ARTICLE (Discoveries)

Parallel adaptations to different host plants despite clonal reproduction in the world most devastating nematode pest

Georgios D. Koutsovoulos[1*], Eder Marques[2], Marie-Jeanne Arguel[3], Andressa CZ Machado[4], Regina M.D.G. Carneiro[2], Philippe Castagnone-Sereno[1], Etienne GJ Danchin[1*], Erika VS Albuquerque[2*]

[1]Institut Sophia Agrobiotech, INRA, CNRS, Université Côte d'Azur, France

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

[3]Institut de pharmacologie moléculaire et cellulaire, CNRS, Université Côte d'Azur, France

[4]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 is surprising given its obligatory clonal reproduction. At the species level, *M. incognita* is polyphagous with thousands of host plants. However, this wide host range is due to an assemblage of different isolates that each have distinct and more restricted host compatibilities. Historically, four 'host races' had been defined as a function of distinct ranges of compatible and incompatible plants. To assess the level of intra-species genome variability underlying biological traits such as the host races and geographical distribution, we sequenced the genomes of 11 isolates across Brazil, covering these four distinct races. By aligning the genomes of the isolates to the *M. incognita* reference genome, we identified SNPs and small-scale insertions/deletions. We showed that there are relatively low point variations between the different isolates, despite their distinct ecological characteristics. Furthermore, these variations neither correlate with the host races, the geographical origin of the samples nor the host plant on which they have been collected. These results suggest that multiple independent adaptations to different hosts and environmental conditions have occurred and illustrate the high adaptive potential of this agricultural pest despite its clonal mode of reproduction. Furthermore, although variations in *M. incognita* host ranges exist, we recommend the discontinuation of the terminology "race" and replacement by "pathotype" due to the lack of underlying phylogenetic signal.

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, which are able to parasitize the vast majority of flowering plants on Earth (Trudgill and Blok 2001), reproduce clonally, without sex and meiosis. 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 extreme polyphagy, *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 broad host range of *M. incognita* as well as the other major RKN species is not be present in every single individual of these species but that different ‘populations’ or ‘isolates’ within one species have more restricted and overlapping host ranges. The assemblage of these different population-specific host ranges constitutes their remarkably wide host range at the species level (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. (Table S1). In *M. incognita*, all

populations tested so far reproduce on tomato, watermelon, and pepper and none infect peanut, but they differ in their response to tobacco and cotton defining four distinct host races (Hartman and Sasser 1985) (Table S1). The genetic features associated with *M. incognita* races remain 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 (Okimoto et al. 1991). While 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 differences underlying the phenotypic plasticity of *M. incognita* isolates in terms of hots preferences 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 adaptations to different host ranges despite the clonal reproduction of *M. incognita*. Theoretically, animal clones have poor adaptability because the efficiency of selection is impaired, advantageous alleles from different individuals cannot be combined in one single genotype 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). However, this apparent lack of phylogenetic signals results from analyses with only a limited set of markers and remain inconclusive.

Therefore, a large-scale whole genome comparative analysis of the different races should more comprehensively inform, on the level of intraspecific genetic polymorphism and phylogenetic signal in *M. incognita*, in relation to variations in biological traits such as the host

preferences. A first version of the genome of *M. incognita* was initially published in 2008 (Abad et al. 2008) and re-sequenced at much higher depth, providing the most complete *M. incognita* reference genome available to date (Blanc-Mathieu et al. 2017). This reference genome comes from a Mexican strain, named Morelos, reared from the offspring of one single female. However, no information about the intra-species variation at the whole genome level was available so far. 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). Because the host-race status was characterised for only one of the eight *M. incognita* strains sequenced in this analysis; no relation between genomic variations and host race or other biological trait was analysed. Furthermore, the variability between isolates at the non-coding level, which represents the vast majority of the genome, was not taken into account in this initial analysis.

Here, we have characterised and sequenced the genomes of 11 isolates representing the four *M. incognita* host-races from populations parasitizing six crops from geographically different sites across Brazil (Figure 1). We used isozymes profiles, SCAR markers, and the NCDHT to define the biological materials before the DNA extraction and high coverage genome sequencing. We classified the *M. incognita* isolates according to their SNP profiles in comparison to the reference genome. We superimposed on this classification the biological traits: their host race, geographical localisation and current host plant. We also measured the level of intra-species divergence in relation to the above-mentioned biological traits.

Utilizing whole genome sequencing we were able to robustly determine whether there is a phylogenetic signal underlying variations in biological traits of agro-economic importance such as the host races in *M. incognita*. 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. This resolution has agricultural and economic implications since crop rotation and control strategies should take into account the adaptive potential of this clonal pest.

Results

We collected 11 *M. incognita* populations from six different states across Brazil and from six different fields (soybean, cotton, coffee, cucumber, tobacco, watermelon). 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 covered the ~183Mb *M. incognita* genome assembly (Blanc-Mathieu et al. 2017) at a depth > 100X (Table S2) for each isolate. Overall, we identified 482,545 positions on the *M. incognita* reference genome that were polymorphic across the different isolates (~0.27% of the total number of non-ambiguous nucleotides). Each isolate shows a different level of divergence from the reference with R1-2 having the highest number of variable positions (377,349) and R2-6 having the least (177,979) variants (Figure 2a). The R3-4 isolate originated from a pool of four populations. However, low number of SNPs compared to the reference indicates that either the genomes of these four populations were very close or that a specific population displaced the other three (Figure 2a). Thus, the R3-4 isolate was analysed exactly as the other isolates. Overall, the level of variation of the genomes of the eleven isolates ranges between 0.10 % and 0.21 % compared to the Morelos reference strain. Similarly, the level of variation in the mitochondrial genome ranges between 0.09% and 0.21%.

Interestingly, only a small proportion (< 100 positions) of the variants appear to be race-specific, except for race 2 (Figure 2b) which shows >4,000 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). All the race-specific variants were checked for possible impact on the predicted gene set. The vast majority (80-90%) of SNPs were outside of coding regions; only 85,413 variable positions fell in coding regions but covered 19,809 out of 43,718 predicted protein-coding genes. In these coding regions, only 10, 2 and 9 SNPs yielded non-synonymous substitution or otherwise disruptive mutations in races 1, 3 and 4, respectively (793 such variants were found in race 2). For each race, the non-synonymous / disruptive variants fell in different genes.

Using principal component analysis on the whole set of SNPs (PCA, methods), we showed that the eleven *M. incognita* isolates formed three distinct clusters, which we named A, B and C (Figure 3). 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 lower variation. The different clusters do not correlate with host race status, there is no possibility to form race-specific clusters based on their SNP patterns. This means that different isolates of the same host race are not more similar to each other than isolates of different host races. Thus, there is no apparent correlation between resemblance in terms of genomic variants and host race status. Furthermore, we observe that the three distinct clusters also correlate neither with geographic origin nor with the original host plant from which the nematodes have been collected. Interestingly, the four different host races are all represented in one single cluster (C). Within this cluster, the total number of variable positions is 176,107. Meaning that the whole range of host-race diversity is present in a cluster that represents only 39% of the total genomic variation.

Using a phylogenetic network analysis based only on SNP present in coding regions we could confirm the same three clusters (Figure 4). According to the two classification methods, isolate R1-2 seems to be the most divergent from the rest of isolates, which is consistent with its

higher total number of SNPs and number of isolate-specific SNPs. Then, a small cluster is composed of isolates R3-2 and R4-3 (equivalent to cluster B of the PCA). Finally, a cluster (equivalent to PCA cluster C) groups the rest of the eight isolates with apparent high similarity regarding SNP pattern, regardless whether considered at the whole genome level (PCA, Figure 3) or only at the coding part level (phylogenetic network, Figure 4).

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

To investigate more widely the diversity of *M. incognita* isolates in a broader analysis, 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 5). All the new isolates from additional and more diverse geographical origins fall 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. Overall, the addition of new geographical isolates (Szitenberg et al. 2017) did not substantially increase the number of detected variable positions in the genome.

Discussion

Our results show that there is no phylogenetic signal at the whole genome level underlying the four *M. incognita* host races. Although this was previously suggested, it had remained inconclusive due to the limited set of markers used (Cenis 1993; Baum 1994; Santos et al. 2012). The host race concept has never been universally accepted, in part because it assured only a small portion of the potential variation in parasitic ability (Moens et al. 2009). In the initial host race concept, even considering that *M. incognita* presented hundreds of known host plants, only six different host standards were used to characterise races. New host races might be defined when including additional hosts in the differential test. Indeed, two additional *M. incognita* races have been characterized in Spain that did not fit into the previously published race scheme (Robertson et al. 2009). Although the terminology 'races' of *Meloidogyne* spp. has been recommended not to be used since 2009 (Moens et al. 2009), some 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 crucial importance for growers and agronomists to distinguish populations that differ in their ability to infest a particular host (that carry specific resistance genes). Indeed, major *Meloidogyne* spp. control strategies such as resistant cultivars and crop rotation are generally specifically deployed against a 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). For this ensemble of reasons, we propose to abandon the term "race" and use instead pathotype as this term does not imply shared common genetic ancestry for the different types of compatible host ranges but still allows to name and differentiate them.

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 divergence across the eleven isolates from Brazil and the reference genome (sampled initially from Mexico) reached ~0.27% of the nucleotides. The maximum divergence between the reference genome and one isolate covers ~0.21% of the nucleotides and addition of the isolates from Africa, the West Indies and the USA did not increase the maximal divergence. In the nematode model *C. elegans*, a RAD-seq analysis covering 8 Mb of the genome revealed 41,188 SNPs over >100 geographical isolates, which corresponds to an estimated ~0.52% variable nucleotides in the genome (Andersen et al. 2012). A more comprehensive genome-wide analysis identified ~600,000 SNPs across 40 *C. elegans* wild isolates from various geographic locations (Thompson et al. 2013). This corresponds to ~0.63% of the nucleotides in the *C. elegans* genome. Thus, intra-species point variations at the genome level seem to happen at a lower magnitude in *M. incognita* than in *C. elegans*. This difference is probably biologically sound and not due to a difference of sensitivity of our method since we covered the *M. incognita* genome at higher depths than all the previous *C. elegans* analyses. This relatively low divergence is rather surprising, considering the variability in terms of distinct compatible host spectra (host races), the geographically distant sampled sites and the different natures of crop species of origin. These observations suggest that the classical mechanism of mutations acquisition and 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 different environmental conditions across the globe. Although all these nematode strains have been collected in cultivated fields, that are by definition anthropized environments, there are substantial differences in temperature range, weather conditions and soil properties across these geographical locations. Furthermore, the six host plants on which the eleven *M. incognita* strains have been collected represent 110 – 124 Million years of evolution (last common ancestor of pentapetalae (Kumar et al. 2017)) with different root systems and defence mechanisms. For instance, in cluster B alone, R3-2 and R4-4 appear to be very closely related in

terms of SNP pattern (Figure 3), however they were collected on two plant species (cotton and soybean) that show >100 Million years of separation.

The relatively low divergence between the *M. incognita* isolates at the nuclear genome was not higher that the mitochondrial level. This is surprising, considering the mitochondrial genome accumulates mutations much faster than the nuclear genome in the nematode model *C. elegans* (Denver et al. 2004; Denver et al. 2009). This suggests that not enough time has passed since the separation of the isolates to observe a significant difference in accumulations of mutations between the mitochondrial and nuclear genomes.

Maybe even more surprising than the low divergence is the fact that none of the biological and ecological traits under consideration matches the SNP-based classification of these isolates. This non-correlation between genetic grouping and bio-ecological traits is confirmed by the two independent methods we employed (PCA and phylogenetic networks), and is consistent with the lack of correlation previously observed with lower resolution markers (Cenis 1993; Baum 1994; Santos et al. 2012). This suggests that the host race phenotypes most likely result from multiple independent adaptations rather than spread of ancestral distinct host races. Consequently, there are cases of higher similarity in SNP patterns between isolates of distinct host races than between isolates of the same host race. The switch from race 1, which are neither able to infect cotton nor tobacco, to race 4, which can infect both, appears particularly challenging. However, these two races are already present in the same dense cluster C that encompasses only 39% of the SNPs. This suggests that *M. incognita* is able to adapt to the challenge represented by two new hosts despite showing relatively few variations regarding the genome sequence. Consistent with this suggestion, 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). Furthermore, *M. incognita* is able to rapidly overcome (5-10 generations) the *Mi1.2* nematode resistance gene in tomato, in controlled lab conditions (Castagnone-Sereno 2006). This also happens in tomato fields (Verdejo-Lucas et al. 2012; Tzortzakakis et al. 2014), although it is unknown if this resistance breaking is as rapid as in controlled lab conditions.

Collectively, our observations indicate that *M. incognita* is particularly versatile and adaptive despite its clonal mode of reproduction. The relatively low divergence at the SNP level suggests 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. Interestingly, the parthenogenetic marbled crayfish has multiplied by more than 100 its original area of repartition across Madagascar despite showing a surprisingly low number of nucleotide variation (only ~400 SNPs on a ~3Gb genome representing a proportion of variable positions of $1.3E^{-7}$ only). This also led the authors to suggest that mechanism other than mutation acquisition 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, and the majority of these gene copies show diverged expression patterns one from the other (Blanc-Mathieu et al. 2017). How the expression patterns of these gene copies vary across different geographical isolates with different host preferences would be interesting to explore.

To date, the whole divergence of Brazilian isolates encompasses the known divergence of all the other sequenced *M. incognita* isolates across the world. Although the sampling is biased towards Brazilian and American isolates with a few coming from Africa, it provides an initial snapshot of the inter-population pattern of variations. The addition of more isolates from other parts of the world will unveil whether populations are part of a gradient transition or are grouped into clusters as shown here. If the former is correct, it means that populations switch host-compatibility gradually, while if the latter is the case, it means that members of a group can quickly transition between different hosts, as suggested by previous observations of switches between races 2 and 3.

M. incognita shows a variability of ranges of compatible host plants and is able to proliferate in various different geographical locations around the globe. This high variability and adaptability is surprising, considering its clonal mode of reproduction and the overall limited variation in terms of SNPs at the whole genome sequence level shown in this study. These observations suggest that *M. incognita* is able to rapidly adapt to other hosts and new environments despite being clonal and showing low variability at the whole genome scale. *M. incognita* constitutes an interesting model system to study adaptive potential in clonal animals.

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). We identified and purified the isolates using individual females and their respective egg–masses (Carneiro & Almeida 2001). The identification was based first on esterase phenotypes (Carneiro et al. 2000). 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). We used the same inoculum to perform 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. 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 eggs 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).

DNA preparation and SCAR test

We extracted and purified the genomic DNA from eggs of each nematode isolate 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 S3 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 the clean reads to the *M. incognita* reference genome (Blanc-Mathieu et al. 2017), using the BWA-MEM software package (Li 2013). We then used SAMtools (Li et al. 2009) to filter alignments with MAPQ lower than 20, sort the alignment file by reference position, and remove duplicate alignments.

We used the FreeBayes variant detection tool (Garrison and Marth 2012) to call SNPs 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.

Principal component analysis

We performed a principal component analysis (PCA) to classify the isolates according to their SNP 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 network 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.

Mitochondrion 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 contigs containing the mitochondrion 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. All the data generated and analysed in this article have been deposited to the NCBI'S GEO repository under accession number GSE116847.

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 gray and 4 in blue). The names of the isolates of the literature are written in white or black.

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

(a) Number of variants per isolate and isolate-specific variants. These isolates show variations in their numbers of total variants and sample specific variants (b) Race-specific variants. The number of race-specific variants are low for almost all races.

Figure 3. 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 4. Phylogenetic network for *M. incognita* isolates based on SNPs present in coding sequences.

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

Figure 5. PCA analysis of all known *M. incognita* genomes.

The isolates were regrouped based on SNP 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).

Author contributions

EGJD, PC-S and EVSA 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, EVSA contributed to the analysis and interpretation of data. EGJD and GDK wrote the manuscript with contributions of EVSA, RMDGC, PC-S and M-JA. All authors approved the final manuscript.

Tables

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

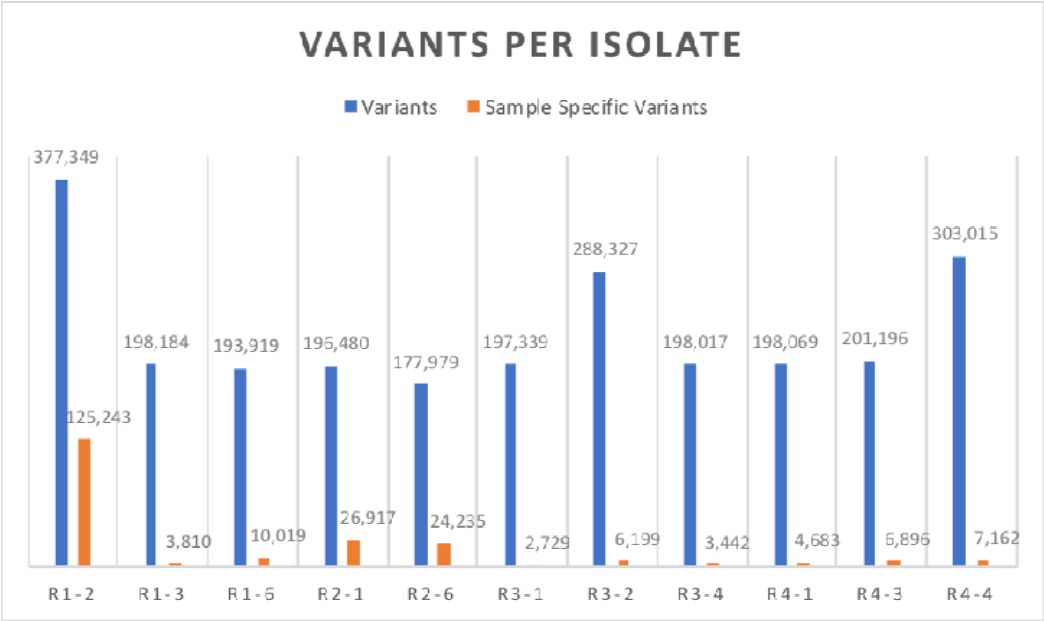
d, host range results for the North Carolina Differential Host Test (NCDHT), numbers represent gall index with 0 = no galls; 1 = 1 to 2; 2 = 3 to 10; 3 = 11 to 30; 4 = 31 to 100; and 5 = more than 100 galls. (tb=tobacco ‘NC95’, tm=tomato ‘Rutgers’, wm=watermelon ‘Charleston Gray’, pt=peanut ‘Florunner’,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

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

a



b

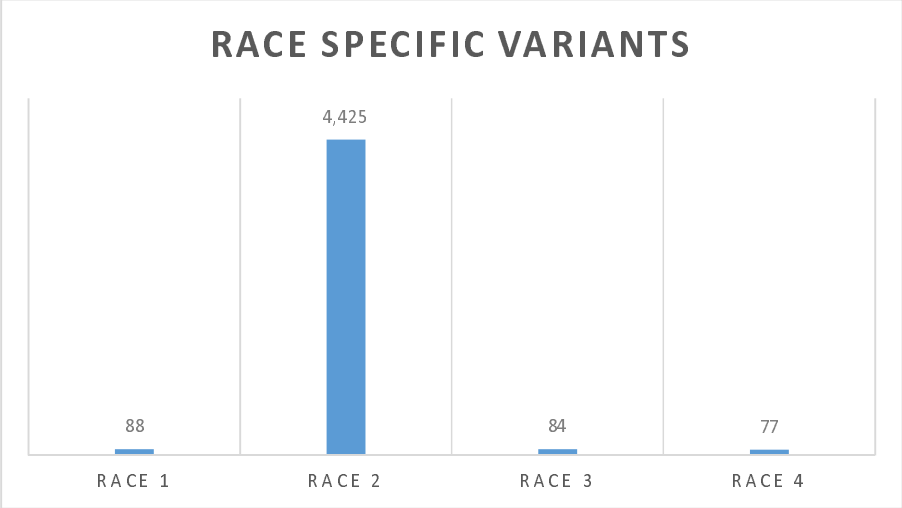


Figure 2. Distribution of the number of variants per race and isolate.

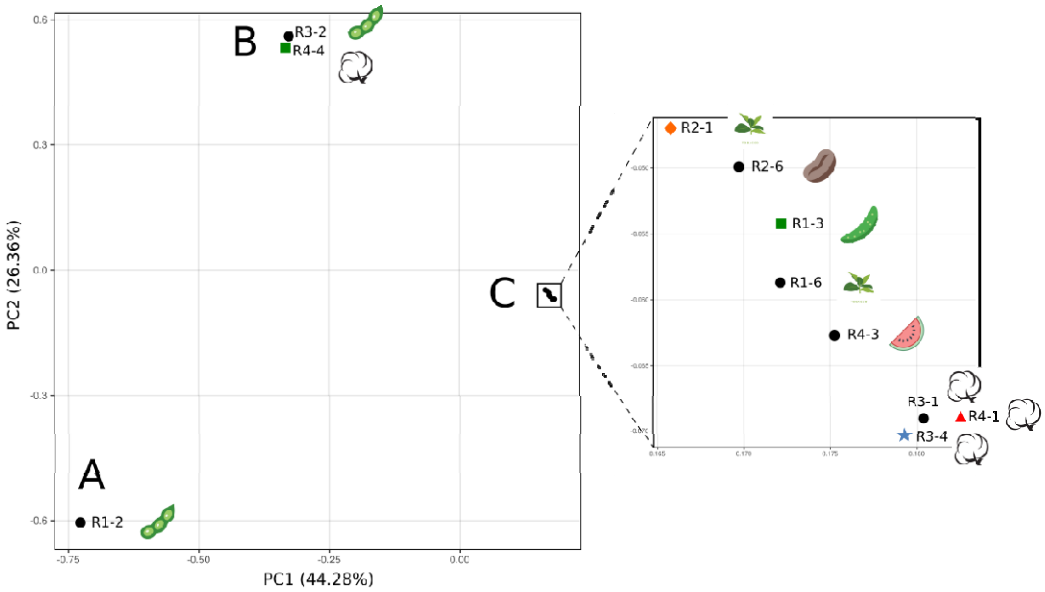


Figure 3. PCA analysis of the different *M. incognita* isolates groups them into three clusters (A, B, and C).

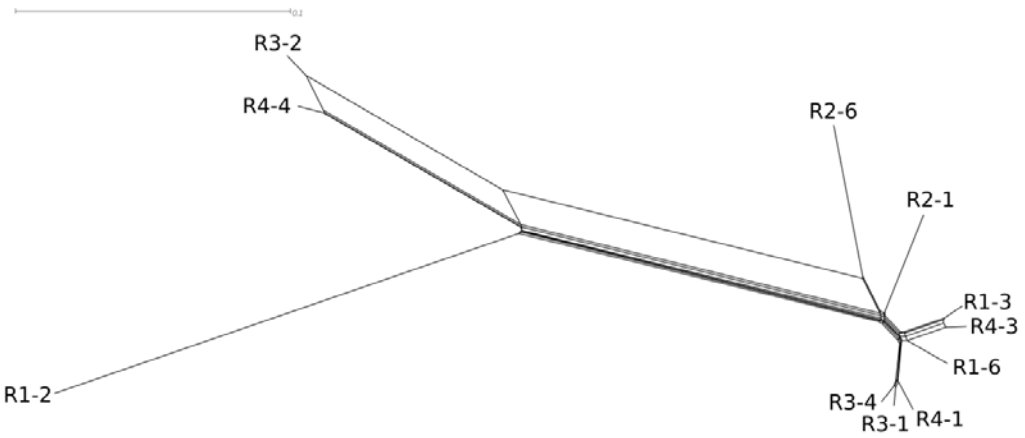
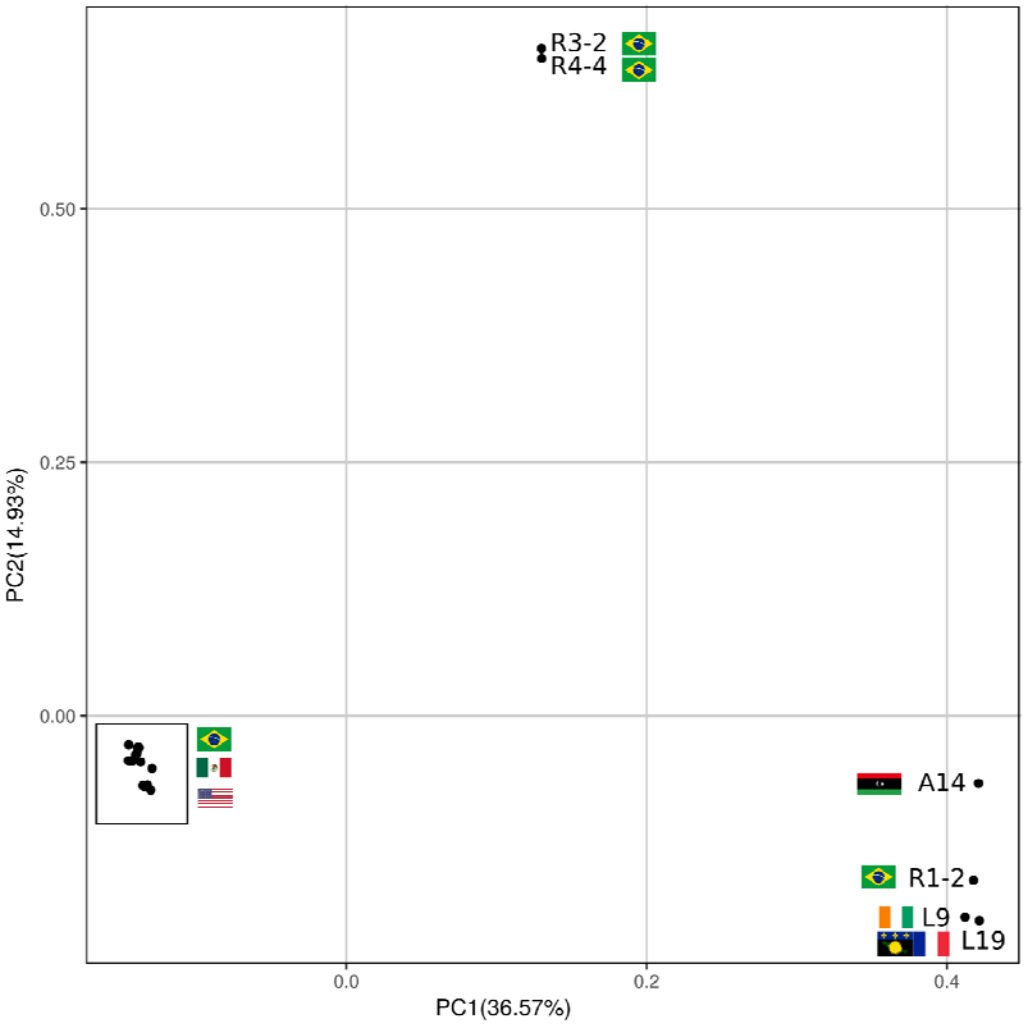
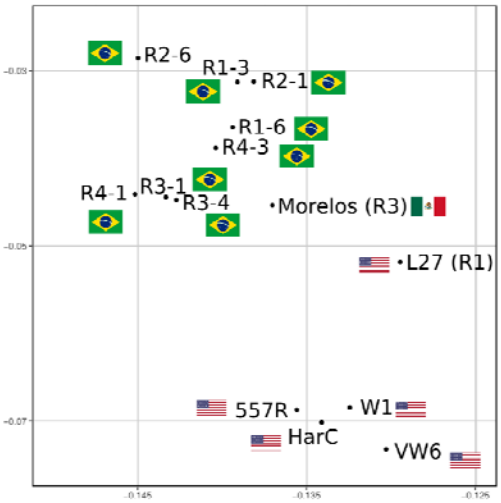


Figure 4. Phylogenetic network for *M. incognita* isolates based on SNPs present in coding sequences.


R3-2
R4-4
0.50
0.25
0.00
PC2(14.93%)
A14
R1-2
L9
L19
0.0
0.2
0.4
PC1(36.57%)


R2-6
R1-3
R2-1
R1-6
R4-3
R4-1
R3-1
R3-4
Morelos (R3)
L27 (R1)
557R
W1
HarC
VW6

464

Figure 5. PCA analysis of all known *M. incognita* genomes.

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.

Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, Félix M-A, Kruglyak L. 2012. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. Nat. Genet. 44:285–290.

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.

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.

Cenis JL. 1993. Identification of Four Major Meloidogyne spp. by Random Amplified Polymorphic DNA (RAPD-PCR). Phytopathology:76–78.

Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. Proc. Natl. Acad. Sci. 106:16310–16314.

Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. Nature 430:679–682.

Eisenback JD, Hunt D. 2009. General morphology. In: R. N. Perry MM, editor. Root-Knot Nematodes. Wallingford: cabi. p. 18–54.

Esbenshade PR, Triantaphyllou AC. 1985. Use of Enzyme Phenotypes for Identification of Meloidogyne Species. J. Nematol. 17:6–20.

Filho JV de A, Machado ACZ, Dallagnol LJ, Aranha Camargo LE. 2016. Root-Knot Nematodes (Meloidogyne spp.) Parasitizing Resistant Tobacco Cultivars in Southern Brazil. Plant Dis. 100:1222–1231.

Fourie H, McDonald AH, Waele DD. 2006. Host suitability of South African and foreign soybean cultivars to Meloidogyne incognita race 2. South Afr. J. Plant Soil 23:132–137.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio [Internet]. Available from: http://arxiv.org/abs/1207.3907

Glémin S, Galtier N. 2012. Genome Evolution in Outcrossing Versus Selfing Versus Asexual Species. In: Anisimova M, editor. Evolutionary Genomics. Vol. 855. Totowa, NJ: Humana Press. p. 311–335. Available from: http://link.springer.com/10.1007/978-1-61779-582-4_11

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.

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.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol. Biol. Evol. 34:1812–1819.

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.

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.

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.

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.

Thompson O, Edgley M, Strasbourger P, Flibotte S, Ewing B, Adair R, Au V, Chaudhry I, Fernando L, Hutter H, et al. 2013. The million mutation project: A new approach to genetics in Caenorhabditis elegans. Genome Res. 23:1749–1762.

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.

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.