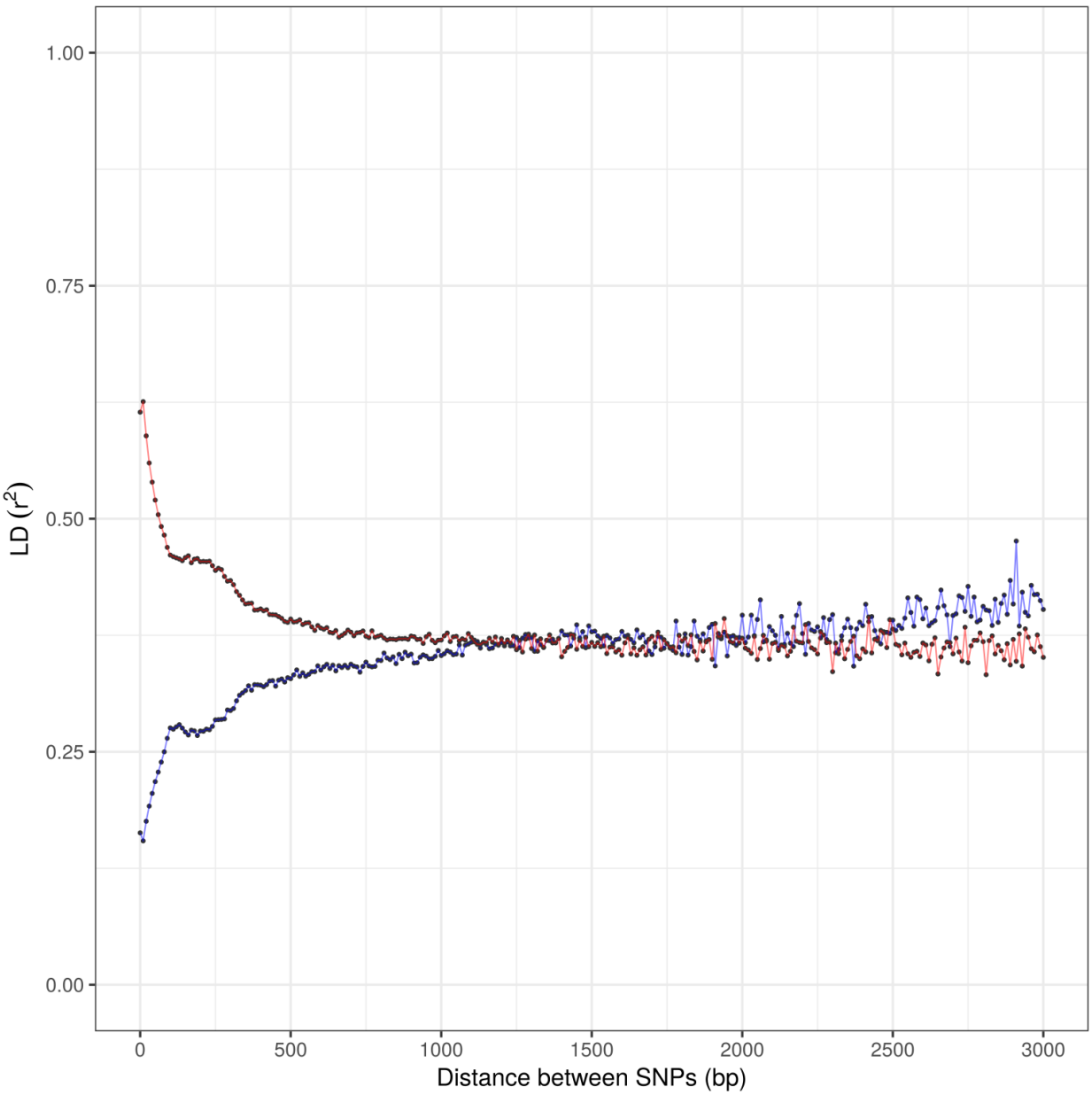


Figure ~~6~~3. Linkage Disequilibrium (red) and 4 gamete test (blue) for phased SN~~P~~Vs in *Globodera rostochiensis* isolates.

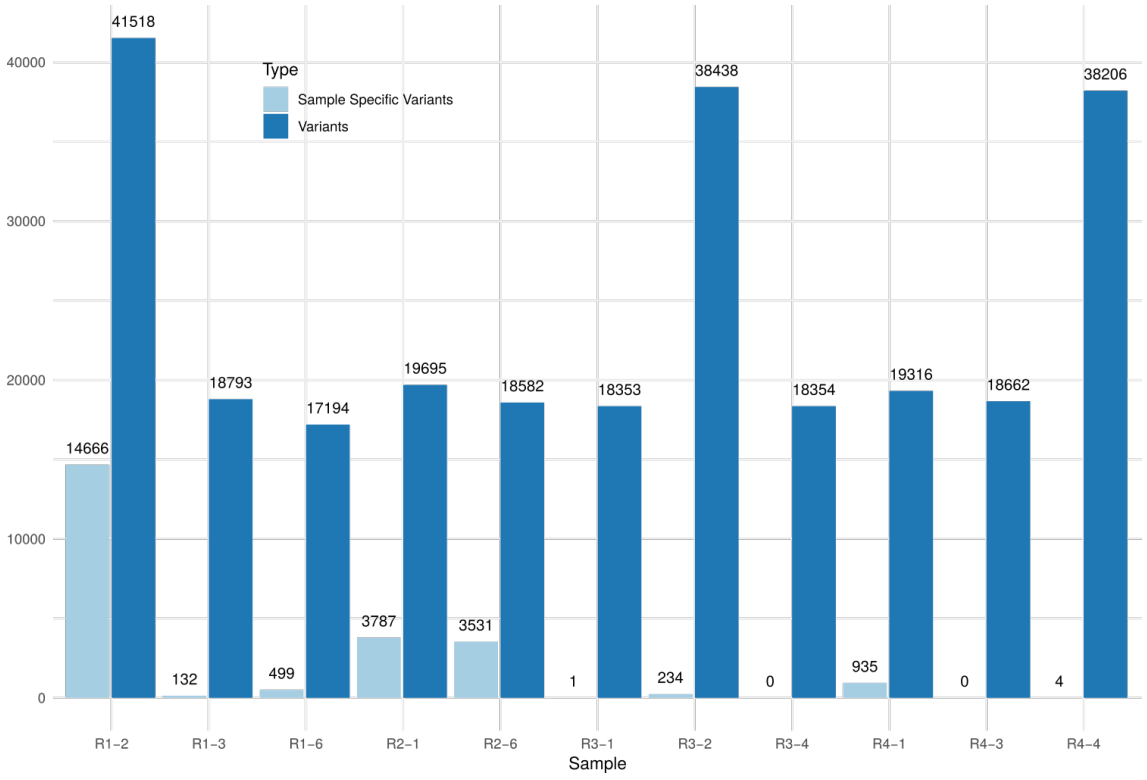


Figure ~~2~~4. Distribution of the number of variants per race and isolate.

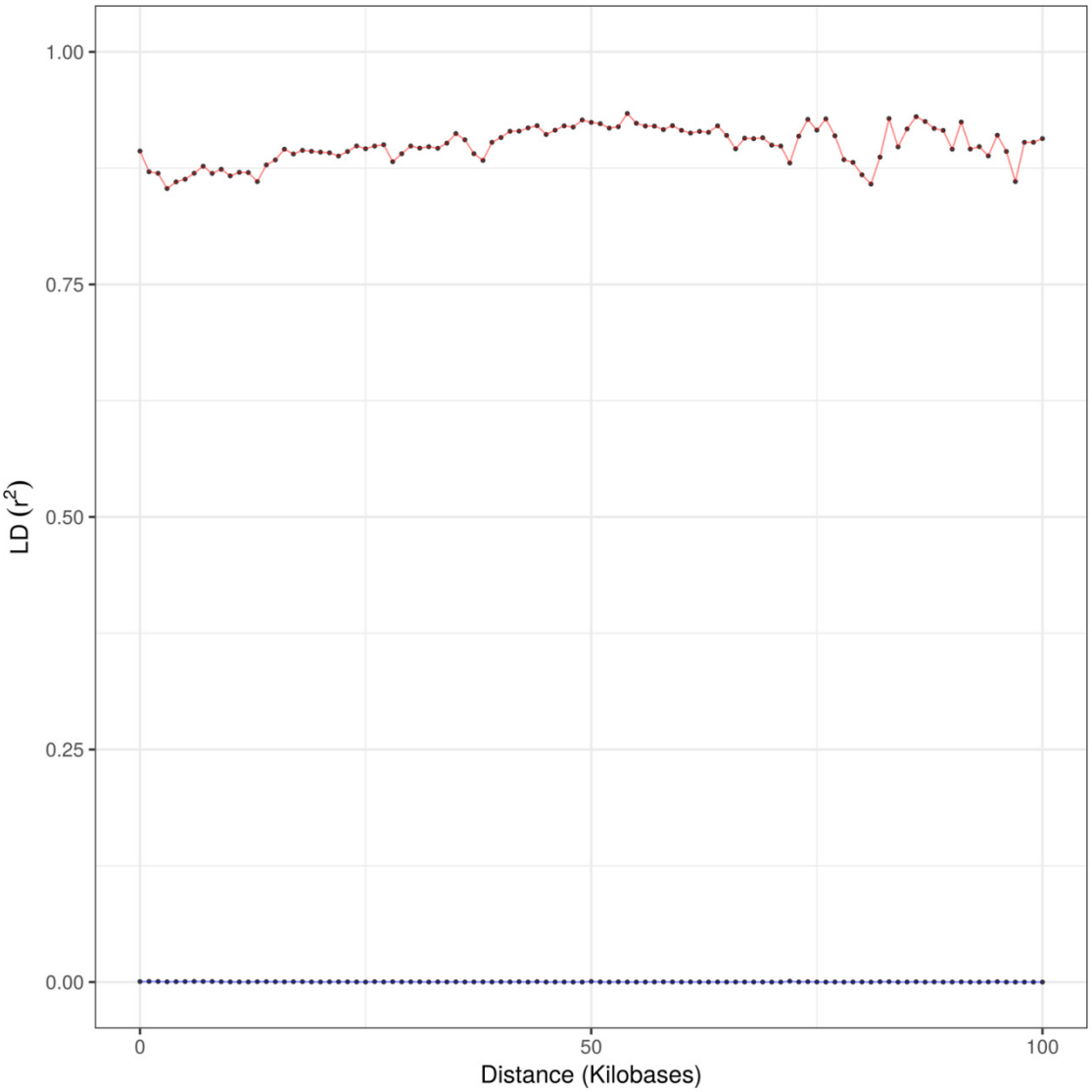


Figure 5. Linkage Disequilibrium (red) and 4-gametes test (blue) of *M. incognita* isolates.

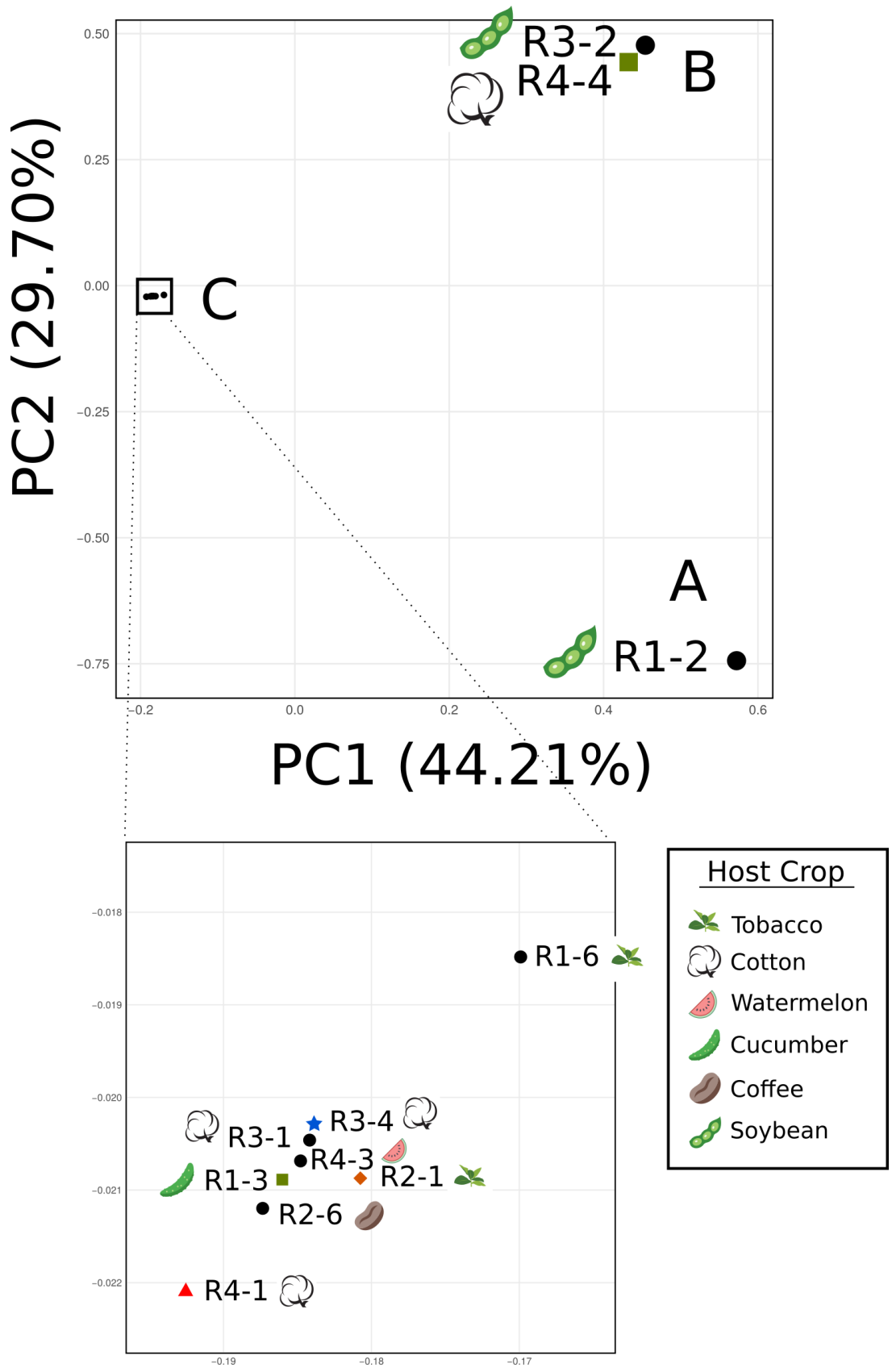
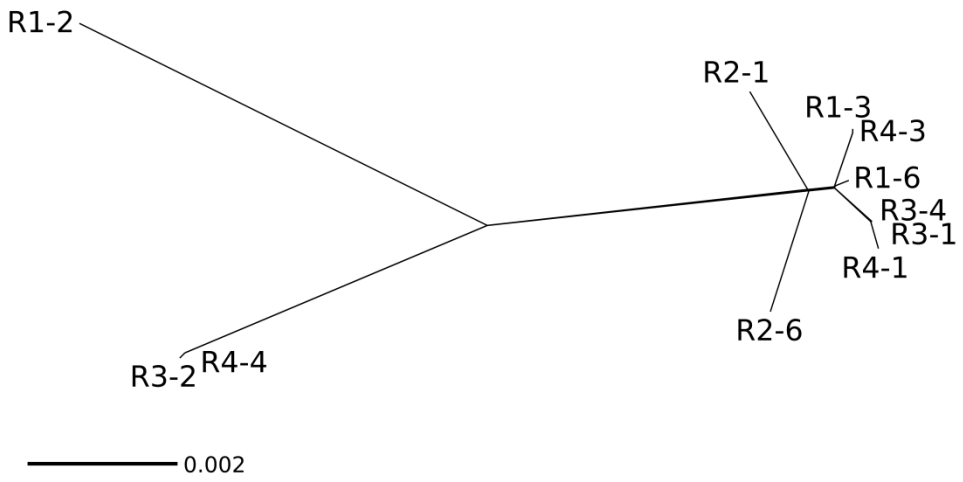


Figure ~~3~~5. PCA analysis of the different *M. incognita* isolates groups them into three clusters (A, B, and C).



777

778 Figure ~~4~~6. Phylogenetic network for *M. incognita* isolates based on ~~SNPs~~ SNV

779 present in coding sequences.

780

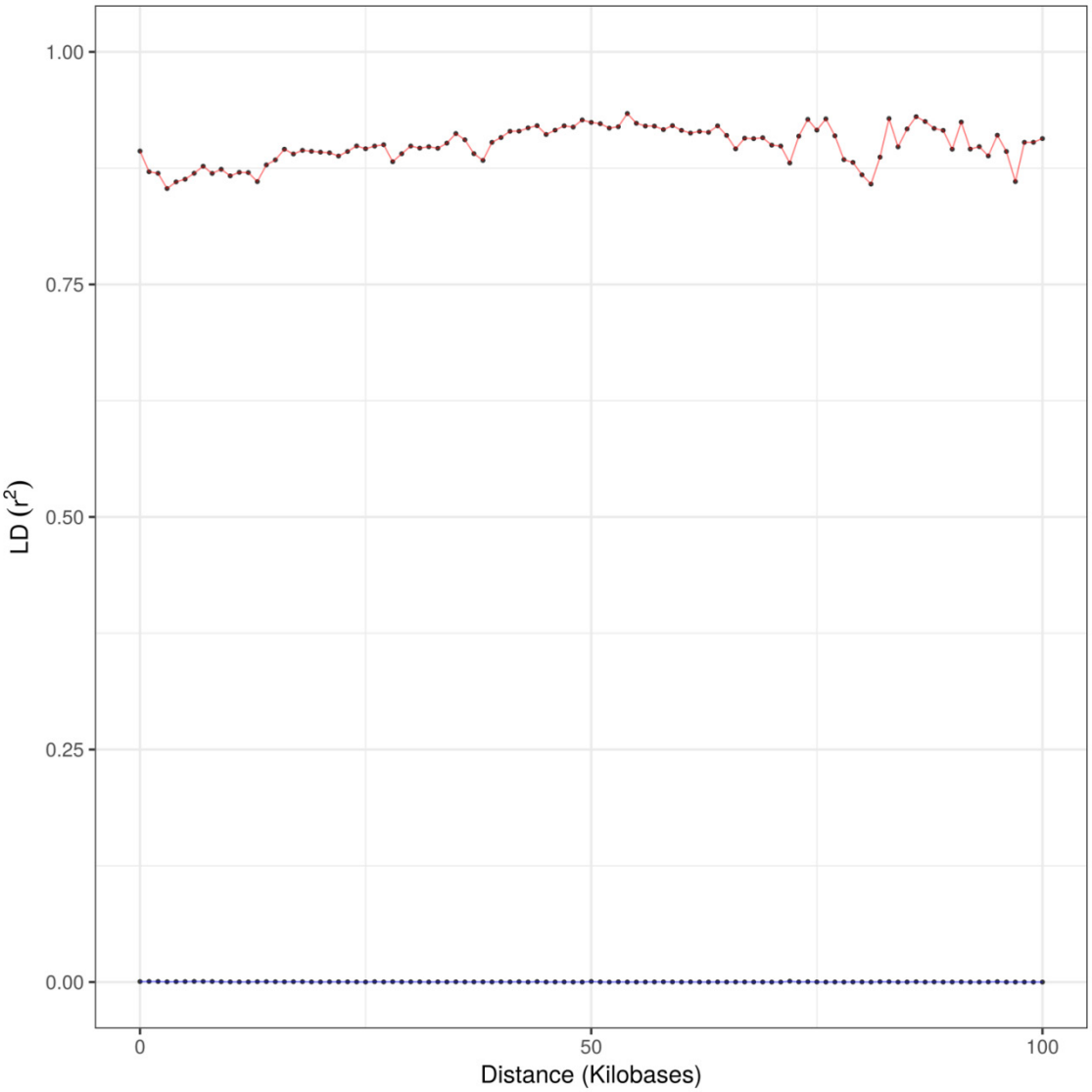
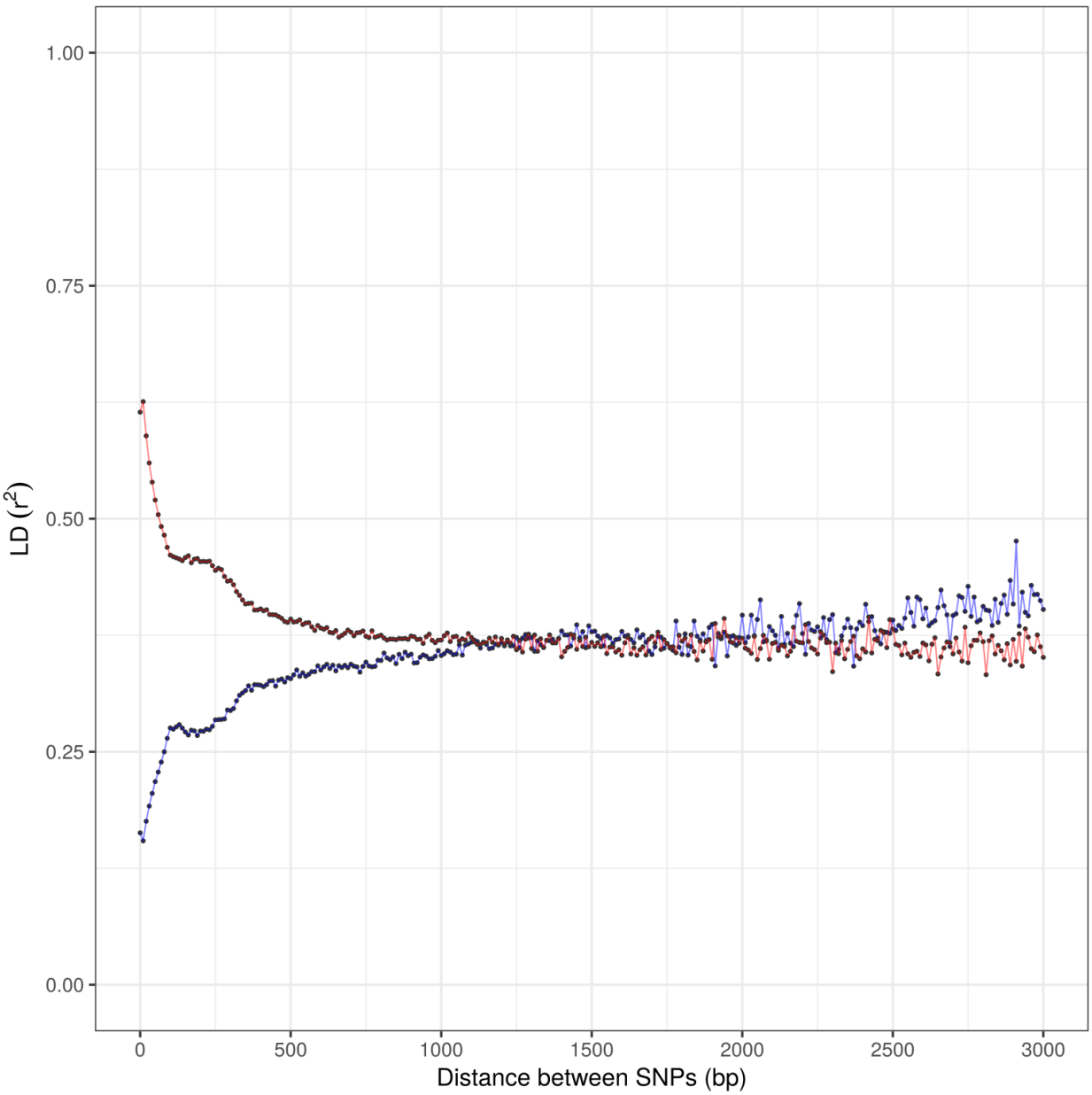
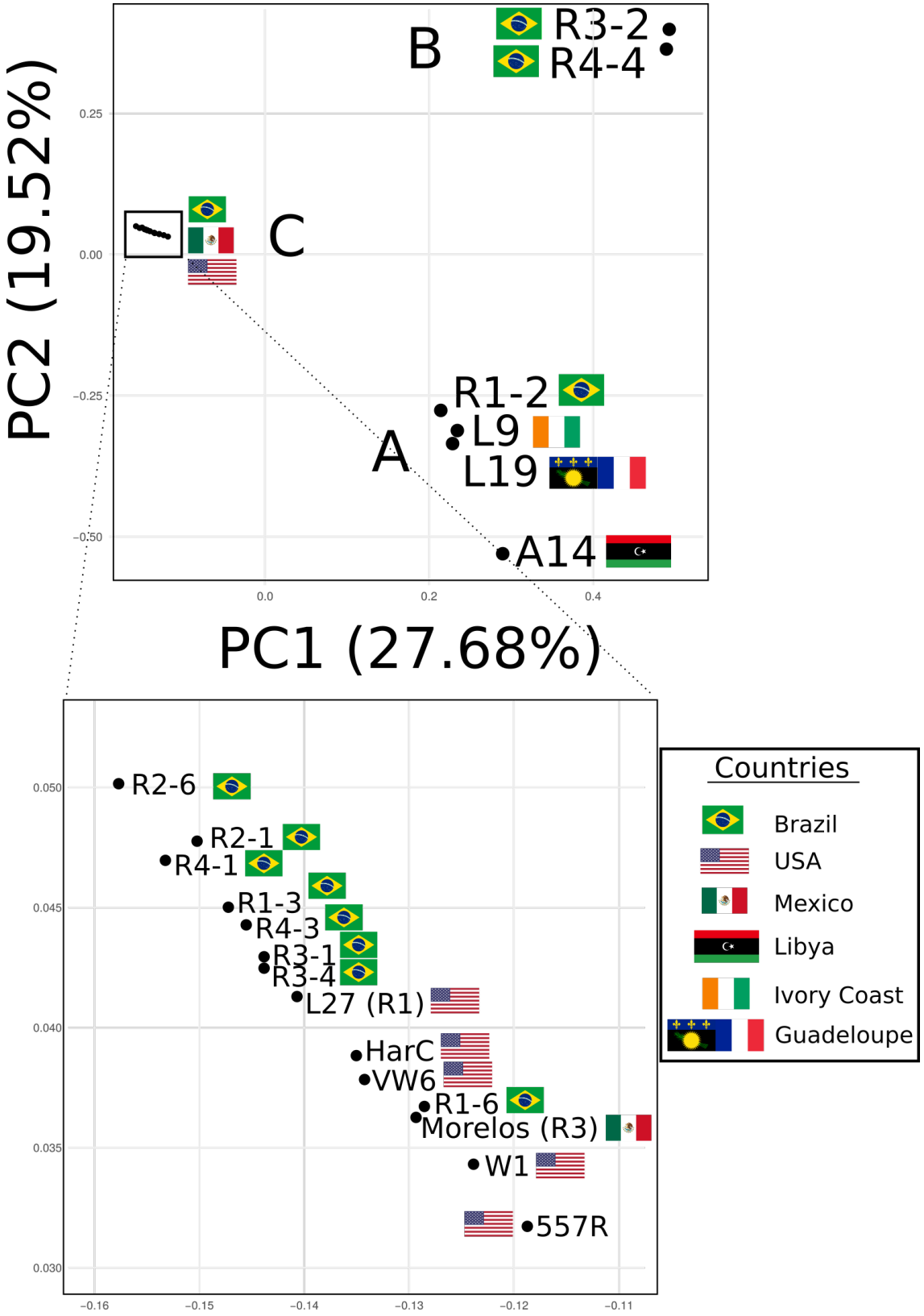


Figure 5. Linkage Disequilibrium (red) and 4-gametes test (blue) of *M. incognita* isolates.



~~Figure 6. Linkage Disequilibrium (red) and 4 gamete test (blue) for phased SNPs in *Globodera rostochiensis* isolates.~~



790

791 Figure 7. PCA analysis of all known *M. incognita* genomes.

792

References

Abad P, Gouzy J, Aury J-M, Castagnone-Sereno P, Danchin EGJ, Deleury E, Perfus-Barbeoch L, Anthouard V, Artiguenave F, Blok VC, et al. 2008. Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nat. Biotechnol. 26:909–915.

Anon. 2018. vcflib: a simple C++ library for parsing and manipulating VCF files, + many command-line utilities. vcflib Available from: https://github.com/vcflib/vcflib

Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics 32:3380–3387.

Barbary A, Djian-Caporalino C, Palloix A, Castagnone-Sereno P. 2015. Host genetic resistance to root-knot nematodes, Meloidogyne spp., in Solanaceae: from genes to the field. Pest Manag. Sci. 71:1591–1598.

Baum TJ. 1994. Characterization and Phylogenetic Analysis of Four Root-Knot Nematode Species Using DNA Amplification Fingerprinting and Automated Polyacrylamide Gel Electrophoresis. Mol. Plant. Microbe Interact. 7:39.

Blanc-Mathieu R, Perfus-Barbeoch L, Aury J-M, Rocha MD, Gouzy J, Sallet E, Martin-Jimenez C, Bailly-Bechet M, Castagnone-Sereno P, Flot J-F, et al. 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. PLOS Genet. 13:e1006777.

Blok VC, Powers TO. 2009. Biochemical and molecular identification. Available from: https://nebraska.pure.elsevier.com/en/publications/biochemical-and-molecular-identification

Carneiro RMDG, Almeida MR, Quénéhervé P. 2000. Enzyme phenotypes of *Meloidogyne* spp. populations. Nematology 2:645–654.

Carneiro RMDG, Almeida MRA. 2001. Técnica de eletroforese usada no estudo de enzimas dos nematóides de galhas para identificação de espécies. Nematol. Bras. 25:35–44.

Castagnone-Sereno P. 2006. Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. Heredity 96:282–289.

Castagnone-Sereno P, Danchin EGJ. 2014. Parasitic success without sex – the nematode experience. J. Evol. Biol. 27:1323–1333.

Castagnone-Sereno P, Mulet K, Danchin EGJ, Koutsovoulos GD, Karaulic M, Da Rocha M, Bailly-Bechet M, Pratx L, Perfus-Barbeoch L, Abad P. 2019. Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita*. Mol. Ecol.:mec.15095.

Castagnone-Sereno P, Wajnberg E, Bongiovanni M, Leroy F, Dalmasso A. 1994. Genetic variation inMeloidogyne incognita virulence against the tomatoMi resistance gene: evidence from isofemale line selection studies. Theor. Appl. Genet. 88:749–753.

828 Cenis JL. 1993. Identification of Four Major Meloidogyne spp. by Random Amplified
829 Polymorphic DNA (RAPD-PCR). Phytopathology:76–78.
830 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,
831 Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics
832 27:2156–2158.
833 Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, Lewis SC, Okamoto
834 K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of Caenorhabditis elegans base-
835 substitution mutation processes. Proc. Natl. Acad. Sci. 106:16310–16314.
836 Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of
837 insertions in the Caenorhabditis elegans nuclear genome. Nature 430:679–682.
838 Eisenback JD, Hunt D. 2009. General morphology. In: R. N. Perry MM, editor. Root-Knot
839 Nematodes. Wallingford: cabi. p. 18–54.
840 Esbenshade PR, Triantaphyllou AC. 1985. Use of Enzyme Phenotypes for Identification of
841 Meloidogyne Species. J. Nematol. 17:6–20.
842 Eves-van den Akker S, Laetsch DR, Thorpe P, Lilley CJ, Danchin EGJ, Da Rocha M, Rancurel
843 C, Holroyd NE, Cotton JA, Szitenberg A, et al. 2016. The genome of the yellow potato cyst
844 nematode, Globodera rostochiensis, reveals insights into the basis of parasitism and virulence.
845 Genome Biol. [Internet] 17. Available from:
846 http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0985-1
847 Filho JV de A, Machado ACZ, Dallagnol LJ, Aranha Camargo LE. 2016. Root-Knot
848 Nematodes (Meloidogyne spp.) Parasitizing Resistant Tobacco Cultivars in Southern Brazil.
849 Plant Dis. 100:1222–1231.
850 Flot J-F, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EGJ, Hejnol A, Henrissat B, Koszul
851 R, Aury J-M, et al. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer
852 Adineta vaga. Nature 500:453–457.
853 Fourie H, McDonald AH, Waele DD. 2006. Host suitability of South African and foreign
854 soybean cultivars to Meloidogyne incognita race 2. South Afr. J. Plant Soil 23:132–137.
855 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing.
856 ArXiv12073907 Q-Bio [Internet]. Available from: http://arxiv.org/abs/1207.3907
857 Glémin S, Galtier N. 2012. Genome Evolution in Outcrossing Versus Selfing Versus Asexual
858 Species. In: Anisimova M, editor. Evolutionary Genomics. Vol. 855. Totowa, NJ: Humana
859 Press. p. 311–335. Available from: http://link.springer.com/10.1007/978-1-61779-582-4_11
860 Guéguen L, Gaillard S, Boussau B, Gouy M, Groussin M, Rochette NC, Bigot T, Fournier D,
861 Pouyet F, Cahais V, et al. 2013. Bio++: Efficient Extensible Libraries and Tools for
862 Computational Molecular Evolution. Mol. Biol. Evol. 30:1745–1750.

Gutekunst J, Andriantsoa R, Falckenhayn C, Hanna K, Stein W, Rasamy J, Lyko F. 2018. Clonal genome evolution and rapid invasive spread of the marbled crayfish. Nat. Ecol. Evol. 2:567–573.

Hartman KM, Sasser JN. 1985. Identification of *Meloidogyne* species on the basis of differential host test and perineal-pattern morphology. In: Barker KR, Carter CC, Sasser JN, editors. Advanced Treatise on Meloidogyne. Vol. II. Raleigh: North Carolina State University. p. 69–77.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071.

Jones JT, Haegeman A, Danchin EGJ, Gaur HS, Helder J, Jones MGK, Kikuchi T, Manzanilla-López R, Palomares-Rius JE, Wesemael WML, et al. 2013. Top 10 plant-parasitic nematodes in molecular plant pathology. Mol. Plant Pathol. 14:946–961.

Kondrashov AS. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336:435–440.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. bioRxiv:447110.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio [Internet]. Available from: http://arxiv.org/abs/1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Lima EA, Furlanetto C, Nicole M, Gomes ACMM, Almeida MRA, Jorge-Júnior A, Correa VR, Salgado SM, Ferrão MAG, Carneiro RMDG. 2015. The Multi-Resistant Reaction of Drought-Tolerant Coffee “Conilon Clone 14” to Meloidogyne spp. and Late Hypersensitive-Like Response in Coffea canephora. Phytopathology 105:805–814.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12.

Martin M, Patterson M, Garg S, Fischer SO, Pisanti N, Klau GW, Schöenhuth A, Marschall T. 2016. WhatsHap: fast and accurate read-based phasing. bioRxiv:085050.

Mattos VS, Furlanetto C, Silva JGP, Santos DFD, Almeida MRA, Correa VR, Moita AW, Castagnone-Sereno P, Carneiro RMDG. 2016. *Meloidogyne* spp. populations from native Cerrado and soybean cultivated areas: genetic variability and aggressiveness. Nematology 18:505–515.

Moens M, Perry RN, Starr JL. 2009. Meloidogyne species: a diverse group of novel and important plant parasites. In: Root-knot Nematodes. CABI Publishing. p. 1–17. Available from: http://hdl.handle.net/1854/LU-866916

Mota FC, Alves GCS, Giband M, Gomes ACMM, Sousa FR, Mattos VS, Barbosa VHS, Barroso P a. V, Nicole M, Peixoto JR, et al. 2013. New sources of resistance to Meloidogyne incognita race 3 in wild cotton accessions and histological characterization of the defence mechanisms. Plant Pathol. 62:1173–1183.

Muller HJ. 1964. The Relation of Recombination to Mutational Advance. Mutat Res 106:2–9.

Neiman Maurine, Taylor Douglas R. 2009. The causes of mutation accumulation in mitochondrial genomes. Proc. R. Soc. B Biol. Sci. 276:1201–1209.

Okimoto R, Chamberlin HM, Macfarlane JL, Wolstenholme DR. 1991. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (Meloidogyne): nucleotide sequences, genome location and potential for host-race identification. Nucleic Acids Res. 19:1619–1626.

Peres ACJ, Salgado SML, Correa VR, Santos MFA, Mattos VS, Monteiro JMS, Carneiro RMDG. 2017. Resistance of genotypes against and under controlled and field conditions. Nematology 19:617–626.

Randig O, Bongiovanni M, Carneiro RMDG, Castagnone-Sereno P. 2002. Genetic diversity of root-knot nematodes from Brazil and development of SCAR markers specific for the coffee-damaging species. Genome 45:862–870.

Robertson L, Díez-Rojo MA, López-Pérez JA, Buena AP, Escuer M, Cepero JL, Martínez C, Bello A. 2009. New Host Races of Meloidogyne arenaria, M. incognita, and M. javanica from Horticultural Regions of Spain. Plant Dis. 93:180–184.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515:261–263.

Santos MFA dos, Furlanetto C, Almeida MRA, Carneiro MDG, Mota FC, Gomes ACMM, Silveira NOR, Silva JGP, Castagnone-Sereno P, Tigano MS, et al. 2012. Biometrical, biological, biochemical and molecular characteristics of Meloidogyne incognita isolates and related species. Eur. J. Plant Pathol. 134:671–684.

Sasser JN. 1952. Identification of root-knot nematodes (Meloidogyne spp.) by host reaction. Plant Dis. Report. 36:84–86.

Sasser JN, Eisenback JD, Carter CC, Triantaphyllou AC. 1983. The International Meloidogyne Project - Its Goals and Accomplishments. Annu. Rev. Phytopathol. 21:271–288.

da Silva E, da Silva Mattos V, Furlaneto C, Giband M, Barroso P, Moita A, Jorge-Junior A, Correa V, Castagnone-Sereno P, Carneiro R. 2014. Genetic variability and virulence of

Meloidogyne incognita populations from Brazil to resistant cotton genotypes. Eur. J. Plant Pathol. 139:195–204.
Singh SK, Hodda M, Ash GJ. 2013. Plant-parasitic nematodes of potential phytosanitary importance, their main hosts and reported yield losses. EPPO Bull. 43:334–374.
Szitenberg A, Salazar-Jaramillo L, Blok VC, Laetsch DR, Joseph S, Williamson VM, Blaxter ML, Lunt DH. 2017. Comparative Genomics of Apomictic Root-Knot Nematodes: Hybridization, Ploidy, and Dynamic Genome Change. Genome Biol. Evol. 9:2844–2861.
Triantaphyllou AC. 1981. Oogenesis and the chromosomes of the parthenogenetic Root Knot Nematode Meloidogyne incognita. J. Nematol.:95–104.
Triantaphyllou AC. 1985. Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes. In: Sasser JN, Carter CC, editors. An advanced treatise on Meloidogyne. Vol. 1. Raleigh: North Carolina State University Graphics. p. 113–126.
Trudgill DL, Blok VC. 2001. Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. Annu. Rev. Phytopathol. 39:53–77.
Tzortzakakis EA, Conceição I, Dias AM, Simoglou KB, Abrantes I. 2014. Occurrence of a new resistant breaking pathotype of Meloidogyne incognita on tomato in Greece. J. Plant Dis. Prot. 121:184–186.
Vakhrusheva OA, Mnatsakanova EA, Galimov YR, Neretina TV, Gerasimov ES, Ozerova SG, Zalevsky AO, Yushenova IA, Arkhipova IR, Penin AA, et al. 2018. Recombination in a natural population of the bdelloid rotifer Adineta vaga. bioRxiv:489393.
Verdejo-Lucas S, Talavera M, Andrés MF. 2012. Virulence response to the Mi.1 gene of Meloidogyne populations from tomato in greenhouses. Crop Prot. 39:97–105.
Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328.
Zijlstra C, Donkers-Venne DTHM, Fargette M. 2000. Identification of *Meloidogyne incognita*, *M. javanica* and *M. arenaria* using sequence characterised amplified region (SCAR) based PCR assays. Nematology 2:847–853.