Multiple independent adaptations to different ranges of host plants indicate high adaptability despite clonal reproduction in the nematode pest *Meloidogyne incognita*

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 authors (equal contribution):

Abstract

The most devastating nematodes to worldwide agriculture are the root-knot nematodes with *Meloidogyne incognita* being considered as the most ubiquitous. At the species level, *M. incognita* is polyphagous with thousands of host plants. However, this wide host range seems to be due to an assemblage of different isolates that each have distinct and more restricted host compatibilities. In this species, four host races have been defined as a function of distinct ranges of compatible and incompatible plants. To assess whether there is an evolutionary signal underlying these host races, we sequenced the genomes of 11 isolates across these four different 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 various different ecological characteristics.

Furthermore, these variations neither correlate with the four 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. Although the recognition of variation in *M. incognita* host range exists, we recommend the discontinuation of the terminology "race" considering there is no genetic signal underlying and supporting the concept of races in *M. incognita.*

35 **Introduction**

36 Nematodes cause severe damages to the world agricultural production every year
37 with the root-knot nematodes (RKN, genus *Meloidogyne*) being the most economically
38 harmful in all temperate and tropical producing areas [1,2]. Curiously, the most
39 polyphagous RKN species, which are able to parasitise the vast majority of flowering
40 plants on Earth [3], have abandoned sex and meiosis. Among these mitotic
41 parthenogenetic RKN, *M. incognita* is the most widespread and is present, at least, in all
42 the countries where the lowest temperature exceeds 3°C. Greenhouses over the world
43 also extend its geographic distribution [4]. *M. incognita* is so widely distributed that it is
44 not even included on the list of regulated pests [5]. Due to its worldwide distribution and
45 extreme polyphagy, *M. incognita* has been deemed the most damaging species of crop
46 pest worldwide [3].

47 However, it has become more and more evident that the broad host range of *M.*
48 *incognita* as well as the other major RKN species might not be present in every single
49 isolate of these species but that different 'populations' or 'isolates' within one species
50 have more restricted and overlapping host ranges. The assemblage of these different
51 population-specific host ranges constitutes their remarkably wide host range at the
52 species level [2]. Variations regarding host range within one given species gave rise to the
53 concept of "host race" as soon as 1952 [6]. The pattern of compatibility/incompatibility of
54 the nematode interaction with a set of different host plants is used to differentiate the
55 host races [7]. The North Carolina Differential Host Test (NCDHT) was initially
56 developed to provide preliminary identification of the four major RKN species (*M.*
57 *incognita*, *M. javanica*, *M. hapla*, and *M. arenaria*) based on the host response [8].
58 Hartman and Sasser [7] later standardised the test to differentiate races within
59 *Meloidogyne* spp. (Table S1). Current identification of species is based on

60 morphological descriptions [9], isozyme phenotypes [10,11], and molecular analysis [12]. The
61 NCDHT is mainly used to determine the host race status of *M. incognita* populations
62 and other main *Meloidogyne* species, given that they are not distinguishable from each
63 other morphologically or biochemically [13]. All *M. incognita* races reproduce on tomato,
64 watermelon, and pepper and none infect peanut, but they differ in their response to
65 tobacco and cotton defining the four host races [7] (Table S1). The genetic features
66 associated with *M. incognita* races remain unknown. Indeed, diversity studies using
67 RAPD and ISSR markers across eight isolates and the four races of *M. incognita* found
68 no correlation between phylogeny and host races [14–16]. Although in one of these works
69 [16] two out of three esterase phenotypes (I1 and S2) were monophyletic in the
70 phylogenetic tree of the *M. incognita* isolates, they did not segregate according to the
71 host races . A different molecular approach to try to differentiate host races was also
72 proposed based on repeated sequence sets in the mitochondrial sequence [17]. While the
73 pattern of repeats allowed differentiating one isolate of race 1, one of race 2 and one of
74 race 4; the study encompassed only one isolate per race, and thus the segregation could
75 be due to differences between isolates unrelated to the host race status itself.
76 Because these previous attempts to differentiate the races were performed only
77 on a few markers, they might have missed other ones, potentially better segregating
78 with the host races. Intriguingly, despite displaying phenotypic plasticity with great
79 evolutionary adaptive impact, no clear underlying genetic differences have been
80 identified so far [18]. Hence, it has remained unclear whether there is a phylogenetic
81 signal underlying the phenotypic variability (including the host races) or whether this
82 results from multiple independent adaptations to the same environment or host plants.
83 Therefore, a large-scale whole genome analysis of the different races should
84 more comprehensively inform on intraspecific genetic polymorphism of

85 *Meloidogyne* spp in relation to a biological trait of high agricultural importance such as
86 their host races. A first version of the genome of *M. incognita* was initially published in
87 2008 [19] and re-sequenced at much higher depth, providing the most complete *M.*
88 *incognita* reference genome available to date [20]. This reference genome comes from a
89 Mexican strain, named Morelos, reared from the offspring of one single female.
90 However, no information about the intra-species variation at the whole genome level
91 was available so far. Recently, a comparative genomics analysis, including different
92 strains of *M. incognita*, showed little variation at the protein-coding genome level
93 between strains collected across different geographical locations [21], confirming previous
94 observations with RAPD and ISSR markers [14–16]. Because the host-race status was
95 characterised for only one of the eight *M. incognita* strains sequenced in this analysis;
96 no relation between genomic variations and host race could be analysed. Furthermore,
97 the variability between isolates at the non-coding level, which represents the vast
98 majority of the genome, was not taken into account in this initial analysis.
99 Here, we have characterised 11 isolates representing the four *M. incognita* host-
100 races from populations parasitising six crops from geographically different sites across
101 Brazil (Figure 1). We used isozymes profiles, SCAR markers, and the NCDHT to
102 define the biological materials before the DNA extraction and high throughput genome
103 sequencing. We classified the *M. incognita* isolates according to their SNP profiles in
104 comparison to the reference genome. We superimposed on this classification
105 information of their host race, geographical localisation and host origin plant. We
106 showed that there is no genealogical signal underlying the host races, suggesting that
107 the capacity of infecting a particular set of plants most likely result from multiple
108 independent and convergent adaptations.

109

110 **Material and Methods**

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

113 The *M. incognita* isolates involved in this study (Table 1) originate from populations
114 collected from different crops and geographically distant origins in Brazil (Figure 1).
115 We identified and purified the isolates using individual females and their respective
116 egg–masses (Carneiro & Almeida 2001). The identification was based first on esterase
117 phenotypes [10]. We reproduced egg-mass isolates on tomato plants (*Solanum*
118 *lycopersicum* L. cv. Santa Clara) under greenhouse conditions at a temperature of 25-28
119 ° C. After three months, we confirmed the *M. incognita* species using esterase
120 phenotypes [22]. We used the same inoculum to perform the North Carolina Differential
121 Host Test (NCDHT) [7] with the following plants: cotton cv. Deltapine 61, tobacco cv.
122 NC95, pepper cv. Early California Wonder, watermelon cv. Charleston Gray, peanut cv.
123 Florunner and tomato cv. Rutgers. We inoculated these plants with 5,000 eggs and J2 of
124 *M. incognita* and maintained them under glasshouse conditions at 25-28° C for three
125 months, with watering and fertilisation as needed. Two months after inoculation, the
126 root system was rinsed with tap water, and eggs masses were stained with Phloxine B [7]
127 to count the number of galls and eggs masses separated for each root system. We
128 assigned a rating index number according to the scale: 0 = no galls or egg masses; 1 =
129 1-2 galls or egg masses; 2 = 3-10 galls or egg masses; 3 = 11-30 galls or egg masses; 4
130 = 31-100 galls or egg masses; and 5 > 100 galls or egg masses per root system (Table
131 1). Host–plants types that have an average gall and egg mass index of 2 or less are
132 designated non-host (-). The other plants (index ≥ 3) are designated hosts (+).

133 We categorised *M. incognita* host races based on their ability to parasitise
134 tobacco and cotton (Table 1). Classically, the index for Rutgers tomato (susceptible
135 control) is higher than 4 (+) (Hartman & Sasser, 1985).
136
137 *DNA preparation and SCAR test*
138 We extracted and purified the genomic DNA from eggs of each nematode isolate
139 with the supplement protocol for nematodes of the QIAGEN Gentra® Puregene®
140 Tissue Kit with the following modifications: incubation at 65 ° C in the cell lysis buffer
141 for 30 min and incubation at 55 ° C with proteinase K for 4h. We verified DNA
142 integrity on 0.8% agarose gel and the DNA quantification on Nanodrop.
143 We confirmed isolate species purity by SCAR-PCR [23,24] using the SCAR
144 primers specified in Table S3 for the RKN *M. javanica, M. paranaensis, M. incognita,*
145 and *M. exigua*.
146
147 *Sequencing library preparation*
148 We assessed input gDNA quantity using Qubit and normalised the samples to
149 20ng/ul as described in TruSeq®DNA PCR-Free Library Prep Reference Guide (#FC-
150 121-3001, Illumina) prior fragmentation to 350bp with Covaris S2. We assessed the
151 quality of fragments after size selection and size control of the final libraries using High
152 Sensitivity DNA Labchip kit on an Agilent 2100 Bioanalyzer.
153
154 *Whole genome sequencing*
155 We quantified isolated sample libraries with KAPA library quantification kit
156 (#7960298001, Roche) twice with two independent dilutions at 1:10,000 and 1:20,000.

157 We calculated the average concentration of undiluted libraries and normalised them to
158 2nM each then pooled them for sequencing step.
159 We generated high-coverage genomic data for the 11 *M. incognita* isolates by
160 2x150 bp paired-end Illumina NextSeq 500 sequencing with High Output Flow Cell
161 Cartridge V2 (#15065973, Illumina) and High Output Reagent Cartridge V2 300 cycles
162 (#15057929, Illumina) on the UCA Genomix sequencing platform, in Sophia-Antipolis,
163 France. We performed two runs to balance the read's representation among the isolates
164 and obtain homogeneity of coverage for the different samples (Table S2).
165
166 *Variant Detection*
167 We trimmed and cleaned the reads from each library with cutadapt tool [25] to
168 remove adapter sequences and bases of a quality inferior to 20. We aligned the clean
169 reads to the *M. incognita* reference genome [20], using the BWA-MEM software package
170 [26]. We then used SAMtools [27] to filter alignments with MAPQ lower than 20, sort the
171 alignment file by reference position, and remove duplicate alignments.
172 We used the FreeBayes variant detection tool [28] to call SNPs and small-scale
173 insertions/deletions, incorporating all the library alignment files simultaneously and
174 produced a variant call file (VCF). We filtered the resulting VCF file with the vcffilter
175 function of vcflib [29] , retaining the positions that had more than 20 Phred-scaled
176 probability (QUAL) and a coverage depth (DP) > 10.
177
178 *Principal component analysis*
179 We performed a principal component analysis (PCA) to classify the isolates
180 according to their SNP patterns and mapped the race characteristics, geographic
181 location, or current host plants on this classification. We used the filtered VCF file as

182 input in the statistical package SNPRelate [30] to perform the PCA with default
183 parameters.
184
185 *Phylogenetic network analysis*
186 Based on the VCF file and the *M. incognita* gene predictions [20], we selected
187 85,413 positions that contained synonymous or non-synonymous mutations (i.e. in
188 coding regions). We aligned these positions and then used them as an input in
189 SplitsTree4 with default parameters.
190
191 *Mitochondrion analysis*
192 We sub-sampled genomic clean reads to 1% of the total library for each *M.*
193 *incognita* isolate. Then, we assembled them independently using the plasmidSPAdes
194 assembly pipeline [31]. We extracted the contigs containing the mitochondrion based on
195 similarity to the *M. incognita* reference mitochondrial genome sequence (NCBI ID:
196 NC_024097). In all cases, the mitochondrion was assembled in one single contig. We
197 identified the two repeated regions (63 bp repeat and 102 bp repeat, described in [17] and
198 we calculated the number of each repeat present in these regions.
199
200

201 **Results**

202 Using the NCDHT (methods), we characterised the host race status of the eleven
203 *M. incognita* Brazilian isolates studied in this paper as well as the Mexican strain
204 (Morelos), corresponding to the reference genome (Table 1). We characterised three
205 Brazilian isolates as race 1, two as race 2, three as race 3, and three as race 4. The *M.*
206 *incognita* reference genome (strain Morelos) was characterised as race 3.

207 We generated high-quality reads (~76 million per isolate) for each of the eleven
208 *M. incognita* isolates. These reads covered the assembled *M. incognita* genome size of
209 ~183Mb [20] at a depth > 100X (Table S2) for each isolate. Overall, we identified 482,545
210 positions on the *M. incognita* reference genome that were polymorphic across the
211 different isolates (~0.27% of the total number of non-ambiguous nucleotides). Each
212 isolate shows a different level of divergence from the reference with R1-2 having the
213 highest number of variable positions (377,349) and R2-6 having the least (177,979)
214 variants (Figure 2a). The R3-4 isolate originated from a pool of four populations.
215 However, low number of SNPs compared to the reference indicates that either the
216 genomes of these four populations were very close or that a specific population
217 displaced the other three (Figure 2a). Thus, the R3-4 isolate was analysed exactly as the
218 other isolates. Overall, the level of variation of the genomes of the eleven isolates
219 ranges between 0.10 % and 0.21 % compared to the Morelos reference strain. Similarly,
220 the level of variation in the mitochondrial genome ranges between 0.09% and 0.21%.

221 Interestingly, only a small proportion (< 100 positions) of the variants appear to
222 be race-specific, except for race 2 (Figure 2b) which shows >4,000 race-specific
223 variations. This is most probably due to the fact that race 2 is represented by only two
224 isolates (vs. 3 for the rest of the isolates). All the race-specific variants were checked for
225 possible impact on the predicted gene set. The vast majority (80-90%) of SNPs were

226 outside of coding regions. In these coding regions, only 10, 2 and 9 SNPs yielded non-
227 synonymous substitution or otherwise disruptive mutations in races 1, 3 and 4,
228 respectively (793 such variants were found in race 2).

229 Using principal component analysis, we showed that the eleven *M. incognita*
230 isolates formed three distinct clusters, which we named A, B and C (Figure 3). Cluster
231 A is represented by Isolate R1-2 alone, which has the highest number of variants.
232 Cluster B is constituted by R3-2 and R4-4. The rest of the isolates fall in a single dense
233 cluster C of overall lower variation. The different clusters do not correlate with host
234 race status, meaning that different isolates of the same host race are not more similar to
235 each other than isolates of different host races. Thus, there is no apparent correlation
236 between resemblance in terms of genomic variants and host race status. Furthermore,
237 we observe that the three distinct clusters also neither correlate with geographic origin
238 nor the original host plant from which the nematodes have been collected. Interestingly,
239 the four different host races are all represented in one single cluster (C). Within this
240 cluster, the total number of variable positions is 176,107. Meaning that the whole range
241 of host-race is present in a cluster that represents only 39% of the total genomic
242 variation.

243 We could confirm the same groups in the phylogenetic network analysis based on SNP
244 present in coding regions only (Figure 4). According to the two classification methods,
245 isolate R1-2 seems to be the most divergent from the rest of isolates, which is consistent
246 with its higher total number of SNPs and number of isolate-specific SNPs. Then, a
247 small cluster is composed of isolates R3-2 and R4-3 (equivalent to cluster B of the
248 PCA). Finally, a cluster (equivalent to PCA cluster C) groups the rest of the eight
249 isolates with apparent high similarity regarding SNP pattern, regardless whether

250 considered at the whole genome level (PCA, Figure 3) or only at the coding part level
251 (phylogenetic network, Figure 4).
252 Consistent with the PCA and phylogenetic network analysis, we also did not
253 observe any correlation between the number of repeats in the two repeat regions in the
254 mitochondrion (63R and 102R) and races, geographical origin or host plant of origin
255 (Table 2).
256 To investigate more widely the diversity of *M. incognita* isolates in a broader
257 analysis, we included whole-genome sequencing data for additional geographical
258 isolates recently published [21]. These genome data included one isolate from Ivory Coast,
259 one from Libya, one from Guadeloupe (French West Indies) and five from the USA [21]
260 (Figure 1). We pooled these eight new isolates with the eleven Brazilian isolates
261 produced as part of this analysis as well as the *M. incognita* Morelos strain (reference
262 genome) and performed a new PCA with the same methodology. Astonishingly, adding
263 these new isolates recovered the same separation in three distinct clusters (A, B and C)
264 (Figure 5). All the new isolates from additional and more diverse geographical origins
265 fall in just two of the previous Brazilian clusters (A and C). Cluster A that previously
266 contained R2-1 alone, now encompasses the Ivory Coast, Libyan and Guadeloupe
267 isolates. Cluster C that previously contained eight of the Brazilian isolates and covered
268 all the host races now includes the five US isolates as well as the Mexican isolate
269 (Morelos, reference genome). Cluster B remains so far Brazilian-specific with only R3-
270 2 and R4-4 in this cluster. Overall, the addition of new geographical isolates [21] did not
271 substantially increase the number of detected variable positions in the genome.
272
273
274

275 **Discussion**

276 Major *Meloidogyne* spp. control strategies such as resistant cultivars and crop
277 rotation are generally specifically deployed against a species or a race. If the identity of
278 a population is unknown, the crop selected for use in a management scheme may cause
279 dramatic increases in nematode populations [7]. This initially led to the formal recognition
280 of 'host races' within *Meloidogyne* species along with a standardised set of differential
281 hosts for distinguishing these host races. The host race concept has never been
282 universally accepted, in part because it assured only a small portion of the potential
283 variation in parasitic ability [2]. In this initial host race concept, even considering that *M.*
284 *incognita* presented hundreds of known host plants, only six different host standards
285 were used to characterise races. Another important aspect is the change of the race of an
286 isolate over time. Isolates of *M. incognita* race 2 and 3, which parasitise tobacco and
287 cotton plants respectively, switched to behaviour similar to race 3 and 2 after staying for
288 eight months on coffee plants. This shows an adaptation of the nematode race in another
289 host (coffee tree) (Rui Gomes Carneiro, personal communication). Also, new host races
290 may be identified when including additional hosts in the differential test. For example,
291 two additional *M. incognita* races have been characterized in Spain that did not fit into
292 the previously published race scheme [32]. The term race is important when used in the
293 context of nematode phenotypic diversity and generally refers to populations of
294 nematodes that differ in their ability to infest a particular host that carry specific
295 resistance genes. It has been shown that even within host races (e.g. 3 and 4), different
296 isolates vary in their virulence against resistant cotton genotypes [33].

297 Besides, important books that described screening protocols like 'Plant
298 Resistance to Parasitic Nematodes' recommended the determination of root-knot
299 nematodes races before resistance screening cultivar tests [34]. Although the terminology

300 ‘races’ of *Meloidogyne* spp. has been recommended not to be used since 2009 [2], some
301 papers related to *M. incognita* diversity or resistance of selecting cultivars on different
302 crops to this species are still using this term : coffee [35,36]; cotton [33,37] or soybean [38]. All
303 these contradictory aspects regarding host race concept and its importance in nematode
304 management in field conditions led us study to the genome-wide variability of the four
305 main races of *M. incognita.*

306 In this study, we found that the cumulative divergence across the eleven isolates
307 from Brazil and the reference genome (sampled initially from Mexico) reached ~0.27%
308 of the nucleotides. The maximum divergence between the reference genome and one
309 isolate covers ~0.21% of the nucleotides and addition of the isolates from Africa, the
310 West Indies and the USA did not increase the maximal divergence. In the nematode
311 model *C. elegans*, a RAD-seq analysis covering 8 Mb of the genome revealed 41,188
312 SNPs over >100 geographical isolates, which corresponds to an estimated ~0.52%
313 variable nucleotides in the genome [39]. A more comprehensive genome-wide analysis
314 identified ~600,000 SNPs across 40 *C. elegans* wild isolates from various geographic
315 locations [40]. This corresponds to ~0.63% of the nucleotides in the *C. elegans* genome.
316 Thus, point variations between genome sequences seem to happen at a lower magnitude
317 in *M. incognita* than in *C. elegans*. This difference is probably biologically sound and
318 not due to a difference of sensitivity of our method since we covered the *M. incognita*
319 genome at higher depths than all the previous *C. elegans* analyses. This relatively low
320 divergence is rather surprising, considering the different compatible host spectra (host
321 races), the geographically distant sampled sites and the different natures of crop species
322 of origin. These observations suggest that the classical mechanism of mutations
323 acquisition and selection of the fittest haplotype is probably not the main or at least not
324 the sole player in the adaptation of this species to different host plants and different

325 environmental conditions across the globe. Although all these nematode strains have
326 been collected in cultivated fields, that are by definition anthropised environments, there
327 are substantial differences in temperature range, weather conditions and soil properties
328 across these geographical locations. Furthermore, the seven host plants on which the
329 eleven *M. incognita* strains have been collected represent 110 – 124 Million years of
330 evolution (last common ancestor of pentapetalae [41]) with different root systems and
331 defence mechanisms. For instance, in cluster B alone, R3-2 and R4-4 appear to be very
332 closely related in terms of SNP pattern (Figure 3), however they were collected on two
333 plant species (cotton and soybean) that show >100 Million years of separation.
334 Curiously, the mitochondrial divergence was not higher than the nuclear genome
335 divergence, suggesting a recent diversification of the *M. incognita* species.
336 Maybe even more surprising is the fact that none of the biological and ecological
337 traits under consideration matches the SNP-based classification of these isolates. This
338 non-correlation between genetic grouping and bio-ecological traits is confirmed by the
339 two independent methods we employed (PCA and phylogenetic networks), and is
340 consistent with the lack of correlation previously observed with lower resolution
341 markers [14–16]. This suggests that the host race phenotypes most likely result from
342 multiple independent adaptations rather than spread of ancestral distinct host races.
343 Indeed, we observed no correlation between host race and SNP pattern at the whole
344 genome scale. This implies that there are cases of higher similarity in SNP patterns
345 between isolates of distinct host races than between isolates of the same host race. The
346 switch from race 1, which are neither able to infect cotton nor tobacco, to race 4, which
347 can infect both, appears particularly challenging. However, these two races are already
348 present in the same dense cluster C that encompasses only 39% of the SNPs. This
349 suggests that *M. incognita* is able to adapt to the challenge represented by two new

350 hosts despite showing relatively few variations regarding the genome sequence. This
351 observation is consistent with the rapid capability (5-10 generations) of *M. incognita* to
352 overcome the *Mi1.2* nematode resistance gene in tomato, in controlled lab conditions [18].
353 *M. incognita* is also able to overcome this resistance gene in the wild on tomato fields
354 [42,43], although it is unknown if this resistance breaking is as rapid as in controlled lab
355 conditions.

356 Collectively, our observations indicate that *M. incognita* is particularly versatile
357 and adaptive despite its clonal mode of reproduction. The relatively low divergence at
358 the SNP level suggests other mechanisms such as epigenetics, copy number variations,
359 movement of transposable elements or large-scale structural variations could be at play
360 in the surprising adaptability of this clonal species. Interestingly, the parthenogenetic
361 marbled crayfish has multiplied by more than 100 its original area of repartition across
362 Madagascar despite showing a surprisingly low number of nucleotide variation (only
363 ~400 SNPs on a ~3Gb genome representing a proportion of variable positions of $1.3E^{-7}$
364 only). This also led the authors to suggest that mechanism other than mutation
365 acquisition and selection of the fittest haplotype must be involved [44].

366 Recently, we have shown that the genome structure of *M. incognita* itself, could
367 participate in its versatility. Indeed, being allopolyploid, *M. incognita* has >90% of its
368 gene in 3 copies, and the majority of these gene copies show diverged expression
369 patterns one from the other [20]. How the expression patterns of these gene copies vary
370 across different geographical isolates with different host preferences would be
371 interesting to explore. Furthermore, the genome of *M. incognita* as well as the two other
372 economically relevant tropical RKNs (*M. javanica* and *M. arenaria*) seem to be
373 enriched in transposable elements (TE) as compared to their sexual relative *M. hapla*,

374 which has a more restricted host range. Whether these TE have different abundances
375 and genomic positions across the isolates is worth investigating in the future.
376 To date, the whole divergence of Brazilian isolates encompasses the known
377 divergence of all the other sequenced *M. incognita* isolates across the world. Although
378 the sampling is biased towards Brazilian and American isolates with a few coming from
379 Africa, it provides an initial snapshot of the inter-population pattern of variations. The
380 addition of more isolates from other parts of the world will unveil whether populations
381 are part of a gradient transition or are grouped into clusters as shown here.
382 If the former is correct, it means that populations switch host-compatibility
383 gradually, while if the latter is the case, it means that members of a group can quickly
384 transition between different hosts. Altogether, our results show that there is no
385 genealogical basis underlying the *M. incognita* host-races. Thus, we recommend the
386 discontinuation of the terminology "race" for this species.
387
388 **Conclusions**
389 *M. incognita* shows a variability of ranges of compatible host plants and is able
390 to proliferate in various different geographical locations around the globe. This high
391 variability and adaptability is surprising, considering their clonal mode of reproduction
392 and the overall low variation in terms of SNPs at the whole genome sequence level
393 shown in this study. This indicates that mechanism other than point mutations
394 accumulation might be major players in the genomic plasticity required to sustain their
395 adaptability (e.g. larger scale variation, including copy number variations and
396 movements of transposable elements). In addition, epigenetics regulation could also
397 provide a transcriptomic plasticity while keeping point mutation-based variability at a
398 low level. Classification of *M. incognita* isolates on the basis of their similarity in SNP

399 patterns at the whole genome level does not correlate with their biological traits (e.g.
400 host compatibility, geographical origin or current host plant). Some isolates highly
401 similar in their pattern of shared SNPs show contrasted ranges of host compatibility.
402 These observations suggest that *M. incognita* is able to rapidly adapt to other hosts and
403 new environments despite being clonal and showing low variability at the whole
404 genome scale. Thus, *M. incognita* constitutes an interesting model system to study
405 adaptive potential in clonal animals.
406

407 **Acknowledgements**
408 We thank the SPIBOC Bioinformatics platform for the data hosting.
409

410 **Author contributions**
411 EGJD, PC-S and EVSA contributed to the design of the project; RMDGC, ACZM, EM
412 contributed to collection of samples. Classification of samples into races was done by
413 RMDGC, ACZM, EM; library preparations were performed by M-JA; GDK, EGJD,
414 EVSA contributed to the analysis and interpretation of data. EGJD and GDK wrote the
415 manuscript with contributions of EVSA, RMDGC, PC-S and M-JA. All authors
416 approved the final manuscript.
417

418 **Financial support**
419 This research was supported by the Agence Nationale de la Recherche program ANR-
420 13-JSV7-0006—ASEXEVOL, the INRA program 'Accueil de Chercheur Etranger'
421 SPE 2016. GK has received the support of the EU in the framework of the Marie-Curie
422 FP7 COFUND People Programme, through the award of an AgreenSkills+ fellowship
423 (under grant agreement n°609398) and ASEXEVOL, EM received a post-doc

fellowship from the Consórcio Pesquisa Café, EVSA received funding from Embrapa for sequencing at UCA Genomix platform.

References

1. Jones, J. T. *et al.* Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol. Plant Pathol.* **14,** 946–961 (2013).
2. Moens, M., Perry, R. N. & Starr, J. L. Meloidogyne species: a diverse group of novel and important plant parasites. in *Root-knot Nematodes* 1–17 (CABI Publishing, 2009).
3. Trudgill, D. L. & Blok, V. C. Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. *Annu. Rev. Phytopathol.* **39,** 53–77 (2001).
4. Sasser, J. N., Eisenback, J. D., Carter, C. C. & Triantaphyllou, A. C. The International Meloidogyne Project - Its Goals and Accomplishments. *Annu. Rev. Phytopathol.* **21,** 271–288 (1983).
5. Singh, S. K., Hodda, M. & Ash, G. J. Plant-parasitic nematodes of potential phytosanitary importance, their main hosts and reported yield losses. *EPPO Bull.* **43,** 334–374 (2013).
6. Sasser, J. N. Identification of root-knot nematodes (Meloidogyne spp.) by host reaction. *Plant Dis. Report.* **36,** 84–86 (1952).
7. Hartman, K. M. & Sasser, J. N. Identification of *Meloidogyne* species on the basis of differential host test and perineal-pattern morphology. in *Advanced Treatise on Meloidogyne* (eds. Barker, K. R., Carter, C. C. & Sasser, J. N.) **II,** 69–77 (North Carolina State University, 1985).
8. Eisenback, J. D., Hirschmann, H., Sasser, J. N. & Triantaphyllou, A. C. *A guide to the four most common species of root-knot nematodes (Meloidogyne app.) with a pictorial key*. (Departments of Plant Pathology and Genetics North Carolina State University and The UYnited States Agency for International Development, 1981).

450 9. Eisenback, J. D. & Hunt, D. General morphology. in *Root-Knot Nematodes.* (ed. R. N.
451 Perry, M. M.) 18–54 (cabi, 2009).
452 10. Carneiro, R. M. D. G., Almeida, M. R. & Quénéhervé, P. Enzyme phenotypes of
453 *Meloidogyne* spp. populations. *Nematology* **2,** 645–654 (2000).
454 11. Esbenshade, P. R. & Triantaphyllou, A. C. Use of Enzyme Phenotypes for Identification of
455 Meloidogyne Species. *J. Nematol.* **17,** 6–20 (1985).
456 12. Blok, V. C. & Powers, T. O. Biochemical and molecular identification. (2009).
457 13. Triantaphyllou, A. C. Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes.
458 in *An advanced treatise on Meloidogyne* (eds. Sasser, J. N. & Carter, C. C.) **1,** 113–126
459 (North Carolina State University Graphics, 1985).
460 14. Baum, T. J. Characterization and Phylogenetic Analysis of Four Root-Knot Nematode
461 Species Using DNA Amplification Fingerprinting and Automated Polyacrylamide Gel
462 Electrophoresis. *Mol. Plant. Microbe Interact.* **7,** 39 (1994).
463 15. Cenis, J. L. Identification of Four Major Meloidogyne spp. by Random Amplified
464 Polymorphic DNA (RAPD-PCR). *Phytopathology* 76–78 (1993).
465 16. Santos, M. F. A. dos *et al.* Biometrical, biological, biochemical and molecular
466 characteristics of Meloidogyne incognita isolates and related species. *Eur. J. Plant Pathol.*
467 **134,** 671–684 (2012).
468 17. Okimoto, R., Chamberlin, H. M., Macfarlane, J. L. & Wolstenholme, D. R. Repeated
469 sequence sets in mitochondrial DNA molecules of root knot nematodes (Meloidogyne):
470 nucleotide sequences, genome location and potential for host-race identification. *Nucleic*
471 *Acids Res.* **19,** 1619–1626 (1991).
472 18. Castagnone-Sereno, P. Genetic variability and adaptive evolution in parthenogenetic root-
473 knot nematodes. *Heredity* **96,** 282–289 (2006).
474 19. Abad, P. *et al.* Genome sequence of the metazoan plant-parasitic nematode Meloidogyne
475 incognita. *Nat. Biotechnol.* **26,** 909–915 (2008).

476 20. Blanc-Mathieu, R. *et al.* Hybridization and polyploidy enable genomic plasticity without
477 sex in the most devastating plant-parasitic nematodes. *PLOS Genet.* **13,** e1006777 (2017).
478 21. Szitenberg, A. *et al.* Comparative Genomics of Apomictic Root-Knot Nematodes:
479 Hybridization, Ploidy, and Dynamic Genome Change. *Genome Biol. Evol.* **9,** 2844–2861
480 (2017).
481 22. Carneiro, R. M. D. G. & Almeida, M. R. A. Técnica de eletroforese usada no estudo de
482 enzimas dos nematóides de galhas para identificação de espécies. *Nematol. Bras.* **25,** 35–44
483 (2001).
484 23. Zijlstra, C., Donkers-Venne, D. T. H. M. & Fargette, M. Identification of *Meloidogyne*
485 *incognita*, *M. javanica* and *M. arenaria* using sequence characterised amplified region
486 (SCAR) based PCR assays. *Nematology* **2,** 847–853 (2000).
487 24. Randig, O., Bongiovanni, M., Carneiro, R. M. D. G. & Castagnone-Sereno, P. Genetic
488 diversity of root-knot nematodes from Brazil and development of SCAR markers specific
489 for the coffee-damaging species. *Genome* **45,** 862–870 (2002).
490 25. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
491 *EMBnet.journal* **17,** 10–12 (2011).
492 26. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
493 *ArXiv13033997 Q-Bio* (2013).
494 27. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–
495 2079 (2009).
496 28. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing.
497 *ArXiv12073907 Q-Bio* (2012).
498 29. *vcflib: a simple C++ library for parsing and manipulating VCF files, + many command-*
499 *line utilities*. (vcflib, 2018).
500 30. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal
501 component analysis of SNP data. *Bioinformatics* **28,** 3326–3328 (2012).

502 31. Antipov, D. *et al.* plasmidSPAdes: assembling plasmids from whole genome sequencing
503 data. *Bioinformatics* **32,** 3380–3387 (2016).
504 32. Robertson, L. *et al.* New Host Races of Meloidogyne arenaria, M. incognita, and M.
505 javanica from Horticultural Regions of Spain. *Plant Dis.* **93,** 180–184 (2009).
506 33. da Silva, E. *et al.* Genetic variability and virulence of Meloidogyne incognita populations
507 from Brazil to resistant cotton genotypes. *Eur. J. Plant Pathol.* **139,** 195–204 (2014).
508 34. Hussey, R. S. & Janssen, G. J. W. Root-knot nematodes: Meloidogyne species. In J. L.
509 Starr, R. Cook, & J. Bridge (Eds.), (pp. 43–70). in *Plant resistance to parasitic nematodes*
510 43–70 (CABI International, 2002).
511 35. Lima, E. A. *et al.* The Multi-Resistant Reaction of Drought-Tolerant Coffee 'Conilon Clone
512 14' to Meloidogyne spp. and Late Hypersensitive-Like Response in Coffea canephora.
513 *Phytopathology* **105,** 805–814 (2015).
514 36. Peres, A. C. J. *et al.* Resistance of genotypes against and under controlled and field
515 conditions. *Nematology* **19,** 617–626 (2017).
516 37. Mota, F. C. *et al.* New sources of resistance to Meloidogyne incognita race 3 in wild cotton
517 accessions and histological characterization of the defence mechanisms. *Plant Pathol.* **62,**
518 1173–1183 (2013).
519 38. Fourie, H., McDonald, A. H. & Waele, D. D. Host suitability of South African and foreign
520 soybean cultivars to Meloidogyne incognita race 2. *South Afr. J. Plant Soil* **23,** 132–137
521 (2006).
522 39. Andersen, E. C. *et al.* Chromosome-scale selective sweeps shape *Caenorhabditis elegans*
523 genomic diversity. *Nat. Genet.* **44,** 285–290 (2012).
524 40. Thompson, O. *et al.* The million mutation project: A new approach to genetics in
525 Caenorhabditis elegans. *Genome Res.* **23,** 1749–1762 (2013).
526 41. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines,
527 Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34,** 1812–1819 (2017).

42. Tzortzakakis, E. A., Conceição, I., Dias, A. M., Simoglou, K. B. & Abrantes, I. Occurrence of a new resistant breaking pathotype of Meloidogyne incognita on tomato in Greece. *J. Plant Dis. Prot.* **121,** 184–186 (2014).
43. Verdejo-Lucas, S., Talavera, M. & Andrés, M. F. Virulence response to the Mi.1 gene of Meloidogyne populations from tomato in greenhouses. *Crop Prot.* **39,** 97–105 (2012).
44. Gutekunst, J. *et al.* Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nat. Ecol. Evol.* **2,** 567–573 (2018).

535 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, Machado, Dallagnol, & Aranha Camargo, 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	

536 a, isolate identification code
537 b, host plant crop (soybean, *Glycine max*; cucumber, *Cucumis sativus*; tobacco, *Nicotiana tabacum*; coffee, *Coffea arabica*; cotton, *Gossypium hirsutum*.; watermelon, *Citrullus vugaris*
538 c, population id
539 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
540 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'

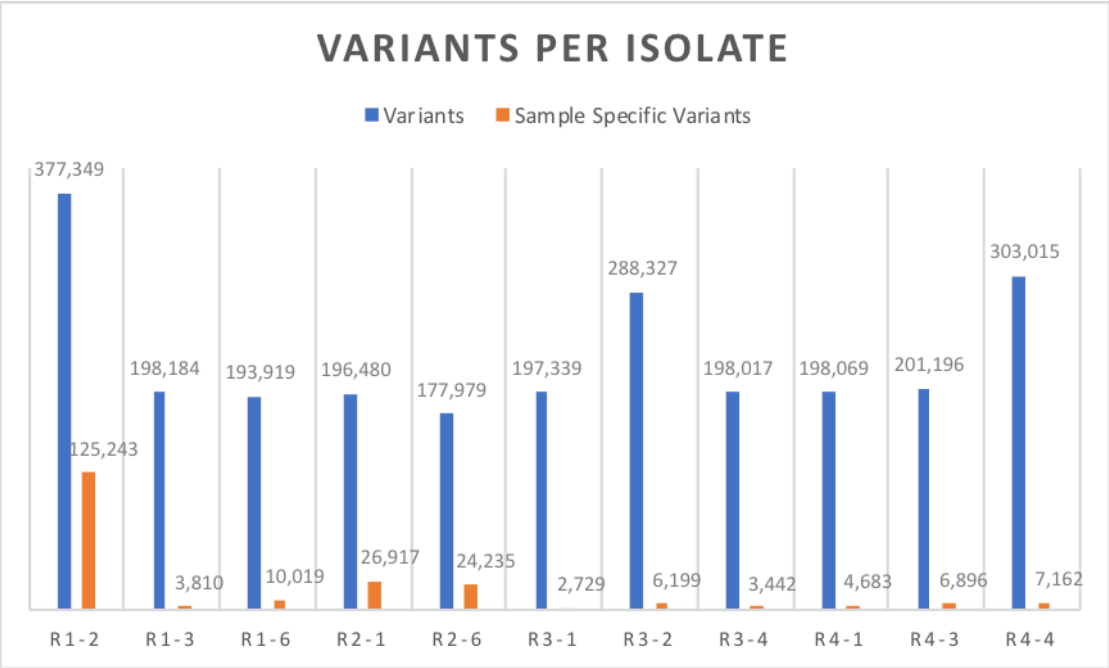
541
542
543 Table 2. Number of repeats per region (63nt and 102nt) in the mitochondrial DNA of each
544 isolate, decimals indicate truncated repeats

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

545

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

a



b

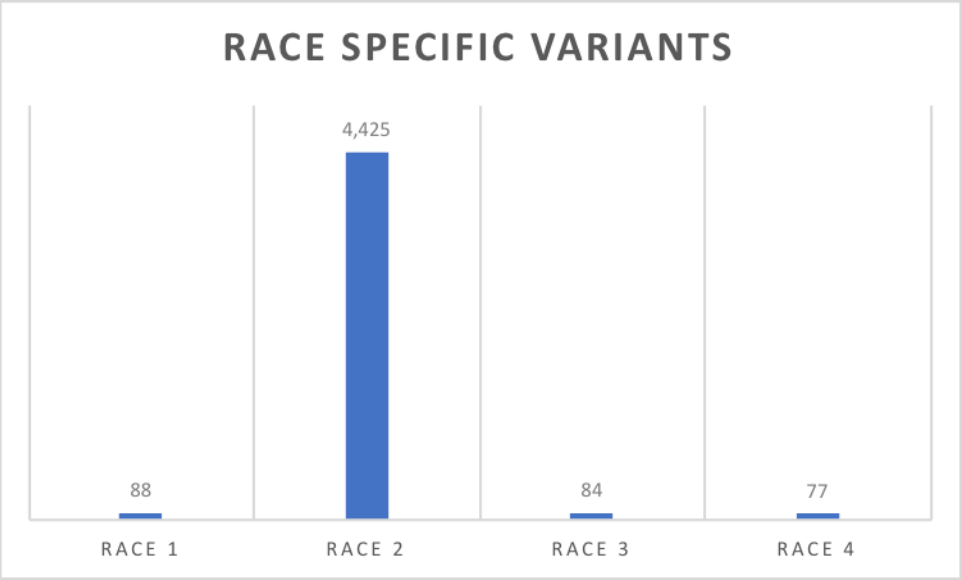


Figure 2. Variant analysis. **a.** Number of variants per isolate, **b.** Common variants in species belonging to the same race.

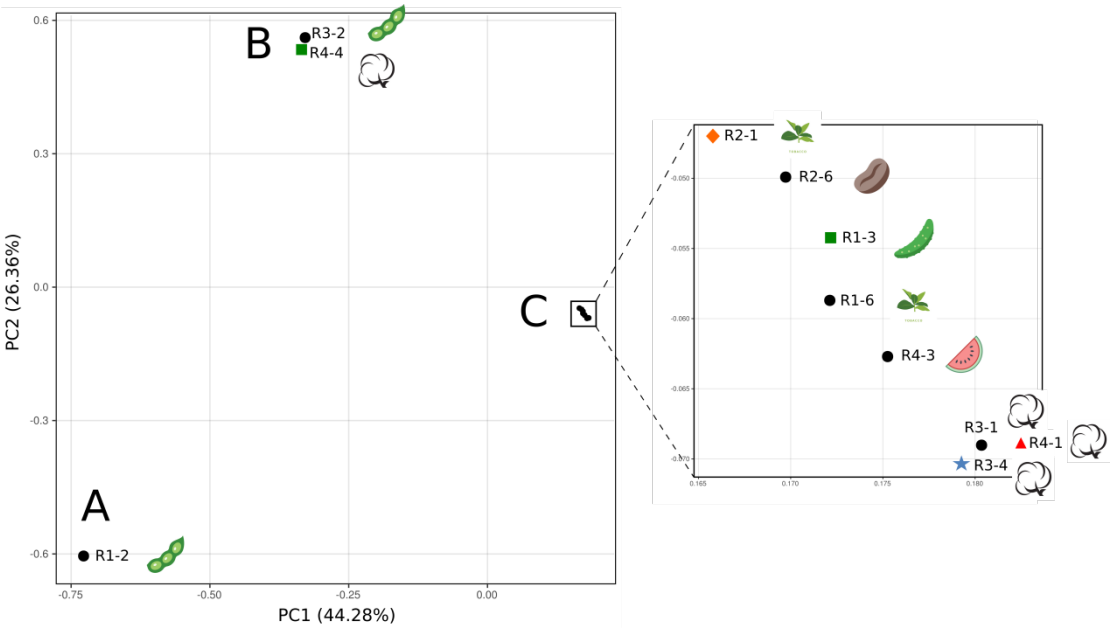


Figure 3. PCA analysis of the different *M. incognita* isolates groups them into three clusters (A, B, and C). The geographic origins are associated to coloured shapes*:* black circle: Paraná, orange diamond: Santa Catarina, green square: São Paulo, red triangle: Mato Grosso, blue star: pool. Host plant 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).

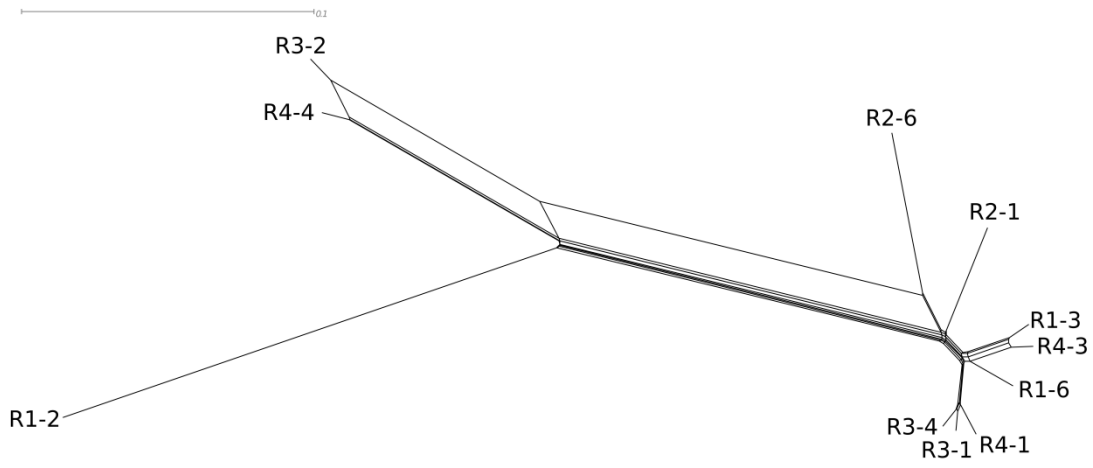
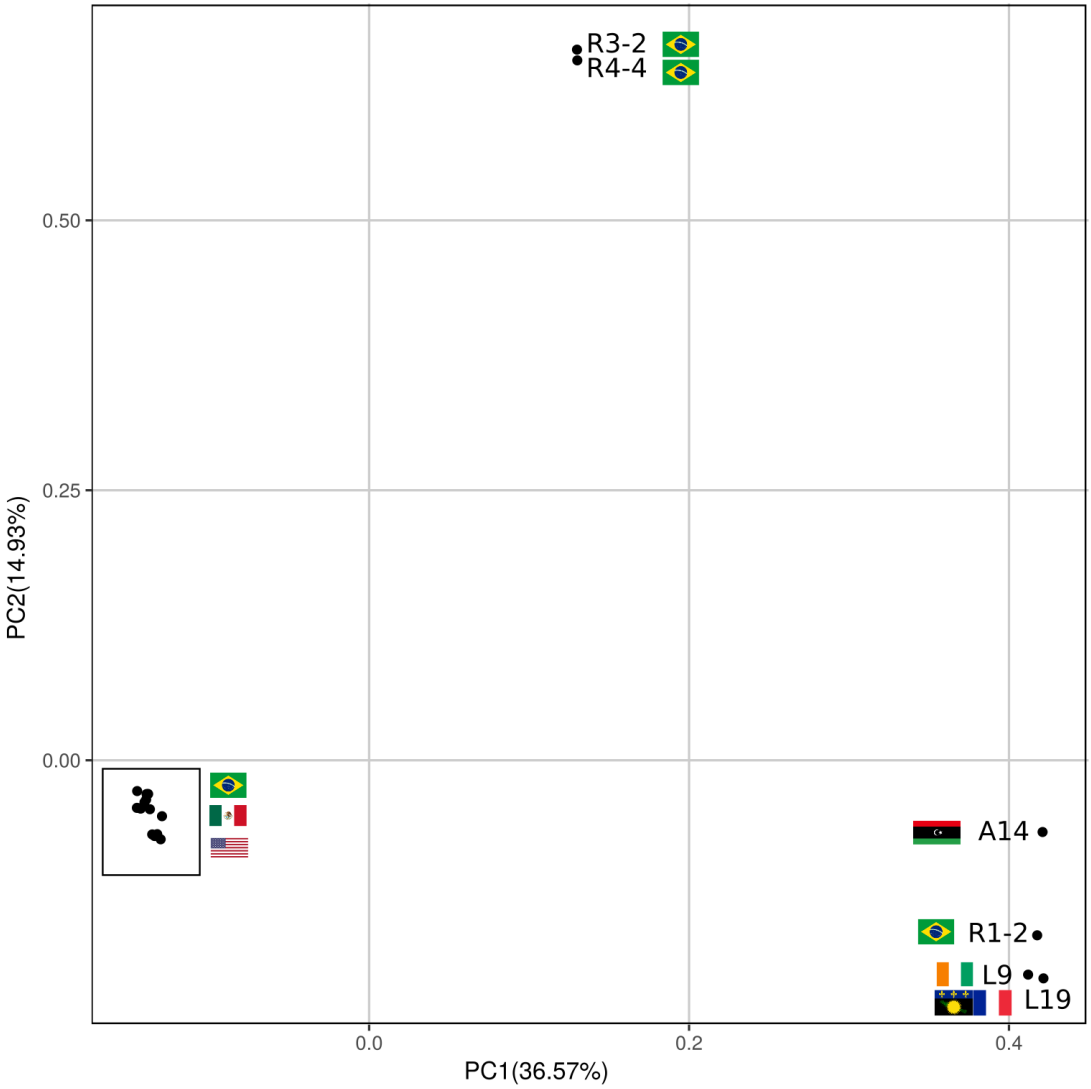
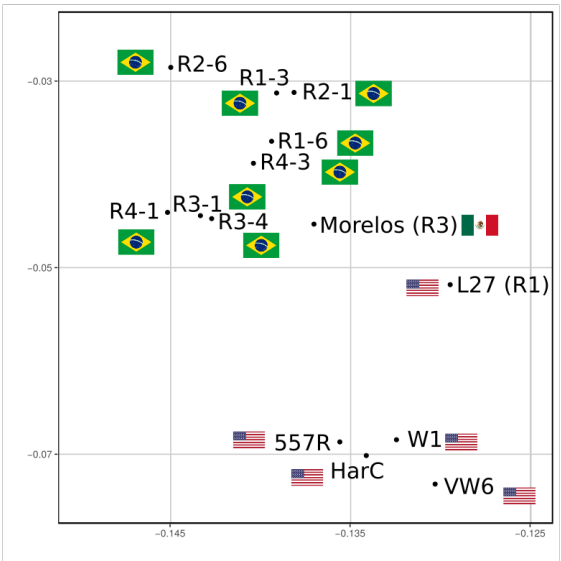


Figure 4. Phylogenetic network for *M. incognita* isolates based on SNPs present in coding sequences.


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
−0.03
−0.05
−0.07
−0.145
−0.135
−0.125

571 Figure 5. PCA analysis of all known *M. incognita* genomes. The isolates were regrouped based
572 on SNP patterns. Origin countries are indicated by flags (Brazil for R1-2, R1-3, R1-6, R2-1, R2-
573 6, R3-1, R3-2, R3-4, R4-1, R4-3, R4-4; USA for L27, 557R, HarC, W1, VW6; Mexico for
574 Morelos; Libya for A14; Ivory Coast for L9; Guadeloupe for L19). Morelos isolate is Race 3
575 and L27 isolate is Race 1.