Evolutionary history of a plant root parasite : Meloidogyne graminicola

Thi Ngan Phan

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie du Développement

École doctorale N°584 GAIA - Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau

Unité de recherche PHIM - Plant Health Institute of Montpellier

HISTOIRE ÉVOLUTIVE D'UN PARASITE RACINAIRE DE PLANTES: *MELOIDOGYNE GRAMINICOLA*

EVOLUTIONARY HISTORY OF A PLANT ROOT PARASITE: *MELOIDOGYNE GRAMINICOLA*

Présentée par Thi Ngan PHAN

Le 05 Mars 2021

Sous la direction de Stéphane BELLAFIORE

Devant le jury composé de

Prof. Godelieve GHEYSEN, Prof., Ghent University	RAPPORTEUR
Dr. Pierre ABAD, DdR (HDR), Institut Sophia Agrobiotech	RAPPORTEUR
Dr. Alain GHESQUIÈRE, DdR (HDR), IRD Montpellier	EXAMINATEUR
Dr. Etienne DANCHIN, DdR (HDR), Institut Sophia Agrobiotech	EXAMINATEUR
Prof. Pierre CZERNIC, Prof. (HDR), Université de Montpellier	PRESIDENT DU JURY
Dr. Stéphane BELLAFIORE, CdR (HDR), IRD Montpellier	DIRECTEUR DE THÈSE
Dr. Guillaume BESNARD, CdR (HDR), CNRS Toulouse	INVITE, ENCADRANT

UNIVERSITÉ DE MONTPELLIER

This PhD was prepared in the PHIM research unit

Institut de Recherche pour le Développement

Centre IRD de Montpellier

911 Avenue Agropolis

34394 Montpellier Cedex 5

France

Abstract in English

Meloidogyne graminicola (*Mg*), commonly known as the rice root-knot nematode, is a devastating pest worldwide, causing severe damages to rice production and other crops. The success of parasitism of this species in different rice-growing agrosystems raises questions about the origin, propagation, reproductive strategies and evolution of the genome in relation to adaptation processes linked to changes in environments and hosts. Using a comparative genomic approach within the species and with other related species, this work aims to trace the molecular evolution and life history of *Mg*. The results of this study can be decomposed into four main parts: **First**, genomic comparative analysis between *Mg* and the phylogenetically related species, *M. oryzae*, showed their close relationship suggesting their common ancestors and a potential hybrid origin. **Second**, a high quality haploid reference sequence of the *Mg* nuclear genome was assembled and annotated. The total DNA content of a somatic cell, probably in the diploid state, is ~82 Mb and the genome has a heterozygosity of ~2%. Transposable elements and horizontal gene transfers were annotated throughout the genome and the potential role of these elements in the parasitism discussed. **Third**, the comparative genomic study of 13 *Mg* isolates collected on a large geographical scale (mainly in South East Asia countries) has allowed to put forward hypotheses on the evolutionary history of the species: i) existence of homologous recombination in *Mg* genome, which is likely to occur during meiosis; ii) loss of heterozygosity (LoH) and copy number variations (CNVs) are potential evolutionary mechanisms that would contribute to bring plasticity in the genome of this meiotic parthenogenesis organism. iii) a recent dispersion of this pathogen in Southeast Asia. **Fourth**, comparative genomic analysis between virulent and avirulent pathotypes revealed SNVs, LoHs and CNVs affected several genes coding for potential effectors in virulent pathotype and thus potentially associated with the process of species adaptation to a resistant rice genotype. **Overall**, these results could open up new avenues of studies in the mechanisms governing the evolution of nematodes and provide the molecular material needed for further in-depth studies such as the evolutionary history of this parasite and the molecular mechanisms governing the adaptation of the parasite during plant-pathogen interactions.

Key words: Plant parasitic nematode, evolutionary history, genome assembly, molecular evolution, comparative genomic, virulence acquisition

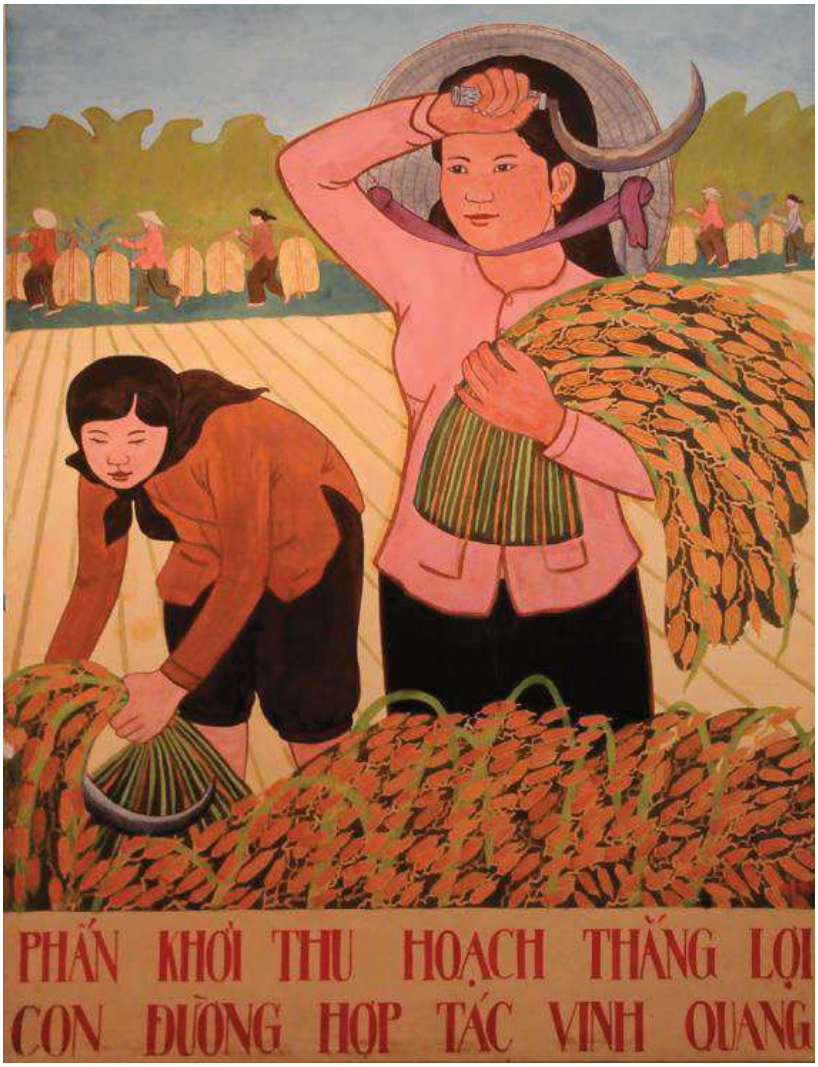
Résumé en Français

Meloidogyne graminicola (*Mg*), communément appelé nématode à galles du riz, est un ravageur dévastateur dans le monde entier, qui cause de graves dommages à la production de riz et à d'autres cultures. Le succès du parasitisme de cette espèce dans différents agrosystèmes de culture du riz soulève des questions sur l'origine, la propagation, les stratégies de reproduction et l'évolution du génome en relation avec les processus d'adaptation liés aux changements d'environnements et d'hôtes. En utilisant une approche de génomique comparative au sein de l'espèce et avec d'autres espèces apparentées, ce travail vise à retracer l'évolution moléculaire et l'histoire de vie de *Mg*. Les résultats de cette étude peuvent être décomposés en quatre parties principales: **Premièrement**, l'analyse génomique comparative entre le *Mg* et une espèce phylogénétiquement apparentée, *M. oryzae*, a montré leur étroite relation suggérant un ancêtre commun et une origine hybride potentielle. **Deuxièmement**, une séquence de référence haploïde de haute qualité du génome nucléaire de *Mg* a été assemblée et annotée. Le contenu total de l'ADN d'une cellule somatique, probablement à l'état diploïde, est de ~82 Mb et le génome a une hétérozygotie de ~2%. Les éléments transposables et les transferts horizontaux de gènes ont été annotés dans l'ensemble du génome et le rôle potentiel de ces éléments dans le parasitisme a été discuté. **Troisièmement**, l'étude par génomique comparative de 13 isolats de *Mg* collectés sur une grande échelle géographique (principalement dans les pays d'Asie du Sud-Est) a permis de formuler des hypothèses sur l'histoire évolutive de l'espèce : i) l'existence d'évènements de recombinaison homologue dans le génome de *Mg*, qui sont susceptibles de se produire lors de la méiose; ii) la perte d'hétérozygotie (LoH) et les variations du nombre de copies (CNV) sont des mécanismes évolutifs potentiels qui contribueraient à apporter de la plasticité dans le génome de cet organisme à parthénogenèse méiotique. ii) une dispersion récente de cet agent pathogène en Asie du Sud-Est. **Quatrièmement**, l'analyse par génomique comparative entre les pathotypes virulents et avirulents a révélé que les SNV, les LoH et les CNV affectaient plusieurs gènes codant pour des effecteurs potentiels dans le pathotype virulent et donc potentiellement associés au processus d'adaptation de l'espèce à un génotype de riz résistant. **Dans l'ensemble**, ces résultats pourraient ouvrir de nouvelles perspectives d'étude dans les mécanismes qui régissent l'évolution des nématodes et fournir le matériel moléculaire nécessaire pour étudier l'histoire évolutive de ce parasite et les mécanismes moléculaires régissant l'adaptation du parasite lors des interactions plante-pathogène.

Mots clés: Nématode parasite des plantes, histoire de l'évolution, assemblage du génome, évolution moléculaire, génomique comparative, acquisition de virulence

"To supply good cultivar for every paddy field"

(Propaganda painting for Vietnamese farmers, by Nam Trân)



"The more productive, the happier"

(Propaganda painting for Vietnamese farmers during 1950-1980)

Acknowledgements

Funding:

I acknowledge that my PhD was funded by French Embassy in Hanoi, Vietnam following the program "Excellence de l'Ambassade de France au Vietnam". I would not have this opportunity to do my Ph.D. in France without their financial support.

To Dr. Stéphane BELLAFIORE

I feel truly privileged to have Stéphane as the greatest and the most influenced teacher in my research life until now. I have worked with him since 2014, and now he is the director and the principal supervisor in my thesis. In him, a combination of the excellent profession, kindness, encouragement, friendliness, patience, positive energy, and jocoseness are always there to guide me through all the difficulties in this PhD journey. His advice on both research and life is always priceless to me on the way to becoming a scientist. My admiration for him not only for his brilliant career but also for his beautiful family. My sincere gratitude to him and his family for all their passionate supports in both my works and my private life. I am very grateful for his countless time, even the weekend and holiday, to improve my thesis story and thesis manuscript.

To Dr. Guillaume BESNARD

I harness the space to express my admiration to my co-supervisor, Guillaume, for his immense expertise, for the speed of his thinking and working, and for the perfection in all the things he engages in. I would like to express my gratefulness for his precious advice which always very importaint for the thesis progress as well as for the publication. I acknowledge his willingness to travel several times between Toulouse and Montpellier to discuss and resolve problems in my thesis. I am also grateful for all of his efforts in editing my publications. I have learned many things from his rigorous way of writing and managing ideas in the manuscripts.

To professor Michel LEBRUN

My grateful attitude to professor Michel for becoming the director of my thesis for almost three years. I have known him since 2015 when he was a director of LMI RICE in Hanoi, Vietnam. He is very friendly, generous, and professional. He is always willing to help students and give them the wings to fly far to their dreams. He is like a venerable grandpa for me and for several other Vietnamese students. We all love him and respect him very much!

To the jury members

I am particularly grateful with Prof. Godelieve GHEYSEN, Dr. Pierre ABAD, Dr. Alain GHESQUIÈRE, Dr. Etienne DANCHIN, Prof. Pierre CZERNIC for having accepted to carry out the evaluation of my thesis

To my thesis committee

I sincerely thank Dr. Antony CHAMPION, Dr. Etienne DANCHIN, Dr. Gilles BENA, Dr. Sophie MANTELIN, Dr Georgios KOUTSOVOULOS, Prof. Pierre CZERNIC for being willing to be part of my thesis committee, for the constructive and interesting discussions we could have, and for all the support they provided me along the course of this work.

To my Nématologie Fonctionnelle et Evolutive (NeFonEv) team

I would like to specially thank to beautiful and lovely girls in our team: Anne-Sophie MASSON, NGUYỄN Thị Huế and, Marieliesse VERMEIRE for their accompanies in both work and life. They are all PhD and postdoc, very busy but always cheerful, kindful, and encourage that make me feel so lucky and comfortable to share the PhD life with them. Always, NeFonEv lab-meeting is always full of interest and excitement for seeing each other and for discussing works.

I thank Rania OUAZAHROU for her nice works during the four-month internship of her master I. Also, I thank Jamel ARIBI for his help in greenhouse works.

To my co-authors for publication

I thank all co-authors who contribute to the two articles in my thesis. I thanks Laetitia PERFUS-BARBEOCH and Djampa KOZLOWSKI for help in some experiments. I thank Julie ORJUELA, Hai HO-BICH, Christophe KLOPP, and Alexis DEREEPER for their help and advice in bioinformatics.

To IRD staffs and managers

I thank Dr. Gilles BENA, vice-director of UMR PHIM, and the secretary team for all their support and kindness when solving administrative tasks. I thank Anne-Sophie PETITOT and Agnès PINEL for their caring when I do my lab-works.

To University of Montpellier

Thanks for being a student of Ecole doctoral GAIA and University of Montpellier, I have chances to follow interesting free scientific and French language courses which are very helpful for my PhD and my living in France.

To my friends

Thank for all my friends, Doâa, Ái Mỹ, Hồng Anh, Rayan, Kade, Teerarat, Nguyệt, Adrian, Cecile, Celine, Elvira, Branly, Kwanho, Coline, Marlene, Juliette, and many other friends in IRD for all the fun we have had during the lunch-time, for cheerful outdoor events (skiing, partying), and for organizing the interesting journal club and great PhD days. I enjoyed my PhD life thanks to all of you.

To my family

Wherever am I, whatever I do, my family is always the greatest motivation and encouragement for me. I am the lucky one who have my husband accompanying me in both work and life in France. I would work hard and take care of them during the rest of my life to express my gratitude to their unconditional love to me.

Table of contents

FREQUENTLY USED ABBREVIATIONS

CN	Cyst nematode
CNV	Copy number variation
ETI	effector-triggered immunity
ETS	effector-triggered susceptibility
HGT	Horizontal gene transfer
HR	Hypersensitive response
J2	Juvenile stage 2
LoH	Loss of heterozygosity
Mg	*Meloidogyne graminicola*
NGS	Next Generation Sequencing
PAMP	Pathogen-Associated Molecular Patterns
PPN	Plant-parasitic nematode
PTI	PAMP Triggered Immunity
RE	Recombination
RKN	Root-knot nematode
ROS	Reactive oxygen species
SNV	Single nucleotide variants
TE	Transposable elements

GENERAL INTRODUCTION

Rice is the staple food for more than half of the world population, a figure that can reach up to 70% for southern regions of Asia continent (Muthayya et al., 2014). Asia is responsible for more than 90% of global rice production, which is estimated at around 705 million tons of paddy rice annually (FAOSTAT, 2018). Rice production in Asia is therefore of essential importance for both local and global food security. Unfortunately, rice production in this region is threatened by root-knot nematodes [RKNs, genus *Meloidogyne* Göeldi, 1892 (Nematoda: Meloidogynidae)] which are responsible for about 15% of the total rice economic losses annually (Mantelin et al., 2017). Different RKN species can affect rice, but only a few have significantly detrimental effects. The most prevalent RKN in Asian rice agrosystems is *M. graminicola* (*Mg*) which was commonly named as rice RKN. *Mg* penetrate the root systems, takes water and nutrient from root cells to develop and reproduce inside the root, that ultimately results in poor growth of the crop and substantial yield losses (Netscher & Erlan, 1993; Plowright & Bridge, 1990; Soriano et al., 2000). In Asia, *Mg* is reported to reduce the yield by 8–70% on upland rice in the Philippines and by 16–97% on lowland rice in India, Nepal, Bangladesh and Indonesia (Bridge et al., 2009; Padgham et al., 2004). Even so, the damage caused by this pathogen is likely underestimated because its above-ground symptoms (such as stunting, chlorosis, and loss of vigour) can lead farmers to wrongly attribute the damage to nutritional and water-associated disorders or to secondary diseases. Besides, *Mg* rapidly increased population and infested a large area in very short time. For instance, in northern Italy, where this pest was recently detected, the total infested area has increased by approximately five times in just one year (from 19 to 90 ha in 2016-2017; EPPO Reports, 2017). Furthermore, changes in agricultural practices in response to environmental (climate change) and socio-economic conditions have led to a dramatic increase in *Mg* populations, particularly in Asia (De Waele & Elsen, 2007). Therefore, this devastating plant pathogen is classified as a quarantine pest in several countries (e.g. Brazil, Madagascar, China) (EPPO, 2017-2018) and was added recently in the EPPO Alert List in Europe after being detected in Northern Italy (Fanelli et al., 2017).

Despite the huge impact of *Mg* on agriculture and on a large geographical scale, the tools available to control it remain limited. The use of nematicides has been the most efficient way to control plant-parasitic nematodes (PPNs) in the field for many years, but have deleterious impact to both farmers and the environment. Alternative strategies such as deep-water cultivating and crop rotation have been tested, however, the efficiency remain limited due to water shortage and farmers' crop production habits. When it comes to pest control, the

more we know about the pest and the factors that influence its development and spread, the easier, more cost-effective, and efficient the control will be. For example, questioning the success of a pathogenic agent in colonizing an environment should make it possible to anticipate the risks of seeing this same agent colonizing new environments. Understanding the evolutionary processes that lead to the emergence of a pathogen is essential to assess the risks posed by potential future emerging pathogen and expansion. Besides, the application of knowledge about pathogen strategies in the adaptation process allows us to predict the impact of invasions and improve the development and implementation of sustainable control strategies. Up to date, little is known about the origin, evolutionary process and population dynamic of *Mg* species.

The lack of nematode fossils has made it difficult to study the evolutionary history of PPNs. However, over the last two decades, the advent of Next Generation Sequencing (NGS) has generated tremendous molecular resources and their exploitation has highlighted the immense diversity of nematodes and the evolutionary history of PPNs and RKNs in particular. Following the sequencing of the model nematode *C. elegans* (The *C. elegans* Sequencing Consortium, 1998), the parasitic nematode *Brugia malayi* (Ghedin et al., 2004) and the phytoparasitic nematodes *Meloidogyne hapla* and *M. incognita*, it appeared that the genomes of the nematodes are very diverse, with species-specific singularities. In *M. graminicola*, due to the absence of genomic sequences, most genetic studies were based on sporadic mitochondrial DNA and rDNA genomic material that did not allow the entire genomic complexity of the species to be understood. Moreover, the level of heterozygosity observed in *Mg* complicated the analysis. A first draft of the *Mg* genome was published in 2018 revealing a 35 Mb genome (Somvanshi, et al., 2018). However, the assembly was very fragmented and incomplete. The generation of a high quality genomic sequence of *M. graminicola* was therefore a prerequisite for any genomic study and more specifically for evolution and life history.

Recent surveys suggest that *Mg* has an almost worldwide distribution (*EPPO Global Database*, 2020). It was firstly discovered in America (Golden & Birchfield, 1965), Asia (Golden & Birchfield, 1968) and recently found in Europe (Fanelli et al., 2017), and Madagascar (Chapuis et al., 2016). Intraspecific diversity analysis among widespread *Mg* isolates using short internally transcribed spacer (ITS) sequence (ca. ~500bp) revealed very few polymorphisms, and was unable to address their phylogenetic relationships (Bellafiore et al., 2015; Pokharel et al., 2010). The apparent worldwide distribution, recent discovery, and low

diversity in ITS sequence of *Mg* raise questions about the origin of the species and its probably recent dispersal. A comparison of the complete mitochondrial and nuclear genomes of populations collected on a global scale will bring solid evidence to answer these questions.

Host-parasite co-evolution occurs under reciprocal selective pressure and leads to adaptive genetic changes. The rice cultivar Zhonghua 11 was described as an *Oryza sativa* variety resistant to *Mg* (Phan et al., 2018). A hypersensitive like-response was observed in Zhonghua 11 in response to *Mg* infection, suggesting a gene-for-gene (*R*/*Avr*) relationship. Recently, my host team has isolated a *Mg* isolate from a Cambodian rice field that bypasses the resistance mechanisms of Zhonghua 11 (unpublished data). It is of particular interest to study how *Mg* could overcome the plant's resistance mechanism and establish a compatible interaction resulting in the development of disease. One of the current challenges in plant nematology is to elucidate the molecular basis of avirulence/virulence, which would allow a better understanding of the complex cascades of events leading to susceptibility or resistance. This knowledge could lead to an understanding of how certain pathotypes bypass the plant's resistance genes and thus contribute to better management of resistance genes. Proteins produced and secreted by pathogens in the host plant cells are known to participate in the sophisticated molecular dialogue between the host and the parasite and thus promote infection. These proteins, called effectors, help optimize the parasite's fitness by acting, for example, on the host's metabolic pathways or defense mechanisms (van Baarlen et al., 2007). But these effectors can also be recognized by the plant and induce an incompatible interaction. In this case, the effector is recognized as an Avr factor interacting directly or indirectly with a R-protein. Thus, it has been suggested that effectors co-evolve with factors produced by the plant, leading to the selection of virulent or avirulent pathotypes during evolution. This evolution of effectors and targets was initially described by the Zig zag model as described by Jones and Dangl (2006). Therefore, once an R protein in the plant participates in immunity, the effector can potentially either trigger the susceptibility (ETS) or trigger the immunity (ETI). In RKNs, it has been suggested that mutations in effector genes could lead to the hijacking of innate plant systems, which would call into question the durability of plant resistance genes (Castagnone-Sereno et al., 1994; Castagnone-Sereno, 2002; Janssen et al., 1998; Petrillo & Roberts, 2005). Besides, gene loss events were associated with the virulent isolate of *M. incognita* and *M. javanica* against resistant tomato plants (Castagnone-Sereno et al., 2019; Gleason et al. 2008).

Mg has an amazing adaptability. Not only can it colonize different environments and be found on most continents, but it is also a polyphagous pathogen capable of infecting more than

120 hosts including cereal and vegetable plants as well as weeds (EPPO Global Database, 2020). This is surprising because *Mg* reproduces mainly by meiotic parthenogenesis reproduction which was thought to limit evolutionary capacities. However, current comparative analysis has revealed that parthenogenetic RKNs can promote genome plasticity and adaptation by several mechanisms including horizontal gene transfer (HGTs) (Danchin et al., 2016; Opperman et al., 2008), insertion of transposable elements (TE), gene conversions, and gene duplications/deletions (i.e. gene copy number variants - CNVs) (Blanc-Mathieu et al., 2017; Castagnone-Sereno et al., 2019; Cook et al., 2012; Szitenberg et al., 2017). Besides, the sexual reproduction is supposed to occur only occasionally in *Mg*, as cytological observations suggest (Triantaphyllou, 1969). If it was the case for sex reproduction in *Mg*, genetic variants might be accumulated during genetic segregation through mating and meiotic recombination. Furthermore, it is also believed that a pathogen with a mixed reproductive mode like *Mg* might combine evolutionary advantages of both parthenogenesis and sexual reproduction that could explain the particularly strong aggressiveness of this pest. However, these hypotheses on the *Mg* reproduction mode are uncertain and have never been confirmed at the molecular level. Up to date, the mechanisms of acquisition of genetic variants associated with *Mg*'s adaptive capacities remain unknown. Comparative genomic analysis should answer this question.

The genus *Meloidogyne* includes species that vary in terms of genetic organization, modes of reproduction (from amphimixis to mitotic parthenogenesis), and host range. The genus *Meloidogyne* thus constitutes a remarkable genetic study tool. Understanding the evolutionary mechanisms leading to the adaptive success of these species are of particular interest. RKNs have been classified into three main lineages (Clades I, II and III) using the 18S ribosomal RNA gene as a marker (Tigano et al., 2005). The recent publication of the parthenogenetic (clade I) and sexual (clade II) mitotic RKN genomes has allowed comparative genomics studies to be carried out, and to discover interesting mechanisms involved in the evolution of genomes. For instance, the genomes of obligate mitotic parthenogenetic (*M. incognita*, *M. arenaria*, *M. javanica*, and *M. enterolobii*) and the meiotic parthenogenetic *M. floridensis* from clade I are polyploidy containing triplicated/duplicated regions with high degrees of sequence divergence (2.6%-8.4%) which results from a hybrid origin (Blanc-Mathieu et al., 2017; Jaron et al., 2020; Szitenberg et al., 2017). Conversely, the genome of facultative sexual *M. hapla* (clade II) is probably diploid and does not have a genome with divergent duplications (Blanc-Mathieu et al., 2017; Opperman et al., 2008). *Mg* belongs to clade III and is phylogenetically close to *M. oryzae (Mo)* and *M. exigua (Mex)* despite of differences

in mode of reproduction, ploidy level and host ranges (Ali et al., 2015; Tigano et al., 2005). *Mg* and *Mex* are facultative meiotic parthenogenetic and diploid (2n=36) but the main hosts are rice and coffee, respectively (Lapp & Triantaphyllou, 1972; Triantaphyllou, 1985). On the other hand, *Mg* and *Mo* are mainly rice pathogens but have different reproduction modes and ploidy levels (obligate mitotic parthenogenesis and 3n=54 for *Mo*) (Esbenshade & Triantaphyllou, 1987). The close phylogenetic relationship and the presence of these three species in South America raises the question of their common origin and potential interspecific hybridization events. Comparative genomics between these three species could allow to determine whether *Mg* underwent hybridization and/or genome duplication during speciation and share a common ancestor with the other species of this clade III.

In this thesis, we used tools from Next Generation Sequencing (NGS) and then performed a comparative genomic analysis to study:

1. Genome content and structure of *Mg*
2. Population structure and spread of *Mg* isolates at a worldwide scale
3. Genetic features (TE, RE, HGT, and CNV) which might promote genome plasticity and evolution
4. Genetic mechanisms (mutation, CNV) associated with virulence behavior of an *Mg* isolate against a resistant rice cultivar
5. Ancestral relationship of *Mg* with other closely related *Meloidogyne* spp.

Following above tasks, four main work packages were organized as below:

Work package 1: Study intraspecific diversity of *Mg* isolates and the ancestral relationship with other closely related species in the mitochondrial genome and some low-copy nuclear contigs

Work package 2: Generate a high-quality genomic sequence and investigate genome features (TE, HGT) associated with *Mg* parasitism

Work package 3: Study the intraspecific diversity of *Mg* isolates and highlight specific evolutionary genomic features (REs, CNVs)

Work package 4: Study of genetic signatures to reveal genetic mechanisms potentially associated with the acquisition of virulence in *Mg*

This manuscript is organized in six parts. The first part is a bibliographical synthesis presenting the current knowledge on the subject. The next four parts report the results obtained from the four work packages described above and are presented in the form of four articles. Finally, the

last part of this manuscript consists of the general discussion and perspectives that emerge from my work.

References

Ali, N., Tavoillot, J., Mateille, T., Chapuis, E., Besnard, G., El Bakkali, A., Cantalapiedra-Navarrete, C., Liébanas, G., Castillo, P., & Palomares-Rius, J. E. (2015). A new root-knot nematode *Meloidogyne spartelensis* n. Sp. (Nematoda: Meloidogynidae) in Northern Morocco. *European Journal of Plant Pathology*, *143*(1), 25–42. https://doi.org/10.1007/s10658-015-0662-3

Bellafiore, S., Jougla, C., Chapuis, É., Besnard, G., Suong, M., Vu, P. N., De Waele, D., Gantet, P., & Thi, X. N. (2015). Intraspecific variability of the facultative meiotic parthenogenetic root-knot nematode (*Meloidogyne graminicola*) from rice fields in Vietnam. *Comptes Rendus Biologies*, *338*(7), 471–483. https://doi.org/10.1016/j.crvi.2015.04.002

Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J.-M., Rocha, M. D., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J.-F., Kozlowski, D. K., Cazareth, J., Couloux, A., Silva, C. D., Guy, J., Kim-Jo, Y.-J., Rancurel, C., Schiex, T., Abad, P., & Danchin, E. G. J. (2017). Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genetics*, *13*(6), e1006777. https://doi.org/10.1371/journal.pgen.1006777

Bridge, J., Luc, M., & Plowright, R. A. (2009). Nematode parasites of rice. In Plant-parasitic nematodes in subtropical and tropical agriculture (pp. 69–108). CAB International.

Castagnone-Sereno, P., Mulet, K., Danchin, E. G. J., Koutsovoulos, G. D., Karaulic, M., Rocha, M. D., Bailly-Bechet, M., Pratx, L., Perfus-Barbeoch, L., & Abad, P. (2019). Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita* . *Molecular Ecology*, *28*(10), 2559–2572. https://doi.org/10.1111/mec.15095

Castagnone-Sereno, P., Wajnberg, E., Bongiovanni, M., Leroy, F., & Dalmasso, A. (1994). Genetic variation in *Meloidogyne incognita* virulence against the tomato *Mi* resistance gene: Evidence from isofemale line selection studies. *Theoretical and Applied Genetics*, *88*(6), 749–753. https://doi.org/10.1007/BF01253980

Castagnone-Sereno, Philippe. (2002). Genetic variability of nematodes: A threat to the durability of plant resistance genes? *Euphytica*, *124*(2), 193–199. https://doi.org/10.1023/A:1015682500495

Chapuis, E., Besnard, G., Andrianasetra, S., Rakotomalala, M., Nguyen, H. T., & Bellafiore, S. (2016). First report of the root-knot nematode *Meloidogyne graminicola* in Madagascar rice fields. *Australasian Plant Disease Notes*, *11*(1), 32. https://doi.org/10.1007/s13314-016-0222-5

Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., Wang, J., Hughes, T. J., Willis, D. K., Clemente, T. E., Diers, B. W., Jiang, J., Hudson, M. E., & Bent, A. F. (2012). Copy Number Variation of Multiple Genes at *Rhg1* Mediates Nematode Resistance in Soybean. *Science*, *338*(6111), 1206–1209. https://doi.org/10.1126/science.1228746

Danchin, G. J. E., Guzeeva, A. E., Mantelin, S., Berepiki, A., & Jones, T. J. (2016). Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. *Molecular Biology and Evolution*, *33*(6), 1571–1579. https://doi.org/10.1093/molbev/msw041

De Waele, D., & Elsen, A. (2007). Challenges in Tropical Plant Nematology. *Annual Review of Phytopathology*, *45*(1), 457–485. https://doi.org/10.1146/annurev.phyto.45.062806.094438

EPPO Global Database. (2020). https://gd.eppo.int/

Esbenshade, P. R., & Triantaphyllou, A. C. (1987). Enzymatic relationships and evolution in the genus *Meloidogyne*(Nematoda: Tylenchida). *Journal of Nematology*, *19*(1), 8–18.

Fanelli, E., Cotroneo, A., Carisio, L., Troccoli, A., Grosso, S., Boero, C., Capriglia, F., & De Luca, F. (2017). Detection and molecular characterization of the rice root-knot nematode *Meloidogyne*

graminicola in Italy. *European Journal of Plant Pathology*, *149*(2), 467–476. https://doi.org/10.1007/s10658-017-1196-7

Ghedin, E., Wang, S., Foster, J. M., & Slatko, B. E. (2004). First sequenced genome of a parasitic nematode. *Trends in Parasitology*, *20*(4), 151–153. https://doi.org/10.1016/j.pt.2004.01.011

Gleason, C. A., Liu, Q. L., & Williamson, V. M. (2008). Silencing a candidate nematode effector gene corresponding to the tomato resistance gene *Mi-1* leads to acquisition of virulence. *Molecular Plant-Microbe Interactions: MPMI*, *21*(5), 576–585. https://doi.org/10.1094/MPMI-21-5-0576

Golden, A. M., & Birchfield, W. (1965). *Meloidogyne graminicola* (Heteroderidae) a new species of root-knot nematode from grass. *Proceedings of the Helminthological Society of Washington*, *32*(2), 228–231

Janssen, G. J. W., Scholten, O. E., van Norel, A., & Hoogendoorn, C. (J.). (1998). Selection of virulence in *Meloidogyne chitwoodi* to resistance in the wild potato *Solanum fendleri*. *European Journal of Plant Pathology*, *104*(7), 645–651. https://doi.org/10.1023/A:1008625404330

Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M., & Schwander, T. (2020). Genomic features of parthenogenetic animals. *BioRxiv*, 497495. https://doi.org/10.1101/497495

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329. https://doi.org/10.1038/nature05286

Lapp, N. A., & Triantaphyllou, A. C. (1972). Relative DNA content and chromosomal relationships of some *Meloidogyne*, *Heterodera*, and *Meloidodera* spp. (Nematoda: Heteroderidae). *Journal of Nematology*, *4*(4), 287–291.

Mantelin, S., Bellafiore, S., & Kyndt, T. (2017). *Meloidogyne graminicola*: A major threat to rice agriculture. *Molecular Plant Pathology*, *18*(1), 3–15. https://doi.org/10.1111/mpp.12394

Muthayya, S., Sugimoto, J. D., Montgomery, S., & Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. *Annals of the New York Academy of Sciences*, *1324*, 7–14. https://doi.org/10.1111/nyas.12540

Netscher, C., & Erlan. (1993). A root-knot nematode, *Meloidogyne graminicola*, parasitic on rice in Indonesia. *Afro-Asian Journal of Nematology*, *3*(1), 90–*95*.

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., & Windham, E. (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences*, *105*(39), 14802–14807. https://doi.org/10.1073/pnas.0805946105

Padgham, J. L., Duxbury, J. M., Mazid, A. M., Abawi, G. S., & Hossain, M. (2004). Yield loss caused by *Meloidogyne graminicola* on lowland rainfed rice in Bangladesh. *Journal of Nematology*, *36*(1), 42–48.

Petrillo, M. D., & Roberts, P. A. (2005). Isofemale line analysis of *Meloidogyne incognita* virulence to cowpea resistance gene *Rk*. *Journal of Nematology*, *37*(4), 448–456.

Phan, N. T., De Waele, D., Lorieux, M., Xiong, L., & Bellafiore, S. (2018). A Hypersensitivity-Like Response to *Meloidogyne graminicola* in Rice (*Oryza sativa*). *Phytopathology*, *108*(4), 521–528. https://doi.org/10.1094/PHYTO-07-17-0235-R

Plowright, R., & Bridge, J. (1990). Effect of *Meloidogyne graminicola* (Nematoda) on the establishment, growth and yield of rice Cv Ir36. *Nematologica*, *36*(1–4), 81–89. https://doi.org/10.1163/002925990X00059

Pokharel, R. R., Abawi, G. S., Duxbury, J. M., Smat, C. D., Wang, X., & Brito, J. A. (2010). Variability and the recognition of two races in *Meloidogyne graminicola*. *Australasian Plant Pathology*, *39*(4), 326–333. https://doi.org/10.1071/AP09100

Somvanshi, V. S., Tathode, M., Shukla, R. N., & Rao, U. (2018). Nematode genome announcement: A draft genome for rice root-knot nematode, *Meloidogyne graminicola*. *Journal of Nematology*, *50*(2), 111–116. https://doi.org/doi:10.21307/jofnem-2018-018.

Soriano, I. R., Prot, J. C., & Matias, D. M. (2000). Expression of tolerance for *Meloidogyne graminicola* in rice cultivars as affected by soil type and flooding. *Journal of Nematology*, *32*(3), 309–317.

Szitenberg, A., Salazar-Jaramillo, L., Blok, V. C., Laetsch, D. R., Joseph, S., Williamson, V. M., Blaxter, M. L., & Lunt, D. H. (2017). Comparative genomics of apomictic root-knot nematodes: Hybridization, ploidy, and dynamic genome change. *Genome Biology and Evolution*, *9*(10), 2844–2861. https://doi.org/10.1093/gbe/evx201

The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, *282*(5396), 2012–2018. https://doi.org/10.1126/science.282.5396.2012

Tigano, M. S., Carneiro, R. M. D. G., Jeyaprakash, A., Dickson, D. W., & Adams, B. J. (2005). Phylogeny of *Meloidogyne* spp. Based on 18S rDNA and the intergenic region of mitochondrial DNA sequences. *Nematology*, *7*, 851–862. https://doi.org/10.1163/156854105776186325.

Triantaphyllou, A. C. (1969). Gametogenesis and the chromosomes of two root-knot nematodes, *Meloidogyne graminicola* and *M. naasi*. *Journal of Nematology*, *1*(1), 62–71.

Triantaphyllou, A. C. (1985). Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes. In J. N. Sasser & C. C. Carter (Eds.), *An advanced treatise on* Meloidogyne. *Biology and control* (Vol. 1, pp. 113–126). Raleigh. http://agris.fao.org/agris-search/search.do?recordID=US8743737

van Baarlen, P., van Belkum, A., Summerbell, R. C., Crous, P. W., & Thomma, B. P. H. J. (2007). Molecular mechanisms of pathogenicity: How do pathogenic microorganisms develop cross-kingdom host jumps? *FEMS Microbiology Reviews*, *31*(3), 239–277. https://doi.org/10.1111/j.1574-6976.2007.00065.x

CHAPTER I

Bibliography synthesis

Summary in English

"If you know the enemy and know yourself, you need not fear the result of a hundred battles."

(Sun Tzu, In "The Art of War")

Rice is a staple food that feeds more than half of the world's population. Constraints exerted by pathogens can seriously compromise food security if sustainable solutions to control them are not rapidly found. The rice root-knot nematode, *Meloidogyne graminicola*, is considered to be a major pathogen, particularly against the species *Oryza sativa*. In conjunction with the search for integrated control methods, it is essential to know the genetic determinants involved in the evolution of the species in order to anticipate the possible mechanisms of adaptation of the parasite to control strategies. Thus, the main objective of this thesis was to acquire knowledge on the life history of this parasite and on the molecular mechanisms that enable it to evolve and adapt to different environments, and in particular to resistant rice varieties.

Before presenting my work, I would like to provide an update on the state of knowledge regarding the parasitism of this pathogen. The knowledge presented in this chapter will inform the discussion in subsequent chapters.

The first part of this manuscript describes the life cycle, habitat and diversity of *M. graminicola*. The life cycle of *M. graminicola* is relatively short for an animal (about 20 days), which allows it to rapidly increase its population as soon as environmental conditions are favourable for its development. This pest has mainly spread around the world by associating itself with rice growing areas where it infects rice as well as weeds that are associated with rice fields. It has an amazing ability to adapt to various types of environment, temperatures, soil types and especially to the reducing conditions observed in flooded rice fields. Identification of the species on morphological criteria is impossible, so enzymatic (e.g. esterase profiles) and molecular (e.g. SCAR) markers had to be developed. The intraspecific morphological variability is also very limited, which raises questions about the molecular variability of the species. I have studied this intraspecific diversity at the molecular level and the results of this work are presented and discussed in chapters II and IV of this manuscript.

The second part deals with the relationship between *M. graminicola* and his host. *M. graminicola* has the ability to invade a wide range of hosts and has a very significant impact on rice cultivation. It is an obligate soil-borne endo-parasite that attacks plants through the root system. It penetrates the root tip, migrates through the apoplastic pathway and targets cells close to the endoderm, which it transforms into giant feeder cells, before reproducing inside the root. To allow infection, this nematode secretes a cocktail of effectors that will allow it to manipulate

the plant's defense systems and metabolism. During the incompatible interaction, the effectors secreted by the parasite into the plant cell are potentially recognized by so-called resistance proteins (*R* genes) according to the "gene for gene" principle. Nematodes can then potentially adapt to no longer be recognized by these R proteins by losing or modifying the effector(s) in question or by acquiring additional effectors whose function will be to inactivate the plant's defense response. Chapter V of this manuscript will focus on the adaptive capacity of *M. graminicola*, which can lead to the genesis of virulent pathotypes.

The third part provides information on *M. graminicola*'s genomic and transcriptomic resources. A draft genome was available before the beginning of the thesis but incomplete and therefore not suitable for comparative genomic analysis. We therefore proceeded with the sequencing and assembly of the nuclear and mitochondrial genomes (Chapter III). The results of this work then provided data for the analyses presented in Chapters IV and V. In addition, and thanks to the transcriptomic analyses available and the reference genomes that we have established, it was possible to discuss the potential role of certain effectors in the resistance bypassing process observed in a virulent pathotype isolated in Cambodia (chapter V).

In the last part, knowledge on the evolutionary histories of species of the genus *Meloidogyne* (RKN) is synthesized. RKNs present surprising singularities. Notably, they include species with an intriguing diversity of reproductive modes ranging from obligatory sexual reproduction (amphimixis) to strict asexual reproduction (apomixis), with intermediate species capable of alternating between sexual (amphimixis) and asexual (automixis) reproduction. The RKNs of clade I (which are mainly composed of asexual mitotic species) have a polyploid and heterozygous genome that most likely comes from a hybrid origin. Mitotic RKNs possess several molecular mechanisms that can affect the plasticity of their genome such as transposable elements, palindromic sequences and copy number variations (CNVs) which could play an important role in their remarkable capacity to adapt to environments and plant resistance. Conversely, the amphimictic *M. hapla* has a more simple genome (in size and ploidy) and has less TE than parthenogenetic species. Finally, numerous horizontal gene transfers (HGTs), mainly of bacterial origin, have been identified in RKN genomes. These HGTs seem to play an important role in the parasitism of these nematodes. Chapter IV of this study presents the genetic characteristics and explores the basis of the molecular evolution of *M. graminicola*.

Résumé en Français

Le riz est un aliment de base qui nourrit plus de la moitié de la population mondiale. Les contraintes exercées par les agents pathogènes peuvent sérieusement compromettre la sécurité alimentaire si des solutions durables permettant de les contrôler ne sont pas rapidement trouvées. Le nématode à galles du riz, *Meloidogyne graminicola*, est considéré comme un agent pathogène majeur en particulier vis-à-vis de l'espèce *Oryza sativa*. Conjointement à la recherche de méthodes de luttes intégrées, il est primordial de connaitre les déterminants génétiques impliqués dans l'évolution de l'espèce afin d'anticiper les possibles mécanismes d'adaptation du parasite aux stratégies de contrôle. C'est ainsi que le principal objectif de cette thèse était d'acquérir des connaissances sur l'histoire de vie de ce parasite et sur les mécanismes moléculaires qui lui permettent d'évoluer et de s'adapter à différents environnements et notamment aux variétés de riz résistantes.

Avant de présenter mes travaux, j'aimerais faire le point sur l'état des connaissances concernant le parasitisme de cet agent pathogène. Les connaissances présentées dans ce chapitre viendront étayer la discussion des chapitres suivants.

La première partie de ce manuscrit décrit le cycle de vie, l'habitat et la diversité de *M. graminicola*. Le cycle de vie de *M. graminicola* est relativement court pour un animal (une vingtaine de jours), ce qui lui permet d'augmenter rapidement sa population dès que les conditions environnementales sont propices à son développement. Ce ravageur s'est surtout répandu dans le monde en s'associant aux zones de riziculture où il infecte le riz ainsi que les graminées et cypéracées qui sont associées aux rizières. Il possède une étonnante capacité d'adaptation à divers types d'environnement, de températures, de types de sol et surtout aux conditions réductrices observées dans les rizières inondées. L'identification de l'espèce sur des critères morphologiques est impossible si bien que des marqueurs enzymatiques (ex : profils d'estérases) et moléculaires (ex : SCAR) ont dû être développés. La variabilité morphologique intraspécifique et elle aussi très limité ce qui interroge sur la variabilité moléculaire de l'espèce. J'ai étudié cette diversité intraspécifique au niveau moléculaire et les résultats de ces travaux sont présentés et discutés au niveau des chapitres II et IV de ce manuscrit.

La deuxième partie porte sur la relation entre *M. graminicola* et son hôte. *M. graminicola* a la capacité d'envahir un large éventail d'hôtes et a un impact très significatif sur la culture du riz. C'est un endo-parasite obligatoire tellurique qui s'attaque aux plantes par le système racinaire. Il pénètre dans l'extrémité racinaire, migre par la voie apoplastique et va cibler des cellules proches de l'endoderme qu'il va transformer en cellules géantes nourricières, avant de se reproduit à l'intérieur de la racine. Pour permettre l'infection, ce nématode sécrète un cocktail d'effecteurs qui va lui permettre de manipuler les systèmes de défense et le

métabolisme de la plante. Durant l'interaction incompatible, les effecteurs sécrétés par le parasite dans la cellule végétale sont potentiellement reconnus par des protéines dites de résistance (*R* genes) selon le principe "gène pour gène". Les nématodes peuvent alors potentiellement s'adapter pour ne plus être reconnu par ces protéines R en perdant ou modifiant le(s) effecteur(s) incriminés ou en acquérant des effecteurs supplémentaires dont la fonction sera d'inactiver la réponse de défense de la plante. Le chapitre V de ce manuscrit s'intéressera à la capacité d'adaptation de *M. graminicola* qui peut notamment aboutir à la genèse de pathotypes virulents.

La troisième partie nous renseigne sur les ressources génomiques et transcriptomiques de *M. graminicola*. Une ébauche de génome était disponible avant le commencement de la thèse mais incomplète et ne se prêtait donc pas à une analyse de génomique comparative. Nous avons donc procédé au séquençage et à l'assemblage des génomes nucléaire et mitochondriale (chapitre III). Les résultats de ces travaux ont ensuite permis d'alimenter en données les analyses présentées dans les chapitres IV et V. En outre, et grâce aux analyses transcriptomiques disponibles et aux génomes de référence que nous avons établis, il a été possible de discuter du rôle potentiel de certains effecteurs dans le processus de contournement de la résistance observé chez un pathotype virulent isolé au Cambodge (chapitre V).

Dans la dernière partie, les connaissances sur les histoires évolutives des espèces du genre *Meloidogyne* (RKN) sont synthétisées. Les RKN présentent des singularités étonnantes. Notamment, ils comprennent des espèces avec une diversité intrigante de modes de reproduction allant de la reproduction sexuée obligatoire (amphimixie) à la reproduction asexuée stricte (apomixie), avec des espèces intermédiaires capables d'alterner entre reproduction sexuée (amphimixie) et asexuée (automixie). Les RKNs du clade I (qui se composent principalement d'espèces asexuées mitotiques) ont un génome polyploïde et hétérozygote qui provient très certainement d'une origine hybride. Les RKN mitotiques possèdent plusieurs mécanismes moléculaires pouvant affecter la plasticité de leur génome tels que des éléments transposables, des séquences palindromiques et des variations du nombre de copies (CNVs) ce qui pourraient jouer un rôle important dans leur capacité remarquable d'adaptation aux environnements et la résistance des plantes. À l'inverse, l'amphimictique *M. hapla* à un génome plus simple (en taille et au niveau de la ploïdie) et présente moins de TE que les espèces parthénogénétiques. Enfin, de nombreux transferts horizontaux de gènes (HGTs), principalement d'origine bactérienne, ont été identifiés dans les génomes des RKNs. Ces HGTs semblent jouer un rôle important dans le parasitisme de ces nématodes. Le chapitre IV de cet ouvrage présente les caractéristiques génétiques et explore les bases de l'évolution moléculaire de *M. graminicola*.

I. *Meloidogyne graminicola* – biology and parasitisms

1. Life cycle

M. graminicola (Mg) is an obligatory endo-parasite that attacks plant through the root system. Following its hatching and while the stage 2 infectious larva (Juvenile stage 2 or J2) is in the soil, it will be attracted to the root of the plant and penetrate it by its extremity generally at the root elongation zone. At the endophytic stage, the J2 migrates intercellularly (apoplastic) through the vascular tissues of the root in search of plant cells close to the endoderm, which it will transform into giant cells whose function as the nematode feeding sites. The formation of giant cells at the end of the roots can be observed within 48 hours after infection. Galls will appear at these sites due to hyperplasia and hypertrophy of the cells located around the feeding site (Roy, 1976). The J2 swells during 3-4 dpi and moults into J3 (5-8 dpi), J4 (9-12 dpi), adult male or female stages (13-15 dpi), and freshly laid eggs (18 -28 dpi) depending on the pathotype and environmental conditions. A female lays about 200 – 500 unembryonated or partially embryonated eggs. Eggs hatch to J2 after 6-9 days under favourable environmental conditions (20-35^{0}C, high moisture), and this extends over several weeks. Compared with other *Meloidogyne* species, *Mg* has a relatively fast life cycle. Its life cycle depends on temperature conditions. It could be as little as 15 days at an ambient temperature of 25-35º C (Bridge & Page, 1982) or prolong to 65 days at low temperature during the winter (Rao & Israel, 1973)

2. Distribution and habitat

M. graminicola was firstly reported in Louisiana, USA on *Echinochloa colonum* L in 1965 (Golden & Birchfield, 1965). Later, this nematode was described as new pest in 30 rice (*Oryza sativa*) varieties in Laos (Golden & Birchfield, 1968). Several reports followed of its association with rice and weeds in many counties in the Americas and Asia (Figure 1, 2). In Americas, *Mg* has been reported in several countries including Brazil (Bellé et al., 2018; Monteiro & Barbosa Ferraz, 1988; Negretti et al., 2017; Oliveira et al., 2018), Colombia (Bastidas & Montealegre, 1994), The Republic of Ecuador (Gilces et al., 2016), and The United States of America (Minton et al., 1987; Windham & Golden, 1990). In Asia, *Mg* invade most of the rice growing countries including China (Long et al., 2017; Song et al., 2017; Tian et al., 2017; Wang et al., 2017; Xie et al., 2019; Zhou et al., 2014), India (Salalia et al., 2017), Vietnam (Cuc & Prot, 1992), Laos (Golden & Birchfield, 1968), Thailand (Toida et al., 1996), Cambodia (Bellafiore et al., 2015), Philippines (Cabasan et al., 2018), Indonesia (Netscher & Erlan, 1993), Malaysia (Zainal-Abidin et al., 1994), Myanmar (CABI Bioscience, 2001), Nepal (Pokharel, 2009), Pakistan (CABI Bioscience, 2001), Singapore (AVA, 2001), Bangladesh (Rahman,

1990), and Sri Lanka (Nugaliyadde et al., 2001). Recently, *M. graminicola* was found in South Africa (Chapuis et al., 2016), and Europe (Fanelli et al., 2017) (Figure 2).

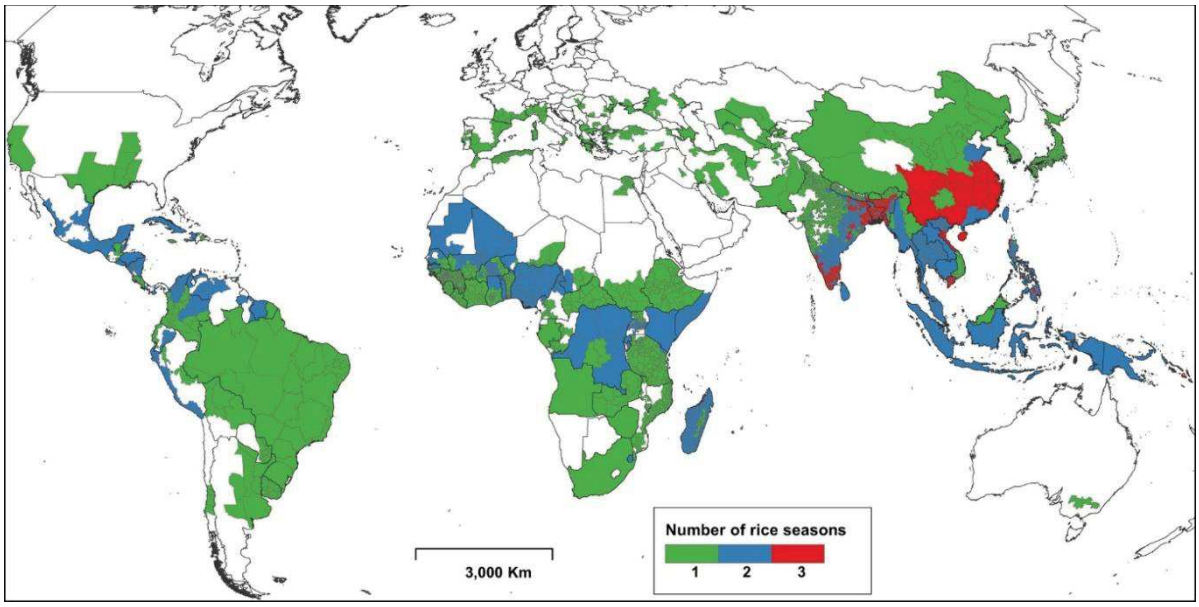


Figure 1. Rice growing area over the world (Laborte et al., 2017)

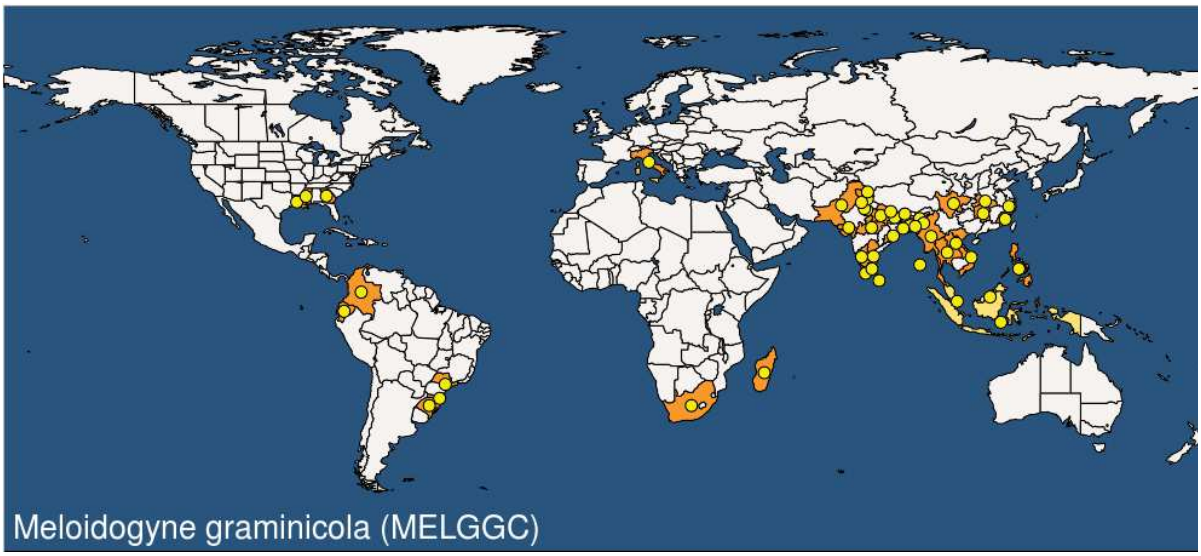


Figure 2. Presence of *Mg* (in yellow) over the world (EPPO database, 2020)

Mg has been reported in a wide range of habitats and environments, from upland to lowland. It infects and reproduces in irrigated rice (Kumar et al., 2014), semi-deep (Prasad et al., 1985) or deepwater rice (Rahman, 1990). *Mg* damages to rice varieties was greater in sandy soils than in clay soils (Soriano et al., 2000).

3. Diversity and identification

Several *Mg* populations in different countries have been isolated and studied for their diversity in terms of morphology, hosts, aggressiveness, and short DNA sequence (ITS). The length of the body and stylet were measured to investigate intraspecific diversity among isolates from Bangladesh, Nepal, India, Thailand and the USA (Jepson, 1983; Pokharel et al., 2010) and among isolates collected from Vietnam (Bellafiore et al., 2015). Different populations from North America and Asia were similar in host range except a Florida isolate suggesting that *Mg* consists of more than one pathotypes (Pokharel et al., 2010). Aggressiveness to the susceptible host has also been shown to differ significantly between some populations collected in Asia (Bellafiore et al., 2015; Pokharel et al., 2007).

Different techniques based on biochemistry (enzyme activity profiles of malate dehydrogenase and esterase) and DNA markers have been used to distinguish *Mg* from other *Meloidogyne* spp. At the molecular level, the mitochondrial and nuclear genomes can be used to define identification markers also called barcodes. These markers can either reveal identification after sequencing (e.g. rDNA, mitochondrial genes) or simply reveal the species by amplification of a specific DNA fragment (e.g. SCAR). Mitochondrial genomes of *Mg* have been sequenced recently and showed specific VNTRs regions that can potentially be used (111R and 94R) (Besnard et al., 2014; Humphreys-Pereira & Elling, 2015; Sun et al., 2014). Genetic markers using polymorphisms in VNTR regions and sequence characterized amplified regions (SCARs) were currently being tested for direct identification of *Mg* (Besnard et al., 2014; Bellafiore et al., 2015; da Silva Mattos et al., 2019; Sun et al., 2014). Nucleotide polymorphisms were found between *Mg* populations using the internal transcribed spacer (ITS) region and sequencing PCR products obtained from amplification of the region in isolated individuals (Bellafiore et al., 2015; Pokharel et al., 2007). However, this apparent polymorphism has been questioned by discovering that the level of heterozygocity in *Mg* is significant (~2%) and that within the same individuals several copies of rDNA can be found (Besnard et al. 2019; Phan et al., 2020). In other words, intraspecific variability may be confused with the polymorphism of rDNA regions present in a single nematode.

II. *Meloidogyne graminicola* – host interaction

1. Host range

M. graminicola was initially found on barnyard grass, *Echinochloa colonum* (Golden and Birchfield, 1965) but it has a wide range of hosts including more than 120 species of mono- and dicotyledonous plants. At the agricultural level, *Mg* is primarily a threat to rice, but it can

also parasite for instance wheat, oat, onion, tomato and bananas (summary in Ravindra et al., 2017).

2. Infection process (penetration, interaction with plant immune system)

RKNs in the invasive stage J2 migrate in the soil, looking for a plant root to penetrate inside through the root tip epidermis or the elongation zone. Once inside the root, the J2 migrates intercellularly through the cortical layer of the root in search of plant cells close to the endoderm to transform these cells into its feeding site. During the penetration process and establishment of feeding site, RKNs need first to pass the physical barrier of the plant root, usually the cell wall. To overcome this obstacle, nematodes use a combination of mechanical penetration with its needle-like stylet together with an arsenal of chemical weapons, including many cell wall-degrading enzymes such as β-1,4-endoglucanase, a functional polygalacturonase or a pectate lyase (Huang et al., 2005; Jaubert et al., 2002; Rosso et al., 1999).

Also at this stage of infection, a war between plant-nematode occurs with two scenarios that can be observed: the nematode successfully infects the plant and establishes the disease (compatible interaction), or the plant successfully combats the nematode (incompatible interaction). First, the host plant detects the presence of the RKNs through molecular signals located on their surfaces called pathogen-associated molecular patterns (PAMPs), or through damage-associated molecular patterns (DAMPs) released by the disrupted host plant tissues to induce the basal immune response (Figure 3, Jones & Dangl, 2006). Ascarosides are the first and the only nematode PAMPs that have been described so far (Manosalva et al., 2015). Ascarosides constitute an important conserved group of proteins produced by RKNs (Manosalva et al., 2015). From the plant side, PAMPs and DAMPs are recognized by pattern recognition receptors (PRRs) located on the cell membrane, resulting in pattern-triggered immunity (PTI) to halt further RKN colonization (Jones & Dangl, 2006). One of the first PRRs described was a leucine-rich repeat (LRR)-RLK encoded by the *nilr1* gene (nematode-induced LRR-RLK 1), firstly identified in the *Arabidopsis thaliana* but widely conserved among dicots and monocots (Mendy et al. 2017).

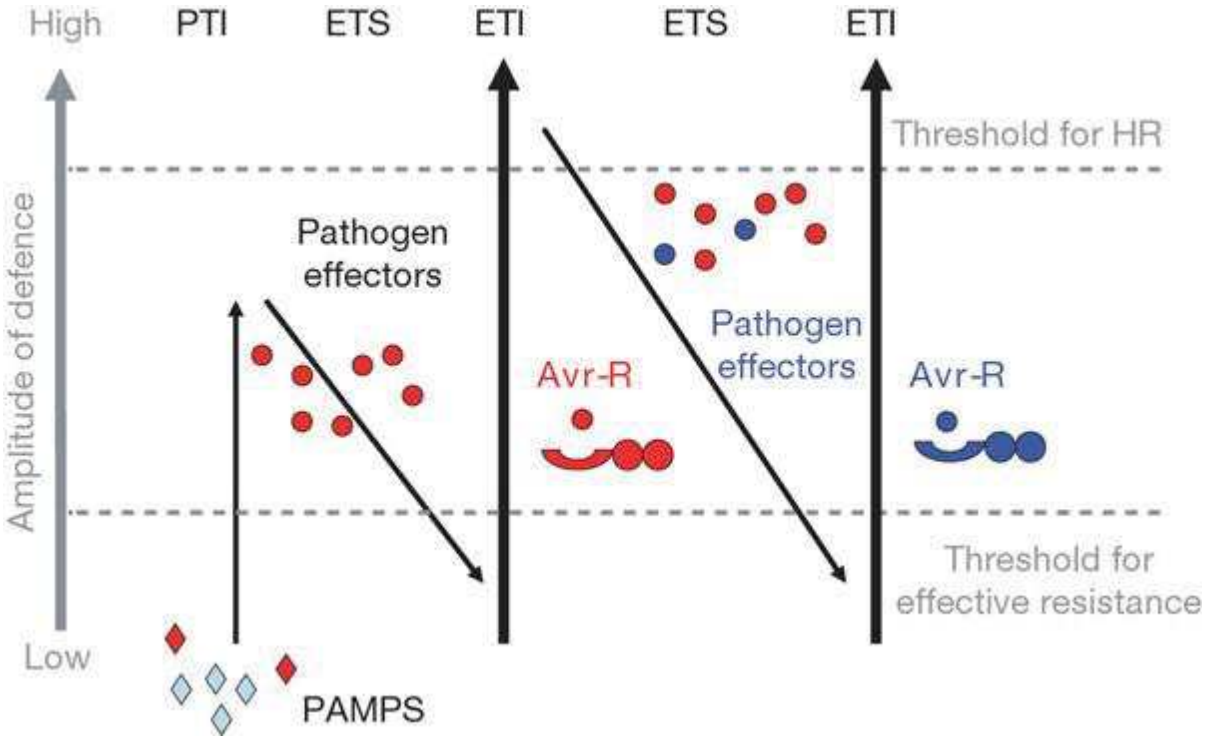


Figure 3. The Zig-zag-zig model of plant – pathogen interaction (Jones and Dangl, 2006).PTI: pathogen (or PAMP) triggered immunity; ETI: effector triggered immunity; PRR: pattern recognition receptor; ETS: effector triggered susceptibility; HR: hypersensitive response.

The war is continuing, RKNs secret hundreds of effectors (mainly proteins) into the host cells, some function in hijacking the plant immune system (reviewed by Haegeman et al., 2012 and Vieira & Gleason, 2019). Several effectors from *Mg* were confirmed to directly manipulate the plant response to promote its infestation, such as MgGPP and MgMO237 (Chen et al., 2017, 2018) (see more on part III.2 in this chapter). When innate immunity is suppressed by pest effectors, effector-triggered susceptibility (ETS) is induced (compatible interaction). Or when a given effector from the pest is recognized by the appropriate proteins in plants, effector-triggered immunity (ETI) is induced, and the PTI response is amplified, the plant presents disease resistance (incompatible interaction) (Jones & Dangl, 2006). In this case, it usually results in a hypersensitive cell death response (HR) at the infection site to block the development of the pathogen, or eventually kill them (Figure 3, Jones & Dangl, 2006). In turn, RKNs evolved to avoid ETI by acquiring additional effectors that may suppress ETI as well as by shedding or diversifying the recognized effector gene (see more on part IV.5 in this chapter and chapter V). The suppression of ETI by effectors is effective in a susceptible host and results in ETS, which allows the nematode to infect the plant (incompatible interaction) (Jones &

Dangl, 2006). It's worth to note that, so far, most of the *Oryza sativa* accessions being used worldwide are susceptible to *Mg* (see more on part II.2 in this chapter).

3. Establishment of feeding site

When the invaded J2 reaching the root tip, they establish giant cells that serve as feeding sites providing nutrients for the nematode to complete its life cycle (Jones, 1981). Giant cells are formed in a similar way for all *Meloidogyne* species. Parenchymal cells surrounding the phloem at the root tip go through sequential mitoses without cytokinesis, thus leading to an increase in the size of the cell and the number of nuclei (Rodiuc et al., 2014). During giant cells formation, RKNs use a cocktail of various effectors to induce the reprogramming of the metabolism of the roots by manipulating the expression patterns of root tip-specific genes or induce the expression of genes that are not usually expressed in the roots (reviewed in Haegeman et al., 2012; Vieira and Gleason, 2019). One key feature of giant cells is their excessive isotropic growth, which requires extensive and coordinated cell wall remodeling (Sobczak et al., 2011). Consistently, cell wall-modifying enzymes, including endoglucanases, extensins, hydrolases, and structural proteins, were found to be enhanced concomitantly with the deposition of newly synthesized cell wall material and the loosening of the cellulose/cross-linking glycan network after *Meloidogyne* spp. infection (Caillaud et al., 2008; Sobczak et al., 2011). In addition, other cellular processes, including nucleotide synthesis, protein and sucrose biosynthesis, transporters, starch production, phospholipid production, photosynthesis and glycolysis, are generally activated in the giant cells from *Mg*-infected rice roots (Ji et al., 2013; Kyndt et al., 2012). Interestingly, epigenetic mechanisms were suggested to play a role in transcriptional reprogramming in infected plant tissues since several genes involved in chromatin remodelling, DNA methylation, small RNA formation and histone modifications were all highly expressed inside giant cells induced by *Mg* in rice (Ji et al., 2013).

4. Interaction with plant's hormone

In order to maintain a compatible interaction, obligate endoparasites such as *Mg* must manipulate the plant's defense and its interacting hormone pathways at the same time. Interestingly, nematodes regulate plant hormones differently during compatible (host plant is susceptible) and incompatible interaction (host plant is resistant). In a compatible interaction, both jasmonate (JA)- and salicylic (SA)-mediated rice defenses are activated during the early stages of *M. graminicola* infection, but these responses are suppressed during the later stages of infection (Kumari et al., 2017; Kumari et al., 2016). Besides, genes of the PR13/thionin gene

family and genes involved in the phenylpropanoid pathway (e.g. *OsPAL*, *OsC4H*, *OsCOMT*, and *OsCAD*) were strongly suppressed in rice gall tissues during giant cell formation of *Mg* infection (Ji et al., 2013, 2015; Kyndt et al., 2012). In addition, callose-degrading enzyme (*OsGNS5*), which degrade callose – a plant defense's related protein, were overexpressed during feeding of *Mg* at 4 dpi (Jacobs et al., 2003; Luna et al., 2011). It is also proposed that plant's hormone balance is important at the frontline in the battle between *Mg* and rice. For example, genes related in brassinosteroid biosynthesis and signaling are generally activated in gall tissue that might antagonize the JA pathway, leading to increased susceptibility to *Mg* (Nahar et al., 2011). In an incompatible interaction, both JA- and SA-mediated host defenses are activated two days postinfection, and are observed in later stages of infection (Kumari et al., 2017; Kumari et al., 2016)

5. Resistant cultivar against *M. graminicola*

Certain varieties from the African rice *O. longistaminata* (WLO2-2 and WLO2-15) and *O. glaberrima* (TOG7235, TOG5674 and TOG5675) species are naturally resistant to *M. graminicola* (Soriano et al., 1999). Interestingly, several cultivars from *O. glaberrima* (TOG5674, TOG5675, CG14, RAM131) are highly resistant to *Mg*, and studies on the kinetics of infection using histological analysis of infected CG14 cultivar suggest that HR-like reactions may occur (Cabasan et al., 2014). Recently, three South-American wild rice *M. glumaepatula* accessions (BGA.012954, BGA.014179, BGA.014210) have been characterized as resistant to *Mg* and present a hypersensitivity response-like reaction phenomenon (Mattos et al., 2019).

Resistance in Asian rice varieties has been actively sought for several decades (Bridge et al., 2005). Recently, four *Oryza sativa* cultivars have been described as resistant to *Mg* including Khao Pahk Maw, LD24 (Dimkpa et al., 2016), Zhonghua 11 (Phan et al., 2018), and two accessions from China (Shenliangyou 1 and Cliangyou 4418) (Zhan et al., 2018). Interestingly, histological analysis showed a HR-like reaction at the penetration site and during the feeding sites formation (Phan et al., 2018). A similar response was observed in *O. glaberrima* (Cabasan et al., 2014) and *O. glumaepatula* accessions (Mattos et al., 2019). Further analysis must be performed to confirm the nature of the resistance observed (HR?) and the genetic determinism of the resistance.

6. Controls

The use of nematicides has been the most efficient way to control nematodes in the field for many years but at a substantial cost to both farmers and the environment. In order to preserve

the environment and to offer sustainable solutions to farmers, several alternative strategies have been developed to control *Mg*. Unfortunately, none of them is sufficient on its own to control this parasite and a combination of different approaches seems necessary. The ability of *Mg* eggs to remain dormant in the soil for several months (Bridge & Page, 1982), as well as their ability to survive in reducing conditions and to reproduce within the rice root in the event of flooding, gives this species exceptional resilience and complicates the possibilities of a control strategy. However, the constant immersion of rice in irrigated fields significantly reduces parasite pressure, but this poses a problem in water management practices (Belder et al., 2004). Rotational cropping systems between rice and non-host plants such as mustard, sesame, cowpea, millet, mung bean (Rahman, 1990; Ventura et al., 1981) can reduce the presence of *Mg* in the rice field, thus mitigating yield losses in rice. This method even increases the rice yield up to 85% in upland rice when grown under crop rotation with cowpea (Soriano & Reversat, 2003). However, because many weeds grow in association with rice and are also host for *Mg*, this method might not be fully efficient in the field if not associated to weed control. In addition, due to the farmers' crop production habits and in the absence of an economically structured market for these plants used in crop rotation (i.e. seed supply, market to collect and sell the production) this method can be difficult to transfer to farmers.

Some chemical and biological compounds such as β-aminobutyric acid (BABA) or benzo-(1,2,3)-thiadiazole-7-carbothioic acid S-methylester (BTH) are used to trigger the plant's defense machinery, resulting in induced resistance to *Mg* (Ton & Mauch-Mani, 2004). However, they have negative effects on plant growth, due to the trade-off between growth and defense. The thiamine treatment has also been tested and it has been shown that its use can reduce the number of nematodes and delay the development of *Mg* in the roots without having negative effects on plant growth (Huang et al., 2015a). The addition of organic compounds such as biochar to the soil also improves plant tolerance to *Mg* infection (Huang et al., 2015b). More details on the use and effects of these compounds were reviewed in Mantelin et al., (2017).

The biological control of *Mg* by endophytic or rhizospheric microorganisms has already been explored and several bacteria and fungi have thus revealed their potential. For instance, under controlled conditions, treatment with Fusarium isolates or Trichoderma species on rice roots reduced the number of galls on rice by 29%–42% and 38%, respectively (Le et al., 2009).

Developing genetic modified (GM) rice plants by transferring resistance genes from other species, overexpressing toxic genes to nematodes (such as OsTH17), or knockdown of nematode effectors by host-mediated gene silencing strategy may be considered to protect crops

against *Mg*. Finally, searching for the natural resistance genes from the rice germplasms promise a sustainable method to deal with *Mg* pathogen.

III. *Meloidogyne graminicola* - Genomic and transcriptomic resources

1. Genomic features and characteristics

Recently, the whole mitochondrial genomes of two *Mg* populations isolated from the Philippines and China were published (Besnard et al., 2014; Sun et al., 2014), providing molecular resources for population genetic studies and life history reconstruction. A first draft of the *M. graminicola* nuclear genome was released, with a genome assembly size of 35 Mb (Somvanshi, et al., 2018). This genome assembly consists of more than 4,300 contigs and an N50 length of 20 kb. A total of 10,196 genes were annotated which covered 84.27% and 73.60% of CEGMA and BUSCO eukaryotic genes. GC content of *Mg* genome is very low compared to other RKNs, with only 23.5% (Table 3). The size of the *Mg* genome had been estimated at 56% ± 7.6% of that of the *M. incognita* genome by Feulgen densitometry (Lapp & Triantaphyllou, 1972). Genome size of *M. incognita* was recently estimated by both genome assembly and flow cytometry at ~180 Mb (Koutsovoulos et al., 2019), thus, *Mg* genome size should range from 84 – 129 Mb according to Feulgen densitometry measurement. Analysis on ITS sequences retrieved from several *Mg* isolates suggests the existence of variant copies of ITS sequences within a single individual (Bellafiore et al., 2015; Pokharel et al., 2010). These sequence variants could come from divergent duplications of the genome and/or show a degree of heterozygocity of the nuclear genome linked to interspecific hybridization at the origin of the species.

2. Putative effectors and parasitism genes

Recently, three studies have explored the *Mg* transcriptome across several life stages of the nematode, from pre-parasitic J2s to established females in rice roots during compatible and incompatible interactions (Haegeman et al., 2013; Petitot et al., 2015, 2020). Initially, the authors of one study identified 52,000 sequences in this transcriptome (Petitot et al., 2015) before refining the number of sequences to 44,137 (Petitot et al., 2020). These transcriptomic studies explored the putative *Mg* secretome and proposed a set of candidate genes involved in parasitism (Table 1). A set of 1,897 sequences have been identified as coding for secreted proteins (Petitot et al., 2015) while Haegeman et al., (2013) detected 499 putative effectors confirming that the *Mg* secretome consists of a complex cocktail of proteins as revealed in other *Meloidogyne* (Bellafiore et al., 2008). The expression in the secretory glands of several of the

putative effectors identified from *Mg* transcriptomes was confirmed by in situ hybridization (Haegeman et al., 2013; Petitot et al., 2015). Most of the predicted effectors were expressed in the nematode dorsal or subventral gland cells, or in the amphids, indicating that the corresponding proteins would certainly be secreted into the plant tissue, thus supporting a role in parasitism.

Table 1. List of major effectors of *M. graminicola* known to date and their potential functions in parasitism

Parasitic process	Functions	Proteins in *M. graminicola*
Plant cell wall modifying enzymes	Pectin degradation	pectate lyase poly-α-d-galacturonosidase
	Cellulose and hemicellulose degradation	β-1,4-endoglucanases xylanase Glycoside Hydrolase family 5 (GH5), GH30
	Non-covalent bonds disruption	expansin-like proteins
Detoxification	ROS detoxification	catalases superoxide dismutases glutathione peroxidases glutathione-S-transferases thioredoxins and peroxiredoxins metallothionein
Plants response/ resistance manipulation	Suppress the activation of defenses	venom allergen-like protein (VAP)
	Manipulate plant basal immunity	MgGPP, MgMO237, MgPDI
	Virulence gene against resistance	*map-1* gene homologous
	Neutralize host defense mechanism	C-type lectins
	Suppress JA pathway	retinol-binding proteins (FAR)
	Mimic function of plant proteins or signalling peptide	14-3-3 proteins annexin-like proteins
Nematode feeding site	Modulate nematode behavior	neuropeptides putative *nlp* or *flp*
	Feeding site establisment	CLE-like peptide

Interestingly, many putative effectors are plant cell wall-modifying enzymes that soften and degrade the plant cell wall, thus facilitating the invasion of root tissues by nematodes. These enzymes potentially participate in the degradation of pectin (pectate lyase, poly-α-d-galacturonosidase), the degradation of cellulose and hemicellulose (β-1,4-endoglucanases, xylanase, Glycoside Hydrolase family 5 (GH5), GH30) and non-covalent bonds disruption (expansin-like proteins). Notably, antioxidant proteins that may be involved in the

detoxification of host-derived ROS have been identified in the *Mg* transciptome. These include catalases, superoxide dismutases, glutathione peroxidases, glutathione-S-transferases, thioredoxins and peroxiredoxins. In addition, metallothioneins, which have a protective role against ROS, appear to be strongly expressed in the early stages of *Mg* infection. Two venom allergen-like protein (VAP)-encoding genes, which may suppress the activation of defenses triggered by DAMPs, were specifically induced in resistant plants and Mg-VAP1 silencing in J2s reduced their ability to colonize roots (Petitot et al., 2020). A family of 10 genes, putatively coding for the neuropeptides *nlp* or *flp* (Petitot et al., 2020), function in modulating nematode behavior, including locomotion and chemo-attraction (Li & Kim, 2008). Homologues of map-1 gene, a small gene family specific to RKNs and only found in avirulent lines of the three *Meloidogyne* species that are controlled by the tomato resistance gene Mi-1 (Castagnone-Sereno et al., 2009; Semblat et al., 2001), have been identified in *Mg* (Haegeman et al., 2013). Three C-type lectin sequences, a class of proteins known to be involved in a multitude of defense processes, have been identified in the transcriptome of *M. graminicola* (Haegeman et al., 2013). One of them (Mg01965) has been shown to function in suppressing the plant's basal immunity and thus promote parasitism of *M. graminicola* (Zhuo et al., 2019). Three other effectors, MgGPP, MgMO237, and MgPDI, were found to interact with multiple host defense-related proteins to manipulate plant basal immunity and promote parasitism (Chen et al., 2017; 2018; Tian et al. 2019). Effectors could also promote parasitism by acting on the signaling pathway of phytohormones (Gheysen and Mitchum, 2019). For example, several transcripts encoding fatty acid and retinol binding proteins (FARs) that can bind to JA precursors have been identified in the *Mg* transcriptome (Haegeman et al., 2013; Petitot et al., 2015). Many putatively secreted 14-3-3 proteins and annexin-like proteins have also been identified in *Mg* transcriptome (Haegeman et al., 2013; Patel et al., 2010). Orthologs of these proteins have also been found in plants, suggesting a plant protein mimicking function in signalling to block, modulate or divert host cellular processes. CLE-like peptides play an important role in the maintenance of plant shoots, roots and vascular meristems. CLE-type effectors have been identified in several CN and RKN genera and could mimic the functions of CLE peptides in plants. Silencing the CLE genes of nematodes or their related plant receptors delays the development of nematodes by inhibiting the formation of feeding site (Gheysen and Mitchum., 2019). It is interesting to note that CLE orthologues were also found in the *Mg* transcriptome (Petitot et al., 2015)

IV. Genome, diversity and molecular evolution from genus view

1. Reproduction mode, cytogenetics, and ploidy

The Root-Knot Nematodes (RKNs) also known as *Meloidogyne* spp. include nearly hundred described species (Álvarez-Ortega et al., 2019; Hunt & Handoo, 2009). RKNs show an intriguing diversity of reproductive modes ranging from obligatory sexual (amphimixis) to fully mitotic asexual reproduction (apomixis) with intermediates capable of alternating between sexual and asexual (automixis or apomixis) reproduction modes (Castagnone-Sereno et al. 2013). Phylogenetic analysis based on nuclear ribosomal DNA could classify *Meloidogyne* spp. between at least 11 clades (Álvarez-Ortega et al., 2019), however, only three clades (I, II, II) were considered as main clades and were validated using other markers such as mitochondrial genes and nucleic sequences (such as ITS, 18S, and D2-D3 sequences). Clade I contains species that reproduce by obligatory mitotic parthenogenesis (apomixis) (eg. *M. incognita*, *M. javanica*, *M. arenaria*, and *M. enterolobii*), and also meiotic parthenogenesis (automixis), (*M. floridensis)*. Clade II consists of amphimictic and/or meiotic parthenogenesis species (eg. *M. hapla*). Clade III includes meiotic parthenogenetic species (e.g., *M. exigua*, *M. graminicola*, *M. chitwoodi*) as well as the apomictic *M. oryzae* (Table 2). Gametogenesis of mitotic parthenogenesis undergoes a single mitotic division without reduction in the chromosome number (Figure 4). Meanwhile, meiotic parthenogenesis in *M. hapla* involves two mature divisions: **1)** During the first division, synaptonemal complexes and recombination nodules occur during the pachytene stage of prophase I (Goldstein & Triantaphyllou, 1978). Bivalents are observed at metaphase I, and homologs seem to separate at the first meiotic division as in a typical meiosis (Figure 4) (Triantaphyllou, 1966). **2)** During the second division, if sperm are present, oocyte maturation occurs to form a pronucleus and two polar bodies. The sperm nucleus fuses with the pronucleus of the haploid egg to form a sexual product. Without fertilization, meiosis is prolonged and the diploid state was reported to be restored by either reunion of the sister chromosomes of a single meiosis or by the duplication of the chromosomes in the nucleus of the egg after the first division (Figure 4) (Triantaphyllou 1966). Meiotic parthenogenesis in species such as *M. graminicola*, *M. chitwoodi*, and *M. fallax* appears to follow a similar pattern to that of *M. hapla*.

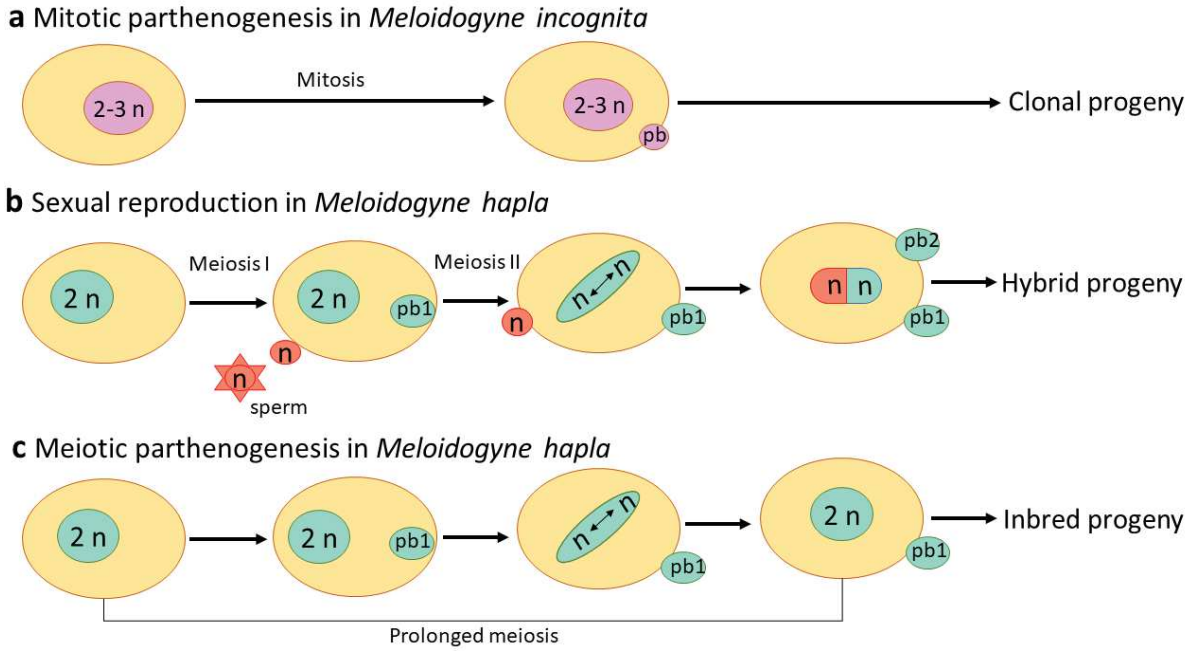


Figure 4. Common reproductive modes of root-knot nematodes (Bird et al., 2009)

Table 2. The three main clades of the *Meloidogyne* species are based on the 18S-rDNA, the chromosome number and the reproduction mode (modified from Castagnone-Sereno et al. 2013).

Clade	Main representative	Chromosome number	Reproduction mode
I	*M. arenaria*	51-56	Obligate mitotic parthenogenesis [1]Meiotic parthenogenesis
	M. enterolobii	46	
	M. incognita	41-46	
	M. javanica	42-48	
	M. floridensis[1]	36	
II	*M. hapla*	32-36	Amphimixis and/or meiotic parthenogenesis
	M. microtyla	-	
	M. spartinae	-	
III	*M. chitwoodi*	36	Meiotic parthenogenesis [2]Facultative meiotic parthenogenesis [3]Obligate mitotic parthenogenesis
	M. fallax	-	
	M. minor	-	
	M. graminicola[2]	36	
	M. naasi	36	
	M. oryzae[3]	51-55	
	M. exigua	36	

The number of chromosomes is varied among *Meloidogyne* spp. suggesting different levels of ploidy. A series of works by Triantaphyllou revealed the cytogenetics of *Meloidogyne* spp. His work determined that the typical number of chromosomes in an RKN haploid cell is

n=16-18 except for the two species *M. spartinae* and *M. kikuyensis* where n=9 (Triantaphyllou, 1987, 1990). The total number of chromosomes in the somatic cells of most RKN varies from 30 to 50, suggesting that they are diploid or triploid. Mitotic RKNs have a higher number of chromosomes than those reproduced by amphimixis or automixis. *M. incognita*, *M. javanica* and *M. enterolobii* likely have similar chromosome number of 41-48, while *M. areanaria* and *M. oryzae* have 51-56 chromosomes (Triantaphyllou, 1981, 1985). Current genomic studies suggest that *M. incognita* is most likely triploid, *M. javanica* tetraploid and *M. arenaria* tetra- or pentaploid (Blanc-Mathieu et al., 2017). In clade II, the facultative meiotic parthenogenetic *M. hapla* has often been considered with a chromosome number 2n=32-36, however, its tetraploid form has also been reported with 68 chromosomes (Triantaphyllou, 1984). The number of chromosomes observed within the same species may vary and may not always be a multiple of 18, suggesting that polyploidization phenomena are accompanied by chromosomal rearrangements (e.g. partial aneuploidization). RKN have holocentric chromosomes like other nematodes (e.g. *C. elegans*), which can lead to aneuploidy/polysomy accumulation, structural rearrangements, chromosome fusions, deletions, duplications, and translocations (Triantaphyllou, 1983). These events result in variations in the number of chromosomes within cells. Cytogenetic diversity as high as that of RKN has not been frequently observed in other groups of animals, making the RKN model a tool of choice to study these chromosomal rearrangements and their impact on the evolution of species.

2. General features of RKN genome

To date, the genomes of eight RKN species have been assembled and most of the sequenced genomes belong to clade I species (*M. incognita*, *M. arenaria*, *M. enterolobii*, *M. javanica*, *M. luci*, and *M. floridensis*) (Table 3). The *M. hapla* and *M. graminicola* genome sequences are the only ones published to date for clade II and III, respectively. Due to the polyploid character, the number of repeated sequences and the absence of genetic maps, it has been very difficult to reconstruct the complete genome of RKNs, especially those reproducing by apomixis such as *M. incognita*, *M. arenaria*, *M. enterolobii*, *M. javanica*, and *M. floridensis*. The assembly of different RKN genomes reveals that mitotic parthenogenetics (e.g. *M. incognita, M. arenaria, M. enterolobii, M. javanica* and *M. luci*) have genomes of 180-284 Mb while those that reproduce by amphimixis or automixis (e.g. *M. floridensis, M. hapla* and *M. graminicola*) have smaller genomes in the range of 38-99 Mb. Total DNA content of five species (*M. incognita*, *M. arenaria*, *M. enterolobii*, *M. javanica*, and *M. hapla*) was measured by flow cytometry, revealing that the four mitotic RKNs have approximatly 2-3 times more

DNA than *M. hapla* (~120 Mb). Among the assembly metrics, *M. luci*'s genome assembly contains the longest scaffolding (N50 the longest and the fewest scaffolds) and the highest degree of CEGMA completeness (95%) (Table 3) but the lack of genetic maps limits the perfect reconstruction of genomes in parthenogenetic species. The development of new techniques such as Hi-C (Lieberman-Aiden et al., 2009) allows the mapping of dynamic conformations of whole genomes. Hi-C allows an unbiased identification of chromatin interactions across a whole genome. The GC content of the RKN genomes sequenced to date ranges from 23 to 30%. Interestingly, *M. graminicola* has both the smallest genome assembly size (38Mb) and lowest GC content (23%) (Somvanshi et al., 2018). Consistent with the size of the genome, the number of protein-coding genes predicted in mitotic RKNs is also 3 to 7 times greater than that observed in sexual/meiotic RKNs (Table 3). The increase in the number of genes in mitotic RKN would be due to the complete or partial duplication of the genome within a species and to allopolyploidization events that have led to the genesis of new species with complex genomes and different modes of reproduction (emergence of meiotic and mitotic parthenogenesis) (Bird et al., 2009; Castagnone-Sereno et al., 2013; Lunt et al., 2014).

Table 3. Genome features of RKN

Species/Feature	Assembly size (Mb)	Nuclear DNA content (Mb)	% complete CEGMA (C) (copies)	% complete BUSCO	N50 (kb)	# contigs/ scaffolds	GC %	# Genes	Reference
M. enterolobii (Swiss)	240	275 ± 19	94.76 (3.30)	87.5	143	4,437	30	59,773	(Koutsovoulos et al., 2020)
M. enterolobii (L30)	162.4	NA	81.45 (2.66)	79.9	9.3	46,09	30.2	31,051	(Szitenberg et al., 2017)
M. incognita (Morelos, 2017)	183.5	189 ± 15	94.76 (2.93)	88.5	38.6	12,091	29.8	45,351	(Blanc-Mathieu et al., 2017)
M. incognita (W1)	122	NA	82.66 (2.34)	80.2	16.5	33,735	30.6	24,714	(Szitenberg et al., 2017)
M. incognita (Morelos, 2008)	86	189 ± 15	74.6 (1.77)	71.3	82.8	2,817	31.4	19,212	(Abad et al., 2008)
M. javanica (Avignon)	235.8	297 ± 27	92.74 (3.68)	90.1	10.4	31,341	30	98,578	(Blanc-Mathieu et al., 2017)
M. javanica (VW4)	142.6	NA	89.52 (2.71)	87.5	14.1	34,394	30.2	26,917	(Szitenberg et al., 2017)
M. arenaria (Guadeloupe)	258.1	304 ± 9	94.76 (3.66)	87.1	16.5	26,196	30	103,001	(Blanc-Mathieu et al., 2017)
M. arenaria (A2-O)	284.05	NA	94.76 (3.57)	87.1	204.6	2,224	30	NA	(Sato et al., 2018)
M. arenaria (HarA)	163.8	NA	91.53 (2.74)	78.2	10.5	46,509	30.3	30,308	(Szitenberg et al., 2017)
M. hapla (VW9)	53.6	121 ± 3	93.55 (1.19)	87.4	83.6	1,523	27.4	14,420	(Opperman et al., 2008)
M. floridensis (JB5)	99.9	NA	56.45 (1.95)	54.1	3.5	81,111	29.7	NA	(Lunt et al., 2014)
M. floridensis (SJF1)	74.9	NA	77.42 (1.71)	76.5	13.3	9,134	30.2	15,327	(Szitenberg et al., 2017)
M. luci (SI-Smartno)	209.2	NA	95.56 (2.92)	87.8	1,712	327	30.2	NA	(Susič et al., 2020)
M. graminicola (IARI)	38.18	NA	84.27 (1.34)	73.6	20.4	4,304	23.05	10,196	(Somvanshi, et al., 2018)

3. Heterozygosity and hybrid origin

The genome of mitotic species such as *M. incognita*, *M. javanica* and *M. arenaria*, or meiotic species such as *M. floridensis* contain collinear duplicated regions with an average nucleotide divergence between pairs at 7-8% (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017). Phylogenetic analysis between orthologous copies showed that the duplicated genome regions in mitotic *Meloidogyne* tend to be more similar across different species than they are to their other copies within the same species. Besides, in phylogenomic analysis, duplicate regions form two or more clades that have different topologies. Thus, apparently duplicate regions within a species might not originate from common ancestral allelic regions, but may have different origins and share different evolutionary histories. Mitochondrial genes are almost identical between *M. incognita*, *M. javanica*, *M. areanaria*, and *M. floridensis* (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017) suggesting their maternal ancestors are closely related or common. Divergent duplicated genomic blocks suggest whole genome duplication (WGD) events via hybridization of these RKN clade I species (Blanc-Mathieu et al., 2017). Besides, the low mitochondrial divergence and the low proportion of gene loss in collinear duplicated blocks suggest the WGD events should have recently occurred. As such, it is interesting to note that the *M. hapla* meiotic genome does not contain divergent duplicate regions and that its level of heterozygosity is most certainly low, although as yet undetermined (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017).

The mapping of reads obtained by genome sequencing on the reference haploid genomes of *M. incognita*, *M. javanica*, *M. arenaria* and *M. enterolobii* revealed 3, 4 and 5 copies, respectively. This suggests that their ploidy is probably triploid (*M. incognita*, *M. enterolobii*) and tetra to penta-ploid (*M. javanica*, *M. arenaria*) (Blanc-Mathieu et al., 2017; Koutsovoulos et al., 2019). The current study on the structures of RKN clade I genome showed that the *M. floridensis* genome is triploid with a predominantly AAB structure, where two of the haploid copies of the genome (A) are almost identical and the last copy (B) is the carrier of the observed heterozygosity of 2.6% (Figure 5) (Jaron et al., 2020). *M. enterolobii* and *M. incognita* have higher levels of heterozygosity and carry both AAB and ABC genome structures (three different haploid genome copies) (Figure 5). The three different haplotypes (ABC) were observed in these three species but at a very low proportion. The genomes of *M. javanica*, and *M. arenaria* are highly heterozygous (~8%) with a tetraploid genome structure of AAAB (three identical copies plus one divergent copy), AABB (two divergent copies shown twice), AABC (three divergent copies, one of which is shown twice), and ABCD (four divergent copies) (Figure 5).

The high degree of ploidy, sometimes observed in genomes with two haplotypes (AAB, AAAB and AABB), can be explained by a single stage of genome fusion. For example, fusion of diploid AA with reduced gamete B of sexual progenitor (BB) resulted in AAB progeny. Meanwhile, the formation of genome structures with more than two alleles (eg. ABC, AABC, and ABCD) should be explained by multiple hybridization steps. For instance, a two-step hybridization process has been suggested to form ABC structures: firstly, homoploid hybridization (hybridization between two diploid AA and BB progenitors without change in chromosome number) took place and led to a diploid AB hybrid; secondly, hybridization between an unreduced AB gamete of the homoploid hybrid with a reduced C gamete of another haploid sexual species led to the presence of three distinct copies (ABC) of the nuclear genomes within the same species (Blanc-Mathieu et al., 2017; Jaron et al., 2020).

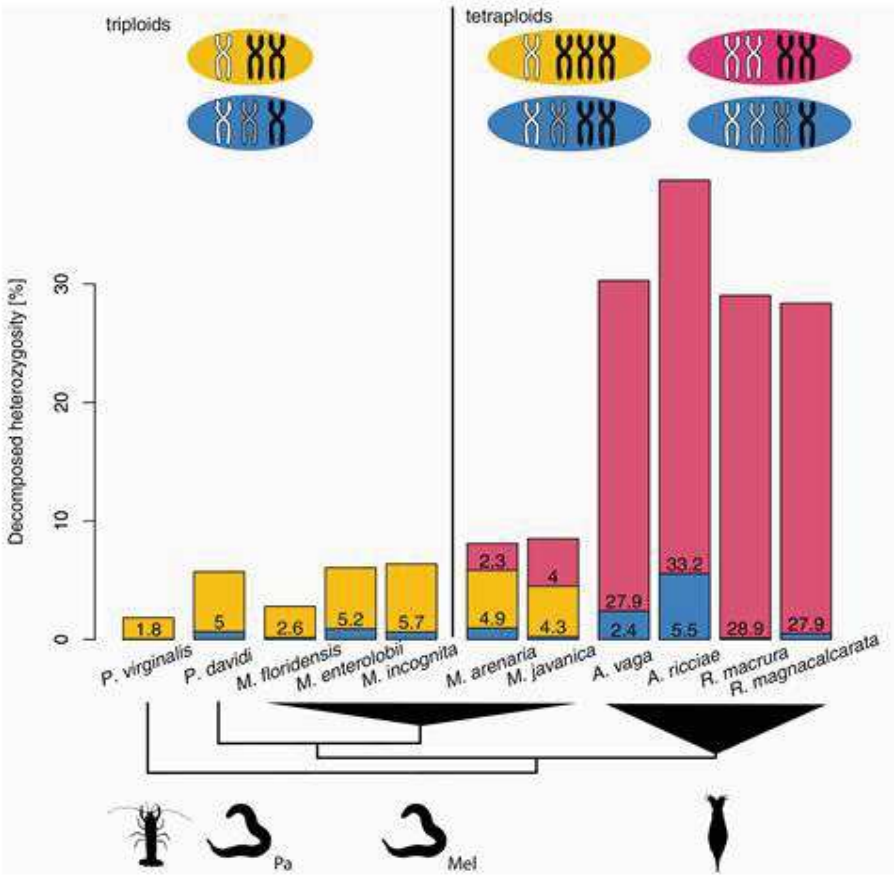


Figure 5. The relative heterozygosity structure in polyploids of parthenogenetic animal. Biallelic loci are indicated in yellow or pink: yellow when the alternative allele is carried by a single haplotype (AAB or AAAB), and pink when both alleles are represented twice (AABB). Loci with more than 2 alleles (ABC, AABC, ABCD) are indicated in blue (Jaron et al., 2020).

4. Molecular evolution of RKN

4.1 Palindromes, gene conversion, and recombination

Recently, four gametes and linkage disequilibrium tests using fixed markers in *M. incognita* populations concluded in the absence of homologous recombination in this mitotic parthenogenetic species (Koutsovoulos et al., 2019). The mitotic RKN genome contains multiple synteny breakpoints and a wide discrepancy between homologous blocks, suggesting that chromosome pairing must be complicated if not impossible.

Palindromes are duplicated regions on a single chromosome in reverse orientation. Palindromes can facilitate gene conversion and therefore help to escape mutational meltdown via Muller' s ratchet. Three and one palindromes, respectively, were observed in the collinear regions of the same scaffolds of *M. arenaria* and *M. incognita* (Blanc-Mathieu et al., 2017; Jaron et al., 2020). Palindromes or blocks of tandem sequences have not been observed in the genome of the meiotic *M. hapla* and *M. floridensis* (Blanc-Mathieu et al., 2017; Jaron et al., 2020). Gene conversions were detected in the mitotic RKN genome without recombination events (Szitenberg et al., 2017). In the amphimictic *M. hapla*, testing molecular markers on F2 generations from hybrid female F1 revealed the existence of recombination events (Liu et al., 2007). Markers flanking recombination events are prevalent in homozygous state, suggesting that recombination preferentially occurs as four-stranded exchanges at similar locations between the both pairs of non-sister chromatids. With these mechanisms, meiotic parthenogenesis in *M. hapla* is expected to result in allele fixation and rapid genomic homozygosity (Liu et al., 2007).

4.2 Transposable elements

Transposable elements (TEs) have played an important role in creating genome diversity through various 'cut-and-paste' and 'copy-and-paste' mechanisms. Sometimes, changing their position creates or reverses mutations, which changes the genotype of the cell and leads to the evolution of the genome of the species (Ayarpadikannan & Kim, 2014; Kazazian, 2004). It has been shown that TE insertions have many effects, such as altering the size and structure of the genome, regulating gene expression, increasing recombination rate, and cross-breeding inequality (Britten, 2010; Tollis & Boissinot, 2012). A hybridization event can lead to a burst in the activity of the transposable element (Arkhipova & Rodriguez, 2013). Indeed, in RKN, TE abundance is more significant in species of hybrid origin (*M. incognita*, *M. javanica*, *M. arenaria*, *M. enterolobii,* and *M. floridensis*) whatever the reproduction mode than in

amphimictic species (e.g. *M. hapla* and *M. chitwoodi*) (Blanc-Mathieu et al., 2017; Szitenberg et al., 2016). If the general meaning of the abundance of TEs as a function of the origin of the species studied persists, due to the different pipelines used in the different publications, the detection of these TEs in the *Meloidogyne* genomes can differ from one study to another. For example, Blanc-Mathieu et al., (2017) reported that TEs covered ~50% of the genomes of *M. incognita*, *M. javanica*, and *M. arenaria* versus only 29% of the genome of *M. hapla*. In parallel, Koutsovoulos et al. (2020) detected that the TE ratios in the genomes of *M. incognita*, *M. javanica, and M. enterolobii* are 7.2%, 6.7%, and 12.1% respectively. In another study, the pipeline of Jaron et al. (2020) showed a ratio of 7.6% (*M. incognita*), 11.5% (*M. javanica*), 11% (*M. arenaria*), 6.5% (*M. enterolobii*) and 7% (*M. floridensis*). Despite the differences in the TE contents that came out from one study to another, a similar conclusion form these studies is that the TE class II (DNA transposons) is more abundant than the class I (retrotransposons) among RKNs. In *M. incognita*, the abundance of DNA transposons and high identity level of TEs copies to their consensus sequence (≥ 94%) suggests they have been at least recently active. Some TEs are inserted in coding regions or in gene regulatory regions suggesting a functional impact (Kozlowski et al., 2020). Interestingly, a transposon has been associated with gene deletion in resistance-breaking pathovar of *M. javanica* suggesting an adaptive impact of TEs on the nematode genome (Gross & Williamson, 2011).

4.3. Horizontal gene transfer

One of the particularly interesting genomic adaptations in RKNs is that they have acquired a number of genes important for host-parasite interactions via Horizontal gene transfer (HGT), mostly from bacteria but also from fungi (Danchin et al., 2010; Danchin et al., 2016; Haegeman et al., 2011). In RKNs, functional genes potentially acquired by HGT have been documented in *M. incognita*, *M. javanica*, *M. floridensis*, and *M. hapla* (Clades I and II) and most of these genes code for proteins involved in plant cell-wall degradation, nematode feeding site establishment, nutrient processing, detoxification, and manipulation of plant defenses (Scholl et al., 2003). The distribution of horizontally acquired genes from different species in clades 12 showed that some HGTs appeared prior the diversification of the PPN groups. In addition, some HGTs were subclade-specific, suggesting that the occurrence of HGT events continued during species emergence (Haegeman et al., 2011). Besides, the presence of these HGT genes as multigenic families could indicate positive selective pressure that would have favored individuals harboring multiple copies of these genes. Because these genes are assumed to

originate from the HGT of bacterial donors (intron less), they then gained introns through duplications that probably started in the common ancestor and resulted in multigenic families. This pattern of duplication and intron gain appears to be common to many genes acquired by HGT in both *M. incognita* and *M. hapla* (Haegeman et al., 2011).

4.4 Copy Number Variation

Copy number variations (CNV) within a population and/or between populations are means of rapid evolution, where repeated gain or loss can lead to a change of function; they can be of particular importance in the evolution of parasite genomes. In the RKN and cyst nematode populations, CNVs have been found to affect several important effector genes, including the gene potentially recognized by the tomato *Mi1.2* resistance gene (MAP-1) (Castagnone-Sereno et al., 2009), the ligand mimics of plant CLE peptides (CLE) (Rutter et al., 2014), an effector that suppresses plant innate immunity (GLAND 18) (Noon et al., 2016), and Hypervariable Apoplastic Effectors (HYPs) (Akker et al., 2014). Interestingly, an outstanding level of gene loss events has been associated with virulent pathovars bypassing the host plant's resistance mechanism (hypersensitive response) in *M. incognita* species (Castagnone-Sereno et al., 2019). Hence, it has been suggested that CNVs may be an adaptive genetic mechanism in response to a challenging new biotic environment.

5. Resistant breaking mechanisms

The RKNs have a huge range of hosts, including most flowering plants (Trudgill & Blok, 2001). Screening of plant germplasms revealed that some domesticated crops and their wild relatives have defense mechanisms that protect them from RKN invasions. Characterization of resistant responses indicated that resistance genes generally act against RKN by inducing a hypersensitive response (HR). This indicates a gene-for-gene relationship in which specific effector proteins produced by the pathogen [avirulence (*Avr*) genes] are recognized by cognate resistance gene (*R*) to trigger a cascade of defense responses. Recently, the durability of key RKN resistance genes has been threatened due to the emergence of virulent pathovars than can overcome the plant resistance genes. Indeed, RKN have developed sophisticated mechanisms for adapting to plant resistance genes in natural populations (Castagnone-Sereno et al., 1994; Janssen et al., 1998; Petrillo & Roberts, 2005).

Tomato plants expressing *Mi1.2* genes are highly resistant to *M. incognita*, *M. arenaria*, and *M. javanica* (Milligan et al., 1998). This gene has been considered a stable resistance resource for over 50 years. However, virulent populations of *M. incognita* have been identified in a

tomato field (Eddaoudi et al., 1997; Kaloshian et al., 1996). In the laboratory, experiments with artificial selection of virulent *M. incognita* pathovars on tomatoes expressing the *Mi1.2* gene proved that virulence can be acquired rapidly in only five to ten generations (Bost & Triantaphyllou, 1982; Jarquin-Barberena et al., 1991). Isofemale lines of a natural avirulent isolate *M. incognita* obtained after two generations in *Mi* resistant tomato plants showed significant variants that were heritable in the following generations (Castagnone-Sereno et al., 1994). The use of 1,550 AFLP markers on avirulent and virulent isolates of *M. incognita*, *M. javanica*, and *M. arenaria* failed to group their avirulence/virulence phenotypes according to their genotype, indicating that the virulent populations do not share a common origin and are likely the result of independent mutational events (Semblat et al., 2000). Therefore, it has been suggested that the acquisition of virulence could be modulated by several genes. The use of AFLP markers in *M. javanica* made it possible to identify the *Cg-1* gene as a potential effector present only in avirulent pathovars (Gleason et al., 2008). It has been shown that silencing the *Cg-1* gene in avirulent *M. javanica* pathovar results in a strain virulent against tomato plant expressing *Mi1.2*, suggesting that *Cg-1* is a key factor in inducing the incompatible interaction mediated by *Mi1.2*. *Map*-1, which is a gene present only in *M. incognita* avirulent lines, has also been suggested as one of the factors in putative avirulence (Semblat et al., 2001) but its direct involvement could not be demonstrated to date . The loss of a DNA fragment comprising the *map-1* gene is associated with the virulence lines of *M. incognita*. Interestingly, different forms of this gene exhibited variations in the number of internal repeat motifs between avirulent and virulent isogenic nematode lines. Intragenic loss of repeat units in the *map-1* gene family may be due to gene duplication and could be the key mechanism for the adaptability of asexual RKN (*M. arenaria*, *M. incognita*, and *M. javanica*) (Semblat et al., 2001).

Recently, a genome wide comparative hybridization between virulent and avirulent *M. incognita* genotypes indicated that gene duplication and loss is associated with the nematode's response to the host resistance selection pressure. It is interesting to note that the abundance of gene loss events appears to be associated with virulent genotypes (Castagnone-Sereno et al., 2019).

The mechanisms of acquisition of genetic changes in RKNs to break the plant resistance mechanism have been studied. The RKN genome has been subjected to important cytological modifications involving, among others, aneuploidy and chromosome rearrangements (Triantaphyllou, 1985). In addition, breakage without reassociation and the loss of certain chromosome fragments are not exceptional in these species. In sexual RKN, genetic variants

could accumulate during genetic segregation by mating and meiotic recombination. In mitotic species, changes in the ploidy of a specific chromosome could lead to phenotypic changes, especially if the individual is heterozygous for the alleles involved. Gene conversion, movement of transposable elements or other mechanisms specialized in genetic changes could also play a role in generating genetic variability.

References

Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M.-C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., … Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, *26*, 909. https://doi.org/10.1038/nbt.1482

Akker, S. E. den, Lilley, C. J., Jones, J. T., & Urwin, P. E. (2014). Identification and Characterisation of a Hyper-Variable Apoplastic Effector Gene Family of the Potato Cyst Nematodes. *PLOS Pathogens*, *10*(9), e1004391. https://doi.org/10.1371/journal.ppat.1004391

Álvarez-Ortega, S., Brito, J. A., & Subbotin, S. A. (2019). Multigene phylogeny of root-knot nematodes and molecular characterization of *Meloidogyne nataliei* Golden, Rose & Bird, 1981 (Nematoda: Tylenchida). *Scientific Reports*, *9*(1), 11788. https://doi.org/10.1038/s41598-019-48195-0

Arkhipova, I. R., & Rodriguez, F. (2013). Genetic and Epigenetic Changes Involving (Retro)transposons in Animal Hybrids and Polyploids. *Cytogenetic and Genome Research*, *140*(2–4), 295–311. https://doi.org/10.1159/000352069

AVA. (2001). *Diagnostic records of the Plant Health Diagnostic Services*. Plant Health Centre,.

Ayarpadikannan, S., & Kim, H.-S. (2014). The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases. *Genomics & Informatics*, *12*(3), 98–104. https://doi.org/10.5808/GI.2014.12.3.98

Bastidas, H., & Montealegre, S. F. A. (1994). General aspects of the new rice disease known as entorchamiento. *Arroz*, *392*(43), 30–35.

Belder, P., Bouman, B. A. M., Cabangon, R., Guoan, L., Quilang, E. J. P., Yuanhua, L., Spiertz, J. H. J., & Tuong, T. P. (2004). Effect of water-saving irrigation on rice yield and water use in typical lowland conditions in Asia. *Agricultural Water Management*, *65*(3), 193–210. https://doi.org/10.1016/j.agwat.2003.09.002

Bellafiore, S., Shen, Z., Rosso, M.-N., Abad, P., Shih, P., Briggs, S.P., 2008. Direct Identification of the *Meloidogyne incognita* Secretome Reveals Proteins with Host Cell Reprogramming Potential. PLOS Pathog. 4, e1000192. https://doi.org/10.1371/journal.ppat.1000192

Bellafiore, S., Jougla, C., Chapuis, É., Besnard, G., Suong, M., Vu, P. N., De Waele, D., Gantet, P., & Thi, X. N. (2015). Intraspecific variability of the facultative meiotic parthenogenetic root-knot nematode (*Meloidogyne graminicola*) from rice fields in Vietnam. *Comptes Rendus Biologies*, *338*(7), 471–483. https://doi.org/10.1016/j.crvi.2015.04.002

Bellé, C., Balardin, R. R., Nora, D. D., Schmitt, J., Gabriel, M., Ramos, R. F., & Antoniolli, Z. I. (2018). First Report of *Meloidogyne graminicola* (Nematoda: Meloidogynidae) on Barley (*Hordeum vulgare*) in Brazil. *Plant Disease*, *103*(5), 1045–1045. https://doi.org/10.1094/PDIS-11-18-2010-PDN

Besnard, G., Jühling, F., Chapuis, É., Zedane, L., Lhuillier, É., Mateille, T., & Bellafiore, S. (2014). Fast assembly of the mitochondrial genome of a plant parasitic nematode *Meloidogyne graminicola* using next generation sequencing. *Comptes Rendus Biologies*, *337*(5), 295–301. https://doi.org/10.1016/j.crvi.2014.03.003

Besnard, G., Thi-Phan, N., Ho-Bich, H., Dereeper, A., Trang Nguyen, H., Quénéhervé, P., Aribi, J., & Bellafiore, S. (2019). On the close relatedness of two rice-parasitic root-knot nematode species and the recent expansion of *Meloidogyne graminicola* in Southeast Asia. *Genes*, *10*(2), 175. https://doi.org/10.3390/genes10020175

Bird, D. M., Williamson, V. M., Abad, P., McCarter, J., Danchin, E. G. J., Castagnone-Sereno, P., & Opperman, C. H. (2009). The genomes of root-knot nematodes. *Annual Review of Phytopathology*, *47*(1), 333–351. https://doi.org/10.1146/annurev-phyto-080508-081839

Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J.-M., Rocha, M. D., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J.-F., Kozlowski, D. K., Cazareth, J., Couloux, A., Silva, C. D., Guy, J., Kim-Jo, Y.-J., Rancurel, C., Schiex, T., Abad, P., … Danchin, E. G. J. (2017). Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genetics*, *13*(6), e1006777. https://doi.org/10.1371/journal.pgen.1006777

Bost, S. C., & Triantaphyllou, A. C. (1982). Genetic Basis of the Epidemiologic effects of Resistance to *Meloidogyne incognita* in the Tomato Cultivar Small Fry. *Journal of Nematology*, *14*(4), 540–544.

Bridge, J., & Page, S. L. J. (1982). The rice root-knot nematode, *Meloidogyne graminicola*, on deep water rice (*Oryza sativa* subsp. Indica). *Revue de Nematologie*, *5*, 225–232.

Bridge, J., Plowright, R. A., & Peng, D. (2005). Nematode parasites of rice. In *Plant Parasitic Nematodes in Subtropical and Tropical Agriculture* (Luc M., Sikora R.A. and Bridge J., pp. 87–130). CABI Bioscience.

Britten, R. J. (2010). Transposable element insertions have strongly affected human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(46), 19945–19948. https://doi.org/10.1073/pnas.1014330107

Cabasan, M. T. N., Kumar, A., Bellafiore, S., & Waele, D. D. (2014). Histopathology of the rice root-knot nematode, *Meloidogyne graminicola*, on *Oryza sativa* and *O. glaberrima*. *Nematology*, *16*(1), 73–81. https://doi.org/10.1163/15685411-00002746

Cabasan, M. T. N., Kumar, A., Bellafiore, S., & Waele, D. D. (2018). Reproductive, pathogenic and genotypic characterisation of five *Meloidogyne graminicola* populations from the Philippines on susceptible and resistant rice varieties. *Nematology*, *20*(4), 299–318. https://doi.org/10.1163/15685411-00003142

CABI Bioscience. (2001). *SQWORM Database*.

Caillaud, M.-C., Dubreuil, G., Quentin, M., Perfus-Barbeoch, L., Lecomte, P., de Almeida Engler, J., Abad, P., Rosso, M.-N., & Favery, B. (2008). Root-knot nematodes manipulate plant cell functions during a compatible interaction. *Journal of Plant Physiology*, *165*(1), 104–113. https://doi.org/10.1016/j.jplph.2007.05.007

Castagnone-Sereno, P., Mulet, K., Danchin, E. G. J., Koutsovoulos, G. D., Karaulic, M., Rocha, M. D., Bailly-Bechet, M., Pratx, L., Perfus-Barbeoch, L., & Abad, P. (2019). Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita* . *Molecular Ecology*, *28*(10), 2559–2572. https://doi.org/10.1111/mec.15095

Castagnone-Sereno, P., Wajnberg, E., Bongiovanni, M., Leroy, F., & Dalmasso, A. (1994). Genetic variation in *Meloidogyne incognita* virulence against the tomato *Mi* resistance gene: Evidence from isofemale line selection studies. *Theoretical and Applied Genetics*, *88*(6), 749–753. https://doi.org/10.1007/BF01253980

Castagnone-Sereno, Philippe, Danchin, E. G. J., Perfus-Barbeoch, L., & Abad, P. (2013). Diversity and evolution of root-knot nematodes, genus *Meloidogyne*: New insights from the genomic era. *Annual Review of Phytopathology*, *51*, 203–220. https://doi.org/10.1146/annurev-phyto-082712-102300

Castagnone-Sereno, Philippe, Semblat, J.-P., & Castagnone, C. (2009). Modular architecture and evolution of the map-1 gene family in the root-knot nematode *Meloidogyne incognita*. *Molecular Genetics and Genomics: MGG*, *282*(5), 547–554. https://doi.org/10.1007/s00438-009-0487-x

Chapuis, E., Besnard, G., Andrianasetra, S., Rakotomalala, M., Nguyen, H. T., & Bellafiore, S. (2016). First report of the root-knot nematode *Meloidogyne graminicola* in Madagascar rice fields. *Australasian Plant Disease Notes*, *11*(1), 32. https://doi.org/10.1007/s13314-016-0222-5

Chen, J., Hu, L., Sun, L., Lin, B., Huang, K., Zhuo, K., & Liao, J. (2018). A novel Meloidogyne graminicola effector, MgMO237, interacts with multiple host defence-related proteins to

manipulate plant basal immunity and promote parasitism. *Molecular Plant Pathology*, *19*(8), 1942–1955. https://doi.org/10.1111/mpp.12671

Chen, J., Lin, B., Huang, Q., Hu, L., Zhuo, K., & Liao, J. (2017). A novel *Meloidogyne graminicola* effector, MgGPP, is secreted into host cells and undergoes glycosylation in concert with proteolysis to suppress plant defenses and promote parasitism. *PLoS Pathogens*, *13*(4), e1006301. https://doi.org/10.1371/journal.ppat.1006301

Cuc, N. T. T., & Prot, J. C. (1992). Root-parasitic nematodes of deep-water rice in the Mekong Delta of Vietnam. *Fundamental and Applied Nematology*, *15*(6), 575–577.

da Silva Mattos, V., Mulet, K., Cares, J. E., Gomes, C. B., Fernandez, D., de Sá, M. F. G., Carneiro, R. M. D. G., & Castagnone-Sereno, P. (2019). Development of Diagnostic SCAR Markers for *Meloidogyne graminicola*, *M. oryzae*, and *M. salasi*associated with irrigated rice fields in Americas. *Plant Disease*, *103*(1), 83–88. https://doi.org/10.1094/PDIS-12-17-2015-RE

Danchin, E. G. J., Rosso, M.-N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., & Abad, P. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17651–17656. https://doi.org/10.1073/pnas.1008486107

Danchin, G. J. E., Guzeeva, A. E., Mantelin, S., Berepiki, A., & Jones, T. J. (2016). Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. *Molecular Biology and Evolution*, *33*(6), 1571–1579. https://doi.org/10.1093/molbev/msw041

Dimkpa, S. O. N., Lahari, Z., Shrestha, R., Douglas, A., Gheysen, G., & Price, A. H. (2016). A genome-wide association study of a global rice panel reveals resistance in *Oryza sativa* to root-knot nematodes. *Journal of Experimental Botany*, *67*(4), 1191–1200. https://doi.org/10.1093/jxb/erv470

Eddaoudi, M., Ammati, M., & Rammah, A. (1997). Identification of the resistance breaking populations of *Meloidogyne* on tomatoes in Morocco and their effect on new sources of resistance. *Fundamental and Applied Nematology*, *20*(3), 285–289.

Fanelli, E., Cotroneo, A., Carisio, L., Troccoli, A., Grosso, S., Boero, C., Capriglia, F., & De Luca, F. (2017). Detection and molecular characterization of the rice root-knot nematode *Meloidogyne graminicola* in Italy. *European Journal of Plant Pathology*, *149*(2), 467–476. https://doi.org/10.1007/s10658-017-1196-7

Gheysen, G., Mitchum, M.G., 2019. Phytoparasitic Nematode Control of Plant Hormone Pathways. Plant Physiol. 179, 1212–1226. https://doi.org/10.1104/pp.18.01067

Gilces, C. T., Santillan, D. N., & Velasco, L. (2016). Plant-parasitic nematodes associated with rice in Ecuador. *Nematropica*, *46*(1), 45–53.

Gleason, C. A., Liu, Q. L., & Williamson, V. M. (2008). Silencing a candidate nematode effector gene corresponding to the tomato resistance gene *Mi-1* leads to acquisition of virulence. *Molecular Plant-Microbe Interactions: MPMI*, *21*(5), 576–585. https://doi.org/10.1094/MPMI-21-5-0576

Golden, A. M., & Birchfield, W. (1965). *Meloidogyne graminicola* (Heteroderidae) a new species of root-knot nematode from grass. *Proceedings of the Helminthological Society of Washington*, *32*(2), 228–231.

Golden, A. M., & Birchfield, W. (1968). Rice root-knot nematode (*Meloidogyne graminicola*) as a new pest of rice. *Plant Disease Reporter*, *52*(6), 243.

Goldstein, P., & Triantaphyllou, A. C. (1978). Occurrence of synaptonemal complexes and recombination nodules in a meiotic race of *Meloidogyne hapla* and their absence in a mitotic race. *Chromosoma*, *68*(1), 91–100. https://doi.org/10.1007/BF00330375

Gross, S. M., & Williamson, V. M. (2011). Tm1: A *mutator*/*foldback* transposable element family in root-knot nematodes. *PLoS ONE*, *6*(9), e24534. https://doi.org/10.1371/journal.pone.0024534

Haegeman, A., Bauters, L., Kyndt, T., Rahman, M. M., & Gheysen, G. (2013). Identification of candidate effector genes in the transcriptome of the rice root knot nematode *Meloidogyne graminicola*. *Molecular Plant Pathology*, *14*(4), 379–390. https://doi.org/10.1111/mpp.12014

Haegeman, A., Jones, J. T., & Danchin, E. G. J. (2011). Horizontal gene transfer in nematodes: A catalyst for plant parasitism? *Molecular Plant-Microbe Interactions*, *24*(8), 879–887. https://doi.org/10.1094/MPMI-03-11-0055

Haegeman, A., Mantelin, S., Jones, J. T., & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene*, *492*(1), 19–31. https://doi.org/10.1016/j.gene.2011.10.040

Huang, G., Dong, R., Allen, R., Davis, E. L., Baum, T. J., & Hussey, R. S. (2005a). Developmental expression and molecular analysis of two *Meloidogyne incognita* pectate lyase genes. *International Journal for Parasitology*, *35*(6), 685–692. https://doi.org/10.1016/j.ijpara.2005.01.006

Huang, Wen-kun, Ji, H., Gheysen, G., Debode, J., & Kyndt, T. (2015b). Biochar-amended potting medium reduces the susceptibility of rice to root-knot nematode infections. *BMC Plant Biology*, *15*(1), 267. https://doi.org/10.1186/s12870-015-0654-7

Huang, Wen-Kun, Ji, H., Gheysen, G., & Kyndt, T. (2015). Thiamine-induced priming against root-knot nematode infection in rice involves lignification and hydrogen peroxide generation. *Molecular Plant Pathology*, *17*(4), 614–624. https://doi.org/10.1111/mpp.12316

Humphreys-Pereira, D. A., & Elling, A. A. (2015). Mitochondrial genome plasticity among species of the nematode genus *Meloidogyne* (Nematoda: Tylenchina). *Gene*, *560*(2), 173–183. https://doi.org/10.1016/j.gene.2015.01.065

Hunt, D. J., & Handoo, Z. A. (2009). Taxonomy, identification and principal species. In R. N. Perry, M. Moens, & J. L. Starr (Eds.), *Root-knot nematodes* (pp. 55–97). CABI. https://doi.org/10.1079/9781845934927.0055

Jacobs, A. K., Lipka, V., Burton, R. A., Panstruga, R., Strizhov, N., Schulze-Lefert, P., & Fincher, G. B. (2003). An Arabidopsis Callose Synthase, GSL5, Is Required for Wound and Papillary Callose Formation. *The Plant Cell*, *15*(11), 2503–2513. https://doi.org/10.1105/tpc.016097

Janssen, G. J. W., Scholten, O. E., van Norel, A., & Hoogendoorn, C. (J.). (1998). Selection of virulence in *Meloidogyne chitwoodi* to resistance in the wild potato *Solanum fendleri*. *European Journal of Plant Pathology*, *104*(7), 645–651. https://doi.org/10.1023/A:1008625404330

Jaron, K.S., Bast, J., Nowell, R.W., Ranallo-Benavidez, T.R., Robinson-Rechavi, M., Schwander, T., 2020. Genomic Features of Parthenogenetic Animals. *J. Hered*. doi.org/10.1093/jhered/esaa031

Jarquin-Barberena, H., Dalmasso, A., Guiran, G. de, & Cardin, M. C. (1991). Acquired virulence in the plant parasitic nematode *Meloidogyne incognita*: 1. Biological analysis of the phenomenon. *Revue de Nématologie*, *14*(2), 299–303.

Jaubert, S., Laffaire, J.-B., Abad, P., & Rosso, M.-N. (2002). A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. *FEBS Letters*, *522*(1–3), 109–112. https://doi.org/10.1016/s0014-5793(02)02906-x

Jepson, S. B. (1983). Identification of Meloidogyne: A general assessment and a comparison of male morphology using light microscopy, with a key to 24 species. *Revue de Nématologie*, *6*(2), 291–309.

Ji, H., Gheysen, G., Denil, S., Lindsey, K., Topping, J. F., Nahar, K., Haegeman, A., De Vos, W. H., Trooskens, G., Van Criekinge, W., De Meyer, T., & Kyndt, T. (2013). Transcriptional analysis through RNA sequencing of giant cells induced by *Meloidogyne graminicola* in rice roots. *Journal of Experimental Botany*, *64*(12), 3885–3898. https://doi.org/10.1093/jxb/ert219

Ji, H., Kyndt, T., He, W., Vanholme, B., & Gheysen, G. (2015). β-Aminobutyric Acid-Induced Resistance Against Root-Knot Nematodes in Rice Is Based on Increased Basal Defense. *Molecular Plant-Microbe Interactions: MPMI*, *28*(5), 519–533. https://doi.org/10.1094/MPMI-09-14-0260-R

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329. https://doi.org/10.1038/nature05286

Jones, M. G. K. (1981). Host cell responses to endoparasitic nematode attack: Structure and function of giant cells and syncytia*. *Annals of Applied Biology*, *97*(3), 353–372. https://doi.org/10.1111/j.1744-7348.1981.tb05122.x

Kaloshian, I., Williamson, V., Miyao, G., Lawn, D., & Westerdahl, B. (1996). "Resistance-breaking" nematodes identified in California tomatoes. *California Agriculture*, *50*(6), 18–19.

Kazazian, H. H. (2004). Mobile Elements: Drivers of Genome Evolution. *Science*, *303*(5664), 1626–1632. https://doi.org/10.1126/science.1089670

Koutsovoulos, G. D., Poullet, M., Ashry, A. E., Kozlowski, D. K., Sallet, E., Rocha, M. D., Martin-Jimenez, C., Perfus-Barbeoch, L., Frey, J.-E., Ahrens, C., Kiewnick, S., & Danchin, E. G. J.

(2019). The polyploid genome of the mitotic parthenogenetic root-knot nematode *Meloidogyne enterolobii*. *BioRxiv*, 586818. https://doi.org/10.1101/586818

Koutsovoulos, G. D., Poullet, M., Elashry, A., Kozlowski, D. K. L., Sallet, E., Da Rocha, M., Perfus-Barbeoch, L., Martin-Jimenez, C., Frey, J. E., Ahrens, C. H., Kiewnick, S., & Danchin, E. G. J. (2020). Genome assembly and annotation of *Meloidogyne enterolobii* , an emerging parthenogenetic root-knot nematode. *Scientific Data*, *7*(1), 324. https://doi.org/10.1038/s41597-020-00666-0

Kozlowski, D. K., Hassanaly-Goulamhoussen, R., Rocha, M. D., Koutsovoulos, G. D., Bailly-Bechet, M., & Danchin, E. G. (2020). Transposable Elements are an evolutionary force shaping genomic plasticity in the parthenogenetic root-knot nematode *Meloidogyne incognita*. *BioRxiv*, 2020.04.30.069948. https://doi.org/10.1101/2020.04.30.069948

Kumar, S., Sharma, G., Mishra, S., & Kushwaha, K. (2014). First record on the occurrence of *Meloidogyne graminicola* on rice in Udham Singh Nagar district of Uttarakhand. *Journal of Hill Agriculture*, *5*(2), 211. https://doi.org/10.5958/2230-7338.2014.00869.6

Kumari, C., Dutta, T. K., Gahoi, S., & Rao, U. (2017). *An insight into the expression profile of defence-related genes in compatible and incompatible* Oryza sativa-Meloidogyne graminicola *interaction*. https://doi.org/10.5958/0975-6906.2017.00006.2

Kumari, Chanchal, Dutta, T. K., Banakar, P., & Rao, U. (2016). Comparing the defence-related gene expression changes upon root-knot nematode attack in susceptible versus resistant cultivars of rice. *Scientific Reports*, *6*, 22846. https://doi.org/10.1038/srep22846

Kyndt, T., Denil, S., Haegeman, A., Trooskens, G., Bauters, L., Van Criekinge, W., De Meyer, T., & Gheysen, G. (2012). Transcriptional reprogramming by root knot and migratory nematode infection in rice. *The New Phytologist*, *196*(3), 887–900. https://doi.org/10.1111/j.1469-8137.2012.04311.x

Laborte, A. G., Gutierrez, M. A., Balanza, J. G., Saito, K., Zwart, S. J., Boschetti, M., Murty, M. V. R., Villano, L., Aunario, J. K., Reinke, R., Koo, J., Hijmans, R. J., & Nelson, A. (2017). RiceAtlas, a spatial database of global rice calendars and production. *Scientific Data*, *4*(1), 170074. https://doi.org/10.1038/sdata.2017.74

Lapp, N. A., & Triantaphyllou, A. C. (1972). Relative DNA content and chromosomal relationships of some *Meloidogyne*, *Heterodera*, and *Meloidodera* spp. (Nematoda: Heteroderidae). *Journal of Nematology*, *4*(4), 287–291.

Le, H. T. T., Padgham, J. L., & Sikora, R. A. (2009). Biological control of the rice root-knot nematode *Meloidogyne graminicola* on rice, using endophytic and rhizosphere fungi. *International Journal of Pest Management*, *55*(1), 31–36. https://doi.org/10.1080/09670870802450235

Li, C., & Kim, K. (2008). Neuropeptides. In *WormBook*. The C. elegans Research Community. http://www.wormbook.org

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

Long, H. B., Sun, Y. F., Feng, T. Z., Pei, Y. L., & Peng, D. L. (2017). First Report of *Meloidogyne graminicola* on Soybean (*Glycine max*) in China. *Plant Disease*, *101*(8), 1554–1554. https://doi.org/10.1094/PDIS-03-17-0334-PDN

Luna, E., Pastor, V., Robert, J., Flors, V., Mauch-Mani, B., & Ton, J. (2011). Callose deposition: A multifaceted plant defense response. *Molecular Plant-Microbe Interactions: MPMI*, *24*(2), 183–193. https://doi.org/10.1094/MPMI-07-10-0149

Lunt, D. H., Kumar, S., Koutsovoulos, G., & Blaxter, M. L. (2014). The complex hybrid origins of the root knot nematodes revealed through comparative genomics. *PeerJ*, *2*, e356. https://doi.org/10.7717/peerj.356

Manosalva, P., Manohar, M., von Reuss, S. H., Chen, S., Koch, A., Kaplan, F., Choe, A., Micikas, R. J., Wang, X., Kogel, K.-H., Sternberg, P. W., Williamson, V. M., Schroeder, F. C., & Klessig, D. F. (2015). Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. *Nature Communications*, *6*(1), 7795. https://doi.org/10.1038/ncomms8795

Mantelin, S., Bellafiore, S., & Kyndt, T. (2017). *Meloidogyne graminicola*: A major threat to rice agriculture. *Molecular Plant Pathology*, *18*(1), 3–15. https://doi.org/10.1111/mpp.12394

Mattos, V. S., Leite, R. R., Cares, J. E., Gomes, A. C. M. M., Moita, A. W., Lobo, V. L. S., & Carneiro, R. M. D. G. (2019). Oryza glumaepatula, a New Source of Resistance to *Meloidogyne graminicola* and Histological Characterization of Its Defense Mechanisms. *Phytopathology®*, *109*(11), 1941–1948. https://doi.org/10.1094/PHYTO-02-19-0044-R

Mendy, B., Wang'ombe, M. W., Radakovic, Z. S., Holbein, J., Ilyas, M., Chopra, D., Holton, N., Zipfel, C., Grundler, F. M. W., & Siddique, S. (2017). Arabidopsis leucine-rich repeat receptor–like kinase NILR1 is required for induction of innate immunity to parasitic nematodes. *PLOS Pathogens*, *13*(4), e1006284. https://doi.org/10.1371/journal.ppat.1006284

Milligan, S.B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P., Williamson, V.M., 1998. The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10, 1307–1319.

Minton, N. A., Tucker, E., & Golden, A. (1987). *First report of <i>Meloidogyne graminicola<\i>a in Georgia*. https://doi.org/10.1094/PD-71-0376C

Monteiro, A. R., & Barbosa Ferraz, L. C. . C. (1988). First record and preliminary information on the host range of *Meloidogyne graminicola* in Brazil. *Nematologia Brasileira*, *12*, 149–150.

Nahar, K., Kyndt, T., De Vleesschauwer, D., Höfte, M., & Gheysen, G. (2011). The Jasmonate Pathway Is a Key Player in Systemically Induced Defense against Root Knot Nematodes in Rice1[C]. *Plant Physiology*, *157*(1), 305–316. https://doi.org/10.1104/pp.111.177576

Negretti, R. R. R. D., Gomes, C. B., Da Silva Mattos, V., Somavilla, L., Manica-Berto, R., Agostinetto, D., Castagnone, P., & Carneiro, R. M. D. G. (2017). Characterisation of a *Meloidogyne* species complex parasitising rice in southern Brazil. *Nematology*, *19*(4), 403–412. https://doi.org/10.1163/15685411-00003056

Netscher, C., & Erlan. (1993). A root-knot nematode, *Meloidogyne graminicola*, parasitic on rice in Indonesia. *Afro-Asian Journal of Nematology*, *3*(1), 90–95.

Noon, J. B., Qi, M., Sill, D. N., Muppirala, U., Akker, S. E. den, Maier, T. R., Dobbs, D., Mitchum, M. G., Hewezi, T., & Baum, T. J. (2016). A Plasmodium-like virulence effector of the soybean cyst nematode suppresses plant innate immunity. *New Phytologist*, *212*(2), 444–460. https://doi.org/10.1111/nph.14047

Nugaliyadde, L., Dissanayake, D. M. N., Herath, H. M. D. N., & Ekanayake, H. M. R. K. (2001). Outbreak of rice root knot nematode, *Meloidogyne graminicola* (Golden and Birchfield) in Nikewaratiya, Kurunegala in maha 2000/2001. *Annals of the Sri Lanka Department of Agriculture*, *3*, 373–374.

Oliveira, S. A. de, Oliveira, C. M. G. de, Maleita, C. M. N., Silva, M. de F. A., Abrantes, I. M. de O., & Wilcken, S. R. S. (2018). First report of *Meloidogyne graminis*on golf courses turfgrass in Brazil. *PLOS ONE*, *13*(2), e0192397. https://doi.org/10.1371/journal.pone.0192397

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., & Windham, E. (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences*, *105*(39), 14802–14807. https://doi.org/10.1073/pnas.0805946105

Patel, N., Hamamouch, N., Li, C., Hewezi, T., Hussey, R. S., Baum, T. J., Mitchum, M. G., & Davis, E. L. (2010). A nematode effector protein similar to annexins in host plants. *Journal of Experimental Botany*, *61*(1), 235–248. https://doi.org/10.1093/jxb/erp293

Petitot, A.-S., Dereeper, A., Agbessi, M., Da Silva, C., Guy, J., Ardisson, M., & Fernandez, D. (2015). Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants. *Molecular Plant Pathology*, *17*(6), 860–874. https://doi.org/10.1111/mpp.12334

Petitot, A.-S., Dereeper, A., Silva, C. D., Guy, J., & Fernandez, D. (2020). Analyses of the Root-Knot Nematode (*Meloidogyne graminicola*) Transcriptome during Host Infection Highlight Specific Gene Expression Profiling in Resistant Rice Plants. *Pathogens*, *9*(8). https://doi.org/10.3390/pathogens9080644

Petrillo, M. D., & Roberts, P. A. (2005). Isofemale line analysis of *Meloidogyne incognita* virulence to cowpea resistance gene *Rk*. *Journal of Nematology*, *37*(4), 448–456.

Phan, N. T., De Waele, D., Lorieux, M., Xiong, L., & Bellafiore, S. (2018). A Hypersensitivity-Like Response to *Meloidogyne graminicola* in Rice (*Oryza sativa*). *Phytopathology*, *108*(4), 521–528. https://doi.org/10.1094/PHYTO-07-17-0235-R

Phan, N. T., Orjuela, J., Danchin, E. G. J., Klopp, C., Perfus-Barbeoch, L., Kozlowski, D. K., Koutsovoulos, G. D., Lopez-Roques, C., Bouchez, O., Zahm, M., Besnard, G., & Bellafiore, S. (2020). Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*). *Ecology and Evolution*, *10*(20), 11006–11021. https://doi.org/10.1002/ece3.6680

Pokharel, R. R. (2009). Damage of root-knot nematode (*Meloidogyne graminicola*) to rice in fields with different soil types. *Nematologia Mediterranea*. https://journals.flvc.org/nemamedi/article/view/87004

Pokharel, R. R., Abawi, G. S., Duxbury, J. M., Smat, C. D., Wang, X., & Brito, J. A. (2010). Variability and the recognition of two races in *Meloidogyne graminicola*. *Australasian Plant Pathology*, *39*(4), 326–333. https://doi.org/10.1071/AP09100

Pokharel, Ramesh R., Abawi, G. S., Zhang, N., Duxbury, J. M., & Smart, C. D. (2007). Characterization of isolates of *Meloidogyne* from rice-wheat production fields in Nepal. *Journal of Nematology*, *39*(3), 221–230.

Prasad, J. S., Panwar, M. S., & Rao, Y. S. (1985). Occurrence of root knot-nematode, *Meloidogyne graminicola*in semideepwater rice. *Curr. Sci*, *54*, 387–388.

Rahman, M. L. (1990). Effect of different cropping sequences on root-knot nematode, *Meloidogyne graminicola*, and yield of deepwater rice. *Nematologia Mediterranea*, 213–217.

Rao, Y. S., & Israel, P. (1973). Life history and bionomics of *Meloidogyne graminicola*, the rice root-knot nematode. *Indian Phytopathology*, *26*, 333–340.

Ravindra, H., Shegal, M., Narasimhamurthy, H. B., Jayalakshmi, K., & Imran Khan, H. S. (2017). Rice Root-Knot Nematode (*Meloidogyne graminicola*) an Emerging Problem. *International Journal of Current Microbiology and Applied Sciences*, *6*(8), 3143–3171.

Rodiuc, N., Vieira, P., Banora, M. Y., & de Almeida Engler, J. (2014). On the track of transfer cell formation by specialized plant-parasitic nematodes. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00160

Rosso, M. N., Favery, B., Piotte, C., Arthaud, L., De Boer, J. M., Hussey, R. S., Bakker, J., Baum, T. J., & Abad, P. (1999). Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. *Molecular Plant-Microbe Interactions*, *12*(7), 585–591. https://doi.org/10.1094/MPMI.1999.12.7.585

Roy, A. (1976). Pathological effects of *Meloidogyne graminicola* on rice and histopathological studies on rice and maize. *Indian Phytopathology*, 352–362.

Rutter, W. B., Hewezi, T., Maier, T. R., Mitchum, M. G., Davis, E. L., Hussey, R. S., & Baum, T. J. (2014). Members of the Meloidogyne avirulence protein family contain multiple plant ligand-like motifs. *Phytopathology*, *104*(8), 879–885. https://doi.org/10.1094/PHYTO-11-13-0326-R

Salalia, R., Walia, R. K., Somvanshi, V. S., Kumar, P., & Kumar, A. (2017). Morphological, Morphometric, and Molecular Characterization of Intraspecific Variations within Indian Populations of *Meloidogyne graminicola*. *Journal of Nematology*, *49*(3), 254–267.

Sato, K., Kadota, Y., Gan, P., Bino, T., Uehara, T., Yamaguchi, K., Ichihashi, Y., Maki, N., Iwahori, H., Suzuki, T., Shigenobu, S., & Shirasu, K. (2018). High-Quality Genome Sequence of the Root-Knot Nematode *Meloidogyne arenaria* Genotype A2-O. *Genome Announcements*, *6*(26). https://doi.org/10.1128/genomeA.00519-18

Scholl, E. H., Thorne, J. L., McCarter, J. P., & Bird, D. M. (2003). Horizontally transferred genes in plant-parasitic nematodes: A high-throughput genomic approach. *Genome Biology*, *4*(6), R39. https://doi.org/10.1186/gb-2003-4-6-r39

Semblat, J. P., Rosso, M. N., Hussey, R. S., Abad, P., & Castagnone-Sereno, P. (2001). Molecular cloning of a cDNA encoding an amphid-secreted putative avirulence protein from the root-knot nematode *Meloidogyne incognita*. *Molecular Plant-Microbe Interactions: MPMI*, *14*(1), 72–79. https://doi.org/10.1094/MPMI.2001.14.1.72

Semblat, J.-P., Bongiovanni, M., Wajnberg, E., Dalmasso, A., Abad, P., & Castagnone-sereno, P. (2000). Virulence and molecular diversity of parthenogenetic root-knot nematodes, *Meloidogyne* spp. *Heredity*, *84*(1), 81–89. https://doi.org/10.1046/j.1365-2540.2000.00633.x

Sobczak, M., Fudali, S., & Wieczorek, K. (2011). Cell Wall Modifications Induced by Nematodes. In J. Jones, G. Gheysen, & C. Fenoll (Eds.), *Genomics and Molecular Genetics of Plant-Nematode Interactions* (pp. 395–422). Springer Netherlands. https://doi.org/10.1007/978-94-007-0434-3_19

Somvanshi, V. S., Tathode, M., Shukla, R. N., & Rao, U. (2018). Nematode genome announcement: A draft genome for rice root-knot nematode, *Meloidogyne graminicola*. *Journal of Nematology*, *50*(2), 111–116. https://doi.org/doi:10.21307/jofnem-2018-018.

Song, Z. Q., Zhang, D. Y., Liu, Y., & Cheng, F. X. (2017). First Report of *Meloidogyne graminicola* on Rice (*Oryza sativa*) in Hunan Province, China. *Plant Disease*, *101*(12), 2153–2153. https://doi.org/10.1094/PDIS-06-17-0844-PDN

Soriano, I. R., Prot, J. C., & Matias, D. M. (2000). Expression of tolerance for *Meloidogyne graminicola* in rice cultivars as affected by soil type and flooding. *Journal of Nematology*, *32*(3), 309–317.

Soriano, Imelda R., Schmit, V., Brar, D. S., Prot, J.-C., & Reversat, G. (1999). Resistance to rice root-knot nematode *Meloidogyne graminicola* identified in *Oryza longistaminata* and *O. glaberrima*. *Nematology*, *1*(4), 395–398. https://doi.org/10.1163/156854199508397

Soriano, I.R., & Reversat, G. (2003). Management of *Meloidogyne graminicola* and yield of upland rice in South-Luzon, Philippines. *Nematology*, *5*(6), 879–884. https://doi.org/10.1163/156854103773040781

Sun, L., Zhuo, K., Lin, B., Wang, H., & Liao, J. (2014). The complete mitochondrial genome of *Meloidogyne graminicola* (Tylenchina): A unique gene arrangement and its phylogenetic implications. *PLoS ONE*, *9*(6), e98558. https://doi.org/10.1371/journal.pone.0098558

Susič, N., Koutsovoulos, G. D., Riccio, C., Danchin, E. G. J., Blaxter, M. L., Lunt, D. H., Strajnar, P., Širca, S., Urek, G., & Stare, B. G. (2020). Genome sequence of the root-knot nematode *Meloidogyne luci*. *Journal of Nematology*, *52*. https://doi.org/10.21307/jofnem-2020-025

Szitenberg, A., Cha, S., Opperman, C. H., Bird, D. M., Blaxter, M. L., & Lunt, D. H. (2016). Genetic Drift, Not Life History or RNAi, Determine Long-Term Evolution of Transposable Elements. *Genome Biology and Evolution*, *8*(9), 2964–2978. https://doi.org/10.1093/gbe/evw208

Szitenberg, A., Salazar-Jaramillo, L., Blok, V. C., Laetsch, D. R., Joseph, S., Williamson, V. M., Blaxter, M. L., & Lunt, D. H. (2017). Comparative genomics of apomictic root-knot nematodes: Hybridization, ploidy, and dynamic genome change. *Genome Biology and Evolution*, *9*(10), 2844–2861. https://doi.org/10.1093/gbe/evx201

Tian, Z. L., Barsalote, E. M., Li, X. L., Cai, R. H., & Zheng, J. W. (2017). First Report of Root-knot Nematode, *Meloidogyne graminicola*, on Rice in Zhejiang, Eastern China. *Plant Disease*, *101*(12), 2152. https://doi.org/10.1094/PDIS-06-17-0832-PDN

Tian, Z., Wang, Z., Maria, M., Qu, N., Zheng, J., 2019. *Meloidogyne graminicola* protein disulfide isomerase may be a nematode effector and is involved in protection against oxidative damage. Sci. Rep. 9, 11949. https://doi.org/10.1038/s41598-019-48474-w

Toida, Y., Tangchitsomkid, N., Keereewan, S., & Mizukubo, T. (1996). Nematode species attacking crops in Thailand with measurements of second-stage juveniles of *Meloidogyne* spp. *JIRCAS Journal*, *No. 3*, 59–68.

Tollis, M., & Boissinot, S. (2012). The evolutionary dynamics of transposable elements in eukaryote genomes. *Genome Dynamics*, *7*, 68–91. https://doi.org/10.1159/000337126

Ton, J., & Mauch-Mani, B. (2004). Beta-amino-butyric acid-induced resistance against necrotrophic pathogens is based on ABA-dependent priming for callose. *The Plant Journal: For Cell and Molecular Biology*, *38*(1), 119–130. https://doi.org/10.1111/j.1365-313X.2004.02028.x

Triantaphyllou, A. C. (1966). Polyploidy and reproductive patterns in the root-knot nematode *Meloidogyne hapla*. *J. Morphol.*, *118*, 403–413.

Triantaphyllou, A. C. (1981). Oogenesis and the chromosomes of the parthenogenetic root knot nematode *Meloidogyne incognita*. *J. Nematol*, *13*, 95–104.

Triantaphyllou, A. C. (1983). Cytogenetic aspects of nematode evolution. *See Ref*, *80*, 55–71.

Triantaphyllou, A. C. (1984). Polyploidy in meiotic parthenogenetic populations of *Meloidogyne hapla* and a mechanism of conversion to diploidy. *Rev. N'ematol.*, *7*, 65–72.

Triantaphyllou, A. C. (1985). Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes. In J. N. Sasser & C. C. Carter (Eds.), *An advanced treatise on* Meloidogyne. *Biology and control* (Vol. 1, pp. 113–126). Raleigh. http://agris.fao.org/agris-search/search.do?recordID=US8743737

Triantaphyllou, A. C. (1987). Cytogenetic status of *Meloidogyne* (*Hypsoperine*) *spartinae* in relation to other *Meloidogyne* species. *Journal of Nematology*, *19*(1), 1–7.

Triantaphyllou, A. C. (1990). Status of *Meloidogyne kikuyensis* in relation to other root-knot nematodes. *Rev. Nematol.*, *13*, 175–180.

Trudgill, D. L., & Blok, V. C. (2001). Apomictic, polyphagous root-knot nematodes: Exceptionally successful and damaging biotrophic root pathogens. *Annual Review of Phytopathology*, *39*, 53–77. https://doi.org/10.1146/annurev.phyto.39.1.53

Ventura, W., Watanabe, I., Castillo, M. B., & Cruz, A. D. la. (1981). Involvement of nematodes in the soil sickness of a dryland rice-based cropping system. *Soil Science and Plant Nutrition*, *27*(3), 305–315. https://doi.org/10.1080/00380768.1981.10431285

Vieira, P., & Gleason, C. (2019). Plant-parasitic nematode effectors—Insights into their diversity and new tools for their identification. *Current Opinion in Plant Biology*, *50*, 37–43. https://doi.org/10.1016/j.pbi.2019.02.007

Wang, G. F., Xiao, L. Y., Luo, H. G., Peng, D. L., & Xiao, Y. N. (2017). First report of *Meloidogyne graminicola* on rice in Hubei province of China. *Plant Disease*, *101*(6), 1056. https://doi.org/10.1094/PDIS-12-16-1805-PDN

Windham, G. L., & Golden, A. M. (1990). First report of *Meloidogyne graminicola* in Mississippi. *Plant Disease*, *74*(12). https://www.cabdirect.org/cabdirect/abstract/19912305461

Xie, J. L., Xu, X., Yang, F., Xue, Q., Peng, Y. L., & Ji, H. L. (2019). First Report of Root-Knot Nematode, *Meloidogyne graminicola*, on Rice in Sichuan Province, Southwest China. *Plant Disease*, *103*(8), 2142–2142. https://doi.org/10.1094/PDIS-03-19-0502-PDN

Zainal-Abidin, A. L., Momen-Abdullah, M. A., & Zawiyah, A. H. (1994). 246–247.

Zhan, L., Ding, Z., Peng, D., Peng, H., Kong, L., Liu, S., Liu, Y., Li, Z., & Huag, W. (2018). Evaluation of Chinese rice varieties resistant to the root-knot nematode *Meloidogyne graminicola*. *Journal of Integrative Agriculture*, *17*(3), 621–630. https://doi.org/10.1016/S2095-3119(17)61802-1

Zhou, X., Liu, G. K., Xiao, S., & Zhang, S. S. (2014). First report of *Meloidogyne graminicola* infecting banana in China. *Plant Disease*, *99*(3), 420–420. https://doi.org/10.1094/PDIS-08-14-0810-PDN

Zhuo, K., Naalden, D., Nowak, S., Huy, N.X., Bauters, L., Gheysen, G., 2019. A *Meloidogyne graminicola* C-type lectin, Mg01965, is secreted into the host apoplast to suppress plant defence and promote parasitism. Mol. Plant Pathol. 20, 346–355. https://doi.org/10.1111/mpp.12759

CHAPTER II

On the relatedness of two rice-parasitic root-knot nematode species and the recent expansion of *Meloidogyne graminicola* in Southeast Asia

Summary in English

In the first chapter, we have provided the general knowledge available at that time to understand the parasitic lifestyle, genomic resources and molecular evolution of *M. graminicola* as well as that of other *Meloidogyne* spp. From chapter II to chapter IV, we will present the results of our study in particular with regard to the evolutionary history of *M. graminicola*.

In this second chapter, we study the intraspecific diversity, genomic structure and origin of the *M. graminicola* species from data obtained by sequencing different isolates with the HiSeq 2000 or 2500 technology (Illumina Inc., San Diego). Intraspecific diversity of *Mg* isolates was addressed using a small rDNA sequence (ITS region of about 500 bp). This sequence, which is easy to amplify by PCR, unfortunately does not provide much useful information in the analysis of intraspecific variability, especially since the polyploid status of the *Mg* species theoretically requires the amplification and sequencing of all the ITS copies present in the genome of an individual to be able to approach a rigorous phylogenetic study. In contrast, the haploid mitochondrial genome, which is known to evolve more rapidly than the nuclear genome in the Nematoda phylum, can be used as a source of molecular markers to study the different isolates. From the reference mitochondrial sequence of *M. graminicola*, we reconstructed the complete mitochondrial genome of ten other *M. graminicola* isolates (exclusively from Southeast Asia, except for one isolate from South America). Comparative analysis of the mitochondrial genomes of 12 isolates (ten isolates from this study plus two reference isolates from the Philippines and China) revealed only fifteen polymorphisms including 11 SNPs, three indels and one inversion. The mitochondrial haplotype network based on these polymorphisms was phylogenetically uninformative for these 12 isolates. The most common ancestral haplotype (from which the other types are derived) is shared by geographically distant populations. To complete this mitogenome-based analysis, two divergent copies (due to the polyploid state of *Mg*) of three nuclear genomic regions (full rDNA and two low copy genomic regions ACC6 and TAA6) were also extracted from the sequence data of the 11 isolates (except the Chinese isolate). No sequence polymorphisms were detected in the nuclear genomic regions among the 11 isolates. We therefore hypothesize that *M. graminicola* has recently spread globally. Because of its aggressiveness towards rice and many grasses, *M. graminicola* has been able to find the necessary hosts for its development after vectors have transported it from infested sites (human and animal migration, water movements, agricultural practices, etc.).

A first incomplete genome and a transcriptome of *M. graminicola* were published and available at the beginning of the thesis. Unfortunately, the published genome was incomplete

and without any description of the structure and organization of the genome. The genomes of *M. graminicola* and the phylogenetically related species *M. oryzae* were sequenced using Illumina sequencing technology. For the first time, the overall genome characteristics of these two species could be studied using k-mer analysis. The genome of *M. graminicola* thus appeared heterozygous with a level of heterozygosity of ~2%. The two k-mer distribution peaks for the *M. graminicola* genome represent either a diploid heterozygous structure or an allotetraploid structure with two divergent genomic copies. The apomictic genome of *M. oryzae* is also heterozygous with possibly three distinct genomic copies (triploid or allohexaploid).

Interestingly, sequence analyses of the mitochondrial and nuclear genomes have shown that, despite a different mode of reproduction, *M. graminicola* (36 chromosomes) and *M. oryzae* (54 chromosomes) are closely related species. The mitochondrial genomes show a high identity (96.8%) as well as strong synthesis. Surprisingly, three divergent copies of three nuclear genomic regions (rDNA, and the two low copy genomic regions ACC6 and TAA6) have also been constructed for *M. oryzae* reinforcing the hypothesis that this species is likely triploid. Phylogenetic analysis using these divergent copies showed that one sequence type for each genomic region was shared between *M. graminicola* and *M. oryzae*. This strongly suggests that these shared sequences would come from a common ancestor of both *M. graminicola* and *M. oryzae*.

The results of this chapter lead to the hypothesis of a recent expansion of *M. graminicola* at least in Southeast Asia. Nevertheless, this hypothesis should be confirmed by a more global analysis of the whole nuclear genome. The polyploid character of *M. graminicola* but with a heterozygosity rate of ~2% reveals that markers should be used with caution if the final objective is to make phylogenetic analyses. The finding that a haplotype was shared between *M. graminicola* and *M. oryzae* tends to support the hypothesis that *M. graminicola* is likely to be derived from interspecific hybridization.

Résumé en Français

Dans le premier chapitre, nous avons apporté les connaissances générales alors disponibles et devant permettre d'appréhender le mode de vie parasitaire, les ressources génomiques et l'évolution moléculaire de *M. graminicola* ainsi que celle d'autres *Meloidogyne* spp. Du chapitre II au chapitre IV, nous présenterons les résultats de notre étude notamment en ce qui concerne l'histoire évolutive de *M. graminicola*.

Dans ce deuxième chapitre, nous étudions la diversité intraspécifique, la structure génomique et l'origine de l'espèces *M. graminicola* à partir de données obtenues par séquençage de différents isolats avec la technologie HiSeq 2000 ou 2500 (Illumina Inc., San Diego). La diversité intraspécifique des isolats de *Mg* avait été abordé en utilisant une petite séquence du rDNA (région ITS d'environ 500 pb). Cette séquence, facile à amplifier n'apporte malheureusement pas beaucoup d'information utile dans l'analyse de la variabilité intraspécifique d'autant que le statut polyploïde de l'espèce *Mg* nécessite théoriquement d'amplifier et de séquencer l'ensemble des copies ITS présente dans le génome d'un individu pour pouvoir aborder une étude phylogénétique rigoureuse. En revanche, le génome mitochondrial haploïde, dont on sait qu'il évolue plus rapidement que le génome nucléaire dans l'embranchement des Nematoda, peut être utilisé comme source de marqueurs moléculaires pour étudier les différents isolats. A partir de la séquence mitochondriale de référence de *M. graminicola*, nous avons reconstruit le génome mitochondrial complet de dix autres isolats de *M. graminicola* (provenant exclusivement d'Asie du Sud-Est, sauf un isolat d'Amérique du Sud). L'analyse comparative des génomes mitochondriaux de 12 isolats (dix isolats de cette étude à laquelle s'ajoute deux isolats de référence provenant des Philippines et de Chine) a révélé seulement quinze polymorphismes dont 11 SNPs, trois indels et une inversion. Le réseau d'haplotypes mitochondrial basé sur ces polymorphismes était phylogénétiquement non informatif pour ces 12 isolats. L'haplotype ancestral le plus courant (dont sont issus les autres types) est partagé par des populations éloignées géographiquement. Pour compléter cette analyse basée sur le mitogénome, deux copies divergentes (due à l'état polyploïde de *Mg*) de trois régions génomiques nucléaires (ADNr complet et deux régions génomiques à faible copie ACC6 et TAA6) ont ainsi également été extraite des données de séquences des 11 isolats (excepté l'isolat de Chine). Aucun polymorphisme de séquence n'a été détecté dans les régions génomiques nucléaires parmi les 11 isolats. Nous émettons donc l'hypothèse que *M. graminicola* s'est récemment propagé au niveau mondial. En raison de son agressivité vis-à-vis du riz et de nombreuses graminées, *M. graminicola* a pu trouver les hôtes nécessaires à son

développement après que des vecteurs l'ai transporté à partir de sites infestés (migration des hommes et des animaux, mouvements d'eau, pratiques agricoles…).

Un premier génome incomplet ainsi qu'un transcriptome de *M. graminicola* étaient publié et disponible au commencement de la thèse. Malheureusement, le génome publié était incomplet et sans aucune description sur la structure et l'organisation du génome. Les génomes de *M. graminicola* et de l'espèce phylogénétiquement proche *M. oryzae* ont été séquencée à l'aide de la technologie de séquençage Illumina. Pour la première fois, les caractéristiques globales du génome de ces deux espèces ont pu être étudié à l'aide de l'analyse k-mer. Le génome de *M. graminicola* est ainsi apparu hétérozygote avec un niveau d'hétérozygotie de ~2%. Les deux pics de distribution de k-mer pour le génome de *M. graminicola* représentent soit une structure hétérozygote diploïde, soit une structure allotétraploïde avec deux copies génomiques divergentes. Le génome apomictique de *M. oryzae* est également hétérozygote avec éventuellement trois copies génomiques distinctes (triploïde ou allohexaploïde).

Il est intéressant de noter que les analyses de séquences des génomes mitochondriaux et nucléaires ont démontré que, malgré un mode de reproduction différent, *M. graminicola* (36 chromosomes) et *M. oryzae* (54 chromosomes) sont des espèces étroitement apparentées. Les génomes mitochondriaux présentent une identité élevée (96,8%) ainsi qu'une forte synténie. De façon surprenante, trois copies divergentes de trois régions génomiques nucléaires (ADNr, et les deux régions génomiques à faible copie ACC6 et TAA6) ont également été construites pour *M. oryzae* renforçant ainsi l'hypothèse que cette espèce est vraisemblablement triploïde. L'analyse phylogénétique en utilisant ces copies divergentes a montré qu'un type de séquences pour chaque région génomique était partagé entre *M. graminicola* et *M. oryzae*. Cela suggère fortement que ces séquences partagées proviendraient d'un ancêtre commun aux deux espèces *M. graminicola* et *M. oryzae*.

Les résultats de ce chapitre conduisent à l'hypothèse d'une expansion récente de *M. graminicola* au moins en Asie du Sud-Est. Néanmoins cette hypothèse devrait être confirmé par une analyse plus globale portant sur l'ensemble du génome nucléaire. Le caractère polyploïde de *M. graminicola* mais avec un taux d'hétérozygotie de ~2% révèle que les marqueurs doivent être utilisés avec précaution si l'objectif final et de faire des analyses phylogénétiques. Le constat qu'un haplotype était partagé entre *M. graminicola* et *M. oryzae* tend à soutenir l'hypothèse que *M. graminicola* proviendrait probablement d'une hybridation interspécifique.

Article

On the relatedness of two rice-parasitic root-knot nematode species and the recent expansion of *Meloidogyne graminicola* in Southeast Asia

Guillaume Besnard [a,1,*], Ngan Thi Phan [b,d,1], Hai Ho Bich [c], Alexis Dereeper [b], Hieu Trang Nguyen [d], Patrick Quénéhervé [b], Jamel Aribi [b] and Stéphane Bellafiore [b,d,*]

[a] CNRS-UPS-IRD, UMR5174, EDB, 118 route de Narbonne, 31062 Toulouse, France

[b] IRD, Cirad, University of Montpellier II, Interactions Plantes Microorganismes Environnement (IPME), 34394 Montpellier, France

[c] IOIT/USTH, VAST and UMMISCO, IRD, 18 Hoang Quoc Viet, Hanoi, Vietnam

[d] IRD, LMI RICE, University of Science and Technology of Hanoi, Agricultural Genetics Institute, Hanoi, Viet Nam

* Correspondence: guillaume.besnard@univ-tlse3.fr (G. Besnard) and stephane.bellafiore@ird.fr (S. Bellafiore).

[1] These authors contributed equally to this work.

Genes (Basel). 2019 Feb; 10(2): 175.

Published online 2019 Feb 25. doi: 10.3390/genes10020175

Abstract

Meloidogyne graminicola is a facultative meiotic parthenogenetic root-knot nematode (RKN) that seriously threatens agriculture worldwide. We have little understanding of its origin, genomic structure and intraspecific diversity. Such information would offer a better knowledge of how this nematode successfully damaged rice in many different environments. Previous studies on nuclear ribosomal DNA (nrDNA) suggested a close phylogenetic relationship between *M. graminicola* and *M. oryzae*, despite their different modes of reproduction and geographical distribution. In order to clarify the evolutionary history of these two species and to explore their molecular intraspecific diversity, we sequenced the genome of 12 *M. graminicola* isolates, representing populations of worldwide-spread origins, and two South American isolates of *M. oryzae*. *k*-mer analysis of their nuclear genome plus the detection of divergent homologous genomic sequences indicate that both species show a high proportion of heterozygous sites (ca. 1-2%), which had never been previously reported in facultative meiotic parthenogenetic RKNs. These analyses also point out to a distinct ploidy level in each species, compatible with a diploid *M. graminicola* and a triploid *M. oryzae*. Phylogenetic analyses of mitochondrial genomes and three nuclear genomic sequences confirm close relationships between these two species, with *M. graminicola* being a putative parent of *M. oryzae*. In addition, comparative mitogenomics of those 12 *M. graminicola* isolates with a Chinese published isolate reveal only 15 polymorphisms that are phylogenetically non-informative. Eight mitotypes were distinguished, the most common one being shared by distant populations from Asia and America. This low intraspecific diversity, coupled to a lack of phylogeographic signal, suggests a recent worldwide expansion of *M. graminicola.*

Keywords: biological invasion; heterozygous genome; *Meloidogyne*; mitogenome; ribosomal DNA; root-knot nematode.

I. Introduction

Nematodes constitute an ancient and diverse animal phylum [1] representing 80% of multicellular animals on Earth [2]. Ubiquitous and found in various environments ranging from marine sediments to arid deserts, these metazoans can be carnivore, omnivore, bacterivore, fungivore or can feed exclusively on plants (15% of the described species). Plant parasitic nematodes (PPNs) cause an annual economic loss of over $US80 billion in worldwide agriculture [3]. Among PPNs, root-knot nematodes (RKNs, genus *Meloidogyne*) and cyst nematodes are the most important crop damaging species. In Asia, RKNs are responsible for about 15% of total economic losses in rice production [4].

Beside the actual sparse knowledge on nematode biodiversity, comprehension of the evolutionary history of this phylum is limited due to the rarity of plant parasitic nematode fossils discovered to date [5]. Interestingly, thanks to the few RKNs currently characterized, different modes of reproduction have been described, from amphimixis to mitotic parthenogenesis with intermediate states where both sexual and asexual (parthenogenetic) reproduction coexist [6–9]. For instance, obligatory cross-fertilization was reported in *M. kikuyensis*, facultative meiosis (automixis) in *M. graminicola*, obligatory mitotic parthenogenesis (apomixis) in *M. oryzae*, and automictic or apomictic parthenogenesis in different polyploid or aneuploid forms, such as *M. arenaria*, *M. incognita* and *M. javanica* [9]. These variations in reproductive modes within RKNs were also related to specific ploidy levels and host range [7,9]. Based on a pioneering work on cytogenetic studies, Triantaphyllou [9] proposed that RKNs have undergone unparalleled extensive cytogenetic diversification. He also concluded that those characteristic features were the establishment of meiotic and mitotic parthenogenesis in association with various degrees of polyploidy and aneuploidy.

During the last 20 years, DNA analysis of PPN genomic sequences has been a great help to taxonomists and evolutionists for analyzing and understanding the huge diversity and life-history of PPNs and particularly RKNs. Indeed, comparative genomics have revealed huge genome diversity in RKNs resulting from hybridization, whole genome duplication, recombination and horizontal gene transfers [10–12].

In RKNs, three main lineages (Clades I, II and III) have been identified using the 18S ribosomal RNA gene as a marker [13]. Clade III includes meiotic parthenogenetic species (e.g., *M. exigua*, *M. graminicola*, *M. chitwoodi*) as well as the apomictic *M. oryzae*. Mainly present in Asia and tropical America, *M. graminicola* has become a worldwide major threat to rice agriculture with recent records in Madagascar and South Europe [14,15]. In contrast, *M. oryzae*, the only described apomictic species belonging to Clade III also feeding on rice, is so far

restricted to northern South America [16,17]. These two rice parasites have been discovered relatively recently, ca. 40-50 years ago [16,18] and they are known to display variable ploidy levels (i.e., 36 chromosomes in *M. graminicola* [6] and 54 in *M. oryzae* [19]). Because of their different levels of ploidy, reproductive patterns and geographic distribution, the comparative study of *M. graminicola* and *M. oryzae* genomes in this narrow Clade III may provide important insights to understand *Meloidogyne*'s evolutionary issues.

Comparative genomics can shed light on the recent history of RKN species: trace back their spread, identify hybridization and genome duplication events, help to understand better the impact of their mode of reproduction on their demography, and help to investigate their adaptive processes to different environmental conditions [10–12]. Compared to other *Meloidogyne* species belonging to Clades I and II, the diversity and genetic structure of Clade III remains poorly understood, except for the facultative meiotic parthenogenetic *M. chitwoodi*, a well-known crop pest of potato [20,21]. Generating genomic data on both distinct species and isolates from various locations is thus necessary to better understand the origin and spread of this clade.

In the present study, we explore the genome structure and diversity of two related RKNs belonging to Clade III: *M. graminicola* and *M. oryzae*. Shotgun sequencing data were generated by HiSeq on 14 isolates. The genome complexity was investigated in both species with a *k*-mer approach and, the complete mitogenome and selected nuclear genomic DNA regions (including the nrDNA cluster and two low-copy genomic regions) were used to investigate the phylogenetic position of the two species within the *Meloidogyne* genus and explore the intraspecific diversity among *M. graminicola* isolates. Our results indicate a high proportion of heterozygous sites (ca. 1-2%) in the genome of the two investigated nematode species. In addition, both species sharing highly similar nuclear sequences, we propose that *M. graminicola* is a potential parent of *M. oryzae*. Finally, the low genomic diversity observed within the studied *M. graminicola* populations strongly suggests a recent expansion of this taxon in SE Asia.

II. Materials and methods

2.1. Nematode sampling, DNA extraction and sequencing

Twelve isolates of *M. graminicola* (*Mg*) and two isolates of *M. oryzae* (*Mo*) (Table S1) were obtained after single nematode infection as described in Bellafiore et al. [22]. Subsequently, each isolate was propagated in a hydroponic solution on the susceptible rice cultivar IR64 during four weeks before proceeding to egg and DNA extractions [22]. Briefly, juveniles and eggs were extracted from roots with a blender in 0.8% hypochlorite solution,

purified by centrifugation in 60:40:20:10% sucrose discontinuous gradient to separate nematodes from different plant debris and limit contamination by other organisms such as bacteria [23], and finally rinsed several times with sterile distilled water and conserved at –80ºC. Total genomic DNA (gDNA) was isolated using proteinase K treatment and phenol-chloroform extraction followed by ethanol precipitation, as described by Besnard et al. [24]. DNA quality and quantification were assessed with a NanoDrop spectrophotometer and PicoGreen®dsDNA quantitation assay. One micro-liter of each gDNA extract was also loaded on 1% agarose gel and subjected to electrophoresis to control the DNA integrity.

For each accession, 300 ng of double-stranded DNA (dsDNA) was used for shotgun sequencing using the Illumina technology at the GeT-PlaGe core facility, hosted by INRA (Toulouse, France). DNAs were sonicated to get inserts of approximately 380 bp. The libraries were constructed using the Illumina TruSeq DNA Sample Prep v.2 kit, following the instructions of the supplier as previously described by Besnard et al. [24]. Libraries were multiplexed with libraries generated in other projects (24 per flow cell), and inserts were then sequenced from both ends on HiSeq 2000 or 2500 (Illumina Inc., San Diego).

2.2. Assembly and phylogenetic analysis of the mitogenome

For each accession, the mitochondrial genome (mitogenome) was *de novo* assembled and annotated using the experimental procedure described by Besnard et al. [24]. As heteroplasmy was observed in a few cases, a consensus sequence was reconstructed for each nematode isolate, only considering sites as ambiguous (using IUPAC codes) when the minor sequence was supported by at least 5% of reads. Reads were then mapped onto the final sequence to assess the sequencing depth of the mitogenomes.

Reference mitogenome sequences for *M. graminicola* (NC_024275.1), *M. chitwoodi* (KJ476150.1), *M. enterolobii* (KP202351.1), *M. arenaria* (KP202350.1), *M. javanica* (KP202352.1), *M. incognita* (KJ476151.1), and *Pratylenchus vulnus* (GQ332425.1) were retrieved from GenBank. Nucleotide sequence of six protein-coding genes (*cox1-cox3-nad4L-nad3-nad4-nad5*) and one rRNA gene (*rrnS*) were extracted from these mitogenomes. In addition, we retrieved these mitochondrial genes from *M. hapla* and *M. floridensis*, using a BLASTN search (http://xyala.cap.ed.ac.uk ; Table S2). For each gene, sequences were aligned with those generated from the *M. graminicola* and *M. oryzae* isolates. All alignments were done using MUSCLE [25], and poorly aligned regions with the outgroup sequence (*P. vulnus*) were manually refined. The phylogenetic analysis of the sequence dataset was conducted and visualized based on Maximum Likelihood (ML) using MEGA v.7 [26] and confirmed using raxmlGUI V1.5 [27]. Initial trees for the heuristic search were obtained automatically by

applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Likelihood approach, and then selecting the topology with the highest log likelihood value. The program JModelTest v.2.1.10 [28,29] was used to select the best-fit nucleotide substitution model by likelihood ratio test using Akaike Information Criterion and Bayesian Information Criterion. A general time reversible with discrete Gamma distribution (GTR + G) was used to model evolutionary rate differences among sites. A bootstrap analysis with 1,000 replicates was performed to estimate the support for each node in the ML tree.

To infer the evolutionary history of *M. graminicola* on a maternally inherited genome, a haplotype network was then reconstructed based on the 13 mitogenome sequences available for this species, i.e., one from Brazil and 12 from South East Asia (including one Chinese isolate; sequence published and retrieved from GenBank [30]). An alignment was performed between all isolates excluding the 111R region (i.e., long region of the control region with 111-bp repetitions). Sequence polymorphisms were used to reconstruct a reduced-median haplotype network with Network v.5 [31]. The presence or absence of indels was coded as 1 and 0, respectively. A short inversion was similarly coded. Mutations leading to heteroplasmic sites were further considered and also positioned on the network.

2.3. Assembly and phylogenetic analysis of nrDNA sequences

The tandemly repeated nature of the nrDNA cluster associated to concerted evolution between adjacent copies is usually an advantage to properly assemble this nuclear genomic region with low-depth sequencing data [32]. Complete nrDNA sequences (including 18S-ITS1-5.8S-ITS2-28S) were thus reconstructed using the approach described in Besnard et al. [24]. Briefly, a sequence containing Internal Transcribed Spacers (ITS) from *M. graminicola* or *M. oryzae* (GenBank nos KF250488 and KY962653) was used as a seed and elongated by recursively incorporating reads identical to at least 90 bp at the ends of the contig (in GENEIOUS). Diverging nrDNA units were however found in each isolate, and needed to be assembled separately for proper phylogenetic analyses. Paired-end reads were thus carefully phased after visual inspection, resulting in the assembly of two homologous sequences in each accession.

Phylogenetic analyses were then conducted to investigate relationships between *Meloidogyne* species and decipher the origin of diverging units in *M. graminicola* and *M. oryzae* genomes. We selected nrDNA sequences from GenBank for seven *Meloidogyne* species that belong to different *Meloidogyne* lineages of Clade III (i.e., *M. trifoliophila*, *M. chitwoodii*, *M. naasi*, *M. fallax*, *M. minor*, *M. graminis* and *M. marylandii*). Two trees were reconstructed using the D2-D3 region (656 bp) from the 28S gene, and the ITS (538 bp). Whole

nrDNA sequences (ca. 8 Kb) of *M. graminicola* and *M. oryzae* were further aligned to investigate phylogenetic relationships between their divergent units. All alignments were done using MUSCLE [25] and cleaned with Gblocks using default settings [33]. Phylogenetic analyses were done using MEGA v.7 [26] and raxmlGUI V1.5 [27]. Phylogenetic trees were reconstructed using the Maximum Likelihood method based on the GTR + G model selected by JModelTest v.2.1.10 [28,29]. Support of nodes was estimated with the rapid bootstrap algorithm (1,000 iterations).

2.4. Nuclear genome structure analysis

A draft genome and a transcriptome of *M. graminicola* were recently published [34, 35]. However, the draft genome of this species is incomplete (84.27% CEGMA, [Core Eukaryotic Genes Mapping Approach] [36]) without any description on genome structure and organization [34]. We further tested the completeness of published draft genome and transcriptome of *M. graminicola* using BUSCO v.3 (Benchmarking Universal Single-Copy Orthologs [37]). The nematode species *Caenorhabditis elegans* was set as species-specific-trained parameters for gene prediction. The BUSCO dataset "Eukaryota *odb9*", which includes 303 Eukaryote single-copy orthologs, was used as the reference.

The sequencing reads of the 14 isolates were cleaned and contaminant sequences (e.g., from bacteria) were removed. FastQC [38] was used for quality control. Preprocessing procedure included adaptor removal, quality trimming, duplication and contaminant removal. In details, Trimmomatic [39] was employed to trim reads of TruSeq adaptors and low quality nucleotides (cut-off score of 20). Duplication levels were reduced to less than 15% by FastUniq [40]. On close inspection of GC content plots, minor second peaks of 50 to 70% were found, indicating potential contaminants. We hence utilized BBTools (https://jgi.doe.gov/data-and-tools/bbtools/) to filter and remove reads that mapped over 80% identity to reference genomes. Detail number of reads for each isolate before and after contamination removal are listed in Table S1. After contamination removal, eight *M. graminicola* (i.e., Mg-VN6, Mg-VN11, Mg-VN18, Mg-VN27, Mg-L1, Mg-C21, Mg-P, Mg-Brazil) and the two *M. oryzae* isolates showed only one peak of GC content in genome sequences. Thus, clean reads from these isolates were then used for further analyses. Four *M. graminicola* isolates (i.e., Mg-L2, Mg-C25, Mg-Java, Mg-Borneo) were removed due to the presence of a second peak in GC content plots.

The overall characteristics of *M. graminicola* and *M. oryzae* nuclear genomes were determined by a *k*-mer analysis. We investigated the Mg-VN18 and Mo-M2 isolates, which have similar sequencing coverage (see below). From the cleaned reads, we extracted canonical *k*-mers (*k* from 15 to 31 with step 2) by Jellyfish [41] to plot *k*-mer distributions of the two

isolates. For each *k*, the abundance histograms and corresponding genome estimates were investigated (Figure S1 shows four different *k* values on genome sequences of eight *M. graminicola* isolates and two *M. oryzae* isolates). Since similar shapes and estimates were found across *k* values, we opted to discuss here the 25-mer analyses of genome structure and complexity. *k* of 25 was also chosen to characterize the genomes of human [42] and plants [e.g., maize [43]], owing to its robustness to repeated region and to sequencing error. More concretely, coverage thresholds of heterozygous and homozygous peaks and repeated regions in the distributions were manually detected, according to Kajitani et al. [44]. Based on those, sequencing coverage, genome size, heterozygous rate and repeated proportions were then calculated as described in Supplementary Methods.

2.5. Analysis of two low-copy nuclear genomic regions among accessions of *M. graminicola* and *M. oryzae*

In order to determine the origin of the high heterozygosity observed in *M. graminicola* and *M. oryzae* (see below), we decided to investigate the diversity of low-copy regions from the nuclear genome. A reference-guided approach was used to *de novo* assemble (in GENEIOUS) two genomic regions containing microsatellite motifs (further referred to TAA6 and ACC6 regions). For each region, two distinct homologous sequences were clearly present in the 12 *M. graminicola* accessions, while two or three variants were detected in *M. oryzae*. Contigs of about 6 Kb were first assembled in Mg-VN18 and Mo-M2 by elongating initial sequence, recursively incorporating reads identical to at least 30 bp at both ends of the contig. Paired-end reads were carefully phased after visual inspection, resulting in the assembly of two or three homologous sequences. Their homology was estimated with MEGA (1 – *p*-distance). For each accession, reads of *M. graminicola* or *M. oryzae* were finally mapped onto the assembled contigs, majority-rule consensus sequences were extracted and sequencing depth was estimated. These final assemblies of each homologous sequence were manually checked to identify potential heterozygous sites (i.e., SNPs supported by at least two distinct reads). To identify coding sequences in these nuclear contigs, BLAST alignments were finally performed, using a database of 66,396 *M. graminicola* transcripts [35], considering each genomic region as a query.

All distinct homologous sequences of these two genomic regions were finally used to investigate the phylogenetic relationship between *M. graminicola* and *M. oryzae*. Sequence alignment and phylogenetic analysis methods were done as described in section 2.3. The GTR + G and GTR + I models were applied for ML analyses of TAA6 and ACC6 genomic regions, respectively.

III. Results

3.1. General mitogenome features of *M. oryzae*

We generated respectively 8,559,374 and 8,316,341 paired-end 100-bp reads on the two *M. oryzae* isolates from French Guiana (M1) and Suriname (M2) (Table S1). Based on these data, two mitogenomes of *M. oryzae* were assembled, the first ever produced so far for this species (GenBank no: MK507908). These sequences are AT-rich (83.2%) as previously reported in other *Meloidogyne* spp. [21,24,45]. Like in *M. graminicola*, tandemly-repeated elements of 111 bp (111R) are found in the control region [24]. In contrast, the other region with repeated elements (namely the 94R region) in the *M. graminicola* mitogenome shows shorter repeated elements in *M. oryzae* (65R). The 94R region corresponds to a 65R region with the addition of 29 bp at the end of the 65-bp elements (Figure S2). Note that the repeated 65-bp element is different from the 63-bp element of the 63R region present in the mitogenomes of the MIG group [*M. incognita* group [46]; Figure S2].

Excluding the 111R region, the mitogenome length of the two *M. oryzae* isolates is estimated at 17,069 bp (Mo-M1) and 17,066 bp (Mo-M2). Thirty-six genes were identified in these *M. oryzae* mitogenomes (12 protein-encoding genes, 22 tRNA genes and two rRNA genes), with the exact same gene arrangement that was observed in *M. graminicola* (Figure S3). Like in *M. graminicola*, one large non-coding region (NCR) was also found in *M. oryzae* between *nad4* and the tRNA gene S_2.

Alignment of the two *M. oryzae* mitogenome sequences without the 111R region revealed 11 nucleotide polymorphisms, including six substitutions (or single nucleotide polymorphisms, SNPs) and five indels (insertion/deletion) (Table S3). Eight of these polymorphisms are located in non-coding regions, while the other three are SNPs located in *apt6*, *nad5* and *cox3*. The two SNPs detected in *apt6* and *nad5* are non-synonymous and the SNP present in *cox3* is a silent mutation. Alignment of *M. oryzae* (Mo-M2) and *M. graminicola* (Mg-P) mitogenomes (excluding the 111R region) revealed a total of 619 SNPs and 58 indels, for an average identity of 96.4% (Table S3).

3.2. Phylogenetic analyses of *Meloidogyne* based on available complete mitogenomes

The phylogeny of the *Meloidogyne* genus reconstructed from seven mitochondrial genes supports the distinction of the three main RKN lineages (Figure 1). The early diverging lineage corresponds to Clade I, with *M. enterolobii* sister to a clade of four closely related species (*M. arenaria*, *M. javanica*, *M. incognita* and *M. floridensis*). Clade II, which is only represented

by *M. hapla*, is placed next to Clade I, but with a low support (bootstrap value of 78). In Clade III, *M. chitwoodii* is sister to *M. oryzae* and *M. graminicola,* with all nodes strongly supported.

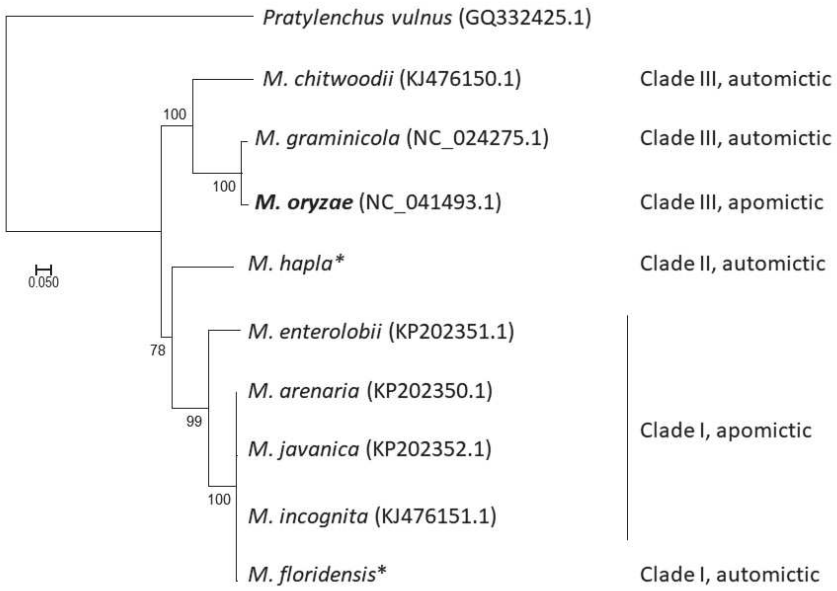


Figure 1. **Maximum Likelihood phylogenetic tree of *Meloidogyne* spp. based on seven mitochondrial genes (*cox1-rrnS-cox3-nad4L-nad3-nad4-nad5*) from nine root-knot nematode species**. The tree was rooted with *Pratylenchus vulnus* as an outgroup. GenBank mitogenome reference sequences are indicated in parenthesis (with sequences from this study in bold font), except for *M. hapla* and *M. floridensis* (*) for which accession numbers corresponding to the different genes used in the phylogeny are given in Table S2. The numbers beside branches represent Maximum Likelihood bootstrap support values >50%. Scale bar represents substitutions per nucleotide position.

3.3. Meloidogyne graminicola haplotype diversity

The mtDNA variability was investigated on 13 *M. graminicola* isolates. Eleven new mitogenome sequences were assembled in our study and were aligned with the two sequences already available in GenBank [NC_024275 (Philippines) and KJ139963 (China)]. These sequences, excluding the 111R region, range from 16,808 bp (Mg-Borneo) to 16,905 bp (Mg-C25 and Mg-VN27; Table S1). Fifteen polymorphisms were detected, including 11 SNPs, three indels and one inversion (Table S3; Figure 2). Eight of these 15 polymorphisms were located in coding regions and seven in the NCR (Table S4). Three substitutions were transversions (A, G ↔ C, T) and eight were transition mutations (A ↔ G, C ↔ T) (Table S4). One indel in *atp6* and four SNPs in *nad5*, *cox1*, *nad1* and *cox3* were non-silent mutations. The 15 polymorphic sites allowed the distinction of eight haplotypes among the 13 isolates (Figure 2). Eight

heteroplasmic sites were also detected in five isolates (Mg-L1, Mg-Borneo, Mg-VN6, Mg-VN11 and Mg-VN18; Table S4; Figure S4). Four of them were non-synonymous and located in *cox1* and *cob*. Taken together, these intraspecific polymorphisms allowed distinction of 12 haplotypes among the 13 isolates analyzed. Although polymorphisms were phylogenetically non-informative (except an insertion of one 94-bp repeat in the 94R region in Mg-C25 and Mg-VN27, that is putatively homoplasic; Table S4), we can assume that the likely ancestral sequence represents the most frequent haplotype in the center of the network. Only Mg-Brazil and Mg-Java2 share this haplotype without any heteroplasmic sites (Figure S4; Table S4).

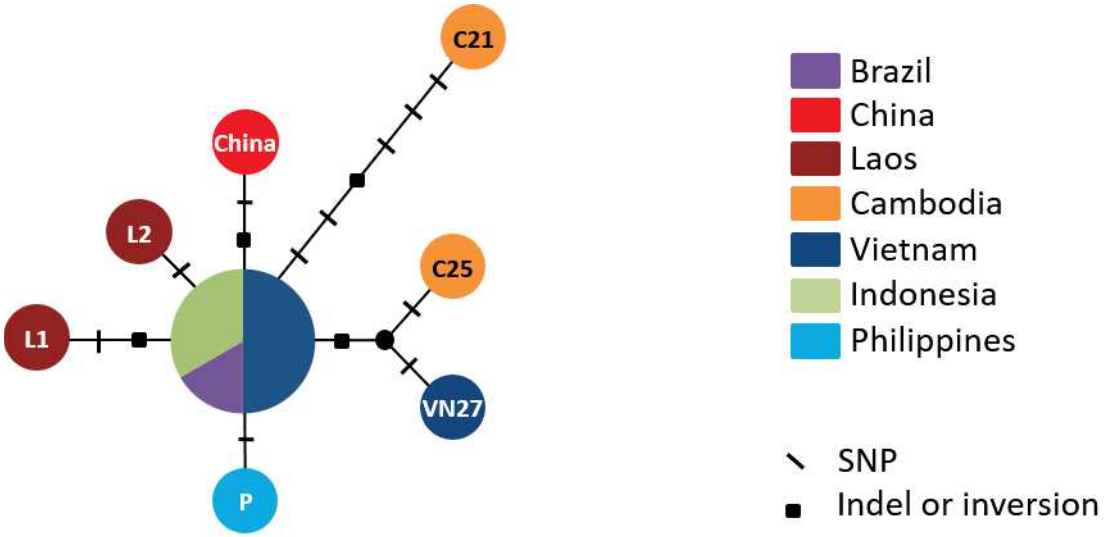


Figure 2. **Reduced-median network of *Meloidogyne graminicola* mitochondrial haplotypes**. The network was reconstructed with Network v.5 [30], using the 13 available mitogenome sequences, excluding the 111R region. Code names of distinct populations are indicated in the circles and geographic origin is displayed by different colors. The number of mutations is shown on the branches with slashes and black squares that respectively indicate SNPs, and indels or inversion sites. See Figure S4 for the network considering heteroplasmic sites.

3.4. Analysis of nuclear ribosomal genes

Complete nrDNA sequences (including NC1-5S-NC2-18S-ITS1-5.8S-ITS2-28S; ca. 8 Kb) were reconstructed for each isolate (GenBank nos: LS974432, LS974433, LS974439, and LS974440). Two divergent nrDNA sequences (or ribotypes) were always detected (with a mean sequence identity of ca. 99%) in all *M. graminicola* isolates. Two main types were also assembled from the *M. oryzae* isolates (with a mean sequence identity of 99%), but a few intra-individual polymorphisms were detected in one of them. In both species, sequence polymorphisms (SNPs and Indels) between two nrDNA types were mostly located in non-coding regions (NC1, NC2 and ITS; Figure S5). No polymorphism was detected between

individuals on each nrDNA type among all *M. graminicola* isolates or among the two *M. oryzae* isolates. For phylogenetic reconstructions, we consequently considered only one pair of nrDNA sequences to represent *M. graminicola* and *M. oryzae*. Since complete nrDNA sequences have not been released for other *Meloidogyne* species belonging to Clade III, only the D2-D3 (656 bp) and ITS (538 bp) regions were independently used to analyze phylogenetic relationships within this root-knot nematode lineage. The two phylogenetic trees reconstructed on these regions were similar (Figures 3A and S6) and sustain relationships that are currently recognized in this *Meloidogyne* clade [47–49]. Three subclades were found where *M. marylandi* and *M. graminis* form an early diverging group. The *M. chitwoodi* group (*M. minor*, *M. chitwoodi* and *M. fallax*) forms a second cluster, sister to the *M. graminicola* group. In this latter, *M. naasi* forms an early-diverging lineage, while *M. trifoliophila* and the two nrDNA types of *M. graminicola* and *M oryzae* were intermingled in a well-supported clade (92% bootstrap). Interestingly, as strongly supported by the phylogenetic analysis from the whole nrDNA sequence, ribotypes I of *M. graminicola* and *M. oryzae* were almost identical, in contrast to their ribotype II that appeared relatively different from each other (Figures 3A-B and S6).

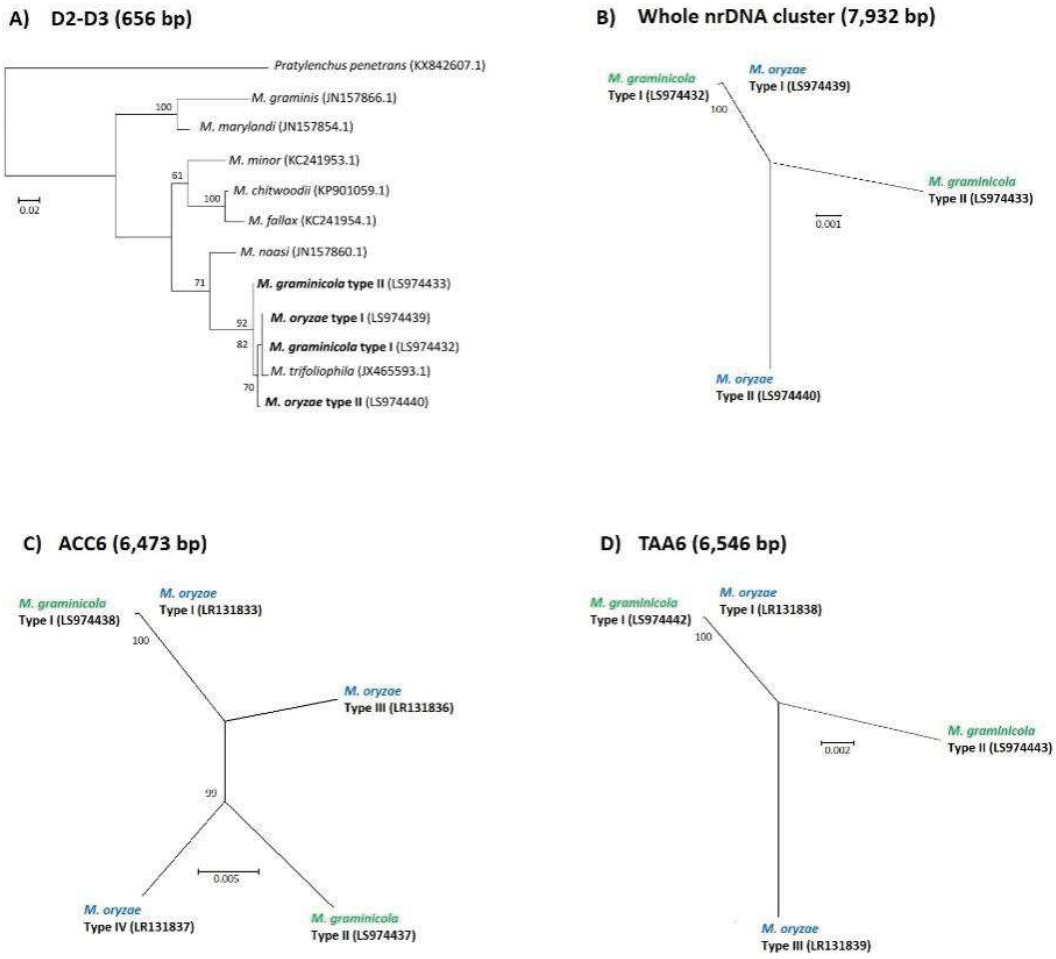


Figure 3. **Maximum Likelihood phylogenetic reconstructions of *Meloidogyne* species belonging to Clade III**, based on (**A**) D2-D3 nrDNA sequence, (**B**) the whole nrDNA cluster, (**C**) the ACC6 genomic region, and (**D**) the TAA6 genomic region, using either the GTR + G (A, B, D) or GTR + I (C) models. The size of each nucleotide alignment is given in parenthesis. Bootstrap values greater than 50% are indicated on nodes of each phylogenetic tree. GenBank sequences are indicated in parenthesis, and sequences from this study are in bold font. The phylogeny (A) is rooted with *Pratylenchus penetrans* as an outgroup. Scale bar represents substitutions per nucleotide position. For increasing the readability of panels B-D, names of *M. graminicola* and *M. oryzae* are distinguished with a distinct color (green and blue, respectively).

3.5. Nuclear genome features of *M. graminicola* and *M. oryzae*

The genome assembly completeness measured with BUSCO (Table S5) showed that only 73.6% of a set of conserved Eukaryote genes are fully present on the published draft genome [34]. More precisely, 15.2% of genes are partially-mapped, while 11.2% are completely missing. In comparison, the transcriptome assembly completeness [35] is higher with the detection of 88.1% of Eukaryote genes (Table S5) but still incomplete.

The Mg-VN18 (*M. graminicola*) and the Mo-M2 (*M. oryzae*) samples contain 8.4 million and 7.6 million of paired-end reads, respectively. The two genomes have similar GC contents, 24% in *M. graminicola* and 26% in *M. oryzae*. The equivalent sequencing coverage for both samples (ca. 21x; Supplementary Methods) allows genome comparison analyses. Under the assumption that genomic *k*-mer profile follows Poisson distribution [50], the peaks can be used to infer about genome ploidy. We noted a bimodal *k*-mer distribution for the *M. graminicola* genome (Figures 4 and S1), a major peak at sequencing coverage of 16x and a minor peak at 32x, which correspond respectively to heterozygous and homozygous peaks. Similar bimodal *k*-mer pattern has been previously described in other species and the two peaks found for *M. graminicola* could theoretically indicate either an allotetraploid genome, or a diploid genome with a high proportion of heterozygous sites (>1%) [44, 51]. In *M. oryzae,* the *k*-mer distribution is more complex with a heterozygous peak at 16x and a heavy tail up to 60x, as expected with an increased ploidy level [19]. In addition, a higher proportion of repeats was found in the *M. oryzae* genome (6.85% versus 5.92% for *M. graminicola*). Based on the hypothesis of a diploid *M. graminicola*, and a triploid *M. oryzae* [6,19], the haploid genome sizes could be estimated at about 35.64 Mb and 37.57 Mb respectively. These sizes are slightly lower than the haploid genome sizes of other *Meloidogyne* spp. (40-50 Mb) but in agreement with previous studies suggesting a smaller genome for *M. graminicola* [6,52].

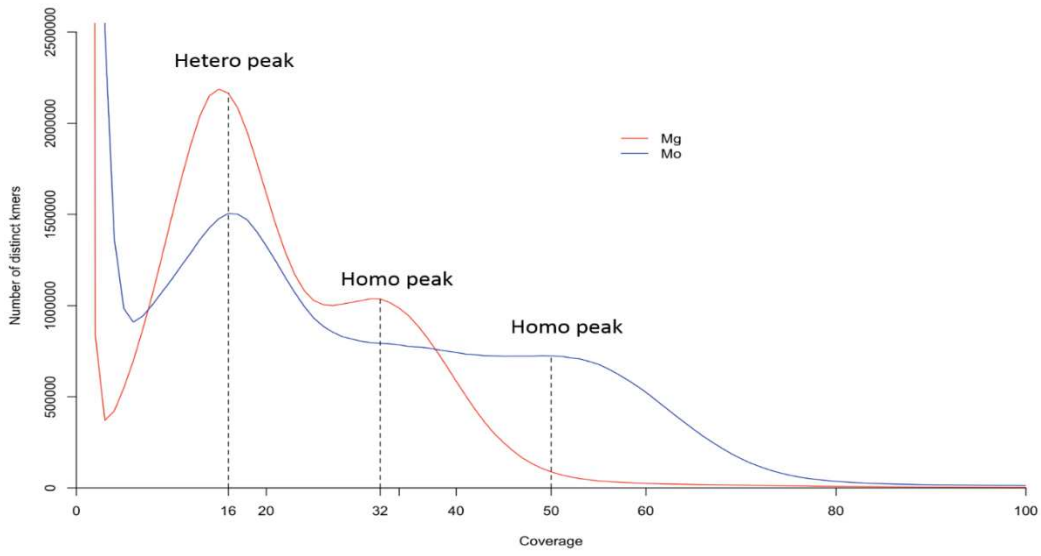


Figure 4. 25-mer distributions of *M. graminicola* (red) and *M. oryzae* (blue) genomes. The dashed vertical lines indicate the possible positions of hetero- and homo-peaks.

3.6. Analysis of two low-copy nuclear regions in *M. graminicola* and *M. oryzae*

Two nuclear low-copy genomic regions of approximately 6.3 Kb were assembled (named ACC6 and TAA6; Genbank nos: LS974437, LS974438, LS974442, LS974443, and LR131833 to LR131839). For both genomic regions, two divergent copies (arbitrarily named Type I and Type II) were detected with an even sequencing depth in all accessions of *M. graminicola*. The sequence identity between the two types was 97.61% and 98.36% for ACC6 and TAA6, respectively (Table S6C-D). Among isolates, the mean sequencing depth of these genomic regions was between 9.4x and 22.2x if we exclude Mg-Java2 (0.9x; Table S7). On average, the sequencing depth of a single-copy region of the nuclear genome was thus 93 ($^{+}$/- 40) times lower than for the mitogenome (Tables S1 and S7). No sequence polymorphism and no recombination between Types I and II were detected on a total of ca. 25.2 Kb among 11 isolates of *M. graminicola* (Mg-Java2 was excluded from this analysis because sequencing depth was too low and some parts were missing). Two open reading frames, encoding GPI ethanolamine phosphate transferase 1 and an unknown protein, were annotated in TAA6, while no gene was detected in ACC6.

Homologous sequences of the ACC6 and TAA6 genomic regions were also found in *M. oryzae* (and their different sequence variants further named Types I, III and IV; see below). Sequence identity between two types of sequences was around 98% [97.81 to 98.14% and 98.07%, for ACC6 and TAA6, respectively (Table S6C-D)]. For TAA6, two homologous sequences were detected with one showing a sequencing depth approximately twice of the other (Table S7). Note that the most abundant type was not the same in Mo-M1 (Type I) as in Mo-

M2 (Type III). In the case of ACC6, two copies were detected in Mo-M1, with Type III twice more abundant than Type I (Table S7), whereas in Mo-M2, three distinct ACC6 copies were assembled, each type with a similar sequencing depth (10.1-11x; Table S7). At each nuclear genomic region, the presence of three copies present in equal abundance or of two copies with one twice as abundant as the other is the expected pattern for a triploid genome. Importantly, in contrast to *M. graminicola* accessions, the genetic profile of the two *M. oryzae* accessions is different (Table S7) indicating genome reshuffling in this species.

For each genomic region, the phylogenetic analysis of the different copies isolated from *M. graminicola* and *M. oryzae* showed that Type I of both species is almost identical (Figure 3C-D; Table S6C-D). On the other hand, Type II is specific to *M. graminicola*, while Types III and IV are only detected in *M. oryzae* isolates (Figure 3D). At the interspecific level, the different types of sequences show an identity comprised between 97.62 and 98.38% (Table S6C-D), so not higher than at the intra-individual level (see above).

IV. Discussion

In the present study, we generated new genomic resources for two species of *Meloidogyne* (*M. graminicola* and *M. oryzae*). Genomes containing distinct nuclear ribosomal DNA units and low-copy regions were reported for the first time in *Meloidogyne* species belonging to Clade III. Our data also provided phylogenetic patterns revealing a close relationship between the two investigated species (with one possibly a parent of the other), as well as evidences for their highly heterozygous genomes and a recent worldwide expansion of *M. graminicola*.

4.1. *Meloidogyne oryzae* and *M. graminicola* are closely related RKN taxa

Sequence analyses of both mitochondrial and nuclear genomes demonstrated that, despite different mode of reproduction, *M. graminicola* and *M. oryzae* are closely related species. In early reports based solely on partial mitochondrial genomic regions [12], *M. graminicola* and *M. oryzae* were already presented as sister species. This was confirmed by our different mitogenomic analyses, where a high identity (96.8%) as well as a highly conserved mitogenome structure were shown between both species (Figure S3). Moreover, analysis of the complete nrDNA cluster together with two low-copy nuclear regions supported not only the existence of two divergent types of sequences in *M. graminicola*, but also at least two or three types in *M. oryzae.* On each of these genomic regions, one type of sequences was shared by *M. graminicola* and *M. oryzae* (Figure 3; Table S6B-C-D) strongly suggesting a closely related ancestor between the two species. Furthermore, this pattern suggests that *M. graminicola* (36 chromosomes) has played the role of parent of *M. oryzae* (54 chromosomes).

4.2. High heterozygosity of *M. graminicola* and *M. oryzae* genomes

With the recent release of RKN nuclear genomes [12,53,54], comparative genomics studies have been done on a few species [12,55], and hybridizations involving whole genome duplications (i.e., allopolyploidization) were underlined in the mitotic parthenogenetic species such as *M. incognita*, *M. arenaria* and *M. javanica* [53,56]. This was also suggested for the obligate meiotic parthenogenetic *M. floridensis* [11] but conversely, the facultative sexual *M. hapla* did not show genome duplication [36,38]. Therefore, allopolyploidization events seem to be associated to apomictic RKNs [11,12,57,58].

Thanks to the presence of two distinct types of homologous sequences (diverging by ca. 2% for low-copy regions) at each nuclear region investigated here, we show strong evidence for a highly heterozygous genome of *M. graminicola*. To date, this has never been reported in facultative meiotic parthenogenetic nematodes. First, two nrDNA types as well as two types of low-copy genomic regions (ACC6 and TAA6) were found in all *M. graminicola* isolates (see above; Figure 3). Secondly, as reported in other species [44,51,59], the *k*-mer distribution in the *M. graminicola* genome (Figure 4) suggests a high proportion of heterozygous sites (>1%). To explain this, homologous regions may harbor two divergent genomic copies that are expected to generate two *k*-mer peaks: a first peak (or hetero peak; here at 16x) and a second one with a two-folds coverage (or homo peak; here at 32x) that respectively result from non-conserved and conserved segments of the genomic region (see [60]). This is important to consider when automatically reconstructing the genome draft sequence of this species [34], because the assembly algorithm may encounter difficulties in regions with a high proportion of heterozygous sites , especially, as here, when using short-read sequences (i.e. HiSeq). In addition, some homologous regions may be separately assembled while others could be merged in a consensus sequence. Thus, this may explain the lower quality assembly of the recently published *M. graminicola* genome [34] than in other *Meloidogyne* spp. genomes (Table S5; [56]). A sequencing strategy based on both long and short reads will be necessary to resolve this issue. Furthermore, we did not detect any polymorphic sites in each homologous sequence of three nuclear regions (nrDNA, TAA6, ACC6) among 11 *M. graminicola* isolates. We also did not detect any recombination events between homologous sequences of the TAA6 and ACC6 regions. The maintenance of the same genetic profile (Table S7) among isolates collected in Asia and America suggests no reassortment of homologous chromosomes during the reproduction since the worldwide spread of *M. graminicola*, probably as a consequence of a preferential asexual reproduction in this species. In addition, our results do not confirm previous works that have reported sequence ITS polymorphisms among *M. graminicola* isolates [22,61].

The observation of two distinct nrDNA types in *M. graminicola* (Figure 3A-B) indicates that this genomic region has to be carefully used when reconstructing phylogenies of *Meloidogyne.* Indeed*,* the coexistence of divergent ribotypes within individuals and chimerical sequences generated by PCR may explain the high ITS variation previously reported [22,61].

Like *M. graminicola,* the apomictic *M. oryzae*, also belongs to Clade III in the nematode classification and shows evidence for a more complex genome (Figure 4). The flattened curve of *k*-mer distribution may reflect the presence of three peaks at ca. 16x, 32x and 48x resulting from the presence of three distinct genomes, due to its triploid status [19]. These results confirm that the *M. oryzae* genome is more complex that the *M. graminicola* genome, due to an additional set of chromosomes (54 vs. 36 chromosomes [8,9]). In addition, two or three types of diverging sequences were detected in the low-copy regions ACC6 and TAA6 (Figure 3C-D; Table S6). Sequencing depth of each sequence type (Table S7) is compatible with the supposed triploid status of *M. oryzae* [8,9], but the distinct genetic profile of the two analyzed accessions is a strong evidence for genome reshuffling in *M. oryzae*. This latter result is surprising since the species is reported as reproducing by mitotic parthenogenesis [9], that should not allow homologous chromosome reassortment. We can thus hypothesize that, following the polyploidy event, at least a few meiosis have been involved in the diversification of this species.

4.3. *M. graminicola* isolates present very low variation and absence of phylogeographic pattern suggesting a recent worldwide expansion

Both nuclear and mitochondrial genomic sequences were analyzed among 12 isolates of *M. graminicola* that were collected from several geographical origins. The mitogenome is known to evolve more rapidly that the nuclear genome in the Nematoda phylum so it can be used to study the populations' history of *Meloidogyne* spp. [20,62]. As mentioned above, no sequence polymorphism was detected within the nuclear genomic regions (25.2 Kb of low-copy regions plus 15.4 Kb of nrDNA regions), but low mitochondrial diversity was also revealed. To explain such a low degree of polymorphism between *M. graminicola* isolates, we propose two hypotheses: the first one is that this species accumulates mutations very slowly due to an uncommon DNA replication apparatus as was suggested for instance in coral [63] and turtle [64]. The second hypothesis is that *M. graminicola* recently spread at worldwide level, as supported by the mitochondrial haplotype network (Figure 2). Indeed, this analysis showed that the most common, ancestral haplotype (from which the other types derived) is shared by distant populations (from Brazil and South East Asia). Interestingly, *M. graminicola* is the only species of Clade III showing a widespread distribution, from America to Asia with recent reports in Madagascar [14] and South Europe [15]. Moreover, this species presents a singular adaptation

to different rice field environments (see map distribution of *Meloidogyne* spp. on https://gd.eppo.int/). We could hypothesize that due to its virulence against rice (*Oryza sativa*), *M. graminicola* was able to rapidly spread over the world due to human activities that allowed long-distance translocations. Considering that rice domestication started between 12,500 and 7,500 BC with three separate diversification centers in Asia [65,66], it is questionable to assert that the pathogen was originally present in today's rice cropping areas. Indeed, if the worldwide distribution of *M. graminicola* was older than rice agricultural development, we could expect a much larger variation in the mitochondrial genome and a phylogeographic structure indicating an ancient expansion of the species. Agricultural practices such as rice transplanting and plantation or banana suckers from infected nurseries seem good ways to spread the pathogen [4,67]. However, the absence of *Meloidogyne* fossils and the lack of evaluation of evolution rates in this species prevent a clear estimation of this event. Resolving the recent expansion of *M. graminicola* populations will probably require an exhaustive nuclear genome analysis, but this will need first to generate a high-quality genome assembly with a clear distinction of homologous sequences that can be used as a reference for detecting SNPs (Phan et al., in progress).

Author Contributions: GB and SB conceived the study and performed lab work. GB, NTP, HHB, HN and SB performed data analyses. PQ provided *M. oryzae* isolates. GB, NTP and SB wrote the paper with the help of all authors.

Funding: This work was funded by the CGIAR Research Program on Rice Agri-food Systems (RICE), the French Ministry of Foreign Affairs, and International Development (grant number #4764 - BIOASIA).

Acknowledgments: GB is member of the Laboratoire Evolution & Diversité Biologique (EDB) part of the LABEX "TULIP" managed by Agence Nationale de la Recherche (ANR-10-LABX-0041) and LABEX "CEBA" (ANR-10-LABX-25-01). The authors are grateful to H. Holota, O. Bouchez (Genopole Toulouse) and K. Lambou (IRIm) for technical assistance, to E.G.J. Danchin for valuable comments, to S. Mantelin for her proof reading and constructive criticisms on an early version of the manuscript, and to colleagues from the "Sunrise" researcher group in SE Asia that provide different RKN isolates: Yuliantoro Baliadi (Indonesian Agency for Agricultural Research and Development, Indonesia), Khun Kim Khuy (Royal University of Agriculture, Cambodia), Malyna Suong & Fidero Kuok (Institute of Technology of Cambodia, Cambodia), Phetsamone Songvilay (Ministry of Agriculture, Forestry and Fisheries, Lao), Dirk

De Waele (University of Leuven, Belgium), Xuyen Ngo Thi (Van Lang University, Vietnam) and Ha Viet Cuong (Vietnam National University of Agriculture, Vietnam), Ma. Teodora Nadong Cabasan (University of Southern Mindanao, Philippines).

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Wang, D.Y.; Kumar, S.; Hedges, S.B. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. R. Soc. B Biol. Sci.* **1999**, *266*, 163–171. doi:10.1098/rspb.1999.0617.

[2] Lorenzen, S. The phylogenetic systematics of free-living nematodes. The Ray Society, London; Andover, UK; **1994;** 383 pages.

[3] Jones, J.T.; Haegeman, A.; Danchin, E.G.J.; Gaur, H.S.; Helder, J.; Jones, M.G.K.; Kikuchi, T.; Manzanilla-López, R.; Palomares-Rius, J.E.; Wesemael, W.M.L.; Perry, R.N. Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol. Plant Pathol.* **2013**, *14*, 946–961. doi:10.1111/mpp.12057.

[4] Mantelin, S.; Bellafiore, S.; Kyndt, T. *Meloidogyne graminicola*: a major threat to rice agriculture. *Mol. Plant Pathol.* **2017**, *18*, 3–15. doi:10.1111/mpp.12394.

[5] Poinar, G.; Kerp, H.; Hass, H. *Palaeonemaphyticum* gen. n.; sp. n. (Nematoda: Palaeonematidae fam. n.), a Devonian nematode associated with early land plants. *Nematology* **2008**, *10*, 9–14. doi:10.1163/156854108783360159.

[6] Lapp, N.A.; Triantaphyllou, A.C. Relative DNA content and chromosomal relationships of some *Meloidogyne*, *Heterodera*, and *Meloidodera* spp. (Nematoda: Heteroderidae). *J. Nematol.* **1972**, *4*, 287–291. PMCID: PMC2619951.

[7] Moens, M.; Perry, R.N.; Starr, J.L. *Meloidogyne* species - a diverse group of novel and important plant parasites. In *Root-knot nematodes*; Perry, R.N.; Moens, M.; Starr, J.L. Eds.; CABI: Wallingford, **2009**; pp. 1–17.

[8] Triantaphyllou, A.C. Gametogenesis and the chromosomes of two root-knot nematodes, *Meloidogyne graminicola* and *M. naasi*. *J. Nematol*. **1969**, *1*, 62–71. PMCID: PMC2617796.

[9] Triantaphyllou, A.C. Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes. In *An advanced treatise on* Meloidogyne. *Biology and Control*; Sasser, J.N.; Carter, C.C. Eds.; Raleigh: North Carolina State University, **1985**; Volume 1, pp. 113–126.

[10] Bird, D.M.; Williamson, V.M.; Abad, P.; McCarter, J.; Danchin, E.G.J.; Castagnone-Sereno, P.; Opperman, C.H. The genomes of root-knot nematodes. *Annu. Rev. Phytopathol.* **2009**, *47*, 333–351. doi:10.1146/annurev-phyto-080508-081839.

[11] Castagnone-Sereno, P.; Danchin, E.G.J.; Perfus-Barbeoch, L.; Abad, P. Diversity and evolution of root-knot nematodes, genus *Meloidogyne*: new insights from the genomic era. *Annu. Rev. Phytopathol.* **2013**, *51*, 203–220. doi:10.1146/annurev-phyto-082712-102300.

[12] Lunt, D.H.; Kumar, S.; Koutsovoulos, G.; Blaxter, M.L. The complex hybrid origins of the root knot nematodes revealed through comparative genomics. *PeerJ* **2014**, *2*, e356. doi:10.7717/peerj.356.

[13] Tigano, M.S.; Carneiro, R.M.D.G.; Jeyaprakash, A.; Dickson, D.W.; Adams, B.J. Phylogeny of *Meloidogyne* spp. based on 18S rDNA and the intergenic region of mitochondrial DNA sequences. *Nematology* **2005**, *7*, 851–862. doi:10.1163/156854105776186325.

[14] Chapuis, E.; Besnard, G.; Andrianasetra, S.; Rakotomalala, M.; Nguyen, H.T.; Bellafiore, S. First report of the root-knot nematode *Meloidogyne graminicola* in Madagascar rice fields. *Australas. Plant Dis. Notes* **2016**, *11*, 32. doi:10.1007/s13314-016-0222-5.

[15] Fanelli, E.; Cotroneo, A.; Carisio, L.; Troccoli, A.; Grosso, S.; Boero, C.; Capriglia, F.; Luca, F.D. Detection and molecular characterization of the rice root-knot nematode *Meloidogyne graminicola* in Italy. *Eur. J. Plant Pathol.* **2017**, *149*, 467–476. doi:10.1007/s10658-017-1196-7.

[16] Maas, P.W.T.; Sanders, H.; Dede, J. *Meloidogyne oryzae* n. sp. (Nematoda, Meloidogynidae) infesting irrigated rice in Surinam (South America). *Nematologica* **1978**, *24*, 305–311. doi:10.1163/187529278X00272.

[17] da Mattos, V.S.; Cares, J.E.; Gomes, C.B.; Mendes Gomes, A.C.M.; da Mata, dos Santos, M.J.; Gomez, G.M.; Castagnone-Sereno, P.; Gomez Carneiro, R.M.D. Integrative taxonomy of *Meloidogyne oryzae* (Nematoda: Meloidogyninae) parasitizing rice crops in Southern Brazil. *Eur. J. Plant Pathol.* **2017**, *151*, 649–662. doi:10.1007/s10658-017-1400-9.

[18] Golden, A.M.; Birchfield, W. *Meloidogyne graminicola* (Heteroderidae), a new species of root-knot nematode from grass. *Proc. Helminthol. Soc. Wash.* **1965**, *32*, 228–231. Available online: http://bionames.org/bionames-archive/issn/0018-0130/32/228.pdf.

[19] Esbenshade, P.R.; Triantaphyllou, A.C. Use of enzyme phenotypes for identification of *Meloidogyne* species. *J. Nematol.* **1985**, *17*, 6–20. PMCID: PMC2618420.

[20] Humphreys-Pereira, D.A.; Elling, A.A. Intraspecific variability and genetic structure in *Meloidogyne chitwoodi* from the USA. *Nematology* **2013**, *15*, 315–327. doi:10.1163/15685411-00002684

[21] Humphreys-Pereira, D.A.; Elling, A.A. Mitochondrial genomes of *Meloidogyne chitwoodi* and *M. incognita* (Nematoda: Tylenchina): comparative analysis, gene order and phylogenetic relationships with other nematodes. *Mol. Biochem. Parasitol.* **2014**, *194*, 20–32.doi:10.1016/j.molbiopara.2014.04.003.

[22] Bellafiore, S.; Jougla, C.; Chapuis, É.; Besnard, G.; Suong, M.; Vu, P.N.; De Waele, D.; Gantet, P.; Thi, X.N. Intraspecific variability of the facultative meiotic parthenogenetic root-knot nematode (*Meloidogyne graminicola*) from rice fields in Vietnam. *C. R. Biol*. **2015**, *338*, 471–483. doi:10.1016/j.crvi.2015.04.002.

[23] Schaad, N.W.; Walker, J.T. The use of density-gradient centrifugation for the purification of eggs of *Meloidogyne* spp. *J. Nematol*. **1975**, *7*, 203–204. PMCID: PMC2620088.

[24] Besnard, G.; Jühling, F.; Chapuis, É.; Zedane, L.; Lhuillier, É.; Mateille, T.; Bellafiore, S. Fast assembly of the mitochondrial genome of a plant parasitic nematode (*Meloidogyne graminicola*) using next generation sequencing. *C. R. Biol*. **2014**, *337*, 295–301. doi:10.1016/j.crvi.2014.03.003.

[25] Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **2004**, *32*, 1792–1797. doi:10.1093/nar/gkh340.

[26] Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. doi:10.1093/molbev/msw054.

[27] Silvestro, D.; Michalak, I. raxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol*. **2012**, *12*, 335–337. doi:10.1007/s13127-011-0056-0.

[28] Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. doi:10.1038/nmeth.2109.

[29] Guindon, S.; Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol*. **2003**, *52*, 696–704. doi:10.1080/10635150390235520.

[30] Wang, G.F.; Xiao, L.Y.; Luo, H.G.; Peng, D.L.; Xiao, Y.N. First report of *Meloidogyne graminicola* on rice in Hubei province of China. *Plant Dis.* **2017**, *101*, 1056. doi:10.1094/PDIS-12-16-1805-PDN.

[31] Bandelt, H.J.; Forster, P.; Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. doi:10.1093/oxfordjournals.molbev.a026036.

[32] Straub, S.C.K.; Parks, M.; Weitemier, K.; Fishbein, M.; Cronn, R.C.; Liston, A. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Am. J. Bot.* **2012**, *99*, 349–364. doi:10.3732/ajb.1100335.

[33] Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. doi:10.1093/oxfordjournals.molbev.a026334.

[34] Somvanshi, V.S.; Tathode, M.; Shukla, R.N.; Rao, U. Nematode genome announcement: A draft genome for rice root-knot nematode, *Meloidogyne graminicola*. *J. Nematol*. **2018**, *50*, 111–116. doi:10.21307/jofnem-2018-018.

[35] Petitot, A.S.; Dereeper, A.; Agbessi, M.; da Silva, C.; Guy, J.; Ardisson, M.; Fernandez, D. Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants. *Mol. Plant Pathol.* **2016**, *17*, 860–874. doi:10.1111/mpp.12334.

[36] Parra, G.,Bradnam, K., Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **2007**, 23, 1061–1067. doi: 10.1093/bioinformatics/btm071.

[37] Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015,** *31*, 3210–3212. doi:10.1093/bioinformatics/btv351.

[38] Andrews, S. FastQC: A quality control tool for high throughput sequence data. **2010**. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

[39] Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. doi:10.1093/bioinformatics/btu170.

[40] Xu, H.; Luo, X.; Qian, J.; Pang, X.; Song, J.; Qian, G.; Chen, J.; Chen, S. FastUniq: A fast de novo duplicates removal tool for paired short reads. *PLoS ONE* **2012**, *7*, e52249. doi:10.1371/journal.pone.0052249.

[41] Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **2011**, *27*, 764–770. doi:10.1093/bioinformatics/btr011.

[42] Li, R.; Zhu, H.; Ruan, J.; Qian, W.; Fang, X.; Shi, Z.; Li, Y.; Li, S.; Shan, G.; Kristiansen, K.; Li, S.; Yang, H.; Wang, J.; Wang, J. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res*. **2010**, *20*, 265–272. doi:10.1101/gr.097261.109.

[43] Liu, S.; Zheng, J.; Migeon, P.; Ren, J.; Hu, Y.; He, C.; Liu, H.; Fu, J.; White, F.F.; Toomajian, C.; Wang, G. Unbiased *k*-mer analysis reveals changes in copy number of highly repetitive sequences during maize domestication and improvement. *Sci. Rep*. **2017**, *7*, 42444. doi:10.1038/srep42444.

[44] Kajitani, R.; Toshimoto, K.; Noguchi, H.; Toyoda, A.; Ogura, Y.; Okuno, M.; Yabana, M.; Harada, M.; Nagayasu, E.; Maruyama, H.; Kohara, Y.; Fujiyama, A.; Hayashi, T.; Itoh, T. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res*. **2014**, *24*, 1384–1395. doi:10.1101/gr.170720.113.

[45] Humphreys-Pereira, D.A.; Elling, A.A. Mitochondrial genome plasticity among species of the nematode genus *Meloidogyne* (Nematoda: Tylenchina). *Gene* **2015**, *560*, 173–183. doi:10.1016/j.gene.2015.01.065.

[46] Janssen, T.; Karssen, G.; Verhaeven, M.; Coyne, D.; Bert, W. Mitochondrial coding genome analysis of tropical root-knot nematodes (*Meloidogyne*) supports haplotype based diagnostics and reveals evidence of recent reticulate evolution. *Sci. Rep.* **2016**, *6*, 22591. doi:10.1038/srep22591.

[47] Ali, N.; Tavoillot, J.; Mateille, T.; Chapuis, É.; Besnard, G.; El Bakkali, A.; Cantalapiedra-Navarrete, C.; Liébanas, G.; Castillo, P.; Palomares-Rius, J.E. A new root-knot nematode *Meloidogyne spartelensis* n. sp. (Nematoda: Meloidogynidae) in Northern Morocco. *Eur. J. Plant Pathol.* **2015**, *143*, 25–42. doi:10.1007/s10658-015-0662-3.

[48] McClure, M.A.; Nischwitz, C.; Skantar, A.M.; Schmitt, M.E.; Subbotin, S.A. Root-knot nematodes in golf course greens of the Western United States. *Plant Dis.* **2011**, *96*, 635–647. doi:10.1094/PDIS-09-11-0808.

[49] Ye, W.; Zeng, Y.; Kerns, J. Molecular characterisation and diagnosis of root-knot nematodes (*Meloidogyne* spp.) from Turfgrasses in North Carolina, USA. *PLoS ONE* **2015**, *10*, e0143556. doi:10.1371/journal.pone.0143556.

[50] Lander, E.S.; Waterman, M.S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **1988**, *2*, 231–239. doi:10.1016/0888-7543(88)90007-9.

[51] Teh, B.T.; Lim, K.; Yong, C.H.; Ng, C.C.Y.; Rao, S.R.; Rajasegaran, V.; Lim, W.K.; Ong, C.K.; Chan, K.; Cheng, V.K.Y.; Soh, P.S.; Swarup, S.; Rozen, S.G.; Nagarajan, N.; Tan, P. The draft genome of tropical fruit durian (*Duriozibethinus*). *Nat. Genet.* **2017**, *49*, 1633–1641. doi:10.1038/ng.3972.

[52] Pableo, E.C.; Triantaphyllou, A.C. DNA complexity of the root-knot nematode (*Meloidogyne* spp.) genome. *J. Nematol.* **1989**, 21, 260–263. PMCID: PMC2618911.

[53] Abad, P.; Gouzy, J.; Aury, J.-M.; Castagnone-Sereno, P.; Danchin, E.G.J.; Deleury, E.; Perfus-Barbeoch, L.; Anthouard, V.; Artiguenave, F.; Blok, V.C.; Caillaud, M.-C.; Coutinho, P.M.; Dasilva, C.; De Luca, F.; Deau, F.; Esquibet, M.; Flutre, T.; Goldstone, J.V.; Hamamouch, N.; Hewezi, T.; Jaillon, O.; Jubin, C.; Leonetti, P.; Magliano, M.; Maier, T.R.; Markov, G.V.; McVeigh, P.; Pesole, G.; Poulain, J.; Robinson-Rechavi, M.; Sallet, E.; Ségurens, B.; Steinbach, D.; Tytgat, T.; Ugarte, E.; van Ghelder, C.; Veronico, P.; Baum, T.J.; Blaxter, M.; Bleve-Zacheo, T.; Davis, E.L.; Ewbank, J.J.; Favery, B.; Grenier, E.; Henrissat, B.; Jones, J.T.; Laudet, V.;

Maule, A.G.; Quesneville, H.; Rosso, M.-N.; Schiex, T.; Smant, G.; Weissenbach, J.; Wincker, P. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* **2008**, *26*, 909–915. doi:10.1038/nbt.1482.

[54] Opperman, C.H.; Bird, D.M.; Williamson, V.M.; Rokhsar, D.S.; Burke, M.; Cohn, J.; Cromer, J.; Diener, S.; Gajan, J.; Graham, S.; Houfek, T.D.; Liu, Q.; Mitros, T.; Schaff, J.; Schaffer, R.; Scholl, E.; Sosinski, B.R.; Thomas, V.P.; Windham, E. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14802–14807. doi:10.1073/pnas.0805946105.

[55] Szitenberg, A.; Salazar-Jaramillo, L.; Blok, V.C.; Laetsch, D.R.; Joseph, S.; Williamson, V.M.; Blaxter, M.L.; Lunt, D.H. Comparative genomics of apomictic root-knot nematodes: Hybridization, ploidy, and dynamic genome change. *Genome Biol. Evol.* **2017**, *9*, 2844–2861. doi:10.1093/gbe/evx201.

[56] Blanc-Mathieu, R.; Perfus-Barbeoch, L.; Aury, J.-M.; Rocha, M.D.; Gouzy, J.; Sallet, E.; Martin-Jimenez, C.; Bailly-Bechet, M.; Castagnone-Sereno, P.; Flot, J.-F.; Kozlowski, D.K.; Cazareth, J.; Couloux, A.; Silva, C.D.; Guy, J.; Kim-Jo, Y.-J.; Rancurel, C.; Schiex, T.; Abad, P.; Wincker, P.; Danchin, E.G.J. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genet*. **2017**, *13*, e1006777. doi:10.1371/journal.pgen.1006777.

[57] Hugall, A.; Stanton, J.; Moritz, C. Reticulate evolution and the origins of ribosomal internal transcribed spacer diversity in apomictic *Meloidogyne*. *Mol. Biol. Evol*. **1999**, *16*, 157–164. doi:10.1093/oxfordjournals.molbev.a026098.

[58] Lunt, D.H. Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. *BMC Evol. Biol*. **2008**, *8*, 194. doi:10.1186/1471-2148-8-194.

[59] Ming, R.; VanBuren, R.; Wai, C.M.; Tang, H.; Schatz, M.C.; Bowers, J.E.; Lyons, E.; Wang, M.-L.; Chen, J.; Biggers, E.; Zhang, Jisen, Huang, L.; Zhang, L.; Miao, W.; Zhang, Jian, Ye, Z.; Miao, C.; Lin, Z.; Wang, H.; Zhou, H.; Yim, W.C.; Priest, H.D.; Zheng, C.; Woodhouse, M.; Edger, P.P.; Guyot, R.; Guo, H.-B.; Guo, H.; Zheng, G.; Singh, R.; Sharma, A.; Min, X.; Zheng, Y.; Lee, H.; Gurtowski, J.; Sedlazeck, F.J.; Harkess, A.; McKain, M.R.; Liao, Z.; Fang, J.; Liu, J.; Zhang, X.; Zhang, Q.; Hu, W.; Qin, Y.; Wang, K.; Chen, L.-Y.; Shirley, N.; Lin, Y.-R.; Liu, L.-Y.; Hernandez, A.G.; Wright, C.L.; Bulone, V.; Tuskan, G.A.; Heath, K.; Zee, F.; Moore, P.H.; Sunkar, R.; Leebens-Mack, J.H.; Mockler, T.; Bennetzen, J.L.; Freeling, M.; Sankoff, D.; Paterson, A.H.; Zhu, X.; Yang, X.; Smith, J.A.C.; Cushman, J.C.; Paull, R.E.; Yu, Q. The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **2015**, *47*, 1435–1442. doi:10.1038/ng.3435.

[60] Evans, B.J.; Upham, N.S.; Golding, G.B.; Ojeda, R.A.; Ojeda, A.A. Evolution of the largest mammalian genome. *Genome Biol. Evol.* **2017**, *9*, 1711–1724. doi:10.1093/gbe/evx113.

[61] Pokharel, R.R.; Abawi, G.S.; Duxbury, J.M.; Smat, C.D.; Wang, X.; Brito, J.A. Variability and the recognition of two races in *Meloidogyne graminicola. Australas. Plant Pathol.* **2010**, *39*, 326–333. doi:10.1071/AP09100.

[62] Whipple, L.E.; Lunt, D.H.; Hyman, B.C. Mitochondrial DNA length variation in *Meloidogyne incognita* isolates of established genetic relationships: utility for nematode population studies. *Fundam. Appl. Nematol.* **1998**, *21*, 265–271.

[63] Hellberg, M.E. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evol. Biol*. **2006**, *6*, 24. doi:10.1186/1471-2148-6-24.

[64] Avise, J.C.; Bowen, B.W.; Lamb, T.; Meylan, A.B.; Bermingham, E. Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. *Mol. Biol. Evol.* **1992**, *9*, 457–473. doi:10.1093/oxfordjournals.molbev.a040735.

[65] Civáň, P.; Craig, H.; Cox, C.J.; Brown, T.A. Three geographically separate domestications of Asian rice. *Nat. Plants* **2015**, *1*, 15164. doi:10.1038/nplants.2015.164.

[66] Huang, X.; Kurata, N.; Wei, X.; Wang, Z.-X.; Wang, A.; Zhao, Q.; Zhao, Y.; Liu, K.; Lu, H.; Li, W.; Guo, Y.; Lu, Y.; Zhou, C.; Fan, D.; Weng, Q.; Zhu, C.; Huang, T.; Zhang, L.; Wang, Y.; Feng, L.; Furuumi, H.; Kubo, T.; Miyabayashi, T.; Yuan, X.; Xu, Q.; Dong, G.; Zhan, Q.; Li, C.; Fujiyama, A.; Toyoda, A.; Lu, T.; Feng, Q.; Qian, Q.; Li, J.; Han, B. A map of rice genome

variation reveals the origin of cultivated rice. *Nature* **2012**, *490*, 497–501. doi:10.1038/nature11532.

[67] Zhou, X.; Liu, G.K.; Xiao, S.; Zhang, S.S. First report of *Meloidogyne graminicola* infecting banana in China. *Plant Dis*. **2015**, *99*, 420–421. doi:10.1094/PDIS-08-14-0810-PDN.

Supplementary Information

On the relatedness of two rice-parasitic root-knot nematode species and the recent expansion of *Meloidogyne graminicola* in Southeast Asia

Guillaume Besnard, Ngan Thi Phan, Hai Ho Bich, Alexis Dereeper, Hieu Trang Nguyen, Patrick Quénéhervé, Jamel Aribi and Stéphane Bellafiore

Supplementary information includes the following items:

Supplementary methods

Supplementary methods 1. Reference genomes used to screen contaminants

Reference genomes used to screen contaminants include (1) NCBI RefSeq bacterial genomes (70,293 entries as of 8 December 2016), (2) NCBI GenBank fungal genomes (2,314 entries as of 4 May 2017), rice genomes (IRGSP-1.0_genome and nippon_ir64 from Schatz laboratory), human genome (GRCh38/hg38) and the mitogenomes presented in this work.

Supplementary methods 2. Estimation of genome size and depth of sequencing

k-mers are nucleotide sequences that can be extracted directly from sequencing reads by a sliding window of length *k*. For novel genome, *k*-mer analysis can reveal some insights in terms of structure and complexity. Hereafter, read length is denoted as *L*, total number of reads as *N*, haploid genome size as *G*, sequencing depth as *S*, and *D* as average homozygous coverage. We excluded *k*-mers with coverage out of range 5 to 1,000, denoted as *B*. Based on the hypothesized ploidy, *D* is determined from the average unique heterozygous λ. In our case, if *M. graminicola* is deemed as diploid and *M. oryzae* as triploid, the average homozygous coverage is 2λ and 3λ, respectively [1]. Repetitive homozygous *k*-mers, used to calculate repeat proportion, have coverage of more than 6 * λ. Considering heterozygosity rate as the probability of a heterozygous nucleotide, we can indirectly estimate it via the number of unique heterozygous *k*-mers. *G* and *S* are calculated as follows:

$$G = \frac{N * (L - K + 1) - B}{D}$$

$$S = \boldsymbol{\lambda} * \frac{L}{L - K + 1}$$

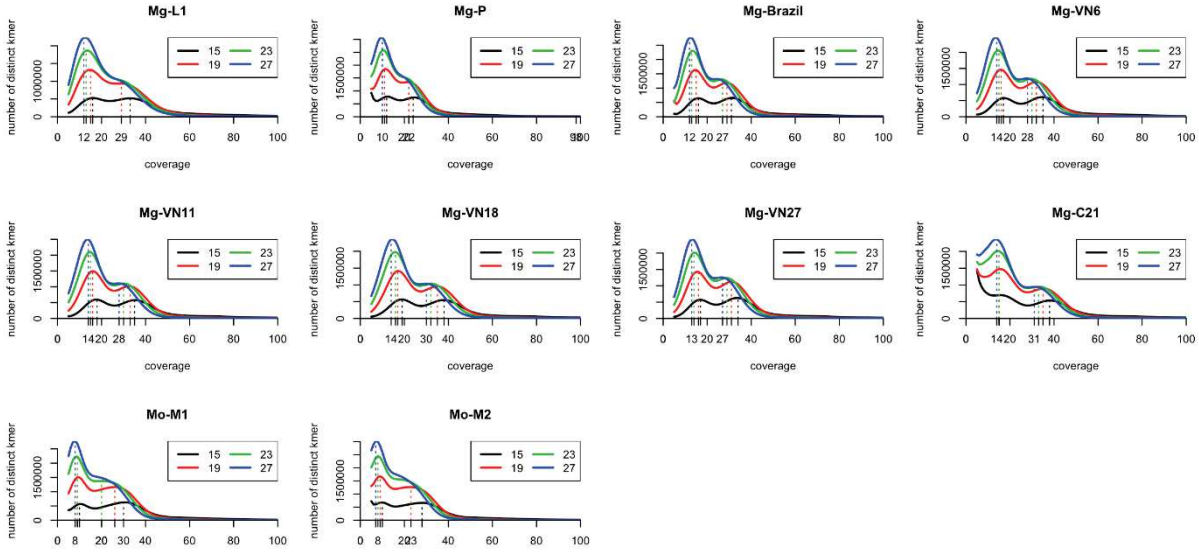


Figure S1. *k*-mers distribution of the *M. graminicola* and *M. oryzae* genomes at *k* = 15, 19, 23 and 27. Histogram curves for *k* = 27 (blue curve), *k* = 23 (green curve), *k* =19 (red curve) and *k* = 15 (black curve).

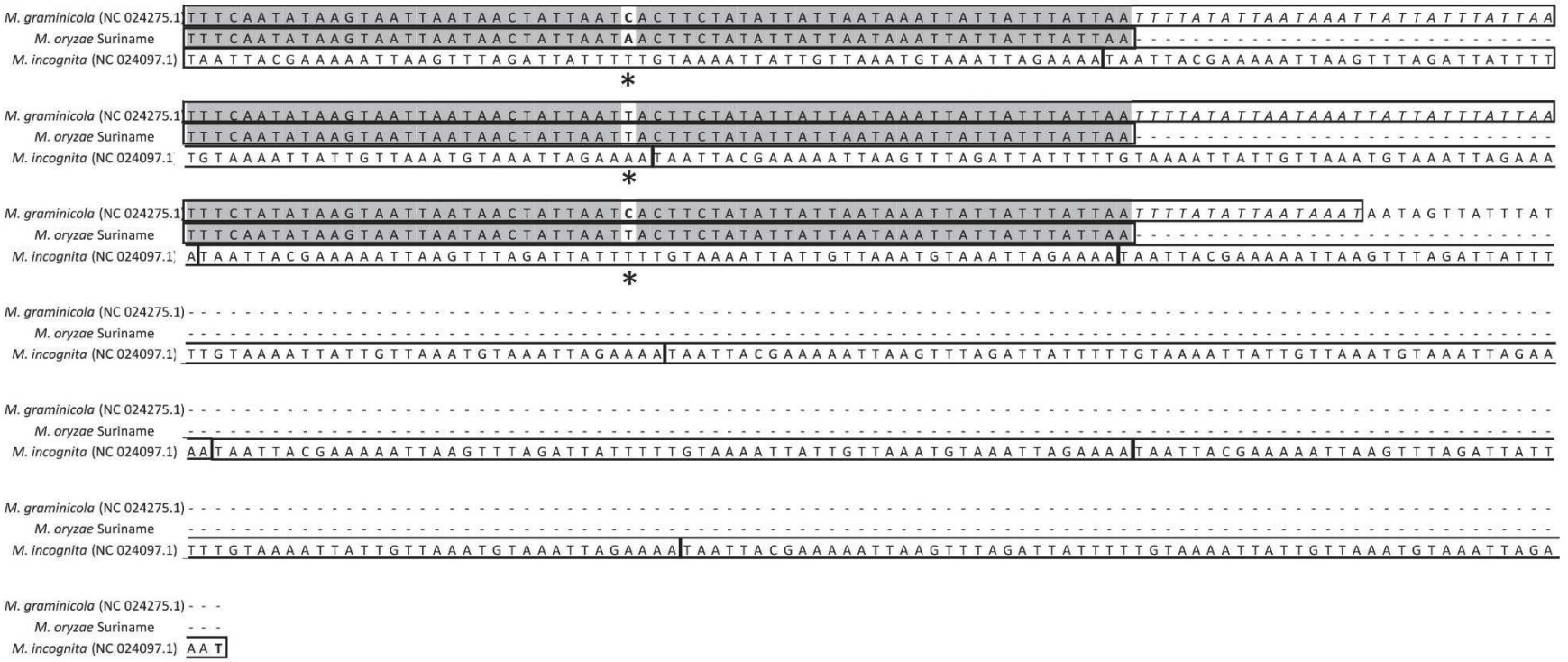


Figure S2. **Comparative alignment of repeated elements composing the 65R/94R/63R mitogenomic region of *M. oryzae* (*Mo*, this study), *M. graminicola* (*Mg*, NC_024275.1) and *M. incognita* (NC 024097.1), respectively**. The boxes indicate the repeated elements in each species. Three repeated 65-bp elements in *M. oryzae* overlap with 94R region in *M. graminicola* which includes two repeated 94-bp elements and one 81-bp element. At position 31 of the 94-bp element, a variable site (C, A or T) was found between repeats for all *Mg* and *Mo* populations. For the three repeats, *Mg* got respectively the nucleotides C-T-C and Mo A-T-T. Therefore at position 31 of the first and third repeats, a specific nucleotide is found for *Mg* and *Mo.* The variable site at position 31 of the 94-bp and 65-bp element is indicated in bold and with a * below the position. The repeated elements of *M. incognita* display very different sequence (that cannot be properly aligned) compared with the two other species.

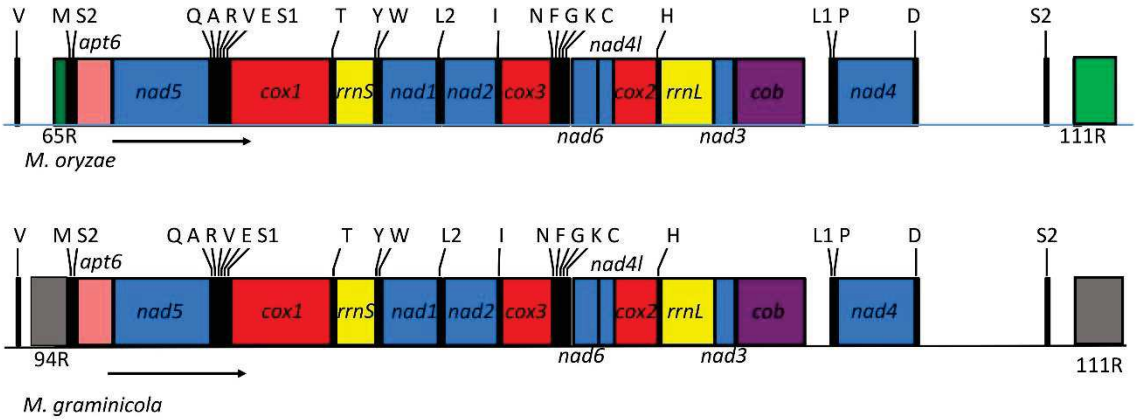


Figure S3. **Linearized representation of mitogenomes of *M. oryzae* (this study) and *M. graminicola* (NC_024275.1)**. Gene and genome size are not drawn to scale. The arrows indicate the direction of the transcription for all genes. The tRNAs are shown by single-letter abbreviation (on the tick marks; see also Besnard et al. [2]).

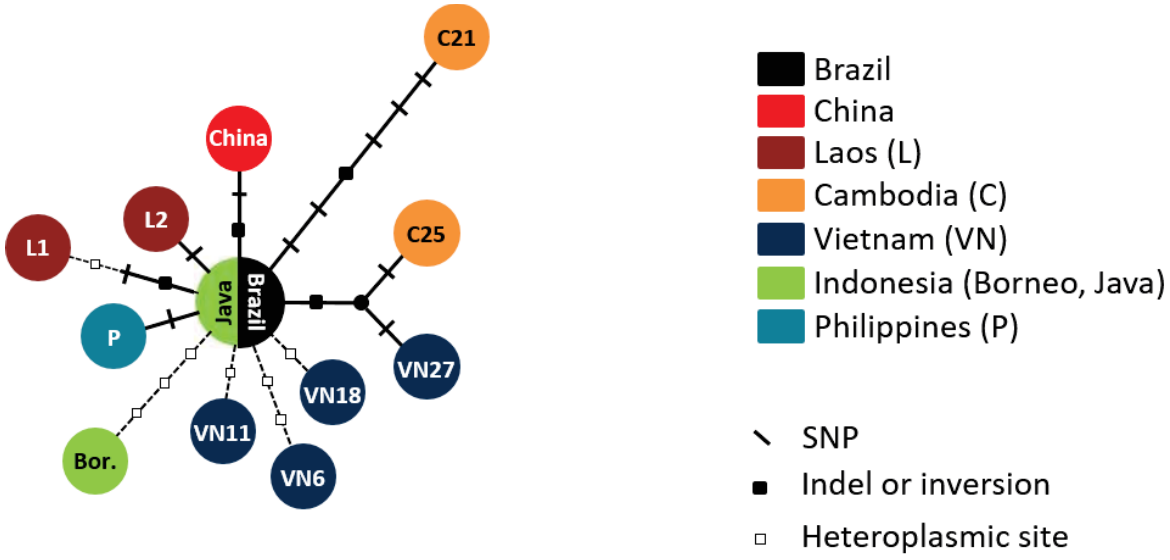


Figure S4. **Reduced-median network of *Meloidogyne graminicola* mitochondrial haplotypes including the information of heteroplasmic sites**. The network was reconstructed with Network v.5 [3], using the 13 available nematode mitogenome sequences, excluding the 111R region. The nematode populations are indicated in the circles and their geographic origin is displayed by different colors. The number of mutations is shown on the branches with slashes, black squares and white squares that respectively indicate SNPs, indels/inversion and heteroplasmic sites.

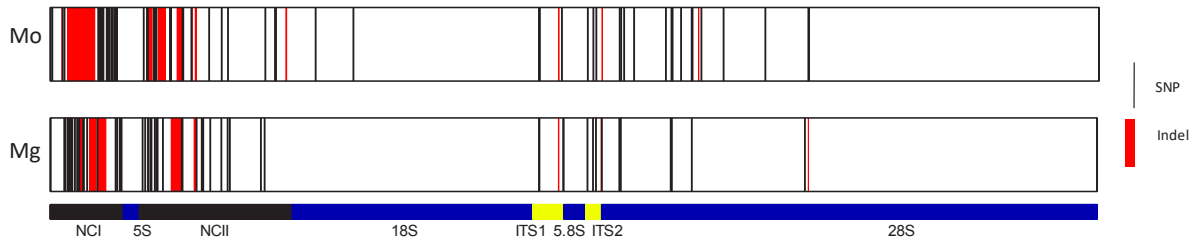


Figure S5. Distribution of polymorphisms between ribotype I and II in the whole nrDNA cluster of *M. graminicola* and *M. oryzae*

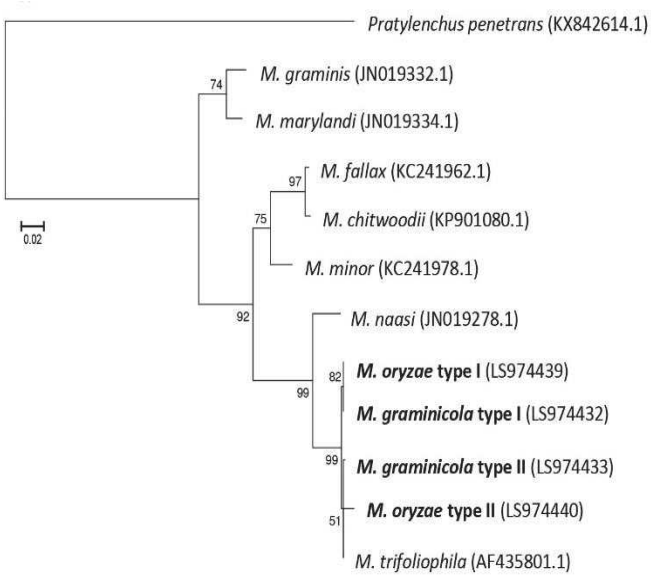


Figure S6. Phylogenetic relationships within *Meloidogyne* species based on ITS using the GTR + G model. Bootstrap values greater than 50% are given on appropriate clades in the Maximum Likelihood tree presented. GenBank sequences are indicated in parenthesis, and sequences from this study are in bold font. The phylogeny is rooted with *Pratylenchus penetrans* as an outgroup. Scale bar represents 0.02 substitutions per nucleotide position.

Table S1. Origin and genome sequencing data summary generated in this study for each nematode isolate

Isolates	Abbreviation	Origin	Number of paired-end reads	Number of cleaned paired-end reads	Read length (bp)	MtDNA sequence depth*	Mitogenome size without 111R (bp)
Meloidogyne graminicola VN6	Mg-VN6	Vietnam	8,187,318	7,591,266	100	1,632× ± 390	`16,809
Meloidogyne graminicola VN11	Mg-VN11	Vietnam	8,613,889	7,873,320	100	1,885× ± 477	16,810
Meloidogyne graminicola VN18	Mg-VN18	Vietnam	9,191,522	8,397,190	100	1,951× ± 513	16,810
Meloidogyne graminicola VN27	Mg-VN27	Vietnam	8,149,454	7,568,959	100	1,742× ± 370	16,905
Meloidogyne graminicola L1	Mg-L1	Laos	6,903,524	6,367,740	125	3,772× ± 764	16,810
Meloidogyne graminicola L2	Mg-L2	Laos	7,639,080	NA	125	197× ± 93	16,811
Meloidogyne graminicola C21	Mg-C21	Cambodia	8,728,853	7,295,797	125	1,571× ± 533	16,813
Meloidogyne graminicola C25	Mg-C25	Cambodia	13,173,663	NA	125	462× ± 211	16,905
Meloidogyne graminicola Java2	Mg-Java	Indonesia	13,657,194	NA	125	74× ± 34	16,811
Meloidogyne graminicola Borneo	Mg-Borneo	Indonesia	8,448,369	NA	125	836× ± 95	16,808
Meloidogyne graminicola P	Mg-P	Philippines	7,762,138	6,146,083	100	1,506× ± 355	16,811
Meloidogyne graminicola Brazil	Mg-Brazil	Brazil	8,440,215	7,347,384	100	1,832× ± 389	16,811
Meloidogyne oryzae M1	Mo-M1	French Guiana	8,559,374	7,900,790	100	794× ± 207	17,069
Meloidogyne oryzae M2	Mo-M2	Suriname	8,316,341	7,662,197	100	1,363× ± 298	17,066

*sequencing depth for the mitogenome assembly with standard deviation

Table S2. **Accession numbers of mitochondrial genes of *M. hapla* and *M. floridensis***

Genes	*M. hapla*[1]	*M. floridensis*[2]
cox1	ABLG01002664.1	nMf.1.0.scaf04464
rrnS	ABLG01002664.1	nMf.1.0.scaf04464
cox3	BM884076.1	nMf.1.0.scaf14978
nad4L	-	nMf.1.0.scaf14978
nad3	L76262.1	nMf.1.0.scaf13075
nad4	-	nMf.1.0.scaf13075
nad5	ABLG01002664.1	nMf.1.0.scaf04464

[1]For *M. hapla*, the accession number of genes/contigs were retrieved from GenBank using Nematode BLAST Server (http://xyala.cap.ed.ac.uk/services/blastserver/)
[2]For *M. floridensis*, the scaffolds containing mitochondrial genes were searched using BLAST in the 959 nematode genome database (http://xyala.cap.ed.ac.uk/downloads/959-nematodegenomes/blast/blast.php)

Table S3. **Diversity of mitochondrial genomes (excluding 111R) between and within the *Meloidogyne* species**

Species	Number of SNPs	Number of indels and inversion
M. graminicola (Mg-P) vs. *M. oryzae* (Mo-M2)	619	58
Between two *M. oryzae* isolates	6	5
Among 13 *M. graminicola* isolates*	11	4

* including a Chinese isolate (KJ139963)

Table S4. **Position of SNPs, indels and heteroplasmic sites within the *M. graminicola* mitogenome among 13 isolates**

	Position site (location)*																					
Isolate	15743	16549	16724	16726	16831-16833	17099	17325	17486	17512	17774	17883	18163	19441	789	912	1245	2386	4144	6728	8632	9224	10642	11809
	(NCR)	(NCR)	(NCR)	(NCR)	(NCR)	(NCR)	(NCR)	(NCR)	(NCR)	(*atp6*)	(*atp6*)	(*atp6*)	(*nad5*)	(*cox1*)	(*cox1*)	(*cox1*)	(*nad1*)	(*cox3*)	(*rrnL*)	(*cob*)	(*nad4*)	(NCR)	(NCR)
Mg-VN6										R (A76; G24)													R (81A; 19G)
Mg-VN11																						Y (T66; C34)	
Mg-VN18																R (74G; 26A)							
Mg-VN27		T					Indel 94R																
Mg-L1					inversion TCC										R (57A; 43G)				C				
Mg-L2				A																			
Mg-C21	indel AT							A	T		T							G			G		
Mg-C25							Indel 94R										G						
Mg-Borneo			K (G53; T47)			K (G54; T46)														Y (T58; C42)			
Mg-Java																							
Mg-P (NC_024275)														C									
Mg-Brazil																							
Mg-China (KJ139963.1)												Indel TA	T										
Consensus sequence	-	C	T	C	GGA	G	-	T	C	G	C	-	A	T	A	A	A	A	T	T	A	C	A
Silent/non silent mutation**	-	-	-	-	-	-	-	-	-	silent	silent	Phe->Leu2	Leu2 -> Phe	Pro->Ser2	Asn->Asp	Ser->Gly	Thr->Ala	Asn->Ser1	-	Phe->Ser2	silent	-	-

The 111R from each population was removed before to proceed to alignment. Therefore, positions 1 and 16,910 of the proposed alignment correspond respectively to the nucleotide number 15,668 and 12,449 from the mitochondrial circular reference genome (NC_024275.1 [2]). The list was established without consideration of variation in mononucleotide stretches (poly A and poly T). The percent of each nucleotide variant for a heteroplasmic position is given in parenthesis.

* in *M. graminicola*-NC_024275

** for protein-coding gene only

Table S5. **Assessment of the assembly completeness of available *Meloidogyne graminicola* genome and transcriptome.** This analysis was performed with BUSCO v.3 (Benchmarking Universal Single-Copy Orthologs [4]). The BUSCO dataset "Eukaryota *odb9*", which includes 303 Eukaryote single-copy orthologs, was used as the reference.

	Draft genome [5]	Transcriptome [6]
Assembly size (Mb)	38.18	61.08
Number of scaffolds/transcripts	4,304	66,396
N50 value (kb)	20.4	0.4
Complete (%)	73.6	88.1
Complete and single-copy (%)	72.6	35.0
Complete and duplicated (%)	1.0	53.1
Fragmented (%)	15.2	7.6
Missing (%)	11.2	4.3

Table S6. **Sequence identity (1 – *p*-distance, in %) between each type of homolog for the three nuclear regions investigated**: **A**) Number of variations (SNPs and Indels) between type I and type II across nrDNA regions, B) nuclear ribosomal DNA cluster or nrDNA (nucleotide alignment of 7,932 bp), **C**) TAA6 region (nucleotide alignment of 6,546 bp), and **D**) ACC6 region (nucleotide alignment of 6,473 bp). *Mg* = *Meloidogyne graminicola*; *Mo* = *M. oryzae*. Intraspecific sequence comparisons are given in bold and italic, for *Mg* and *Mo*, respectively. Interspecific comparisons are in black

A)

nrDNA regions	*Mg*		*Mo*	
	SNPs	Indels	SNPs	Indels
NCI	30	4	20	3
5S	0	0	0	0
NCII	22	3	24	6
18S	0	0	2	0
ITS1	4	0	4	1
5.8S	0	1	0	0
ITS2	3	1	3	0
28S	7	1	17	2
Total	66	10	70	12

B)

nrDNA	***Mg* Type II**	*Mo Type I*	*Mo Type III*
***Mg* Type I**	**99.29%**	100.00%*	99.25%
***Mg* Type II**		99.38%	98.99%
Mo Type I			*99.34%*

C)

TAA6	***Mg* Type II**	*Mo Type I*	*Mo Type III*
***Mg* Type I**	**98.36%**	99.98%*	98.06%
***Mg* Type II**		98.38%	97.84%
Mo Type I			*98.07%*

D)

ACC6	***Mg* Type II**	*Mo Type I*	*Mo Type III*	*Mo Type IV*
***Mg* Type I**	**97.61%**	99.95%*	98.12%	97.79%
***Mg* Type II**		97.62%	97.65%	97.99%
Mo Type I			*98.14%*	*97.81%*
Mo Type III				*97.91%*

***** Note that sequences of Type I from *M. graminicola* and *M. oryzae* are almost identical for the three genomic regions

Table S7. **Sequencing depth (with standard deviation) of each type of homolog for the nuclear genomic regions TAA6 and ACC6 among 12 *M. graminicola* and two *M. oryzae* isolates**. The size of each sequence type (in bp) is also given for each species.

Taxa	TAA6			ACC6			
	Type I	**Type II**	**Type III**	**Type I**	**Type II**	**Type III**	**Type IV**
M. graminicola	**(6,361 bp)**	**(6,399 bp)**	**-**	**(6,151 bp)**	**(6,310 bp)**	**-**	**-**
- Mg-VN6	16.4× ± 5.1	16.5× ± 5.9	-	19.2× ± 4.6	18.8× ± 4.5	-	-
- Mg-VN11	18.4× ± 5.7	18.2× ± 5.3	-	20.8× ± 4.1	20.4× ± 4.9	-	-
- Mg-VN18	20.6× ± 6.2	21.7× ± 7.0	-	21.5× ± 5.0	21.9× ± 4.8	-	-
- Mg-VN27	16.4× ± 4.4	17.4× ± 5.2	-	19.4× ± 5.4	19.6× ± 4.9	-	-
- Mg-L1	22.2× ± 6.7	21.7× ± 7.7	-	19.9× ± 5.2	19.9× ± 5.7	-	-
- Mg-L2	10.2× ± 4.7	9.5× ± 4.1	-	11.6× ± 3.5	9.4× ± 2.9	-	-
- Mg-C21	20.9× ± 6.9	20.8× ± 7.2	-	19.9× ± 4.9	20.3× ± 4.9	-	-
- Mg-C25	12.7× ± 5.6	13.6× ± 6.6	-	14.8× ± 4.1	13.7× ± 3.7	-	-
- Mg-Java	1.3× ± 1.3	1.7× ± 1.4	-	0.9× ± 1.0	0.9× ± 1.0	-	-
- Mg-Borneo	11.6× ± 4.0	12.0× ± 4.1	-	9.4× ± 2.9	10.0× ± 3.2	-	-
- Mg-P	14.1× ± 4.6	14.3× ± 4.6	-	15.4× ± 4.2	14.3× ± 4.0	-	-
- Mg-Brazil	16.6× ± 4.8	17.3× ± 4.5	-	17.4× ± 4.5	17.4× ± 4.6	-	-
M. oryzae	**(6,357 bp)**	**-**	**(6,367 bp)**	**(6,153 bp)**	**-**	**(6,317 bp)***	**(6,309 bp)**
- Mo-M1	11.2× ± 4.3	-	24.2× ± 5.8	11.6× ± 3.1	-	22.9× ± 4.5	-
- Mo-M2	19.2× ± 6.3	-	11.8× ± 4.0	10.6× ± 4.1	-	11.0× ± 3.5	10.1× ± 3.0

* Excluding a repeated element (1,032 bp) in Mo-M2

References

[1] Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. doi:10.1093/bioinformatics/btx153.

[2] Besnard, G.;Jühling, F.; Chapuis, É.; Zedane, L.;Lhuillier, É.; Mateille, T.; Bellafiore, S. Fast assembly of the mitochondrial genome of a plant parasitic nematode (*Meloidogyne graminicola*) using next generation sequencing. *C. R. Biol*. **2014**, *337*, 295–301. doi:10.1016/j.crvi.2014.03.003.

[3] Bandelt, H.J.; Forster, P.;Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. doi:10.1093/oxfordjournals.molbev.a026036.

[4] Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015,** *31*, 3210–3212. doi:10.1093/bioinformatics/btv351.

[5] Somvanshi, V.S.;Tathode, M.; Shukla, R.N.; Rao, U.Nematode genome announcement: A draft genome for rice root-knot nematode, *Meloidogyne graminicola*. *J. Nematol*. **2018**, *50*, 111–116. doi:10.21307/jofnem-2018-018.

[6] Petitot, A.S.; Dereeper, A.; Agbessi, M.; da Silva, C.; Guy, J.; Ardisson, M.; Fernandez, D. Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants. *Mol. Plant Pathol.* **2016**, *17*, 860–874. doi: 10.1111/mpp.12334.

CHAPTER III

Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*)

Summary in English

In Chapter II, we investigated intraspecific diversity at the mitogenome level and some nuclear contiguity of *M. graminicola* isolates from distant geographic regions. We hypothesized a recent expansion of *M. graminicola* throughout Southeast Asia and probably globally. Nevertheless, this hypothesis needs to be validated by studying more isolates from diverse and distant ecosystems but also by exploring genetic diversity at the whole genome level. The discovery of a complex genomic organization (heterozygous), with probably a hybrid origin, has raised our curiosity about the plasticity of the genome and the mechanisms of molecular evolution that may be involved in the adaptation of *M. graminicola* to different ecosystems. It is indeed surprising to note that this parasite, reproducing by meiotic parthenogenesis, is visibly capable of rapidly and successfully colonizing very different environments. These interesting hypotheses motivated us to assemble the high-quality genome of *M. graminicola* to address this evolutionary issue.

In this second chapter, we have presented a complete and quality assembly of the nuclear genome of *M. graminicola* that reveals important genomic characteristics that could potentially contribute to the success of the parasite's parasitism. A first draft of the genome had been published, with a genome assembly size of 35 Mb. However, the assembly was highly fragmented, totalisingmore than 4,300 contiguous units with an N50 length of 20 kb, and a low level of completeness (84.27% and 73.60% respectively for CEGMA and BUSCO). The low GC content (23.5%) of the *M. graminicola* genome makes it extremely fragile and favours breakage during DNA extraction. Moreover, its genome is heterozygous (heterozygosity ~2%), and an assembly with all haplotypes is difficult to achieve from short readings. Some homologous regions have a sufficient rate of polymorphism to assemble them separately while others are fused into a consensus sequence. To overcome these difficulties, we opted for a hybrid genome sequencing strategy, combining long readings (Oxford Nanopore Technologies) with short, high-precision Illumina readings. Genome assembly was performed using different software and strategies, and the one offering the best statistical results was finally selected. This new assembly gives a size of 41.5 Mb with a length N50 (294 kb) exceptionally large for a *Meloidogyne* (second highest N50 observed to date). In addition, this consensus assembly (haplotype fusion) appears relatively complete with CEGMA (95.97%) and BUSCO (88.8%) scores. The length of the haploid genome calculated by k-mer analysis from the Illumina readings is between 41.1 and 41.6 Mb, which is very similar to the final genome assembly we obtained. In addition, the flow cytometric analysis of the DNA contained in the nucleus ranges from 81.5 to 83.8 Mb, which again corresponds to a haploid genome size between 40.7 and

41.9 Mb. These measurements tend to confirm that the genome assembly we have obtained is relatively complete and corresponds to a haploid genome where the haplotypes have been fused for most genomic regions.From this genome assembly and thanks to the published transcriptomic data, we were able to automatically annotate the genes present in the *M. graminicola* genome. 10,284 genes were thus predicted, which corresponds to an average of 247.51 genes per Mb of genome. The genes of *M. graminicola* appear particularly fragmented with an average of 11.2 exons per gene compared to an average of 6 exons per gene found in mitotic RKNs and 8.8 exons per gene found in *Globodera rostochiensis*.

It is interesting to note that we have identified 67 genes coding for 68 proteins that have an "alien index" > 14, indicating a possible acquisition by horizontal transfer of these genes (mainly of bacterial origin). These putative HGTs represent 31 different gene families based on their pfam domains and could provide important functions for the parasitic lifestyle of *M. graminicola.* For example, 44% of these genes code for plant cell wall degradation enzymes such as polygalacturonase, xylanase, arabinanase, pectate lyase, expansin-like proteins and cellulases, which could facilitate the migration of the parasites into the root tissue and the establishment of feeding sites. Also included in this list of 67 genes are genes coding for chorismate mutase, isochorismatase and carboxylesterase, which are potentially involved in the detoxification of plant defense products, such as reactive oxygen species (ROS), or the manipulation of jasmonate pathways known to cause plant susceptibility to nematodes. Finally, other of these genes could be involved in nematode metabolism, including the biosynthesis of vitamin B7, glutamine and carbohydrates, the degradation of galactose and sucrose, and sucrose transport.

The *M. graminicola* genome that we have assembled was also used to study transposable elements (TE). Thus, 575 canonical TEs were annotated, extending over 1.08 Mb or 2.61% of the genome. DNA transposons were slightly more abundant than retrotransposons (1.49% versus 1.12% of the genome, respectively). In addition, non-autonomous TEs appeared to be more abundant (54.6% of TEs). We also found that some TEs associated with HGT genes, suggesting that some TEs may have been transferred from bacteria to the *M. graminicola* genome.

The novel *M. graminicola* genome sequence and interesting genomic features (gene annotation, HGT and TE) will potentially have immediate and important implications for research on the evolutionary biology of this pathogen.

Résumé en Français

Dans le chapitre II, nous avons étudié la diversité intraspécifique au niveau du mitogénome et de certains contigus nucléaires d'isolats de *M. graminicola* provenant de régions géographiques éloignées. Nous avons avancé l'hypothèse d'une expansion récente de *M. graminicola* sur l'ensemble de l'Asie du Sud Est et probablement à l'échelle mondiale. Néanmoins cette hypothèse doit être validé en étudiant d'avantage d'isolats provenant d'écosystèmes divers et éloignés mais également en explorant la diversité génétique au niveau du génome entier. La découverte d'une organisation génomique complexe (hétérozygote), avec vraisemblablement une origine hybride, a suscité notre curiosité concernant la plasticité du génome et les mécanismes d'évolution moléculaire pouvant participer à l'adaptation de *M. graminicola* à différents écosystèmes. Il est en effet surprenant de constater que ce parasite, se reproduisant par parthénogénèse méiotique, est visiblement capable de coloniser rapidement et avec succès des environnements très différents. Ces hypothèses intéressantes nous ont motivés à faire un assemblage du génome de haute qualité de *M. graminicola* pour aborder cette question évolutive.

Dans ce deuxième chapitre, nous avons présenté un assemblage complet et de qualité du génome nucléaire de *M. graminicola* permettant de révéler des caractéristiques génomiques importantes pouvant potentiellement contribuer au succès du parasitisme du parasite. Une première ébauche du génome avait été publiée, avec une taille d'assemblage du génome de 35 Mb. Cependant, l'assemblage était très fragmenté, totalisant plus de 4300 contigs avec une longueur N50 de 20 kb, et un faible niveau de complétude (84,27% et 73,60% respectivement pour CEGMA et BUSCO). La faible teneur en GC (23,5 %) du génome de *M. graminicola* le rend extrêmement fragile et favorise les cassures lors de l'extraction de l'ADN. De plus, son génome est hétérozygote (hétérozygotie ~2%), et un assemblage présentant tous les haplotypes est difficile à réaliser à partir de lectures courtes. Certaines régions homologues ont un taux de polymorphisme suffisant permettant de les assembler séparément tandis que d'autres sont fusionnées en une séquence consensus. Pour surmonter ces difficultés, nous avons opté pour une stratégie hybride de séquençage du génome, combinant des lectures longues (Oxford Nanopore Technologies) avec des lectures courtes Illumina de haute précision. L'assemblage du génome a été réalisé à l'aide de différents logiciels et stratégies, et celui offrant les meilleurs résultats statistiques a finalement été sélectionné. Ce nouvel assemblage donne une taille de 41,5 Mo avec une longueur N50 (294 kb) exceptionnellement grande pour un *Meloidogyne* (second plus haut N50 observé à ce jour). En outre, cet assemblage consensus (fusion des haplotypes) semble relativement complet avec des scores CEGMA (95,97%) et BUSCO (88,8%). La longueur du génome haploïde calculée par analyse k-mer à partir des lectures Illumina se situe entre 41,1 et 41,6 Mb, ce qui est très similaire à l'assemblage final du génome que nous avons obtenu. De plus, l'analyse par cytométrie de flux de l'ADN contenu dans le noyau varie de 81,5 à 83,8 Mb, ce qui correspond là aussi à une taille du génome haploïde

comprise entre 40,7 et 41,9 Mb. Ces mesures tendent à confirmer que l'assemblage du génome que nous avons obtenu est relativement complet et correspond à un génome haploïde où les haplotypes ont été fusionnés pour la plupart des régions génomiques.

A partir de cet assemblage du génome et grâce aux données transcriptomiques publiées, nous avons pu annoter automatiquement les gènes présents dans le génome de *M. graminicola*. 10 284 gènes ont été ainsi prédit ce qui correspond à une moyenne de de 247,51 gènes pour 1 Mb de génome. Les gènes de *M. graminicola* apparaissent particulièrement fragmentés avec une moyenne de 11,2 exons par gène en comparaison des 6 en moyenne trouvés chez les RKN mitotiques et des 8,8 chez *Globodera rostochiensis*.

Il est intéressant de noter que nous avons identifié 67 gènes codant pour 68 protéines qui ont un « alien index" > 14, indiquant une possible acquisition par transfert horizontal de ces gènes (principalement d'origine bactérienne). Ces HGT présumés représentent 31 familles de gènes différentes en se basant sur leurs domaines pfam et pourraient assurer des fonctions importantes pour le mode de vie parasitaire de *M. graminicola*. A titre d'exemple, 44% de ces gènes codent pour des enzymes de dégradation de la paroi cellulaire des plantes telles que la polygalacturonase, la xylanase, l'arabinanase, la pectate lyase, des protéines de type expansine et des cellulases, ce qui pourrait faciliter la migration des parasites dans le tissu racinaire et la mise en place des sites nourriciers. Dans cette liste de 67 gènes on retrouve également des gènes codants pour la chorismate mutase, l'isochorismatase et la carboxylestérase qui sont potentiellement impliquées dans la détoxication des produits de défense des plantes, tels que les espèces réactives de l'oxygène (ROS), ou la manipulation des voies jasmonates connues pour entrainer la sensibilité des plantes aux nématodes. Enfin, d'autres de ces gènes pourraient être impliquées dans le métabolisme du nématode, et notamment la biosynthèse des vitamines B7, de la glutamine et des glucides, la dégradation du galactose et du saccharose, et le transport du saccharose.

Le génome de *M. graminicola* que nous avons assemblé a également été utilisé pour étudier les éléments transposables (TE). Ainsi, 575 TE canoniques ont été annotés et qui s'étendaient sur 1,08 Mb soit 2,61% du génome. Les transposons d'ADN étaient légèrement plus abondants que les rétrotransposons (1,49% contre 1,12% du génome, respectivement). De plus, les TE non autonomes sont apparus plus abondants (54,6 % des TE). Nous avons également constaté que certaines TE s'associent à des gènes HGT, ce qui suggère que certains TE pourraient avoir été transférées des bactéries vers le génome de *M. graminicola*.

La nouvelle séquence du génome de *M. graminicola* et les caractéristiques génomiques intéressantes (annotation des gènes, HGT et TE) auront potentiellement des implications immédiates et importantes pour la recherche sur la biologie évolutive de cet agent pathogène

Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*)

Ngan Thi Phan[1], Julie Orjuela[1], Etienne G.J. Danchin[2], Christophe Klopp[3], Laetitia Perfus-Barbeoch[2], Djampa K. Kozlowski[2], Georgios D. Koutsovoulos[2], Céline Lopez-Roques[4], Olivier Bouchez[4], Margot Zahm[3], Guillaume Besnard[5]*, Stéphane Bellafiore[1]*

[1]University of Montpellier II, IRD, Cirad, Interactions Plantes Microorganismes Environnement (IPME), 34394 Montpellier, France;

[2]Institut Sophia Agrobiotech, INRAE, Université Côte d'Azur, CNRS, 06903 Sophia Antipolis, France;

[3]Plateforme BioInfo Genotoul, INRAE, UR875, 31326 Castanet-Tolosan cedex, France;

[4]INRAE, US 1426, GeT-PlaGe, Genotoul, 31326 Castanet-Tolosan, France;

[5]CNRS-UPS-IRD, UMR5174, EDB, 118 route de Narbonne, Université Paul Sabatier, 31062 Toulouse, France.

* co-senior authors for correspondence: guillaume.besnard@univ-tlse3.fr (G.B.); stephane.bellafiore@ird.fr (S.B.)

Ecology and Evolution. 2020;10:11006–11021

First published: 13 September 2020 https://doi.org/10.1002/ece3.6680

Abstract:

Discovered in the 1960s, *Meloidogyne graminicola* is a root-knot nematode species considered as a major threat to rice production. Yet, its origin, genomic structure, and intraspecific diversity are poorly understood. So far, such studies have been limited by the unavailability of a sufficiently complete and well-assembled genome. In this study, using a combination of Oxford Nanopore Technologies and Illumina sequencing data, we generated a highly contiguous reference genome (283 scaffolds with an N50 length of 294 kb, totaling 41.5 Mb). The completeness scores of our assembly are among the highest currently published for *Meloidogyne* genomes. We predicted 10,284 protein-coding genes spanning 75.5% of the genome. Among them, 67 are identified as possibly originating from horizontal gene transfers (mostly from bacteria), which supposedly contribute to nematode infection, nutrient processing, and plant defense manipulation. Besides, we detected 575 canonical transposable elements (TEs) belonging to seven orders and spanning 2.61% of the genome. These TEs might promote genomic plasticity putatively related to the evolution of *M. graminicola* parasitism. This high-quality genome assembly constitutes a major improvement regarding previously available versions and represents a valuable molecular resource for future phylogenomic studies of *Meloidogyne* species. In particular, this will foster comparative genomic studies to trace back the evolutionary history of *M. graminicola* and its closest relatives.

Keywords: reference genome, root-knot nematode (RKN), pest, cereals, transposable element, horizontal gene transfer (HGT).

I. Introduction

Meloidogyne graminicola, commonly called the rice Root-Knot Nematode (rice RKN), is a prevalent pest at a global scale, causing severe damages to cereals (Dutta et al., 2012) and infecting more than 100 plant species (EPPO Global Database, 2019). This pest was first described in Louisiana (Golden & Birchfield, 1965) and Laos (Golden & Birchfield, 1968), before being found attacking several rice agrosystems (from upland to lowland, and irrigated to deep-water fields) in many countries from America, Africa, Europe and especially Asia. While Asia provides 90% of the global rice production, a 15%-yield loss due to RKNs was estimated in this area, and this is probably an underestimate because of the lack of specific above-ground symptoms (Mantelin et al., 2017).

Meloidogyne graminicola is mainly reproducing through facultative meiotic parthenogenesis with a very short life-cycle (Narasimhamurthy et al., 2018). A freshly hatched juvenile can develop into an adult female laying 250 to 300 eggs after only 25-28 days. Such reproductive abilities may explain its rapid population increase and spread. For instance, in northern Italy, where this pest was recently detected, the total infected area has increased by approximately five folds in just one year (from 19 to 90 ha in 2016-2017; EPPO Reports, 2017). This nematode is therefore classified as a quarantine pest in several countries (e.g. Brazil, Madagascar, China; EPPO, 2017-2018) and was added recently to the EPPO Alert List in Europe (Fanelli et al., 2017). Despite the huge impact of *M. graminicola* on agriculture worldwide, its evolutionary history as well as adaptive behavior in variable environments are still poorly documented. Therefore, control of this pathogen remains limited.

Root-Knot Nematode species (RKNs; *Meloidogyne* spp.) exhibit a striking diversity of reproductive modes, chromosome counts and hosts (Castagnone-Sereno et al., 2013). Those with obligate sexual reproduction have fewer chromosomes and a narrow host spectrum [e.g., *M. spartinae,* n = 7 (Triantaphyllou, 1987)], compared to those with facultative sexual reproduction [e.g. *M. graminicola*, *M. hapla*, *M. chitwoodii*; n = 13-19 (Triantaphyllou, 1985)] which have a broader host range and larger geographic distribution. Curiously, the most damaging RKNs to worldwide agriculture, owing to the diversity of infected hosts and most extensive global distribution are reproducing asexually by obligatory mitotic parthenogenesis (Castagnone-Sereno & Danchin, 2014). These species are polyploid with numerous chromosomes [e.g., *M. javanica,* 3n = 42-48 (Triantaphyllou, 1985)]. During the last fifteen years, advances in next-generation genome sequencing have provided new insights into the considerable diversity and life history of plant-parasitic nematodes (PPNs), particularly RKNs.

According to phylogenetic studies based on nuclear ribosomal DNA (nrDNA), RKNs can be classified in three main clades (De Ley et al., 1999), with most of the knowledge recently accumulated on species belonging to Clade I (e.g. *M. incognita*, *M. floridensis, M. javanica*, *M. arenaria* and *M. enterolobii*) and Clade II (e.g. *M. hapla*). Comparative genomics on some mitotic parthenogenesis RKN species of Clade I provided relevant data on the origin and evolution of their polyploid genomes. Highly diverged genome copies and lack of recombination events were reported in these species, indicating hybrid origins and clonal reproduction (Blanc-Mathieu et al., 2017; Koutsovoulos et al., 2019; Lunt et al., 2014; Szitenberg et al., 2017). Besides, their genomes contain numerous transposable elements (TEs), while the meiotic facultative sexual diploid *M. hapla* (Clade II) does not show diverged genome copies and seems to have a lower TE load (Blanc-Mathieu et al., 2017; Szitenberg et al., 2017; Bird et al., 2009). Horizontal Gene Transfers (HGTs) originating from bacteria and fungi have probably played an important role in the evolution of plant parasitism in RKNs, as well as in other nematode groups (Danchin & Rosso, 2012; Danchin et al., 2010; Haegeman et al., 2011). In RKNs, functional genes potentially acquired via HGT have been documented in *M. incognita*, *M. javanica*, *M. floridensis* and *M. hapla* (Clades I and II) for proteins involved in plant cell wall degradation, nutrient processing, detoxification, and manipulation of plant defenses (Scholl et al., 2003). Compared to other mitotic parthenogenetic and sexual RKNs, the diversity and genetic structure of facultative meiotic parthenogenetic species of Clade III remain however poorly understood. In *M. graminicola*, most of the genetic studies were based on mitochondrial DNA and nrDNA. These sequences revealed very low polymorphism and lack of phylogeographic signal among the isolates sampled at a global scale, suggesting a recent spread of this pathogen (Besnard et al., 2019). Divergent low-copy nuclear homologous sequences were also found indicating either a potential hybrid origin or high heterozygosity in this species. These hypotheses, based on sporadic pieces of evidence, need to be better documented. Generating a high-quality genome sequence of *M. graminicola* integrating close relatives is thus necessary for further comparative genomic analyses, especially to trace back their origin and global spread. Moreover, this will allow a better understanding of the impact of reproduction strategies and genome evolution in adaptive processes linked to different environmental conditions.

A first draft of the *M. graminicola* genome was released, with a genome assembly size of 35 Mb (Somvanshi, et al., 2018). However, the assembly was highly fragmented, totaling more than 4,300 contigs and an N50 length of 20 kb. In addition, compared to other RKN

genomes, including the only other meiotic facultative sexual *M. hapla*, gene completeness (assessed on widely conserved single-copy eukaryotic genes) was relatively low in this genome. For instance, respectively 84.27% and 73.60% of CEGMA and BUSCO eukaryotic genes were found in complete length in the *M. graminicola* genome vs respectively 93.55% and 87.40% for *M. hapla* (Koutsovoulos et al., 2019). This means that some genomic regions were probably not captured in the assembly or too fragmented. Therefore, the quality of this draft genome currently limits further sensitive studies such as comparative genomics of RKNs or population genomics studies at the species level. The reconstruction of the *M. graminicola* genome is challenged by two main features. Firstly, the *M. graminicola* genome is GC-poor (GC content = 23.5%) which makes it extremely fragile and favors breaks during DNA extraction. Secondly, the genome is heterozygous (heterozygosity = ca. 2%), and its assembly is made difficult by the presence of divergent haplotypes, especially when using short reads (Besnard et al., 2019). For instance, some divergent homologous regions may be separately assembled, while others could be merged in a unique consensus sequence (Besnard et al., 2019).

To overcome these difficulties, we opted for a hybrid genome sequencing strategy, combining long reads (Oxford Nanopore Technologies; ONT) with high-accuracy Illumina short reads to obtain a more complete and contiguous genome assembly. Genome assembly was performed with different softwares and strategies, and the one having the best biological and statistical metrics was finally selected. We annotated the genome for protein-coding genes, TEs and potential HGTs. Total DNA content of *M. graminicola* cells was also measured by flow cytometry to validate genome size. So far, this genome assembly is the most complete and contiguous available for *Meloidogyne* of Clade III, and this reference will assist a range of genetic, genomic and phylogenetic studies to uncover the life history of *M. graminicola* and related RKNs.

II. Material & methods

2.1 Nematode DNA extraction

The *M. graminicola* isolate Mg-VN18 was isolated from rice roots collected in a high-land field of the Lao Cai province, Vietnam (Bellafiore et al., 2015). Mg-VN18 was cultivated from a single juvenile on the root system of the susceptible rice cultivar IR64. Eggs and juveniles were extracted from roots two months after infection using a hypochlorite extraction method and a blender (McClure et al., 1973) with minor modifications from Bellafiore et al. (2015). Roots were treated for 15 min in 0.8% hypochloride at room temperature to eliminate bacteria and fungi. After washing these nematodes carefully with water, the mixture was purified using discontinuous sucrose gradient as described in Schaad & Walker (1975) to remove potential remaining sources of DNA contaminants such as rice root tissues, bacteria, and fungi. After purification, the fresh eggs and juveniles were used directly for DNA extraction without freezing to avoid DNA fragmentation.

Getting high-molecular-weight DNA is a crucial step to benefit from the full potential of Oxford Nanopore Technologies (ONT) sequencing. Two different DNA extraction protocols were tested [i.e. protocol of Epicentre's MasterPure Complete DNA Purification Kit (Lucigen, USA) and a modified phenol-chloroform based method (Sambrook et al., 1989)]. The phenol protocol method yielding good quality DNA with an average fragment length of 39 kb for a total of 8.2 µg, which is suitable for ONT sequencing (Sambrook et al., 1989). Following this protocol, 260 µL of extraction buffer (0.1 M Tris, pH 8, 0.5 M NaCl, 50 mM EDTA, 1% SDS) and 40 µL of proteinase K (20 mg/mL; Qiagen, Germany) were added into the tube containing 0.1 mL of fresh eggs and juveniles. Nematodes were then crushed by twisting with an autoclaved mircropestle for about 30 s. The solution was incubated at 55°C for 24 hrs. Then, 10 µL of RNAse A (10 mg/mL; Qiagen, Germany) were added and the mix was incubated at room temperature for 50 min. Genomic DNA (gDNA) was recovered by a phenol-chloroform step (Sambrook et al., 1989). The chloroform-free phase was treated with NH_4OAC (for a final concentration of 0.75 M) before ethanol precipitation. To reduce DNA fragmentation, no freezing nor vortexing steps were performed. All the mixing steps were done by three meticulous tube inversions, and final gDNAs were stored at 4°C for less than one week before sequencing. For Illumina sequencing, gDNA was extracted following the manual of the Epicentre's MasterPure Complete DNA Purification Kit (Lucigen, USA). For all gDNA samples, double stranded DNA concentration was assessed using the Qubit dsDNA HS Assay Kit (Life Technologies, US). DNA purity was checked using the Nanodrop (Thermo Fisher

Scientific, US). Distribution and degradation of DNA fragment sizes were assessed using the Fragment analyzer (AATI) High Sensitivity DNA Fragment Analysis Kit (Thermo Fisher Scientific, US). DNA integrity was also checked by electrophoresis, loading 1 µL on a 1%-agarose gel.

2.2 Whole-genome sequencing, read processing, and k-mer analysis

Long reads sequencing: Library preparation and sequencing were performed at the GeT-PlaGe core facility, INRA Toulouse, according to the manufacturer's instructions "1D gDNA selecting for long reads (SQK-LSK109)". Aiming at covering the *M. graminicola* genome at > 70× with long reads, sequencing was done on one ONT flowcell. Genomic DNA was purified using AMPure XP beads (Beckman Coulter, US). Eight µg of purified DNA were sheared at 20 kb using the megaruptor system (Diagenode, Belgium). A "one-step" DNA damage repair + END-repair + dA tail of double-stranded DNA fragments was performed on 2 µg of DNA. Adapters were ligated to the library that was then loaded (0.03 pmol) onto an R9.4.1 revD flowcell. It was sequenced on the GridION instrument for 48 hrs. Final reads were base-called using Guppy v.1.8.5-1 (Oxford Nanopore).

After sequencing, adapters of raw ONT reads were trimmed using Porechop (Wick, 2019). Only reads with a Q-score value greater or equal to seven were selected using NanoFilt v.1.1.0 (De Coster et al., 2018). Minimap2 (Li, 2018) was used to map long reads to the *M. graminicola* mitogenome (GenBank no. HG529223) and Samtools Fasta - f 0x4 (Li et al., 2009) were used to sort out long reads that mapped to this reference.

Short reads sequencing: High-depth short reads sequencing was performed at the GeT-PlaGe core facility, INRA Toulouse. DNA-seq libraries have been prepared according to the Illumina's protocol "TruSeq Nano DNA HT Library Prep Kit" (Illumina sequencing technology, US). Briefly, tree µg of gDNA were fragmented by sonication. Then, DNA fragments were selected by size (mean insert size = approx. 380 bp) using SPB beads (kit beads), and then ligated to adaptors. Quality of libraries was assessed using a Fragment Analyzer (Advanced Analytical), and DNA quantity was measured by qPCR using the Kapa Library Quantification Kit (Roche, Switzerland). Sequencing was performed on an Illumina HiSeq-3000 using a paired-end read length of 2 x 150 bp with the Illumina HiSeq 3000 Reagent Kits.

Illumina raw reads were trimmed and cleaned from contamination. First, the short-reads were processed for quality control using FastQC (Andrews, 2010). Second, Skewer (Jiang et

al., 2014) was used to trim reads considering a minimum quality score of 30 and a minimum read length of 51 bp. Third, the trimmed reads were pre-assembled using Platanus (Kajitani et al., 2014). Subsequently, the pre-assembled contigs were blasted against the NCBI's nucleotide (nt) database using Blastn (Altschul et al., 1990) for contamination screening on BlobTools (Kumar et al., 2013; Laetsch & Blaxter, 2017). A group of pre-assembled contigs annotated as proteobacteria at low coverage (< 10×) was considered as contaminants. Therefore, the reads that belonged to these contigs were removed from the pool of short reads, resulting in a cleaned Illumina dataset. The cleaned reads that aligned to the mitogenome of *M. graminicola* (GenBank no. HG529223) were also removed using Bowtie2 (Langmead & Salzberg, 2012). Finally, the reads were error-corrected using Musket (Liu et al., 2013).

Jellyfish (Marçais & Kingsford, 2011) was used to extract and count canonical *k*-mers ($k = 17$, 21, 27, and 47 nucleotides) from cleaned Illumina reads. For each *k* value, GenomeScope (Vurture et al., 2017) was used to estimate haploid genome length, heterozygosity, and repeats content from the *k*-mer counts. The parameter *MaxCov* was set at 900,000, as recommended by Mgwatyu et al. (2020).

2.3 Quantification of nuclear DNA content

To assess the nuclear genome size of Mg-VN18, two independent flow-cytometry runs were done for five replicates, which were collected at different time points. Eggs and juveniles from each replicate were extracted and purified using the same method described above, then stored at -82ºC. Besides, two species with known genome size, *Caenorhabditis elegans* strain Bristol N2 [200 Mb, diploid (The *C. elegans* Sequencing Consortium, 1998)] and *Drosophila melanogaster* Canton-S strain [350 Mb, diploid (Bosco et al., 2007)] were used as internal standards. In each run, nuclei extraction, nuclei stain and DNA content measurements were done using the same protocol as previously described (Perfus-Barbeoch et al., 2014; Blanc-Mathieu et al., 2017) for both samples and internal standards. In short, 0.1 mL of fresh eggs and juveniles were ground carefully for 7 min in 2 mL of the lysis buffer (1 mM KCl, 30 mM NaCl, 10 mM $MgCl_2$, 0.2 mM EDTA, 30 mM Tris, 300 mM sucrose, 5 mM sodium butyrate, 0.1 mM PMSF, 0.5 mM DTT, 40 µL Igepal), and then, 8 mL suspension buffer (same as lysis buffer except for sucrose, 1.2 M, and without Igepal) was overlaid on top of lysis buffer. Subsequently, the tube was centrifuged to separate nuclei from other cell debris. After removing the supernatant, the pellet of nuclei was re-suspended in 1 mL of staining buffer containing propidium iodide (final concentration of 75 µg/mL) and DNAse-free RNAse (final concentration of 50 µg/mL) at 37ºC for 30 min. Each sample was first measured independently

and then mixed with standard controls in the same tube. Flow cytometry analysis was then performed using the LSRII/Fortessa (BD Biosciences) flow cytometer operated with the FACSDiva v.6.1.3 software (BD Biosciences). For each measurement, the fluorescence cytograms were analyzed on Kaluza v.1.2 (Beckman Coulter). For each species, fluorescent peaks corresponding to three phases of the cell cycle (G0/G1, S and G2/M) were obtained (Ormerod, 2008). Only mean fluorescence intensity of the G0/G1 phase (first peak) was taken into account, and *M. graminicola* DNA content was then estimated using the following equation:

Total DNA content of *M. graminicola* sample = (G0/G1 peak value of sample × Whole genome size of internal control *i*) / (G0/G1 peak value of internal control *i*)

with *i* being either *C. elegans* or *D. melanogaster*.

2.4 Genome assembly, completeness assessment and haplotigs purging

Five popular assemblers were first tested to assemble the *M. graminicola* genome: Flye v.2.4.1 (Kolmogorov et al., 2019), Ra v.0.2.1 (Vaser & Šikić, 2019), MaSuRCA v.3.2.4 (Zimin et al., 2013), Canu v.1.8 (Koren et al., 2017) and Miniasm v.2.2.16 (Li, 2016). Flye, Ra, Canu and Miniasm use long reads only to build contigs, while MaSuRCA combines both long (ONT) and short (Illumina) reads. Subsequently, Racon (Vaser et al., 2017) and Pilon (Walker et al., 2014) were used to correct bases and homopolymer lengths. To scaffold the genome, a set of 66,396 transcripts (Petitot et al., 2015) was blasted to the genome assemblies. Then, the Perl script SCUBAT v.2 (Koutsovoulos, 2018) was used to identify transcripts that were split over multiple contigs. This information was then used to concatenate the contigs. After obtaining corrected and concatenated contigs, assemblies statistics were computed using QUAST (Gurevich et al., 2013) and compared. The genome completeness was assessed using both CEGMA [Core Eukaryotic Genes Mapping Approach (Parra et al., 2007)] and BUSCO v.3 [Benchmarking Universal Single-Copy Orthologs (Simão et al., 2015)]. For CEGMA, the provided core set of 248 eukaryotic orthologs was used as a reference, and genes were predicted using default parameters (e.g. maximum intron length of 5 kb and gene flanks of 2 kb). For BUSCO, the provided nematoda data set is not appropriate for RKNs because it contains orthologous genes of eight nematode species belonging to only three (2, 8 and 9) out of the 12 described nematoda clades (Megen et al.,2009) and no species from clade 12, to which RKNs belong. Meanwhile, the eukaryotic dataset is a pool of single-copy orthologs from 65 eukaryote species, including the nematoda dataset. Therefore, the "Eukaryota_odb9" library including 303

Eukaryote single-copy orthologs was preferred and used as the reference. The species-specific-trained parameters of the nematode species *C. elegans* were used for gene prediction and BUSCO was run in '-long' mode for AUGUSTUS optimization. We used both the median length of scaffolds (N50) and genome completeness (i.e. the percentage of fully assembled conserved eukaryote genes) to select the best genome assembly for further analyses.

Heterozygous regions can severely complicate genome assembly with regions of higher heterozygosity being assembled separately, while regions of lower heterozygosity being collapsed in one consensus region. This may cause issues with genome size estimation, spurious annotation, variant discovery or haplotype reconstruction. An ideal haploid representation (primary contigs) would consist of one allelic copy of all heterozygous regions in the two haplomes, as well as all hemizygous regions from both haplomes. Purge Haplotig (Roach et al., 2018) was used to identify contigs that were likely to be allelic contigs and retained only the primary contig. Briefly, in a first step, the program created a read-depth histogram using the mapped long-reads to the assembly. If the histogram shows only one read-depth peak, there is no need to purge haplotigs because the entire genome contains collapsed haplotype contigs. Otherwise, if two peaks are observed, one being at half the coverage of the second, both allelic contigs and collapsed haplotype contigs are present in the assembly. For collapsed haplotypes, the reads from both alleles will map to the same contig, resulting in one read-depth peak. In contrast, if the alleles are assembled as separate contigs, the reads will be split over the two contigs, resulting in another peak at half the read-depth ("0.5 unit" read-depth peak). The half read-depth contigs will be assigned as suspect contigs (or supposedly uncollapsed contigs). In the second step, these suspect contigs are aligned against the entire genome to identify synteny with its allelic companion contig. Contigs with an alignment score greater than the cut-off (by default $\geq 70\%$) are marked for reassignment as haplotigs and removed from the assembly. In addition, the contigs with an abnormally low long-read depth ($\leq 10\times$) are likely to be assembly artifacts, while unusually high read-depth ($\geq 195\times$) are likely to be collapsed repeats, organellar DNA contigs or contaminants. Such contigs were thus also removed from the rest of the assembly. Finally, the program will produce three FASTA format files: contigs reassigned as haplotigs, the abnormally covered contigs reassigned as artifacts, and the curated contigs that represent the haploid assembly.

The purged-haplotig genome (curated contigs) was then blasted to the NCBI nt database using Blastn (Altschul et al., 1990) for contamination screening on BlobTools (Kumar et al., 2013; Laetsch & Blaxter, 2017). Contigs with short read-depth inferior to 100× showing highest

similarity to non-nematoda sequences, were considered as potential contaminants and thus removed from the assembly.

To investigate the heterozygous regions on the genome, the short reads were mapped against the curated genome assembly to call single nucleotide variants (SNV) using TOGGLE's configuration file *SNPdiscoveryPaired.config.txt* (Tranchant-Dubreuil et al., 2018). The reads from the two divergent haplotype copies will map on a single collapsed region in the reference genome, resulting in heterozygous SNVs. SNV positions with mapping quality ≥ 30 and sequencing depth ≥ 10× were selected. The number of heterozygous variants per 10-kb window was then calculated using BEDOPS (Neph et al., 2012). The above short-read mapping file was also used to calculate short-read depth per window using BEDtools *multicov* (Quinlan & Hall, 2010). Long reads were mapped onto the genome using Minimap2 (Li, 2018) to generate a long-read mapping file. The mapping file was sorted using Samtools *sort* and used for the calculation of long-read depth per genome window using BEDtools. GC content per sliding window of 1 kb was calculated using BEDtools *nuc* (Quinlan & Hall, 2010). The distribution of heterozygous variants, short-read depth, long-read depth, and GC content was shown on the genome scaffolds per 10-kb sliding window using CIRCOS (http://circos.ca/).

2.5 Gene prediction, annotation and detection of putative horizontal gene transfers

Protein-coding genes were predicted with the MAKER v.2.31.9 genome annotation pipeline (Holt & Yandell, 2011). To improve homology search during the annotation process, low complexity regions, satellites and simple sequence repeats (SSR) were soft-masked with lower-case letters in the genome using RepeatMasker v.4.0.7 (http://www.repeatmasker.org). A transcriptome of *M. graminicola* at juvenile stage (Petitot et al., 2015) was used as source of evidence for gene predictions. A *de novo* transcriptome assembly was obtained using Trinity v.2.5.1 (Grabherr et al., 2011). For a given locus of the Trinity output, only the contigs with the longest ORF were kept. Hisat2 v2.1 (Kim et al., 2015) and StringTie v.1.3.4 0 (Pertea et al., 2015) were used to obtain a guided assembly of transcripts. Finally, four datasets were thus used as references: (i) the available dataset of 66,396 ESTs (Petitot et al., 2015), (ii) the longest transcripts among their isoforms assembled by Trinity, (iii) the whole transcripts assembled by StringTie, and (iv) the *EST_nematoda* UniProt database. MAKER was run in two steps. The first step was based on pieces of evidence from the transcriptomes (*est2genome*) and protein sequences from UniProt and Trembl databases (*protein2genome*). In the second step, MAKER predicted genes by reconciling evidence alignments and *ab initio* gene predictions using SNAP v.2013-11-29 (Korf, 2004). Functional annotation for predicted genes was done by searching

homology to UniProt/Swiss-Prot databases. In addition, InterProScan v.5.19-58.0 (Zdobnov & Apweiler, 2001) was used to examine conserved protein domains, signatures, and motifs present in the predicted protein sequences. Gene sequences with annotation edit distance (AED) values of less than one with domain content were retained using the Perl script *quality_filter.pl* (Campbell et al., 2014). The higher the AED value was, the higher sequence divergence was detected between the predicted protein and the sources of evidence. The statistics of the gene prediction and annotation were retrieved using the Python script Genome Annotation Generator *gag.py* (Hall et al., 2014). Further, to infer the completeness of the predicted protein coding genes, the BUSCO score was calculated using the parameters described above for the genomic sequence. The number of genes per sliding genome window of 10 kb was calculated using BEDOPS (Neph et al., 2012). Distribution and density of genes on genome scaffolds were visualized using CIRCOS (http://circos.ca/).

The coding genes were then used to detect candidate horizontal gene transfers (HGTs) of non-metazoa origin in the *M. graminicola* genome using Alienness (Rancurel et al., 2017). Basically, Alienness identifies genes in *M. graminicola* that are substantially more similar to non-metazoan than metazoan homologs. In a first step, all the predicted proteins were compared to the NCBI's nr library using BLASTp with an E-value threshold of $1E^{-3}$ and no filtering for low complexity regions. Because we were looking for genes of non-metazoan origin in a metazoan, we selected 'Metazoa' as taxonomic recipient group. To avoid self-hits to RKNs and other related plant-parasitic nematodes, we excluded the sub-order 'Tylenchina'. Beside Bacteria, two additional taxonomic groups - Viridiplantae and Fungi - were used to classify the potential donors. Then, based on the taxonomy identity and the E-value for each blast hit, Alienness calculates an Alien Index (AI) for each query protein as following: AI = ln(best metazoan E-value + $1E^{-200}$) − ln(best non-metazoan E-value + $1E^{-200}$). An AI > 0 indicates a better hit to the donor (non-metazoan) than recipient (metazoan) taxa and a putative HGT of non-animal origin. Higher AI represents a higher gap of E-values between candidate donor and recipient and a more likely HGT. According to the 70% rule (Ku & Martin, 2016), all *M. graminicola* proteins returning an AI > 0 with a 70% identity to a putative donor were discarded from the rest of the analyses to eliminate possible assembly or annotation artifacts. As recommended by Rancurel et al. (2017), an AI threshold > 14 represents the right balance between recall and precision of the method, at least in RKNs. With an AI > 26, the accuracy (proportion of candidate genes supported as HGT by phylogenies) is even higher, but the recall rate is lower (Rancurel et al., 2017). Therefore, in our study, we used both values as thresholds

to detect putative HGTs and highly likely HGTs. Location of these genes on the whole genome was finally represented using CIRCOS (http://circos.ca/).

2.6 Annotation of transposable elements

The assembled genome of *M. graminicola* was finally used to investigate transposable elements (TEs) using the REPET meta-pipeline, which includes TEdenovo and TEannot (Flutre et al., 2011). The TE prediction and annotation protocols followed in this study are described in details in Koutsovoulos et al. (2019). In brief, all the unresolved regions (Ns) of the genome longer than 11 nucleotides were first removed. Then, genomic sequences shorter than the L99 (5,010 bp) were discarded. Remaining sequences were used as input for the TEdenovo pipeline to *de novo* build a TEs consensus library. The obtained sequence library was then automatically filtered doing a minimal genome annotation with TEannot and only retaining consensuses with at least one Full-length copy (FLC) annotated on the genome. The filtered consensus TEs library was then used in the TEannot pipeline to perform a full annotation of the whole *M. graminicola* genome. Finally, strict filters were applied to only retain annotations conform to two main criteria: i) conserved TE annotations must be classified as retro-transposons or DNA-transposons and be longer than 250 bp; and ii) TE copies must share 85% identity with their consensus and cover more than 33% of its length. Distribution of TEs on the genome was visualized using CIRCOS (http://circos.ca/).

III. Results

3.1 Whole-genome sequence and total DNA content of *M. graminicola*

In total, 3.9 Gb of raw reads were produced by the Oxford Nanopore technology (N50 length = 8.9 kb), while Illumina sequencing technology generated 122 million reads with a total volume of 17.4 Gb. After cleaning, 3.5 Gb of long reads with an N50 length of 9.4 kb and 87 million short reads (11.98 Gb) were retained (Table S1). The *k*-mer analysis on cleaned short reads allowed us to estimate the haploid genome length at different *k* values, from 41.1 to 41.6 Mb with average heterozygosity varying from 1.69 to 1.90 % (Table S2). In contrast, the repeats content of the genome dramatically depended on the *k* value used, although the highest values (k = 27 and 47) rendered similar results (7.8 Mb; Table S2).

The cleaned long and short reads were used for the genome assembly. After polishing, the assembly length obtained with the five methods ranged from 39 (Ra) to 56 Mb (Canu) with a GC content of 23-24% (Table S3). The contig-scaffolding process allowed reducing the number of contigs and increasing the N50 length with no effect on genome GC content and

CEGMA score, except for Miniasm (Table S3). Among the five methods, the Miniasm assembler returned the lowest number of contigs and the longest contig (~2 Mb) as well as the largest N50 length (425 kb). However, the completeness measured on eukaryotic BUSCO genes was the second worst (78.6%; Figure 1), casting doubt on the per-base quality of the assembly. The three assemblies MaSuRCA, Ra and Canu returned a BUSCO completeness score greater than 87% and were then selected for further steps (Figure 1).

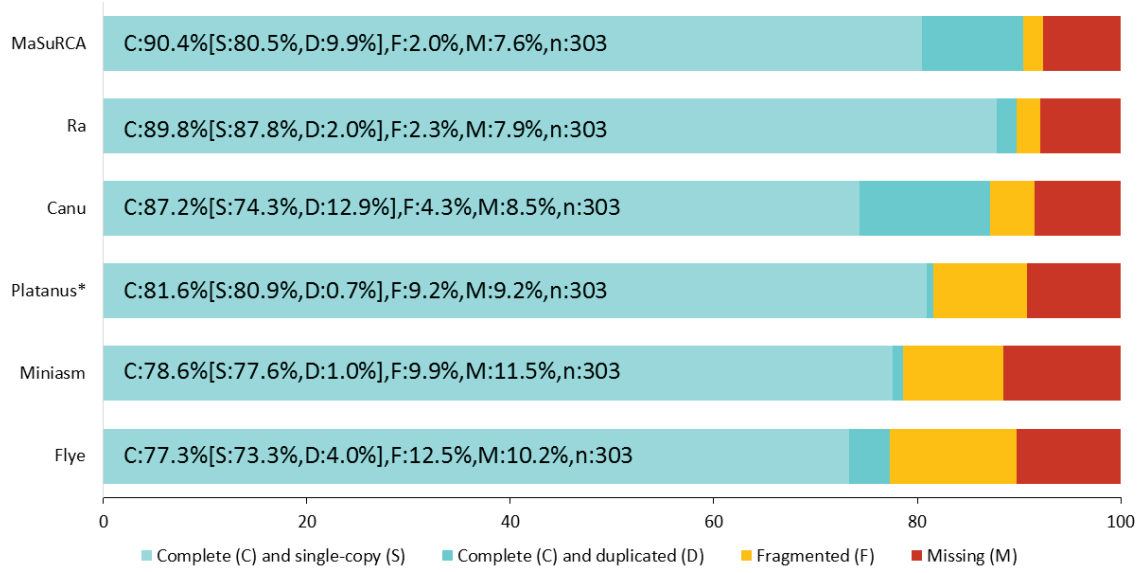


Figure 1. BUSCO completeness of genome assemblies generated with different assemblers. Five assemblies were generated in our study and are compared to the published assembly (Somvanshi et al., 2018) that was reconstructed with Platanus (indicated by the asterisk).

Read-depth analysis of MaSuRCA, Ra and Canu assemblies showed a bimodal distribution (Figure S1 - A, B, C). The half-coverage read-depth peak on Ra assembly seemed smaller than on MaSuRCA and Canu suggesting that Ra tended to create mostly collapsed haplotype contigs. After purging haplotigs and potential artifacts, the genome assembly sizes were reduced from 47.4 – 39.7 – 57.2 Mb (MaSuRCA – Ra – Canu) to 40.9 – 38.9 – 42.7 Mb, respectively (Table S4) and the peak at half coverage was almost totally absent (Figure S1 – D, E, F). At this stage, Canu showed the best assembly metrics with the longest scaffolds: 1.4 Mb for the largest contig, an N50 length of 292 kb, the smallest number of contigs (i.e. 357), and the lowest number of mismatches (i.e. 300; Table S4). A higher number of reads (long, short and RNA-seq) were mapped on the Canu assembly, suggesting a higher efficiency of the Canu

software. The genome completeness of the three assemblies remained high with a total BUSCO completeness score superior or equal to 87%. Compared to the initial assembly, the total BUSCO completeness of purged-haplotig genome slightly increased in the Canu assembly from 87.2 to 88.1%, while it decreased in the two others, from 90.4 to 89.2% in MaSuRCA and from 89.8 to 87.5% in Ra (Figure 1, Table S4). Besides, the haplotigs-purging process allowed a significant increase (+10.8%) of the completeness of single-copy genes in the Canu genome, while there was a marginal gain in genomes assembled with Ra (+2.0%) and MaSuRCA (+3.7%) (Figure 1, Table S4). In parallel, the completeness of duplicated genes in the Canu genome was strongly decreased (-9.9%) after purging haplotigs, while those were slightly reduced in Ra (-0.3%) and MaSuRCA (-4.9%). The Canu haplotyped-purged assembly, which had longer scaffolds and higher completeness, was finally selected as the reference. For the Canu assembly, artifacts (726 kb) were removed by haptotigs-purging process. Furthermore, contamination screening detected 74 contigs (total of 1.2 Mb) which had read-depth inferior to 100× and showed highest similarity (identity ≥ 70%) to Chordata phylum; therefore, these potential contaminant contigs were filtered out. After removing potential artifacts and contaminations, this final assembly was 41.5 Mb long, with 283 contigs, and an N50 length of 294 kb (Table 1). Figure S2 compares the GC content (peaking at 23%) and read-coverage of all contigs. Most of them have a sequencing depth superior to 100× [only two short contigs (i.e., mg287, 11 kb; mg295, 3 kb) with "no-hit" in the nt database showed a lower depth (68 and 83×)]. One hundred twenty-one contigs (covering ca. 29.8 Mb; 71.9% of the genome) contain genomic regions that were identified as belonging to the nematode phylum (identity ≥ 70%; Figure S2). The BUSCO and CEGMA completeness scores for the final assembly were 88.8% and 95.97%, respectively (Table 1). Reads were evenly mapped over most of the scaffolds with a mean coverage of 228× for short reads and 38× for long reads (Figure 2, Figure S4). The number of heterozygous SNVs varied from 0 to 407 per sliding window, which corresponds to nucleotide divergence ranging from 0 to 5% with a mean value of 1.36 ± 0.78% (Figures 2 and S4).

Table 1. Compared statistics of the haplotype–fused genome assemblies for *M. graminicola* obtained in our study (with Canu; Koren et al., 2017) and in Somvanshi et al. (2018).

Assembly features	Canu	Somvanshi et al. (2018)
# contigs	283	4,304
Largest contig (bp)	1,433,372	145,493
Total length (bp)	41,549,413	38,184,958
N50	294,907	20,482
N75	185,679	9,797
L50	43	522
L75	78	1,189
GC (%)	23.28	23.49
Mismatches	300	715,992
CEGMA completeness (n:248)	C:95.97%	C: 84.27%
BUSCO completeness (n:303)	C:88.8% [S:85.8%, D:3.0%]	C: 81.6 %

Flow cytometry outputted clearly G0/G1 peaks for each sample and both internal controls (Figure S3). Thanks to the presence of two internal controls, the reference DNA-content of one of them could be used as a standard to estimate the DNA-content and then the genome size of the other. The calculated genome sizes ranged from 203.9 to 221.6 Mb for *C. elegans*, and from 315.8 to 343.3 Mb for *D. melanogaster*. These estimates are relatively close to their expected genome sizes (Table S5). The genome size of *C. elegans* being closer to *M. graminicola* than *D. melanogaster*, *C. elegans* was therefore used as a standard to calculate the final DNA content of *M. graminicola* samples. The total nuclear genome size for four independent measurements of the Mg-VN18 isolate ranged from 81.5 to 83.8 Mb (average 82.6 Mb), although a fifth estimate was higher and highly suspect (103.9 Mb; Table S5; Figure S3).

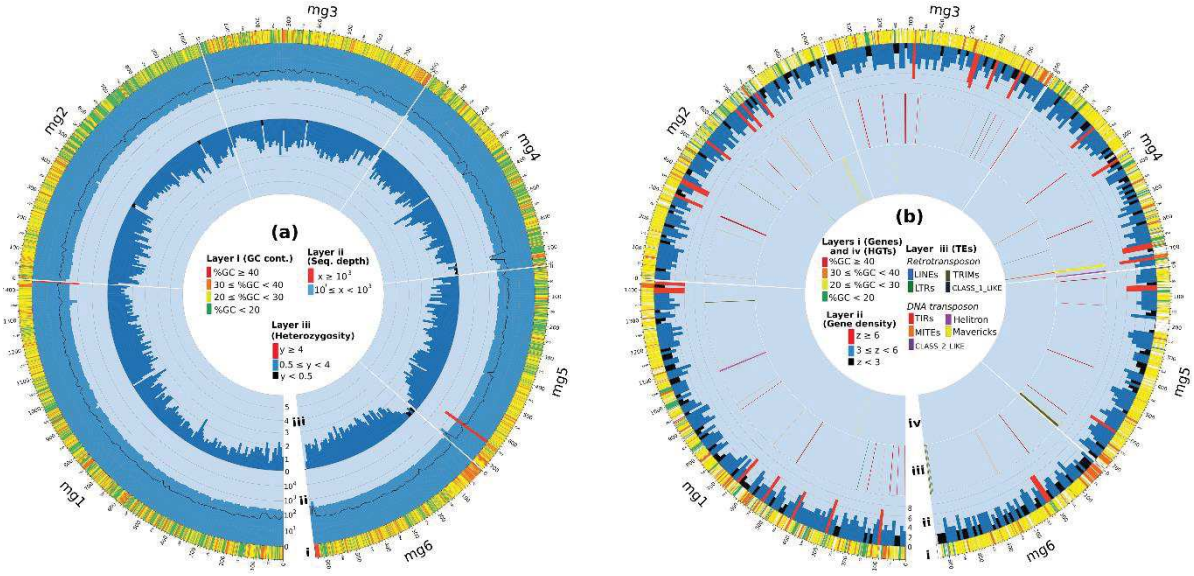


Figure 2. Genomic features along the six longest scaffolds (mg1 to mg6) with total size of 5.4 Mb. The scaffolds were sorted by length, following clockwise from the longest to the smallest one. **Circle A** shows three layers: (i) Scaffolds with length and GC content per 1-kb sliding window; (ii) Short read-depth (x, histogram) and long read-depth (black line) per 10-kb sliding window; and (iii) Histogram of heterozygous SNVs density (y) per 10-kb genome window. **Circle B** shows five layers: (i) Scaffolds with length and gene distribution on scaffold, each gene was displayed by a color representing its GC content; (ii) Histogram of gene density (z) per 10-kb genome window; (iii) Transposable elements (TEs) distribution on scaffolds with a specific color for each TE family; and (iv) Horizontal gene transfers (HGTs) distribution on scaffold, with color representing GC content of each HGT. Meaning of coded colors in each layer is also given in the middle of each circle.

3.2 Protein-coding gene annotation

A total number of 10,331 protein-coding genes were predicted with the Maker2 pipeline, of which 10,284 were selected with AED less than 1 and/or had *Pfam* and InterPro evidence (Table 2). On average, 247.5 protein-coding genes were thus annotated per Mb. The full genes and their coding sequences (CDS) spanned 75.5% (31.4 Mb) and 28.4% of the total genome length, respectively. Among them, 268 genes showed alternative splice forms, leading to the prediction of 10,631 mRNA with a total length of 32.2 Mb (Table 2). Number of exons per protein-coding gene varied from 1 to 152 with an average of 11.2 per gene and 8.4 per CDS (Figure S5A). Number of exons per gene was related to gene length (Figure S5B). On average, genes had 4.3 exons per kb, similar to that reported in four cloned genes of *M. graminicola* (on average 4.6 exons/kb in Mg01965, MgM0237, Mg16820, and MgPDI; Chen et al., 2018;

Naalden et al., 2018; Tian, Wang, Maria, Qu, & Zheng, 2019; Zhuo et al., 2019). Raw RNA-seq reads were mapped on 20 eukaryote ortholog genes, which were completely annotated by BUSCO on the *M. graminicola* genome sequence. RNA reads mapped on multiple regions on most of orthologous genes confirming dense distribution of exons in genes of *M. graminicola* (Figure S6). Intronic regions represented 31.9% of the genome, with an average of 10.2 introns per protein-coding gene. More than 60% of all introns are shorter than 60 nucleotides. Overall, the proportion of canonical splice sites is 94.65% including GT-AG (92.39%) and CT-AC (2.26%) for reversed genes. Noncanonical splice sites account for 5.34% consisting of TT-AG (0.51%), GC-AG (0.45%), and other minor splice sites (4.39%). The 5′-UTR and 3′-UTR spanned 7.6% and 9.7% of the genome, respectively. The GC content of protein-coding gene was 29.03%, and thus higher than in the whole genome. The length of the 10,631 annotated proteins ranged from ~300 to ~6,000 amino acids (Table 2). The BUSCO completeness of the predicted protein dataset was 86.5% (Table 2). Genes were located in most scaffolds (262 out of 283), and only 21 short scaffolds (<30 kb) did not bear any annotated gene (Figures 2 and S7).

Table 2. **General characteristics of protein-coding genes in the *M. graminicola* genome reconstructed with Canu** (Koren et al., 2017). CDS = coding sequence; UTR = untranslated region.

Statistics	**Protein-coding gene**	**mRNA**	**CDS**	**Exon**	**Coding exon**	**Intron**	**5'-UTR**	**3'-UTR**
Total number	10,284	10,631	10,654	115,769	88,994	105,138	6,756	6,467
Total length (kb)	31,387	32,191	11,808	19,014	11,747	13,282	3,148	4,057
% genome	75.5	77.5	28.4	45.8	28.2	31.9	7.6	9.7
Mean length (bp)	3,052	3,028	1,110	164	132	126	466	627
Longestl ength (bp)	38,328	38,328	18,477	5,281	5,821	12,794	17,158	14,255
# per protein-coding gene				11.2	8.4 per CDS	10.2		
# per Mb genome	247.51							
GC	29.03%							
BUSCO	C:86.5% [S:81.5%, D:5.0%]							

3.3 Identification and function of horizontal gene transfers

We identified 67 genes encoding 68 proteins that returned an AI > 14, indicating a possible acquisition via HGT from non-metazoan origin. All these proteins had predicted *pfam* domains, which allowed classifying them in 31 different gene families (Tables 3 and S6). Among them, 54 genes (80.9%) had strong support with AI > 26. A total of 28 genes from six families encode for several plant cell-wall modification and degradation enzymes such as polygalacturonase, xylanase, arabinase, pectate lyase, expansin-like proteins, andcellulases. Fourteen genes are possibly involved in nutrient processing (including biosynthesis of vitamins B7, glutamine, and carbohydrate), galactose and sucrose degradation, and transportation of sucrose and sugar moieties. Nine putative HGTs encode for chorismate mutase, isochorismatase, and carboxylesterase that are involved in the detoxification and modulation of plant defense. Other six HGT candidates are related to different pathways such as metabolic processes of nucleosides, amino acids, keto acids, and fatty acids. Two genes encoding peptidase and two others encoding integrase were also identified as HGTs. Other six putative HGTs encode membrane component, carbohydrate-binding module, thaumatin, unknown protein binding domain, and lysozyme (Table 3). For 92.5% of HGT candidates (62/67), the most similar sequence was of bacterial origin. For the five remaining HGTs, the most similar sequence indicated a potential origin from fungus, archaea, virus and viridiplantae (Table S6). In addition, a gene encoding cyanate lyase, which contributes to the detoxification process, was detected as an HGT with a low AI score of 4.0. Proteins related to induction of feeding site (candidate

Table 3. **Summary of putative horizontal gene transfers (HGT) in the *Meloidogyne graminicola* genome**. Putative HGTs are classified according to the general process in which they are involved. For each gene family, their supposed function(s) and the number of copies (N) are also given. The HGT detection thresholds (Alien index) are 14 or 26. More details on each gene (i.e., Alien index, genome location, and accession number) are given in Table S6. *One gene copy encodes two different proteins (see Table S6)

General Process	Gene/gene family	Function(s)	N (AI > 14)	N (AI > 26)
Plant cell wall degradation	GH28 Polygalacturonase	Pectin decorations degradations	3	3
	GH30 Xylanase	Xylan degradation	2	2
	GH43 candidate Arabinanase	Pectin decorations degradation	1	1
	PL3 Pectate lyase	Pectin degradation	10	10
	Expansin-likeproteins	Softening of non-covalent bonds	4	2
	GH5_2 Cellulases	Cellulose degradation	8	6
Plant defense manipulation	Candidate Isochorismatase	Catalyses the conversion of isochorismate	1	1
	Chorismate Mutase	Conversion of Chorismate into SA	1	1
	pnbA Carboxylesterase	Hydrolysis of ester and amide bonds	7	6
Nutrient processing	bioB Biotin synthase	Vitamin B7 biosynthesis	1	0
	Candidate GS1 Glutamine Synthetase	Nitrogen assimilation	1	1
	galM Candidate galactose mutarotase	Galactose metabolism	1	1
	GH2 β-galactosidase	Galactose degradation	1	1
	GH32 invertase	Sucrose degradation	2*	2*
	Sugar transporter (MFS) family	Transport of carbohydrates, organic alcohols and acids	4	4
	rfaG Glycosyltransferases Group 1	Catalyses the transfer of sugar moieties	4	3
Not known	Phosphoribosyl transferase	Nucleoside metabolic process	1	1
	tdk Thymidine kinase	Nucleoside metabolic process	1	0
	Candidate L-threonine aldolase	Cellular amino acid metabolic process	1	1
	Gamma-glutamyl cyclotransferase	Degrade gamma-glutamylamines to amino acid	1	1
	FAD-dependent oxidoreductase	Catalyses D-amino acids into keto acids	1	1
	HADH	Enzyme involved in fatty acid metabolism	1	0
	DJ-1/PfpI family cysteine peptidase	Degrade intracellular protein	1	1
	FtsH peptidase	Degrade membrane-embedded and soluble protein	1	1
	Integrase	Integrates the viral genome into a host chromosome	2	0
	Collagen	Cuticle and basement membrane collagen	1	1
	Phlebovirus glycoprotein G2	Component of Golgi complex membrane	1	0
	Thaumatin-like protein	Sweet-tasting protein	1	0
	Domain DUF1772	unknown	1	1
	Laminin_G_3 family	Carbonhydrate binding module	1	1
	GH25 Lys1-like	Bacteria cell wall lytic enzyme	1	1

acetyltransferase) and biosynthesis of vitamin B1 (VB1 thiD) were present in the *M. graminicola* genome. Still, none was detected as putative HGT (AI > 0). The GC content of putative HGTs (Table 3) ranged from 14 to 36% with an average value of 24%. Short read-coverage over these 67 genes ranged from 100 to 540× (with a mean value of 297×). The value close to the whole sequencing depth suggests putative HGTs were actually part of the *M. graminicola* genome. The 67 putative HGTs were located on 47 scaffolds with no apparent hotspot of foreign genes integration (Figures 2 and S7). Besides, average coverage of RNA-seq reads (at J2 stage) on 67 candidate HGTs was 729×, while the average coverage of these RNA-seq data on gene set at whole-genome level was 212×. Among them, putting aside the three genes encoding for putative integrase and glycoprotein (< 10×), 64 genes had a RNA-seq coverage superior to 30, and more interestingly, six of them encoding for putative cellulase, xylanase, and pectinase had a RNA-seq coverage superior to 1,000×.

3.4 Diversity and distribution of transposable elements

One hundred sixteen consensus sequences of repetitive elements were first identified and used as a reference library. This allowed us to annotate 4,513 loci in the genome (16.45% of the genome spanned) among which 575 presented canonical signatures of TEs. Canonical TE annotations spanned 1.08 Mb in total, representing 2.61% of the genome (Table 4). Only canonical TE annotations were then analysed in detail. DNA-transposons were slightly more abundant than retrotransposons, as they respectively covered 1.49 and 1.12 % of the genome. Three retrotransposon orders were found, including LINEs (Long Interspersed Nuclear Elements), LTRs (Long Terminal Repeats), and TRIMs (Terminal Repeat retrotransposon In Miniatures). The four detected DNA-transposons consisted of TIRs (Terminal Inverted Repeats), MITEs (Miniature inverted-repeat transposable elements), Helitrons, and Mavericks (Table 4). Interestingly, the non-autonomous TEs present in the genome (TRIMs, MITEs) accounted for 54.6% of TEs, which corresponded to 1.13% of the total genome assembly (Table 4). TEs were distributed in 195 scaffolds (Figures 2 and S7) with the highest number on scaffolds mg96 (i.e., 22 TEs, density of 1.5 TEs per 10 kb). Two of the three Maverick TEs overlapped with two putative HGT events bearing integrase core domain on scaffolds mg4 and mg32 (Figures 2 and S7).

Table 4. **Abundance and diversity of transposable elements (TEs) in the *Meloidogyne graminicola* genome**.

TE family	Number	Total length (bp)	% genome	Minimum length (bp)	Maximum length (bp)
Class I (total)	133	463,595	1.12	-	-
LINEs	5	15,552	0.04	591	5,500
LTRs	26	96,561	0.23	556	8,035
TRIMs	97	340,975	0.82	420	9,959
CLASS_1_LIKE	5	10,507	0.03	417	4,462
Class II (total)	442	621,066	1.49	-	-
TIRs	202	366,687	0.88	387	10,090
MITEs	217	129,768	0.31	258	1,440
Helitrons	16	89,046	0.21	2,842	7,557
Mavericks	3	23,131	0.06	4,441	9,346
CLASS_2_LIKE	4	12,434	0.03	1,506	6,703
Total	575	1,084,661	2.61	-	-

IV. Discussion

4.1 A highly complete and contiguous genome revealed peculiar features in *M. graminicola*

By optimizing DNA extraction methods and utilizing the advantages of long-read sequencing, the genome assembly of *M. graminicola* is here greatly improved compared with the previously published version (Somvanshi et al., 2018). This new genome presents better completeness and a larger genome size with ten times fewer scaffolds. This new assembly yields the second largest N50 length (294 kb) among all *Meloidogyne* genomes publicly available to date (summarized in Susič et al., 2020). The removal of haplotigs and potential contaminants on genome sequence provides a clean genetic material, reducing errors in downstream analyses. Finally, this haplotype-merged assembly is highly complete regarding CEGMA and BUSCO scores when compared to available RKN genomes (summarized in Koutsovoulos et al., 2019). A higher number of exons per gene (11.2) was detected in *M. graminicola* compared with other

PPN species [e.g., ~6 in mitotic RKN (Blanc-Mathieu et al., 2017); 8.8 in *Globodera rostochiensis* (Akker et al., 2016)]. Frequent noncanonical splice sites (5.34%) were detected in predicted genes of *M. graminicola*, as similarly reported in other nematode species belonging to sister genera [e.g., 3.47% in *G. rostochiensis* (Akker et al., 2016); 4.29% in *Heterodera glycine* (Masonbrink et al., 2019)]. In contrast, a quasiabsence of noncanonical splice sites was reported in RKN species (Akker et al., 2016), but this may be due to restrictive settings during gene annotation in this group. Interestingly, while mainly GC-AG introns were found as noncanonical in cyst nematode species, several other minor noncanonical splice sites were detected in *M. graminicola*. Such a diversity could be related to an extremely low GC content (23%). In plants and worms, AT content has been demonstrated to represent an important determinant of intron recognition (Aroian et al., 1993; Luehrsen & Walbot, 1994). Notably, nematodes have unique features (e.g., *trans*-splicing, diverse spliced leader) allowing them to develop specific ways of constructing and altering their genome expression (Barnes et al., 2019; Davis, 1996). Besides, it has been demonstrated that spliceosome mutation of *C. elegans* can lead to recognition of variant sequences at both ends of introns (Aroian et al., 1993). Therefore, we can hypothesize that the *M. graminicola* spliceosome has evolved toward small introns and flexible noncanonical sites recognition, but anyhow further studies are required to support this assumption.

The haploid genome length calculated by *k*-mer analysis using Illumina reads ranges from 41.1 to 41.6 Mb, which is very similar to the final genome assembly (41.5 Mb). Furthermore, the experimentally measured total DNA content over four replicates ranges from 81.5 to 83.8 Mb, which corresponds to a haploid genome size ranging between 40.7 and 41.9 Mb. These measures suggest our genome assembly is almost complete and corresponds to a haploid genome with merged haplotypes on most genomic regions. This is similar to the facultative sexual *M. hapla*, which indicates a canonical sexual diploid genome (Blanc-Mathieu et al., 2017). The heterozygosity between haplotypes ranges from 1.69% to 1.90%, according to the *k*-mer analysis and is 1.36% ± 0.78 based on the SNV analysis. In *M. hapla*, meiotic parthenogenesis occurs via terminal fusion (fusion of the terminal products after the two meiotic divisions), which is supposed to homogenize the genome and eventually yield low heterozygosity (Castagnone-Sereno et al., 2013; Triantaphyllou, 1985). In that perspective, the relatively high heterozygosity in *M. graminicola* is unexpected. It suggests either a different mechanism (i.e., the central fusion of the products of the first division of meiosis) or more

frequent outcrossing events. The exact reproductive mode of *M. graminicola* thus still needs more investigation, particularly for documenting the process of genome segregation during meiosis.

4.2 Evidence of horizontal gene transfers in the *M. graminicola* genome

We identified several robust horizontal gene transfer (HGT) candidates in the *M. graminicola* genome (i.e., 54 genes with AI > 26; Table 3). Many of these genes are predicted to play a role in the degradation of the plant cell wall (44%), which represents a crucial role in parasitism by allowing the migration of parasites in the root tissue. In addition, other HGTs are potentially involved in nutrient biosynthesis and processing, detoxification, and hijack of host plant defenses (Haegeman et al., 2011). A comparison of HGTs discovered in this study with those already known in other RKNs reveals common characteristics, in particular 12 gene families that were phylogenetically supported as HGTs in other PPNs (Table S6). Among them, HGTs encode six plant cell-wall degradation enzymes, two nutrients processing enzymes, two plant defenses manipulation enzymes, and two unknown proteins, which are all described in details in Appendix S1. In addition, new HGT candidates, not previously described so far in other *Meloidogyne* and with comparably high AI values, are here identified. Specificities of those putative HGTs in *M. graminicola* are following summarized by considering the process they are supposedly involved in:

4.2.1 Plant defense manipulation and detoxification

As in other PPNs, candidate HGT genes encoding for chorismate mutase, isochorismate synthase, and cyanate lyases are also found in *M. graminicola*. In addition, seven genes encoding carboxylesterases are firstly reported as HGTs in *M. graminicola*. These carboxylesterases might help this parasite to detoxify ester-containing xenobiotics that are present in phytoalexins secreted by plants in response to nematode infection (Gillet, Bournaud, de Souza, Júnior, & Grossi-de-Sa, 2017; Hatfield et al., 2016; Shukla et al., 2017).

4.2.2 Nutrient processing

Some HGTs involved in biosynthesis and process of nutrients have been previously reported in PPNs (Danchin, Guzeeva, Mantelin, Berepiki, & Jones, 2016). Unlike other PPNs, *M. graminicola* has more putative HGTs involved in the metabolism linked to the carbohydrate pathways and fewer genes linked to the biosynthesis of vitamins. For instance,

only the GH32 gene family related to sucrose degradation has been reported as a HGT in PPNs (Danchin et al., 2016), but we here reveal that 11 *M. graminicola* genes involved in carbohydrate metabolism, galactose degradation, and sugar transport should result from horizontal transfers. Notably, multigenic families encoding for sugar transporters and glycosyltransferase present a high Alien Index (>300) strongly supporting their foreign origin. Interestingly, sugar transporters carry sucrose into the syncytium made by cyst nematodes (*Heterodera* spp.) at the early stage of infection before the establishment of plasmodesmatal connections between the feeding site and the phloem (Zhao et al., 2018). Therefore, such sugar transporters must play a critical role at the early stage of parasitism. In contrast, while nine HGTs involved in the synthesis or salvage of the four vitamins B1, B5, B6, B7 are found in cyst nematode (Craig, Bekal, Niblack, Domier, & Lambert, 2009), *M. graminicola* only acquired a single gene encoding vitamin B7 from bacteria. This HGT was not detected in *M. incognita*, which, however, acquired HGTs for two other genes encoding vitamins (i.e., B1 and B5; Craig et al., 2009).

4.2.3 Other functions

Novel presumed HGTs with a potential contribution to nematode infection are also detected in *M. graminicola* for the first time: (a) Firstly, *M. graminicola* has a candidate GH25 lysozyme likely acquired by HGT and this enzyme could participate in cell division and cell-wall remodeling in bacteria (Vollmer, Joris, Charlier, & Foster, 2008) and bacteriophages (Fastrez, 1996). Consequently, this gene is suspected of playing a role in the invasion of root tissue (Paganini et al., 2012), but its precise function still remains unknown. (b) Secondly, the HGT candidate with the highest AI (i.e., 370) encodes a protein bearing laminin_G_3 domain belonging to the concanavalin A-like lectin/glucanases superfamily. This gene is suggested to contribute to cell-wall degradation process because it acts as a carbohydrate-binding module and contributes for the hydrolysis activity of arabinofuranosidase (Sakka, Kunitake, Kimura, & Sakka, 2019). (c) Thirdly, in addition to two HGTs putatively involved in the nucleoside metabolic process (candidate phosphoribosyltransferase) and amino acid metabolism (candidate L-threonine aldolase) previously reported among other PPNs (Danchin et al., 2016), two other genes (encoding candidate gamma-glutamylcyclotransferase and thymidine kinase) possibly involved in these processes are found for the first time as putative HGTs in *M. graminicola*. (d) Fourthly, *M. graminicola* has potentially laterally acquired genes for protein degradation and keto acid and fatty acid metabolism. Although there is no clearly

defined nematode requirements for these nutriments (i.e., amino acids, fatty acids, keto acids, nucleosides, and acid amins), they are thought to be necessary for PPN development (Goheen, Campbell, & Donald, 2013). Therefore, these HGTs are suspected to contribute to nematodes living inside root tissues. v) Finally, two genes coding for integrase enzymes, which may promote the integration of HGTs into the host chromosome, are also identified as HGTs. Interestingly, these genes are associated with TEs (see Results on "Diversity and distribution of transposable elements") that potentially created more copies of these genes in the genome. Therefore, they could have themselves contributed to the HGT events observed in *M. graminicola*.

Most of these putative HGTs found in the *M. graminicola* genome may thus play a crucial role in nematode infection, nutrition requirements, and suppression of plant defenses as already shown in other PPNs (Craig et al., 2009; Danchin et al., 2016; Haegeman et al., 2011). Therefore, these HGTs acquired by *M. graminicola* during its evolution have likely contributed to its successful parasitism. Most of these genes, however, have not yet been subjected to functional validation and detailed phylogenetic analysis, so additional studies are still required to identify putative donors and precise the timing of their acquisition and spread.

4.3 Diversity and abundance of transposable elements in *M. graminicola*

Transposable elements (TEs) are DNA sequences with the ability to move and to make copies within the genome causing changes in its structure and organization, contributing among other things to the evolution of species (Bonchev & Parisod, 2013; Serrato-Capuchina & Matute, 2018). More than half of the *M. graminicola* TEs are nonautonomous transposons that have lost their transposition machinery. TEs have been annotated in the genomes of other RKNs, including mitotic and meiotic parthenogenetic species (Blanc-Mathieu et al., 2017; Koutsovoulos et al., 2019). However, as the software version used to annotate the genomes and filters to retrieve canonical TEs was different in each study, the abundance of TEs detected in the facultative meiotic parthenogenetic *M. graminicola* is not directly comparable to other species. The TEs load seems to be higher in mitotic parthenogenetic RKNs than in the facultative sexual *M. hapla* (Blanc-Mathieu et al., 2017). In *M. enterolobii,* a mitotic parthenogenetic RKN, more nonautonomous TEs were detected (3.12% genome size) than in *M. incognita* and *M. javanica* (2.27% and 1.63%, respectively; Koutsovoulos et al., 2019). Considering TE diversity, certain retrotransposon families previously detected in mitotic

parthenogenetic RKNs, such as DIRS, SINE, and LARD, are not found in *M. graminicola*. Interestingly, the *Cg-1* gene, whose deletion is associated with resistance-breaking strains of *M. javanica*, has been identified within one transposon (Tm1) belonging to the TIR superfamily suggesting an adaptive impact of TEs on nematode genomes (Gross & Williamson, 2011). Notably, homologs of the Tm1 transposon are also found in the *M. graminicola* genome but not in *M. hapla*. We also found that two copies of Mavericks bear a HGT encoding DNA integrase, suggesting that some TEs might have been laterally transferred from bacteria to the *M. graminicola* genome.

V. Conclusion and perspectives

This new and more complete genome sequence of *M. graminicola* has immediate and important implications for research on the evolutionary biology of this pathogen and on other broader studies of phytoparasitic nematodes. Notably, the high contiguity of the genome presented here enabled us to produce important genetic information, including gene structure and TE/HGT content. This decisive step allows a diversity of investigations at both intra- and interspecies levels to decipher geographic origin and diffusion of *M. graminicola*, to investigate genome evolution of RKNs associated with their adaptation to different environmental conditions and hosts, and to understand deeper of their evolutionary history.

Data Accessibility

All genomic raw sequence reads are accessible as NCBI BioProject with the id [currently in submission “Genome submission: SUB7179509”] and the genome assembly data have been deposited at GenBank under Accession no PRJNA615787. Procedural information concerning the genome assembly and analysis presented in this paper can be found at the GitHub repository at https://github.com/PhanNgan/Genome_Assembly_MG.

Acknowledgements

This research was funded by the Consultative Group for International Agricultural Research Program on rice-agrifood systems (CRP-RICE, 2017–2022). Thi Ngan Phan was supported by a PhD fellowship from French Embassy in Vietnam. Guillaume Besnard is member of the EDB

laboratory supported by the excellence projects Labex CEBA (ANR-10-LABX-25-01) and Labex TULIP (ANR-10-LABX-0041), managed by the French ANR. The authors want to thank Jamel Aribi (IRD-IPME, France), Michel Lebrun and Thi Hue Nguyen ("Rice Functional Genomics and Plant Biology" International Joint Laboratory, Hanoi, Vietnam) for their technical support in nematology. We also want to thank Ndomassi Tando and the IRD itrop ''Plantes Santé'' bioinformatic platform for providing HPC resources and support for our research project.

References

Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M.-C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., … Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, *26*, 909. https://doi.org/10.1038/nbt.1482

Akker, S. E. den, Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G. J., Rocha, M. D., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M.-O., Boyes, I., Marvin, J. M. C., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., … Jones, J. T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis* , reveals insights into the basis of parasitism and virulence. *Genome Biology*, *17*(1), 124. https://doi.org/10.1186/s13059-016-0985-1

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aroian, R. V., Levy, A. D., Koga, M., Ohshima, Y., Kramer, J. M., & Sternberg, P. W. (1993). Splicing in *Caenorhabditis elegans* does not require an AG at the 3' splice acceptor site. *Molecular and Cellular Biology*, *13*(1), 626–637. https://doi.org/10.1128/mcb.13.1.626

Barnes, S. N., Masonbrink, R. E., Maier, T. R., Seetharam, A., Sindhu, A. S., Severin, A. J., & Baum, T. J. (2019). *Heterodera glycines* utilizes promiscuous spliced leaders and demonstrates a unique preference for a species-specific spliced leader over C. elegans SL1. *Scientific Reports*, *9*(1), 1356. https://doi.org/10.1038/s41598-018-37857-0

Bellafiore, S., Jougla, C., Chapuis, É., Besnard, G., Suong, M., Vu, P. N., De Waele, D., Gantet, P., & Thi, X. N. (2015). Intraspecific variability of the facultative meiotic parthenogenetic root-knot nematode (*Meloidogyne graminicola*) from rice fields in Vietnam. *Comptes Rendus Biologies*, *338*(7), 471–483. https://doi.org/10.1016/j.crvi.2015.04.002

Besnard, G., Thi-Phan, N., Ho-Bich, H., Dereeper, A., Trang Nguyen, H., Quénéhervé, P., Aribi, J., & Bellafiore, S. (2019). On the close relatedness of two rice-parasitic root-knot nematode species and the recent expansion of *Meloidogyne graminicola* in Southeast Asia. *Genes*, *10*(2), 175. https://doi.org/10.3390/genes10020175

Bird, D. M., Williamson, V. M., Abad, P., McCarter, J., Danchin, E. G., Castagnone-Sereno, P. , & Opperman, C. H. (2009). The genomes of root-knot nematodes. *Annual Review of Phytopathology*, 47: 333-51.https://doi.org/10.1146/annurev-phyto-080508-081839

Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J.-M., Rocha, M. D., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J.-F., Kozlowski, D. K., Cazareth,

J., Couloux, A., Silva, C. D., Guy, J., Kim-Jo, Y.-J., Rancurel, C., Schiex, T., Abad, P., … Danchin, E. G. J. (2017). Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genetics*, *13*(6), e1006777.

Bonchev, G., & Parisod, C. (2013). Transposable elements and microevolutionary changes in natural populations. *Molecular Ecology Resources*,*13*(5) 765–775.https://doi.org/10.1111/1755-0998.12133

Bosco, G., Campbell, P., Leiva-Neto, J. T., & Markow, T. A. (2007). Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics*, *177*(3), 1277–1290. https://doi.org/10.1534/genetics.107.075069

Campbell, M. S., Holt, C., Moore, B., & Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *48*, 4.11.1-4.11.39. https://doi.org/10.1002/0471250953.bi0411s48

Castagnone-Sereno, P., & Danchin, E. G. J. (2014). Parasitic success without sex – the nematode experience. *Journal of Evolutionary Biology*, *27*(7), 1323–1333. https://doi.org/10.1111/jeb.12337

Castagnone-Sereno, P., Danchin, E. G. J., Perfus-Barbeoch, L., & Abad, P. (2013). Diversity and evolution of root-knot nematodes, genus *Meloidogyne*: new insights from the genomic era. *Annual Review of Phytopathology*, *51*, 203–220. https://doi.org/10.1146/annurev-phyto-082712-102300

Chen, J., Hu, L., Sun, L., Lin, B., Huang, K., Zhuo, K., & Liao, J. (2018). A novel *Meloidogyne graminicola* effector, MgMO237, interacts with multiple host defence-related proteins to manipulate plant basal immunity and promote parasitism. *Molecular Plant Pathology*, *19*(8), 1942–1955. https://doi.org/10.1111/mpp.12671

Craig, J. P., Bekal, S., Niblack, T., Domier, L., & Lambert, K. N. (2009). Evidence for horizontally transferred genes involved in the biosynthesis of vitamin B1, B5, and B7 in *Heterodera glycines*. *Journal of Nematology*, *41*(4), 281–290.

Danchin, E. G. J., & Rosso, M.-N. (2012). Lateral gene transfers have polished animal genomes: lessons from nematodes. *Frontiers in Cellular and Infection Microbiology*, *2*, 27. https://doi.org/10.3389/fcimb.2012.00027

Danchin, E. G. J., Rosso, M.-N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B., & Abad, P. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17651–17656. https://doi.org/10.1073/pnas.1008486107

Danchin, G. J. E., Guzeeva, A. E., Mantelin, S., Berepiki, A., & Jones, T. J. (2016). Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. *Molecular Biology and Evolution*, *33*(6), 1571–1579. https://doi.org/10.1093/molbev/msw041

Davis, R. E. (1996). Spliced leader RNA trans-splicing in metazoa. Parasitology Today (Personal Ed.), 12(1), 33–40. https://doi.org/10.1016/0169-4758(96)80643-0

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

De Ley, I. T., Karssen, G., De Ley, P., Vierstraete, A., Waeyenberge, L., Moens, M., & Vanfleteren, J. (1999). Phylogenetic analyses of internal transcribed spacer region sequences within *Meloidogyne*. *Journal of Nematology*, *31*(4), 530–531.

Dutta, T. (2012). Global status of rice root-knot nematode, *Meloidogyne graminicola*. *African Journal of Microbiology Research*. https://www.academia.edu/10533467/Global_status_of_rice_root-knot_nematode_Meloidogyne_graminicola

EPPO Global Database.European and Mediterranean Plant Protection Organization. https://gd.eppo.int/

Fanelli, E., Cotroneo, A., Carisio, L., Troccoli, A., Grosso, S., Boero, C., Capriglia, F., & De Luca, F. (2017). Detection and molecular characterization of the rice root-knot nematode *Meloidogyne graminicola* in Italy. *European Journal of Plant Pathology*, *149*(2), 467–476. https://doi.org/10.1007/s10658-017-1196-7

Fastrez, J. (1996). Phage lysozymes. *EXS*, *75*, 35–64. https://doi.org/10.1007/978-3-0348-9225-4_3

Flutre, T., Duprat, E., Feuillet, C., & Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PLoS ONE*, *6*(1), e16526. https://doi.org/10.1371/journal.pone.0016526

Gillet, F.-X., Bournaud, C., Antonino de Souza Júnior, J. D., & Grossi-de-Sa, M. F. (2017). Plant-parasitic nematodes: towards understanding molecular players in stress responses. *Annals of Botany*, *119*(5), 775–789. https://doi.org/10.1093/aob/mcw260

Goheen, S. C., Campbell, J. A., & Donald, P. (2013). Nutritional requirements of soybean cyst nematodes. *Soybean - Pest Resistance*. https://doi.org/10.5772/54247

Golden, A. M., & Birchfield, W. (1965). *Meloidogyne graminicola* (Heteroderidae) a new species of root-knot nematode from grass. *Proceedings of the Helminthological Society of Washington*, *32*(2), 228–231.

Golden, A. M., & Birchfield, W. (1968). Rice root-knot nematode (*Meloidogyne graminicola*) as a new pest of rice. *Plant Disease Reporter*, *52*(6), 423.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Gross, S. M., & Williamson, V. M. (2011). Tm1: A *mutator*/*foldback* transposable element family in root-knot nematodes. *PLoS ONE*, *6*(9), e24534. https://doi.org/10.1371/journal.pone.0024534

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haegeman, A., Jones, J. T., & Danchin, E. G. J. (2011). Horizontal gene transfer in nematodes: A catalyst for plant parasitism? *Molecular Plant-Microbe Interactions*, *24*(8), 879–887. https://doi.org/10.1094/MPMI-03-11-0055

Hall, B., DeRego, T., & Geib, S. (2014). *GAG: the Genome Annotation Generator (Version 1.0) [Software]*. http://genomeannotation.github.io/GAG.

Hatfield, M. J., Umans, R. A., Hyatt, J. L., Edwards, C. C., Wierdl, M., Tsurkan, L., Taylor, M. R., & Potter, P. M. (2016). Carboxylesterases: General detoxifying enzymes. *Chemico-Biological Interactions*, *259*(Pt B), 327–331. https://doi.org/10.1016/j.cbi.2016.02.011

Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491. https://doi.org/10.1186/1471-2105-12-491

Jiang, H., Lei, R., Ding, S.-W., & Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, *15*(1), 182. https://doi.org/10.1186/1471-2105-15-182

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., Kohara, Y., Fujiyama, A., Hayashi, T., & Itoh, T. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, *24*(8), 1384–1395. https://doi.org/10.1101/gr.170720.113

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59. https://doi.org/10.1186/1471-2105-5-59

Koutsovoulos, G. (2018). *Scaffolding contigs with transcripts* [Python]. https://github.com/GDKO/SCUBAT2 (Original work published 2014)

Koutsovoulos, G. D., Poullet, M., Ashry, A. E., Kozlowski, D. K., Sallet, E., Rocha, M. D., Martin-Jimenez, C., Perfus-Barbeoch, L., Frey, J.-E., Ahrens, C., Kiewnick, S., & Danchin, E. G. J. (2019). The polyploid genome of the mitotic parthenogenetic root-knot nematode *Meloidogyne enterolobii*. *BioRxiv*, 586818. https://doi.org/10.1101/586818

Ku, C., Martin, F. W. (2016). A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70 % rule. *BMC Biology*, 14(1): 89. https://doi.org/10.1186/s12915-016-0315-9.

Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., & Blaxter, M. (2013). Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Frontiers in Genetics*, *4*, 237. https://doi.org/10.3389/fgene.2013.00237

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287. https://doi.org/10.12688/f1000research.12232.1

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, *32*(14), 2103–2110. https://doi.org/10.1093/bioinformatics/btw152

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, Y., Schroeder, J., & Schmidt B. (2013). Musket: a multistage k-mer spectrum based error corrector for Illumina sequence data. *Bioinformatics*, *29*(3), 308–315. https://doi.org/10.1093/bioinformatics/bts690

Liu, Q.L., Thomas, V.P., Williamson, V.M., 2007. Meiotic parthenogenesis in a root-knot nematode results in rapid genomic homozygosity. Genetics 176, 1483–1490. https://doi.org/10.1534/genetics.107.071134

Luehrsen, K. R., & Walbot, V. (1994). Intron creation and polyadenylation in maize are directed by AU-rich RNA. *Genes & Development*, *8*(9), 1117–1130. https://doi.org/10.1101/gad.8.9.1117

Lunt, D. H., Kumar, S., Koutsovoulos, G., & Blaxter, M. L. (2014). The complex hybrid origins of the root knot nematodes revealed through comparative genomics. *PeerJ*, *2*, e356. https://doi.org/10.7717/peerj.356

Mantelin, S., Bellafiore, S., & Kyndt, T. (2017). *Meloidogyne graminicola*: a major threat to rice agriculture. *Molecular Plant Pathology*, *18*(1), 3–15. https://doi.org/10.1111/mpp.12394

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics (Oxford, England)*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Masonbrink, R., Maier, T. R., Muppirala, U., Seetharam, A. S., Lord, E., Juvale, P. S., Schmutz, J., Johnson, N. T., Korkin, D., Mitchum, M. G., Mimee, B., den Akker, S. E., Hudson, M., Severin, A. J., & Baum, T. J. (2019). The genome of the soybean cyst nematode (*Heterodera glycines*) reveals complex patterns of duplications involved in the evolution of parasitism genes. *BMC Genomics*, *20*(1), 119. https://doi.org/10.1186/s12864-019-5485-8

McClure, M. A., Kruk, T. H., & Misaghi, I. (1973). A method for obtaining quantities of clean *Meloidogyne* eggs. *Journal of Nematology*, *5*(3), 230.

Megen, H. van, van den Elsen,S., Holterman, M., Karssen, G., Mooyman, P., Bongers, T., Holovachov, O., Bakker, J. &Helder, J. (2009). A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology*, 11(6): 927–950.

Mgwatyu, Y., Stander, A. A., Ferreira, S., Williams, W., & Hesse, U. (2020). Rooibos (*Aspalathus linearis*) Genome size estimation using flow cytometry and *k*-mer analyses. *Plants*, *9*(2), 270. https://doi.org/10.3390/plants9020270

Naalden, D., Haegeman, A., de Almeida-Engler, J., Birhane Eshetu, F., Bauters, L., & Gheysen, G. (2018). The *Meloidogyne graminicola* effector Mg16820 is secreted in the apoplast and cytoplasm to suppress plant host defense responses. *Molecular Plant Pathology*, *19*(11), 2416–2430. https://doi.org/10.1111/mpp.12719

Narasimhamurthy, H. B., Ravindra, H., Sehgal, M., Rani, N., Ekabote, S. D., &Ganapathi. (2018). Biology and life cycle of rice root-knot nematode (*Meloidogyne graminicola*). *Journal of Entomology and Zoology Studies*, *6*(1), 477–479.

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., & Stamatoyannopoulos, J. A. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics*, *28*(14), 1919–1920. https://doi.org/10.1093/bioinformatics/bts277

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., & Windham, E. (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences*, *105*(39), 14802–14807. https://doi.org/10.1073/pnas.0805946105

Ormerod, G. M. (2008). *Flow Cytometry - A basic introduction*.Ormerod MG (Ed.), Null edition, 126 p.

Paganini, J., Campan-Fournier, A., Da Rocha, M., Gouret, P., Pontarotti, P., Wajnberg, E., Abad, P., & Danchin, E. G. J. (2012). Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes. *PLoS ONE*, *7*(11), e50875. https://doi.org/10.1371/journal.pone.0050875

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, *23*(9), 1061–1067. https://doi.org/10.1093/bioinformatics/btm071

Perfus-Barbeoch, L., Castagnone-Sereno, P., Reichelt, M., Fneich, S., Roquis, D., Pratx, L., Cosseau, C., Grunau, C., & Abad, P. (2014). Elucidating the molecular bases of epigenetic inheritance in non-model invertebrates: the case of the root-knot nematode *Meloidogyne incognita*. *Frontiers in Physiology*, *5*, 211. https://doi.org/10.3389/fphys.2014.00211

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Petitot, A.-S., Dereeper, A., Agbessi, M., Da Silva, C., Guy, J., Ardisson, M., & Fernandez, D. (2016). Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants. *Molecular Plant Pathology*, *17*(6), 860–874. https://doi.org/10.1111/mpp.12334

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rancurel, C., Legrand, L., & Danchin, E. G. J. (2017). Alienness: Rapid Detection of Candidate Horizontal Gene Transfers across the Tree of Life. *Genes*, *8*(10), 248. https://doi.org/10.3390/genes8100248

Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, *19*(1), 460. https://doi.org/10.1186/s12859-018-2485-7

Sakka, M., Kunitake, E., Kimura, T., & Sakka, K. (2019). Function of a laminin_G_3 module as a carbohydrate-binding module in an arabinofuranosidase from *Ruminiclostridium josui*. *FEBS Letters*, *593*(1), 42–51. https://doi.org/10.1002/1873-3468.13283

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Molecular cloning: a laboratory manual. *Molecular Cloning: A Laboratory Manual.*, *Ed. 2*. https://www.cabdirect.org/cabdirect/abstract/19901616061

Schaad, N. W., & Walker, J. T. (1975). The use of density-gradient centrifugation for the purification of eggs of *Meloidogyne* spp. *Journal of Nematology*, *7*(2), 203–204.

Scholl, E. H., Thorne, J. L., McCarter, J. P., & Bird, D. M. (2003). Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biology*, *4*(6), R39. https://doi.org/10.1186/gb-2003-4-6-r39

Serrato-Capuchina, A., & Matute, D. R. (2018). The role of transposable elements in speciation. *Genes*, *9*(5), 254. https://doi.org/10.3390/genes9050254

Shukla, N., Yadav, R., Kaur, P., Rasmussen, S., Goel, S., Agarwal, M., Jagannath, A., Gupta, R., & Kumar, A. (2017). Transcriptome analysis of root-knot nematode (*Meloidogyne incognita*)-infected tomato (*Solanum lycopersicum*) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses. *Molecular Plant Pathology*, *19*(3), 615–633. https://doi.org/10.1111/mpp.12547

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Somvanshi, V. ., Tathode, M., Shukla, R. N., & Rao, U. (2018). Nematode genome announcement: A draft genome for rice root-knot nematode, *Meloidogyne graminicola*. *Journal of Nematology*, *50*(2), 111–116. https://doi.org/doi:10.21307/jofnem-2018-018.

Susič, N., Koutsovoulos, G. D., Riccio, C., Danchin, E. G. J., Blaxter, M. L., Lunt, D. H., Strajnar, P., Širca, S., Urek, G., & Stare, B. G. (2020). Genome sequence of the root-knot nematode *Meloidogyne luci*. *Journal of Nematology*, 52, 1–5. https://doi.org/10.21307/jofnem-2020-025

Szitenberg, A., Salazar-Jaramillo, L., Blok, V. C., Laetsch, D. R., Joseph, S., Williamson, V. M., Blaxter, M. L., & Lunt, D. H. (2017). Comparative genomics of apomictic root-knot nematodes: Hybridization, ploidy, and dynamic genome change. *Genome Biology and Evolution*, *9*(10), 2844–2861. https://doi.org/10.1093/gbe/evx201

The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, *282*(5396), 2012–2018. https://doi.org/10.1126/science.282.5396.2012

Tian, Z., Wang, Z., Maria, M., Qu, N., & Zheng, J. (2019). *Meloidogyne graminicola* protein disulfide isomerase may be a nematode effector and is involved in protection against oxidative damage. *Scientific Reports*, *9*(1), 11949. https://doi.org/10.1038/s41598-019-48474-w

Tranchant-Dubreuil, C., Ravel, S., Monat, C., Sarah, G., Diallo, A., Helou, L., Dereeper, A., Tando, N., Orjuela-Bouniol, J., & Sabot, F. (2018). TOGGLe, a flexible framework for easily building complex workflows and performing robust large-scale NGS analyses. *BioRxiv*, 245480. https://doi.org/10.1101/245480

Triantaphyllou, A. C. (1985). Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes. In J. N. Sasser & C. C. Carter (Eds.), *An advanced treatise on* Meloidogyne*. Biology and control*(Vol. 1, pp. 113–126). Raleigh. http://agris.fao.org/agris-search/search.do?recordID=US8743737

Triantaphyllou, A. C. (1987). Cytogenetic status of *Meloidogyne* (*Hypsoperine*) *spartinae* in relation to other *Meloidogyne* species. *Journal of Nematology*, *19*(1), 1–7.

Vaser, R., & Šikić, M. (2019). Yet another de novo genome assembler. *BioRxiv*, 656306. https://doi.org/10.1101/656306

Vaser, R., Sovic, I., Nagarajan, N., & Sikic, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746. https://doi.org/10.1101/gr.214270.116

Vollmer, W., Joris, B., Charlier, P., & Foster, S. (2008). Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiology Reviews*, *32*(2), 259–286. https://doi.org/10.1111/j.1574-6976.2007.00099.x

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Wick, R. (2019). *Porechop* [C++]. https://github.com/rrwick/Porechop (Original work published 2017)

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics (Oxford, England)*, *17*(9), 847–848. https://doi.org/10.1093/bioinformatics/17.9.847

Zhao, D., You, Y., Fan, H., Zhu, X., Wang, Y., Duan, Y., Xuan, Y., & Chen, L. (2018). The role of sugar transporter genes during early infection by root-knot nematodes. *International Journal of Molecular Sciences*, *19*(1), 302. https://doi.org/10.3390/ijms19010302

Zhuo, K., Naalden, D., Nowak, S., Xuan Huy, N., Bauters, L., & Gheysen, G. (2019). A *Meloidogyne graminicola* C-type lectin, Mg01965, is secreted into the host apoplast to suppress plant defence and promote parasitism. *Molecular Plant Pathology*, *20*(3), 346–355. https://doi.org/10.1111/mpp.12759

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, *29*(21), 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Supporting Information

Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*)

Thi Ngan Phan, Julie Orjuela, Etienne G.J. Danchin, *et al.*

Supporting Information contains:

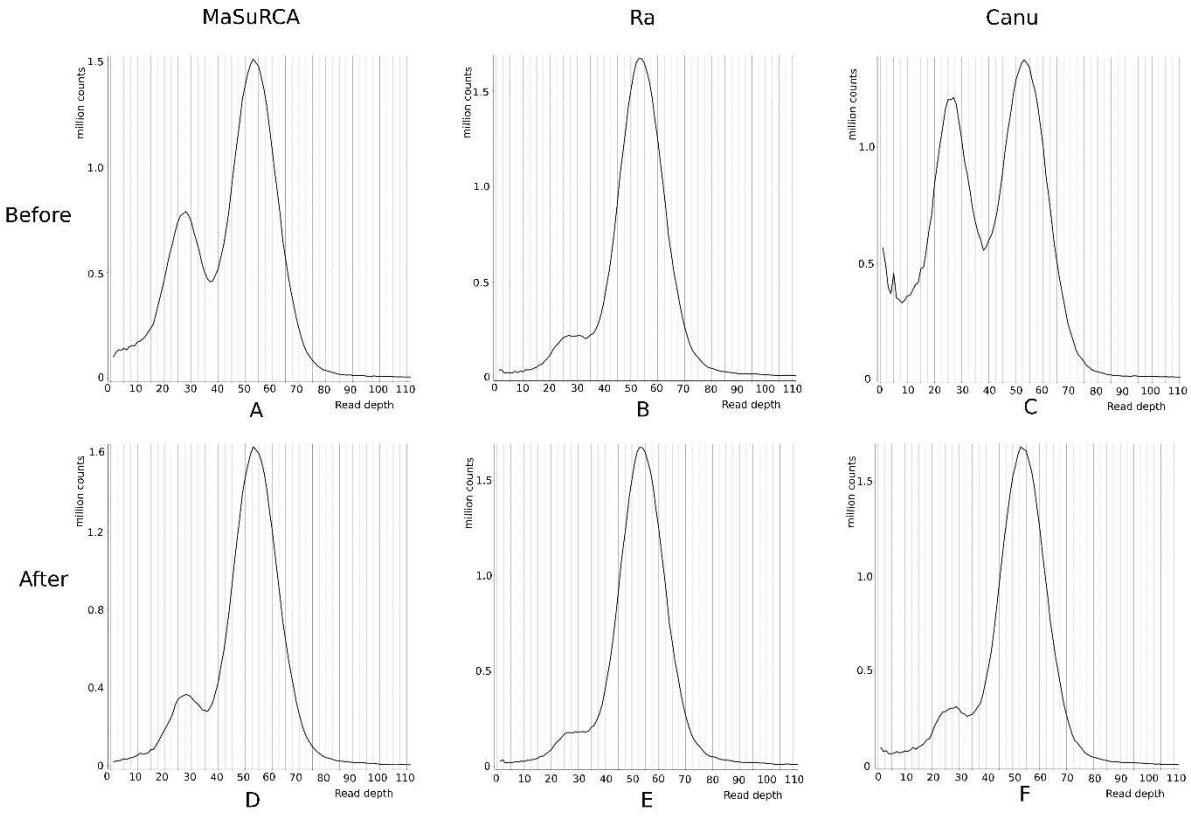


Figure S1. Long read-coverage analysis of MaSuRCA, Ra and Canu assemblies before (A, B, C) and after (D, E, F) purging haplotigs using long reads

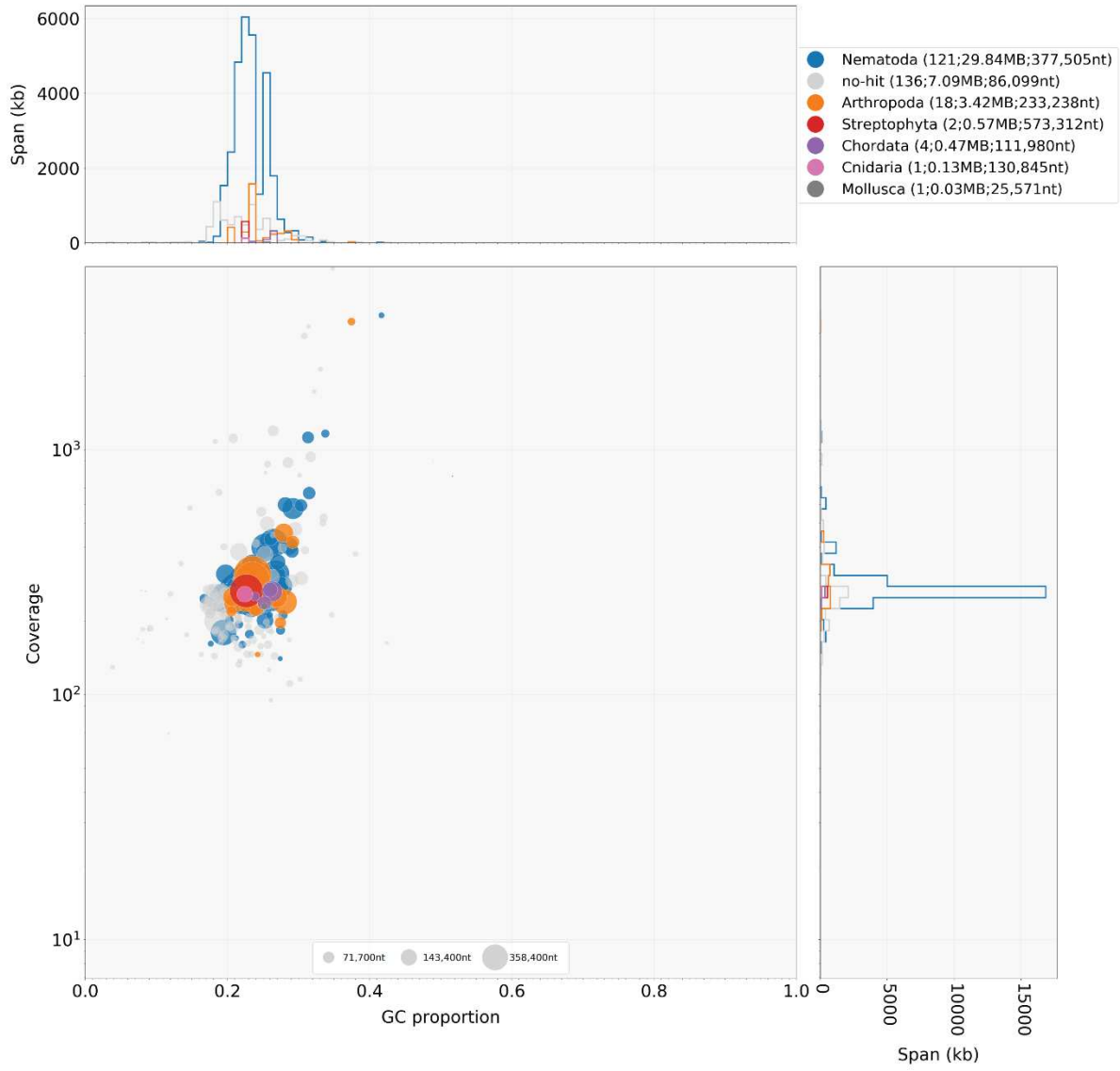


Figure S2. Screening of contaminant contigs in the cleaned haplotype-fused genome assembly

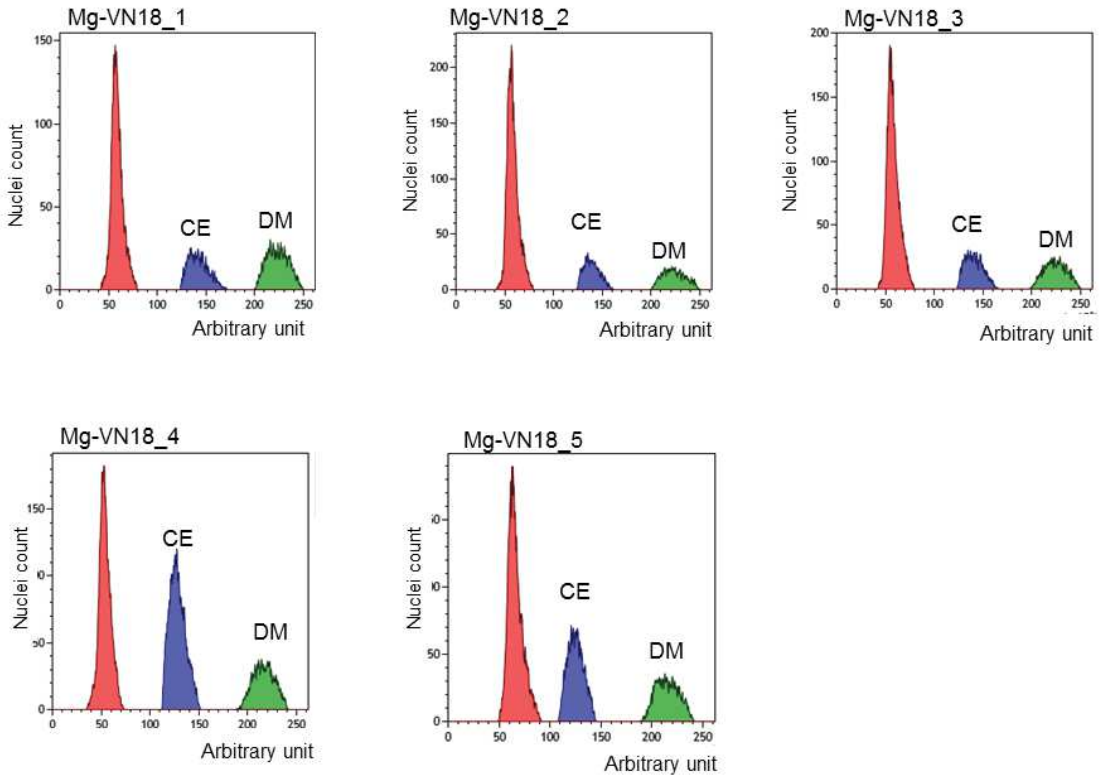


Figure S3. **Cytometry analyses**: Relative DNA staining of nuclei at the G0/G1 phase of *M. graminicola* (Red peak), *Drosophila melanogaster* (DM; green) and *Caenorhabditis elegans* (CE; blue). Cytograms showed fluorescence measurement (arbitrary unit) and number of G0/G1 nuclei of five *M. graminicola* isolates of Mg-VN18, mixed with nuclei of *D. melanogaster* (DM; 350 Mb, diploid) and *C. elegans* (CE; 200 Mb, diploid).

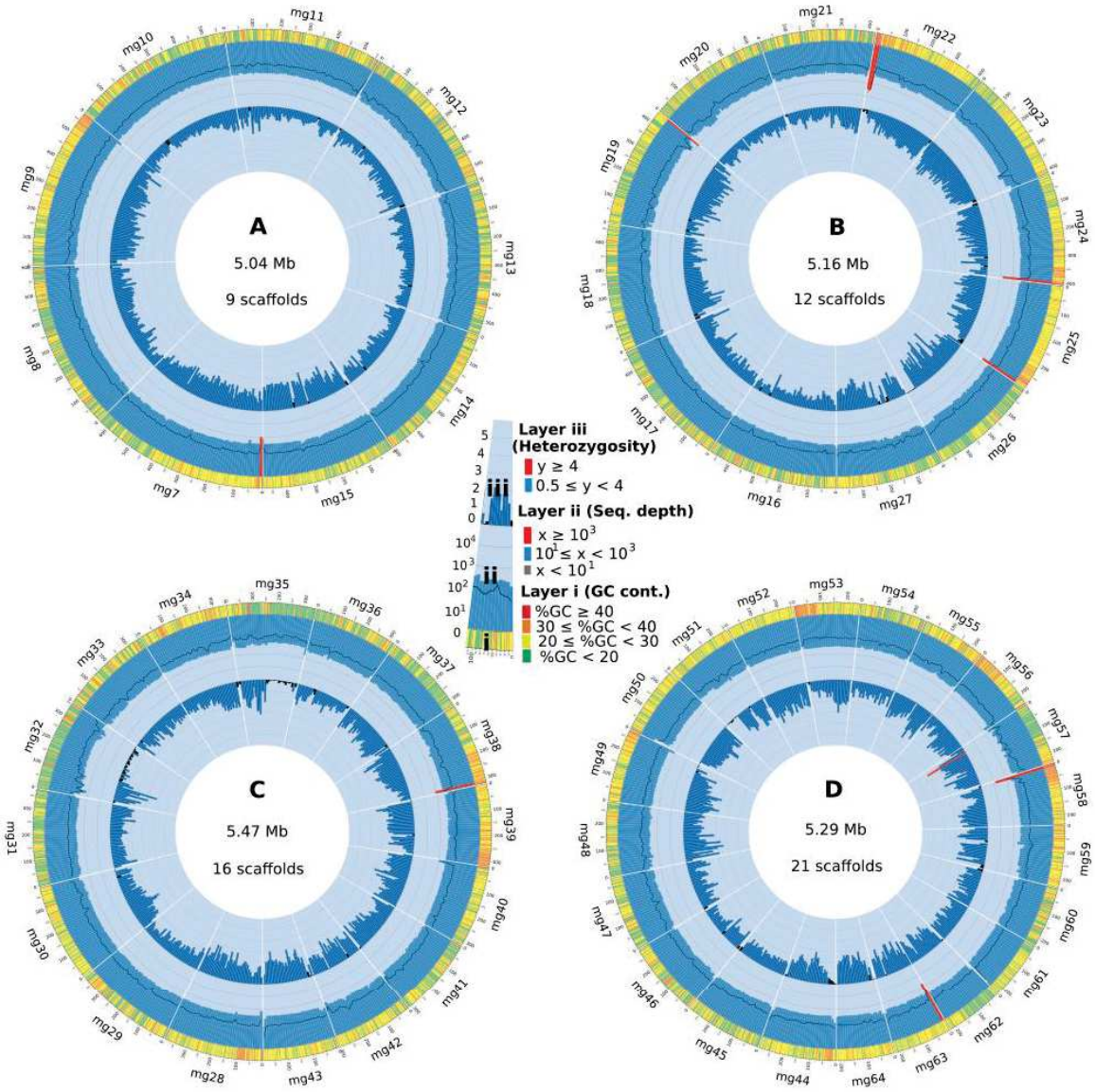


Figure S4. **Read-depth and heterozygosity along 277 haplotype-purged scaffolds of the *Meloidogyne graminicola* genome**. Number of scaffolds and total length are indicated in the middle of each circle. The scaffolds are sorted by length following clockwise from the longest to the smallest one. In each circle, three layers represent: (i) the GC content per 1-kb sliding window; (ii) short read-depth (x, histogram) and long read depth (black line) per 10-kb sliding window; and (iii) number of single nucleotide variants (y, histogram) per 10-kb genome window. Meaning of coded colors in each layer is detailed in the middle of the figure.

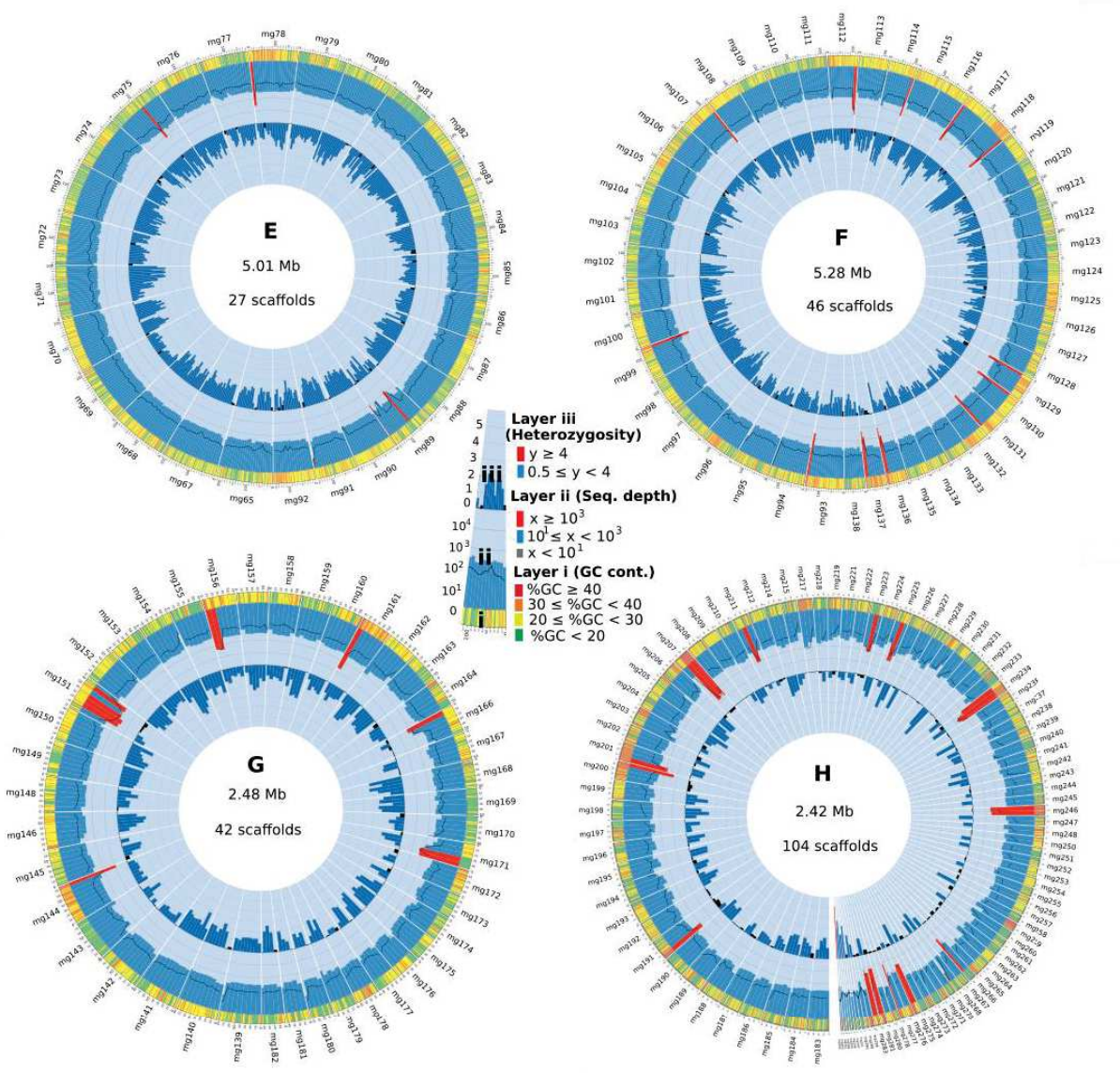


Figure S4, end.

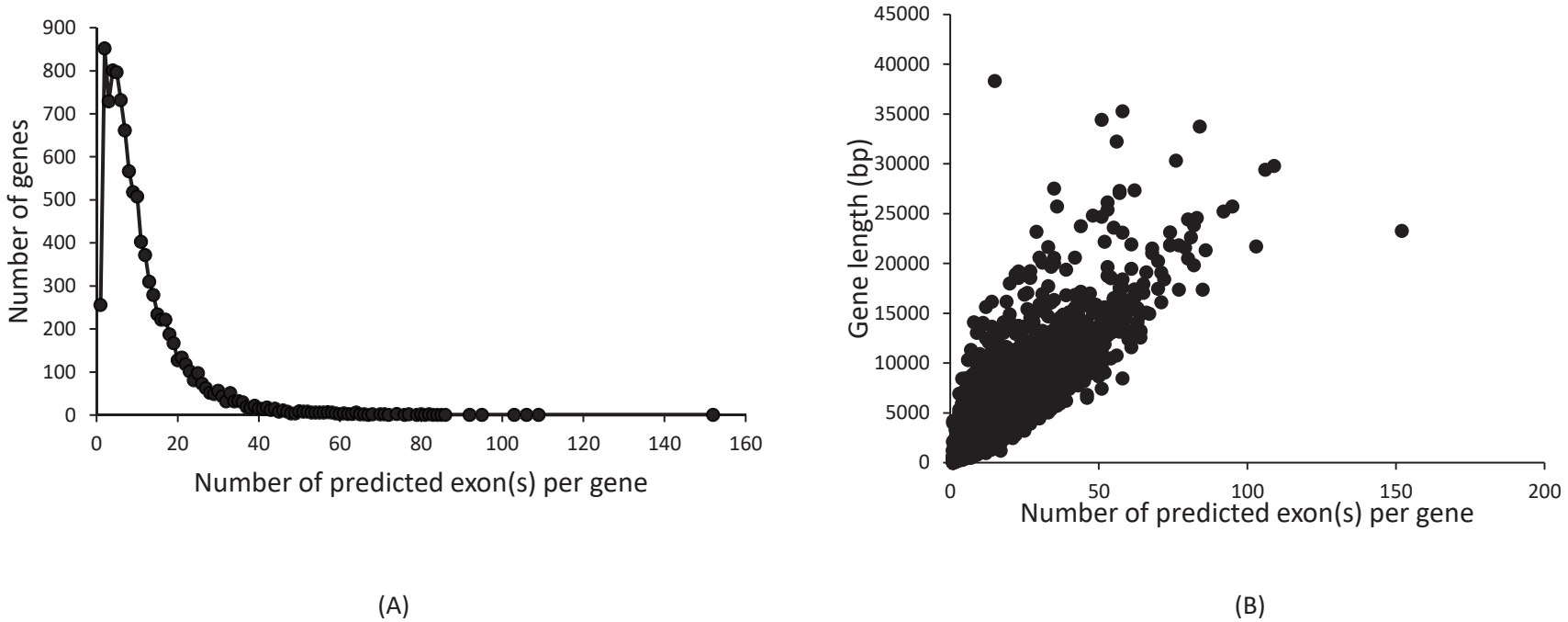


Figure S5. **Exon distribution in *Meloidogyne graminicola* genes**: (A) Distribution of the number of exons among all predicted protein coding genes, and (B) number of exons as a function of gene length.

Figure S6. **Per base coverage of raw RNA-seq reads across 20 BUSCO Eukaryote orthologous genes**. Each horizontal bar represents a gene. Y-axis indicates the depth of RNA-seq in a log(10) scale

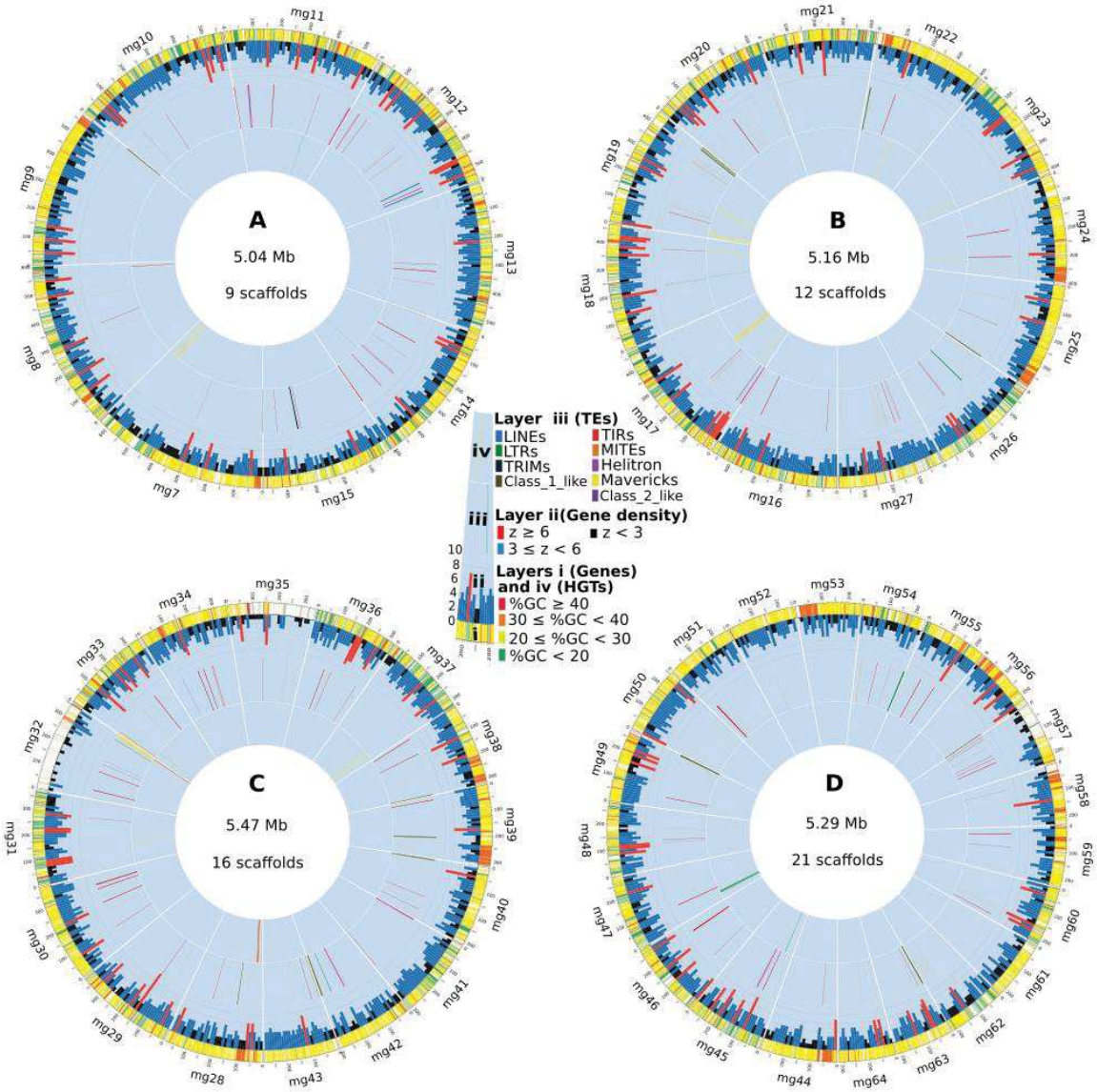


Figure S7. **Distribution of annotated genes, transposable elements, and putative horizontal gene transfers on 277 scaffolds of the *Meloidogyne graminicola* genome**. Number of scaffolds and total length are indicated in the middle of each circle. The scaffolds are sorted by length following clockwise from the longest to the smallest one. In each circle, four layers indicate: (i) location of genes on scaffolds, with a color representing their GC content; (ii) gene density (z, histogram) per 10-kb genome window; (iii) distribution of transposable elements (TEs) on scaffolds, with a specific color for each TE family; (iv) Distribution of horizontal gene transfers (HGTs) on scaffolds, with color representing GC content of each HGT. Meaning of coded colors in each layer is also given in the middle of the figure.

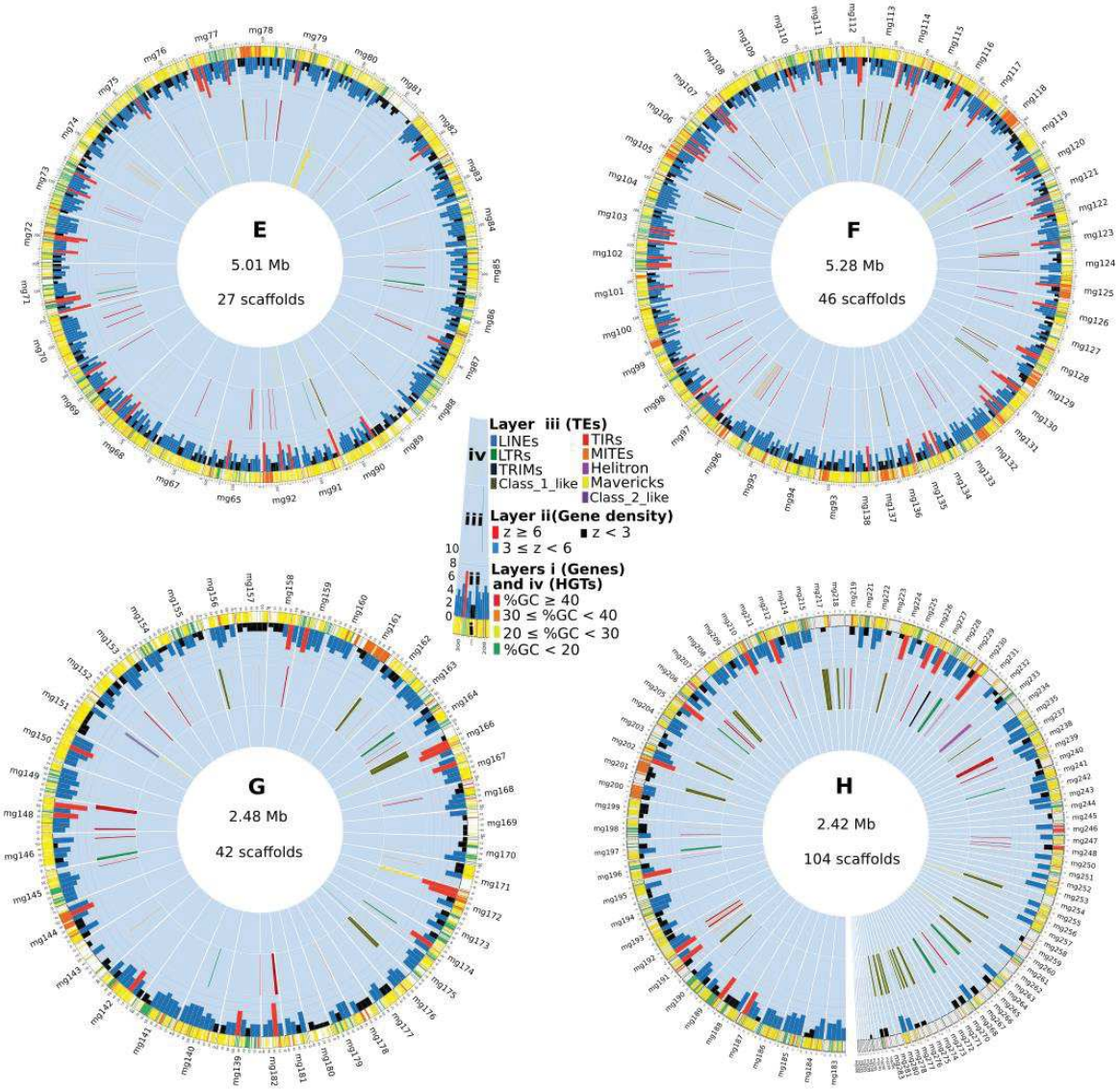


Figure S7, end.

Table S1. **General summary of long and short reads from Oxford Nanopore and Illumina sequencing technology**

	Nanopore data		Illumina data	
	Raw reads	Cleaned reads	Raw paired-end reads	Cleaned paired-end reads
Number of reads	869,830	727,017	122 890 657	86,991,142
Mean read length	4,533	4,844	150	139
Median read length	2,558	2 ,799	-	-
Median read quality	10.78	11.2	38	39
Read N50 length	8,943	9,387	-	-
Total bases	3,943,624,234	3,522,271,162	17 460 298 153	11,981,155,793
Estimated coverage*	94×	84×	415×	288×

* considering assembly genome size of 41.5 Mb

Table S2. **Haploid genome size, repeats content and heterozygosity** estimated for *M. graminicola* using four *k*-mer values with GenomScope (Vulture et al., 2017).

Genomic feature	*k* value (bp)			
	17	21	27	47
Haploid genome length (Mb)	41.1	41.4	41.3	41.6
Repeat content (Mb)	19.4	11.3	7.8	7.8
Heterozygosity (%)	1.69	1.85	1.9	1.72

Table S3. **Assembly metrics generated by five different methods after polishing and scaffolding processes**

Process	Assembler	# contigs	Largest contig (bp)	Total length (bp)	N50 (bp)	L50	GC (%)	CEGMA (%)
Assembly and polishsing	MaSuRCA	814	864,857	47,404,012	181,255	77	23.76	95.56
	Ra	429	498,246	39,709,036	138,010	85	23.17	93.55
	Canu	1,141	1,433,372	57,182,528	186,718	77	24.37	95.97
	Miniasm	322	2,083,370	42,398,816	425,868	28	23.49	94.76
	Flye	610	1,342,289	46,902,925	197,395	56	23.59	92.74
After scaffolding	MaSuRCA	808	864,857	47,404,612	189,278	76	23.76	95.56
	Ra	420	498,246	39,709,936	146,723	82	23.17	93.55
	Canu	1,138	1,433,372	57,182,828	194,506	75	24.37	95.97
	Miniasm	322	2,083,370	42,398,816	425,868	28	23.49	94.76
	Flye	587	1,342,289	46,905,225	215,369	56	23.59	92.74

Table S4. **Assembly metrics and BUSCO completeness after purging haplotigs on MaSuRCA, Ra, and Canu assemblies**

	Assemblers								
	MaSuRCA			Ra			Canu		
Metrics	Haplotype-purged contigs	Haplotigs	Artefacts	Haplotype-purged contigs	Haplotigs	Artefacts	Haplotype-purged contigs	Haplotigs	Artefacts
# contigs	363	419	25	390	27	3	357	678	103
Largest contig (bp)	865,251	79,346	9,809	498,246	94,725	25,679	1,433,372	124,629	32,469
Total length (bp)	40,997,469	6,320,255	114,560	38,894,616	765,352	49,968	42,778,169	13,677,693	726,966
N50	221,462	20,313	4,655	152,068	31,028	25,679	292,908	21,446	10,990
N75	110,482	12,406	3,645	76,988	23,145	15,694	147,791	16,457	7,536
L50	61	100	9	80	9	1	45	211	24
L75	126	200	16	171	17	2	96	393	44
GC (%)	23.55	24.9	35.65	23.21	20.96	27.23	23.94	25.37	31.14
Mismatches (Ns)	32,744	4450	0	900	0	0	300	0	0
Mapped-short reads (%)	97.75			98.50			98.55		
Mapped-long reads (%)	87.22			95.64			97.06		
Mapped-RNAseq reads (%)	90.72			89.41			90.93		
BUSCO completeness (n: 303)*	S:84.2% D:5.0%			S:85.8% D:1.7%			S:85.1% D:3.0%		

* BUSCO analysis was done only on haplotype-purged contigs

Table S5. **Total nuclei DNA content of five replicates of *Meloidogyne graminicola* (MG; isolate Mg-VN18) measured by flow cytometry** (see Figure S3). Nuclei of *Drosophila melanogaster* (DM; 350 Mb, diploid) and *Caenorhabditis elegans* (CE; 200 Mb, diploid) were mixed in the same tube as internal controls and served as references to estimate the MG genome size for each replicate.

Samples	Collecting date	G0/G1 peak value			CE genome size based on DM (350Mb)	Genome size based on CE (200 Mb)	
		CE	DM	MG		DM	MG
Mg-VN18_1	18/04/2019	141.2	223.0	58.2	221.6	315.8	82.4
Mg-VN18_2	18/04/2019	138.5	222.5	57.4	217.8	321.4	82.9
Mg-VN18_3	18/04/2019	140.2	223.8	57.1	219.2	319.3	81.5
Mg-VN18_4	25/02/2019	127.3	216.1	53.3	206.2	339.5	83.8
Mg-VN18_5	18/01/2019	125.0	214.5	64.9	203.9	343.3	103.9

Table S6. Details of putative HGTs detected in the *Meloidogyne graminicola* genome

Process	Gene/gene family	*Pfam* domain	Functions[1]	AI[2]	*M. graminicola* protein accession no	Species with best hit[3]	*M. inc.*[4]	*G. ros.*[4]
Plant cell wall degradation	GH28 Polygalacturonase	PF00295 Glycosyl hydrolases family 28	Pectin decorations degradation	326.9	Mgra_00000599-RA	(Bac) *Ralstonia pseudosolanacearum*	351.6 (4)	
				325.3	Mgra_00005314-RA	(Bac) *Ralstonia pseudosolanacearum*		
				264.4	Mgra_00010244-RA	(Bac) *Methylibium* sp. CF059		
	GH30 Xylanase	PF02055; PF17189 Glycosyl hydrolase family 30 TIM-barrel domain; GH30 beta sandwich domain	Xylan degradation	216	Mgra_00003019-RA	(Bac) *Clostridium* sp. DL-VIII	259.49 (6)	
				198.2	Mgra_00008402-RA	(Bac) *Clostridium* sp. DL-VIII		
	GH43 candidate Arabinase	PF04616 Glycosyl hydrolases family 43	Pectin decorations degradation	198.4	Mgra_00002782-RA	(Bac) *Streptomyces* sp. WAC 00631	69.07 (2)	
	PL3 Pectate lyase	PF03211 Pectate lyase	Pectin degradation	145.1	Mgra_00003596-RA	(Bac) *Streptomyces kanasensis*	137.46 (31)	137.06 (3)
				113.9	Mgra_00002412-RA	(Bac) *Streptomyces viridochromogenes*		
				91.4	Mgra_00008269-RA	(Bac) *Streptomyces griseoviridis*		
				67	Mgra_00008277-RA	(Bac) *Jonesia quinghaiensis*		
				66.	Mgra_00010266-RA	(Bac) *Frankia* sp. EAN1pec		
				64.9	Mgra_00004469-RA	(Bac) *Frankia* sp. EAN1pec		
				62.5	Mgra_00000523-RA	(Bac) *Paraburkholderia monticola*		
				51.7	Mgra_00002896-RA	(Bac) *Myxococcales bacterium*		
				40.9	Mgra_00004553-RA	(Bac) *Streptomyces sparsogenes*		
				34.3	Mgra_00002990-RA	(Bac) *Jonesia quinghaiensis*		
	Expansin-like protein	PF03330 Lytic trans glycolase	Softening of non-covalent bonds	53.2	Mgra_00008314-RA	(Bac) *Streptomyces luteoverticillatus*	86.11 (8)	29.93 (7)
				41.7	Mgra_00009665-RA	(Bac) Streptomycetaceae		
				15.8	Mgra_00009039-RA	(Bac) *Streptomyces davaonensis* JCM 4913		
				14.2	Mgra_00007051-RA	(Bac) *Saccharothrix* sp. ST-888		
	GH5_2 Cellulases	PF00150 Cellulase (glycosyl hydrolase family 5)	Cellulose degradation	46.2	Mgra_00008344-RA	(Bac) *Rufibacter tibetensis*	39.14 (23)	198.94 (11)
				39.3	Mgra_00007596-RA	(Bac) *Salinimicrobium xinjiangense*		
				32.8	Mgra_00008505-RA	(Bac) *Hymenobacter terrenus*		
				32.2	Mgra_00008504-RA	(Bac) *Rufibacter tibetensis*		
				29.5	Mgra_00009334-RA	(Bac) *Flaviramulus basaltis*		

Process	Gene/gene family	*Pfam* domain	Functions[1]	AI[2]	*M. graminicola* protein accession no	Species with best hit[3]	*M. inc.*[4]	*G. ros.*[4]
				28.7	Mgra_00000930-RA	(Bac) *Aureispira* sp. CCB-QB1		
				15.9	Mgra_00000972-RA	(Bac) *Gramella* sp. SH35		
				14.5	Mgra_00008510-RA	(Bac) *Leeuwenhoekiella polynyae*		
Plant defense	Candidate Isochorismatase	PF00857 Isochorismatase family	Catalyses the conversion of isochorismate	84.0	Mgra_00000516-RA	(Bac) *Nitratireductor* sp. OM-1	91.41 (1)	66.08 (1)
	Chorismate Mutase	PF01817 Chorismatemutase type II	Conversion of Chorismate into SA	62.5	Mgra_00002988-RA	(Bac) *Streptomyces* sp.	15.02 (1)	42.36 (2)
	pnbA Carboxylesterase	PF00135 Carboxylesterase family	Esterase	213.2	Mgra_00007278-RA	(Bac) *Acidobacteria bacterium*		
				113.4	Mgra_00001447-RA	(Fgi) *Saitozyma podzolica*		
				76.5	Mgra_00001438-RA	(Bac) *Duganella* sp.		
				69.6	Mgra_00004379-RA	(Bac) *Sphingobium* sp. AP49		
				58.8	Mgra_00007277-RA	(Fgi) *Glonium stellatum*		
				39.4	Mgra_00007109-RA	(Bac) *Niveispirillum cyanobacteriorum*		
				22.3	Mgra_00004380-RA	(Bac) *Acidobacteria bacterium*		
Nutrient processing	GH32 invertase	PF00251 Glycosyl hydrolases family 32 N-terminal domain	Sucrose degradation	328.6	Mgra_00002716-RA	(Bac) *Sinorhizobium saheli*	154.42 (1)	241.26 (11)
				185.2	Mgra_00008226-RA	(Bac) *Rhizobium grahamii*		
				169.7	Mgra_00008226-RB	(Bac) *Rhizobium grahamii*		
	Candidate GS1 Glutamine Synthetase	PF00120 Glutamine synthetase, catalytic domain	Nitrogen assimilation	306.6	Mgra_00008592-RA	(Bac) *Aquamicrobium aerolatum*	35.59 (4)	29.24 (1)
	Candidate galactose mutarotase	PF01263 Aldose 1-epimerase	Galactose metabolism	31.3	Mgra_00010300-RA	(Bac) *Chelativorans* sp. BNC1		
	GH2 β-galactosidase	PF02836 Glycosyl hydrolases family 2, TIM barrel domain	Galactose degradation	293.5	Mgra_00007323-RA	(Bac) *Ferrovibrio* sp.		
	Sugarporter (MFS transporter family)	PF00083 Sugar (and other) transporter	Transmembrane sugar transporter	353.9	Mgra_00007663-RA	(Bac) *Arthrobacter* sp. ZGTC212		
				285.6	Mgra_00003518-RA	(Bac) *Arthrobacter crystallopoietes*		
				252.1	Mgra_00004274-RA	(Bac) *Sporolactobacillus* sp. THM19-2		
				36.4	Mgra_00002779-RA	(Bac) *Arthrobacter crystallopoietes*		
	Candidate glycosyl Transferases Group 1	PF00534 Glycosyl transferases group 1; PF13439 Glycosyltransferase family 4	Catalyse the transfer of sugar moieties	307.4	Mgra_00003366-RA	(Bac) *Cupriavidus gilardii*		
				238.9	Mgra_00007025-RA	(Bac) *Proteobacteria bacterium*		

Process	Gene/gene family	*Pfam* domain	Functions[1]	AI[2]	*M. graminicola* protein accession no	Species with best hit[3]	*M. inc.*[4]	*G. ros.*[4]
				29	Mgra_00008526-RA	(Bac) *Rhizobiales bacterium*		
				16.6	Mgra_00008649-RA	(Bac) *Rhizobium leguminosarum*		
	bioB Biotin synthase	PF06968 Biotin and Thiamin Synthesis associated domain; PF04055 Radical SAM superfamily	Vitamin B7 biosynthesis	22.8	Mgra_00002952-RA	(Bac) Cardinium endosymbiont of *Sogatella furcifera*		
Suspect function in nematode living	Candidate L-threonine aldolase	PF01212 Beta-eliminating lyase	Amino acid transport and metabolism	341.2	Mgra_00005491-RA	(Bac) *Chelativorans* sp. BNC1	? (1)	164.69 (3)
	Gamma-glutamyl cyclo transferase	PF06094 Gamma-glutamyl cyclo transferase, AIG2-like	Gamma-glutamyl amine degradation	49.6	Mgra_00008680-RA	(Bac) *Bradyrhizobium diazoefficiens*		
	FAD dependent oxidoreductase	PF01266 FAD dependent oxidoreductase	Keto acid metabolism	91.5	Mgra_00007719-RA	(Bac) *Alphaproteobacteria bacterium* 16-39-46		
	DJ-1/PfpI family cysteine peptidase	PF01965 DJ-1/PfpI family	Degrade intracellular protein	49.7	Mgra_00007789-RA	(Bac) *Candidatus Cardinium hertigii*		
	FtsH peptidase	PF01434 Peptidase family M41	Degrade membrane-protein & soluble protein	40.1	Mgra_00007321-RA	(Bac) *Faecalibacterium* sp. An121		
	HADH	PF00725 3-hydroxyacyl-CoA dehydrogenase; PF02737 C-terminal domain and NAD binding domain	Enzyme involved in fatty acid metabolism	23.6	Mgra_00000083-RA	(Bac) *Pararhizobium haloflavum*		
	Phosphoribosyl transferase	PF00156 Phosphoribosyl transferase domain	Nucleoside metabolic process	200.3	Mgra_00009552-RA	(Bac) *Chelatococcus* sp.	202.63 (1)	198.13 (2)
	tdk Thymidine kinase	PF00265 Thymidine kinase	Nucleoside metabolic process	25.2	Mgra_00001990-RA	(Bac) *Candidatus Cardinium hertigii*		
Suspect function in cell wall degradation	GH25_Lys1-like	PF01183 Glycosyl hydrolases family 25	Lysozyme activity	33.1	Mgra_00008696-RA	(Arc) Archaeon		
	Laminin_G 3 family	PF13385 Concanavalin A-like lectin/glucanases superfamily; PF05426 Alginate lyase	Carbohydrate-binding module for hydrolysis	370.9	Mgra_00004780-RA	(Bac) *Collimonas* sp. OK412		
Support for HGT event	Integrase	PF00665 Integrase core domain	Integrate the foreigner DNA into host genome	23.3	Mgra_00004369-RA	(Bac) *Gammaproteobacteria bacterium*		
				22.6	Mgra_00001012-RA	(Bac) *Gammaproteobacteria bacterium*		
Unknown	Collagen	PF01391 Collagen triple helix repeat (20 copies)	Cuticle and basement membrane collagen	30.3	Mgra_00003940-RA	(Bac) *Cellulophaga lytica*		
	Glycoprotein G2	PF07245 Phlebovirus glycoprotein G2	Component of Golgi complex membrane	14.4	Mgra_00001607-RA	(Viru) Sugarbeet cyst nematode virus 2		
	Thaumatin-like protein	PF00314 Thaumatin family	An intensely sweet-tasting protein	14.2	Mgra_00000598-RA	(Vird) *Durio zibethinus*		

Process	Gene/gene family	*Pfam* domain	Functions[1]	AI[2]	*M. graminicola* protein accession no	Species with best hit[3]	*M. inc.*[4]	*G. ros.*[4]
	Unknown	PF08592 Domain of unknown function (DUF1772)	unknown	31.1	Mgra_00000676-RA	(Bac) *Rhodospirillaceae bacterium* SYSU D60006		

[1] Function of genes was deduced from previous studies performed either on nematodes and/or other groups of organisms (see details, below).

[2] AI =Alien Index.

[3]Bac: Bacteria; Fgi: Fungi; Viru: Virus; Vird: Viridiplantae; Arc: Archaeon.

[4]For each gene/gene family, the highest AI and number of HGTs (AI > 0) found in *Meloidogyne incognita* (*M. inc.*; Abad et al., 2008) and *Globodera rostochiensis* (*G. ros.*; Akker et al., 2016) are given.

Appendix S1. Summary of putative function(s) of detected HGTs

Putative function(s) of detected HGTs (Table S6), classified by different processes they are supposedly involved in, are here detailed:

1. Plant cell wall degradation. First, putative HGTs encoding cellulases, polygalacturonases, pectate lyases, arabinanase, xylanases, and expansin-like proteins were detected in *M. graminicola.* Such enzymes involved in the degradation, modification and softening of the plant cell wall, are usually absent from animal genomes, but have been already reported in a wide range of tylenchomorph PPN species (Danchin, 2011). It was also shown that nematode genes encoding these enzymes also exhibit high similarity to bacterial genes (Danchin, 2011). Such genes may be involved in the root tissue colonization by nematodes, by allowing the degradation of the protective barrier of the plant cell wall, which is constituted mainly of cellulose, hemicelluloses, pectin, and its branched decorations (Malinovsky et al., 2014). It has been already documented that encoded proteins of such HGTs are expressed in the subventral gland cells of nematodes and assist them to penetrate into the roots (Haegeman et al., 2011). More in details, we found in the *M. graminicola* genome:

- a) Eight HGTs encoding cellulases (glycosyl hydrolase family 5). Such genes were the first HGTs to be identified in several cyst nematodes (*G. tabacum*, *G. rostochiensis*) (Akker et al., 2016; Goellner et al., 2000; Smant et al., 1998), *Pratylenchus* spp. (*P. penetrans*) (Momota et al., 2001), and RKNs (*M. incognita*, *M. hapla*) (Opperman et al., 2008; Rosso et al., 1999). Knockout of the gene encoding cellulase (β-1,4, endoglucanases) resulted in the reduction of invasion ability in *G. rostochiensis* (Chen et al., 2005) and the decrease of propagation and dispersal ability in the pine wood nematode *Bursaphelenchus xylophilus* (Ma et al., 2011).

- b) Three genes encoding polygalacturonases (glycoside hydrolase GH28 family) and a multigenic family of ten members encoding pectate lyases (PL3s; 10 genes). These enzymes participate in pectin degradation. Such HGTs have been also reported in other RKNs. GH28 polygalacturonases have thus been isolated and biochemically characterized in *M. incognita* (Jaubert et al., 2002) and were identified only in RKNs until now. Furthermore, all pectate lyases characterized in plant-parasitic nematodes belong to the polysaccharide lyase family 3, and several copies have been also found [i.e., 22 and 31in *M. hapla* and *M. incognita*, respectively (Abad et al., 2008; Opperman et al., 2008)].

- c) Two genes encoding GH30 xylanases (degrading xylan - the major component of hemicelluloses), one gene encoding a candidate GH43 arabinanase (hydrolyzing beta-1,4-galactan in the hairy regions of pectin), and four genes encoding expansin-like proteins (weakener of the non-covalent interactions between cellulose and hemicellulose). Such HGTs were also found in other RKNs (Abad et al., 2008; Opperman et al., 2008).

2. Plant defense manipulation and detoxification. We detected in the *M. graminicola* genome several HGTs encoding enzymes potentially involved in the parasitism.

First, putative HGTs encoding for chorismate mutase (1 gene), isochorismatase (1 gene), and cyanate lyase (1 gene, AI = 4, not shown in the table) were detected. Such genes were already reported as HGTs in other PPNs, and may play an important role as detoxifying agents or as true suppressors of host defense signaling (Haegeman et al., 2011). They may help obligate endoparasites to modulate host defenses for the duration of their life cycle that lives

inside root tissues. Chorismate mutase and isochorismatase, which most closely resemble bacterial enzymes, may reduce the pool of chorismate available for conversion to the plant defense-signaling compound salicylic acid, thus preventing normal activation of host defenses (Wildermuth et al., 2001). It is known that bacterial cyanate lyases can detoxify cyanate, a compound secreted by the plant in response to herbivore attack (Johnson & Anderson, 1987; Jones, 1998). Candidate cyanate lyase has been identified in the genome of *Meloidogyne* spp., which are biotrophic and not herbivorous, but it is possible that the cyanate lyase produced by the nematode has a similar role. However, it should be confirmed by the biochemical function.

In addition, HGTs encoding other enzymes putatively involved in plant defense were here detected for the first time in PPNs. Seven HGTs encoding carboxylesterases were detected in the *M. graminicola* genome. Carboxylesterasesare thought to act as a mechanism to detoxify ester-containing xenobiotics, which are toxic properties of phytoalexins, secreted by plants in response to nematode infections (Gillet et al., 2017; Hatfield et al., 2016). In *M. incognita*, one gene (not reported as HGT) encoding a carboxylesterase was upregulated in the compatible interaction with susceptible tomato variety at the latterly stage (13-28 days-after-inoculation) (Shukla et al., 2017).

3. Nutrition processing. As a parasite, *M. graminicola* feeds on plant cells by using carbohydrates, amino acids, and vitamins as a source of nutrients for its development. HGTs may have contributed to the optimization of these metabolic processes.

As in other PPNs, we detected HGTs encoding invertases (GH32; 2 loci) and glutamine synthetase (GS1; 1 gene). Invertases (beta-fructofuranosidases) catalyze the hydrolysis of sucrose into fructose and glucose, which can be used by nematodes as a source of energy. A GH32 gene encoding invertase was already phylogenetically supported of bacterial origin in PPNs (Danchin et al., 2016). Similarly, a HGT encoding a glutamine synthetase involved in ammonium assimilation, as part of the nitrogen-fixation pathway in rhizobia, was also reported in PPNs. Furthermore, while nine HGT genes involved in the synthesis or salvage of the four vitamins B1, B5, B6, B7 are found in cyst nematode (Craig et al., 2009), *M. graminicola* only acquired a single gene encoding vitamin B7 from bacteria. This HGT was not detected in *M. incognita*, which however acquired HGTs for two other genes encoding vitamins (i.e., B1 and B5; Craig et al., 2009).

In addition, several genes potentially involved in the nutrition process were here newly reported as candidate HGTs in the *M. graminicola* genome including genes encoding β-galactosidase (GH2; 1 gene), candidate galactose mutarotase (1 gene), glycosyl transferases (4 genes), and sugar transporters (MFS family; 4 genes). β-galactosidase (GH2) catalyzes the hydrolysis of β-galactosides into monosaccharides (*CAZypedia.org*, 2020). In bacteria, galactose mutarotaseis involved in the first step of the galactose metabolism by catalyzing the conversion of β-d-galactose to α-d-galactose (Bouffard et al., 1994; Thoden et al., 2003). Glycosyl transferases of group 1 catalyze the formation of the glycosidic linkage to form a glycoside. These enzymes utilize 'activated' sugar phosphates (including glycogen, fructose-6-phosphate and lipopolysaccharides) as glycosyl donors, and catalyze glycosyl group transfer to a nucleophilic group (Campbell et al., 1997). Sugar porters are responsible for the binding and transport of various carbohydrates, organic alcohols, and acids in a wide range of prokaryotic and eukaryotic organisms, including PPNs, in which sugar transporters are specifically

expressed and active in syncytia, indicating a profound role in inter- and intra-cellular transport processes (Hofmann et al., 2009).

4. Other suspect and unknown function. The function of a few putative HGTs remains speculative. Here, we briefly summarize the state of current knowledge on these different classes of genes. We detected:

a) One HGT encoding a phosphoribosyl transferase. Such HGTs have been already reported in PPNs, but their functions remained unknown in nematode (Paganini et al., 2012; Scholl et al., 2003). Phosphoribosyl transferase catalyzes the displacement of the alpha-1'-pyrophosphate of 5-phosphoribosyl-alpha 1-pyrophosphate by a nitrogen-containing nucleophile (https://pfam.xfam.org). This domain is found in a range of diverse phosphoribosyl transferase enzymes and regulatory proteins of the nucleotide synthesis and salvage pathways (Sinha & Smith, 2001).

b) One HGT encoding a L-threonine aldolase (TA). TAs represent a family of homologous pyridoxal 5'-phosphate-dependent enzymes found in bacteria and fungi, and catalyze the reversible cleavage of several l-3-hydroxy-α-amino acids (di Salvo et al., 2014). HGTs encoding L-threonine aldolases have been already detected in *Meloidogyne* (Scholl et al., 2003), and these genes (as well as those encoding glutamine synthetase and *nodL*) were probably acquired from a *Mesorhizobium* species (Scholl et al., 2003).

c) One HGT encoding gamma-glutamylamine cyclotransferase. This ubiquitous enzyme is found in bacteria, plants, and metazoans from *Dictyostelium* through to humans. It catalyzes the conversion of epsilon-(L-gamma-glutamyl)-L-lysine to free lysine and 5-oxo-L-proline as well as the release of free amines and the formation of 5-oxo-L-proline from a variety of other L-gamma-glutamylamines (Oakley et al., 2010). The free lysine and/or amines would be utilized by nematodes for their living.

d) One HGT encoding FAD dependent oxidoreductase. This *M. graminicola* gene shows 31% identity with the D-aspartate oxidase (DDO) gene and 25% identity with D-amino-acid oxidase (DAO) genes of *C. elegans* (Katane et al., 2010). DAO catalyzes the oxidation of neutral and basic D-amino acids into their corresponding keto acids while DDO, structurally related to DAO, catalyzes the same reaction but is active only toward dicarboxylic D-amino acids. In higher animals, DAO and DDO regulate endogenous d-Ser and d-Asp levels, respectively, as well as mediate the elimination of accumulated exogenous d-amino acids in various organs (Katane et al., 2010).

e) One HGT for PfpI. The function of this gene is unclear in nematodes, however, it was characterized in bacteria as endopeptidase which degrades intracellular proteins to free acid amines (Halio et al., 1996; Zhan et al., 2014).

f) One HGT for FtsH, an ATP-dependent integral membrane protease. It plays a crucial role in quality control of integral membrane proteins by degrading unneeded or damaged membrane proteins, but it also targets cytoplasmic soluble proteins (Bieniossek et al., 2006).

g) One HGT for 3-hydroxyacyl-CoA dehydrogenase (EC) (HCDH). This enzyme is involved in fatty acid metabolism by catalyzing the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA. There are two major regions of similarities in protein sequences of the

HCDH family, the first one located in the N-terminal, corresponds to the NAD-binding site, while the second one is located in the center of the sequence, which represents the C-terminal domain (Birktoft et al., 1987).

h) One HGT for Thymidine kinase. This ubiquitous enzyme catalyzes the ATP-dependent phosphorylation of thymidine, and has been already characterized in *C. elegans* (Skovgaard & Munch-Petersen, 2006).

i) Two genes coding for integrase enzymes. Integrases may be essential for the integration of HGTs into the host chromosome, and are also identified as HGTs in *M. graminicola*. Interestingly, these genes were associated with TEs that potentially created more copies of these genes in the genome. Therefore, they could have themselves contributed to the HGT events observed in *M. graminicola*

j) One HGT for collagen triple helix repeat. Collagens are generally extracellular structural proteins involved in formation of connective tissue structure. The alignment contains 20 copies of the G-X-Y repeat that forms a triple helix. The collagen genes of nematodes encode proteins that have a diverse range of functions. Among their most abundant products are the cuticular collagens, which include about 80% of the proteins present in the nematode cuticle (Fetterer & Rhoads, 1993).

k) One HGT for glycoprotein G2. Glycoprotein G2 is component of the viral envelop in the Bunyaviruses family (Andersson & Pettersson, 1998). This gene in *M. graminicola* shows 40% identity with a glycoprotein precursor that has been found in the sugar beet cyst nematode virus (Lin et al., 2018).

l) One HGT encoding a thaumatin-like protein (TLP). TLPs are polypeptides of about 200 residues synthesized by plants in response to fungal infection (Zhang et al., 2018). The antifungal function of this gene was also proposed in *Schistocerca gregaria* and *C. elegans* (Brandazza et al., 2004; Dierking et al., 2016), however, the detailed mechanism is not yet completely understood.

References

Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V. C., Caillaud, M.-C., Coutinho, P. M., Dasilva, C., De Luca, F., Deau, F., Esquibet, M., Flutre, T., Goldstone, J. V., Hamamouch, N., … Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, *26*, 909. https://doi.org/10.1038/nbt.1482

Akker, S. E. den, Laetsch, D. R., Thorpe, P., Lilley, C. J., Danchin, E. G. J., Rocha, M. D., Rancurel, C., Holroyd, N. E., Cotton, J. A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M.-O., Boyes, I., Marvin, J. M. C., Jones, L. M., Yusup, H. B., Lafond-Lapalme, J., … Jones, J. T. (2016). The genome of the yellow potato cyst nematode, *Globodera rostochiensis* , reveals insights into the basis of parasitism and virulence. *Genome Biology*, *17*(1), 1–23. https://doi.org/10.1186/s13059-016-0985-1

Andersson, A. M., & Pettersson, R. F. (1998). Targeting of a short peptide derived from the cytoplasmic tail of the G1 membrane glycoprotein of *Uukuniemi virus* (Bunyaviridae) to the Golgi complex. *Journal of Virology*, *72*(12), 9585–9596.

Bieniossek, C., Schalch, T., Bumann, M., Meister, M., Meier, R., & Baumann, U. (2006). The molecular architecture of the metalloprotease FtsH. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(9), 3066–3071. https://doi.org/10.1073/pnas.0600031103

Birktoft, J. J., Holden, H. M., Hamlin, R., Xuong, N. H., & Banaszak, L. J. (1987). Structure of L-3-hydroxyacyl-coenzyme A dehydrogenase: preliminary chain tracing at 2.8-A resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(23), 8262–8266. https://doi.org/10.1073/pnas.84.23.8262

Bouffard, G. G., Rudd, K. E., & Adhya, S. L. (1994). Dependence of lactose metabolism upon mutarotase encoded in the gal operon in *Escherichia coli*. *Journal of Molecular Biology*, *244*(3), 269–278. https://doi.org/10.1006/jmbi.1994.1728

Brandazza, A., Angeli, S., Tegoni, M., Cambillau, C., & Pelosi, P. (2004). Plant stress proteins of the thaumatin-like family discovered in animals. *FEBS Letters*, *572*(1–3), 3–7. https://doi.org/10.1016/j.febslet.2004.07.003

Campbell, J., Davies, G., Bulone, V., & Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *The Biochemical Journal*, *326*(Pt 3), 929–939. https://doi.org/10.1042/bj3260929u

CAZypedia.org. (2020, January 23). Glycoside Hydrolase Family 2. CAZypedia.org

Chen, Q., Rehman, S., Smant, G., & Jones, J. T. (2005). Functional analysis of pathogenicity proteins of the potato cyst nematode *Globodera rostochiensis* using RNAi. *Molecular Plant-Microbe Interactions*, *18*(7), 621–625. https://doi.org/10.1094/MPMI-18-0621

Craig, J. P., Bekal, S., Niblack, T., Domier, L., & Lambert, K. N. (2009). Evidence for horizontally transferred genes involved in the biosynthesis of vitamin B1, B5, and B7 in *Heterodera glycines*. *Journal of Nematology*, *41*(4), 281–290.

Danchin, É. G. J. (2011). What nematode genomes tell us about the importance of horizontal gene transfers in the evolutionary history of animals. *Mobile Genetic Elements*, *1*(4), 269–273. https://doi.org/10.4161/mge.18776

Danchin, G. J. E., Guzeeva, A. E., Mantelin, S., Berepiki, A., & Jones, T. J. (2016). Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. *Molecular Biology and Evolution*, *33*(6), 1571–1579. https://doi.org/10.1093/molbev/msw041

di Salvo, M. L., Remesh, S. G., Vivoli, M., Ghatge, M. S., Paiardini, A., D'Aguanno, S., Safo, M. K., & Contestabile, R. (2014). On the catalytic mechanism and stereospecificity of *Escherichia coli* L-threonine aldolase. *The FEBS Journal*, *281*(1), 129–145. https://doi.org/10.1111/febs.12581

Dierking, K., Yang, W., & Schulenburg, H. (2016). Antimicrobial effectors in the nematode *Caenorhabditis elegans*: an outgroup to the Arthropoda. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1695). https://doi.org/10.1098/rstb.2015.0299

Fetterer, R. H., & Rhoads, M. L. (1993). Biochemistry of the nematode cuticle: relevance to parasitic nematodes of livestock. *Veterinary Parasitology*, *46*(1–4), 103–111. https://doi.org/10.1016/0304-4017(93)90051-n

Gillet, F.-X., Bournaud, C., Antonino de Souza Júnior, J. D., & Grossi-de-Sa, M. F. (2017). Plant-parasitic nematodes: towards understanding molecular players in stress responses. *Annals of Botany*, *119*(5), 775–789. https://doi.org/10.1093/aob/mcw260

Goellner, M., Smant, G., De Boer, J. M., Baum, T. J., & Davis, E. L. (2000). Isolation of beta-1,4-endoglucanase genes from *Globodera tabacum* and their expression during parasitism. *Journal of Nematology*, *32*(2), 154–165.

Haegeman, A., Jones, J. T., & Danchin, E. G. J. (2011). Horizontal gene transfer in nematodes: A catalyst for plant parasitism? *Molecular Plant-Microbe Interactions*, *24*(8), 879–887. https://doi.org/10.1094/MPMI-03-11-0055

Halio, S. B., Blumentals, I. I., Short, S. A., Merrill, B. M., & Kelly, R. M. (1996). Sequence, expression in *Escherichia coli*, and analysis of the gene encoding a novel intracellular protease (PfpI) from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Journal of Bacteriology*, *178*(9), 2605–2612.

Hatfield, M. J., Umans, R. A., Hyatt, J. L., Edwards, C. C., Wierdl, M., Tsurkan, L., Taylor, M. R., & Potter, P. M. (2016). Carboxylesterases: General detoxifying enzymes. *Chemico-Biological Interactions*, *259*(Pt B), 327–331. https://doi.org/10.1016/j.cbi.2016.02.011

Hofmann, J., Hess, P. H., Szakasits, D., Blöchl, A., Wieczorek, K., Daxböck-Horvath, S., Bohlmann, H., van Bel, A. J. E., & Grundler, F. M. W. (2009). Diversity and activity of sugar transporters in nematode-induced root syncytia. *Journal of Experimental Botany*, *60*(11), 3085–3095. https://doi.org/10.1093/jxb/erp138

Jaubert, S., Laffaire, J.-B., Abad, P., & Rosso, M.-N. (2002). A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. *FEBS Letters*, *522*(1–3), 109–112. https://doi.org/10.1016/s0014-5793(02)02906-x

Johnson, W. V., & Anderson, P. M. (1987). Bicarbonate is a recycling substrate for cyanase. *The Journal of Biological Chemistry*, *262*(19), 9021–9025.

Jones, D. A. (1998). Why are so many food plants cyanogenic? *Phytochemistry*, *47*(2), 155–162. https://doi.org/10.1016/s0031-9422(97)00425-1

Katane, M., Saitoh, Y., Seida, Y., Sekine, M., Furuchi, T., & Homma, H. (2010). Comparative characterization of three D-Aspartate oxidases and one D-amino acid oxidase from *Caenorhabditis elegans*. *Chemistry & Biodiversity*, *7*(6), 1424–1434. https://doi.org/10.1002/cbdv.200900294

Lin, J., Ye, R., Thekke-Veetil, T., Staton, M. E., Arelli, P. R., Bernard, E. C., Hewezi, T., Domier, L. L., & Hajimorad, M. R. (2018). A novel picornavirus-like genome from transcriptome sequencing of sugar beet cyst nematode represents a new putative genus. *The Journal of General Virology*, *99*(10), 1418–1424. https://doi.org/10.1099/jgv.0.001139

Ma, H. B., Lu, Q., Liang, J., & Zhang, X. Y. (2011). Functional analysis of the cellulose gene of the pine wood nematode, *Bursaphelenchus xylophilus*, using RNA interference. *Genetics and Molecular Research*, *10*(3), 1931–1941. https://doi.org/10.4238/vol10-3gmr1367

Malinovsky, F. G., Fangel, J. U., & Willats, W. G. T. (2014). The role of the cell wall in plant immunity. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00178

Momota, Y., Uehara, T., & Kushida, A. (2001). PCR-based cloning of two β-1,4-endoglucanases from the root-lesion nematode *Pratylenchus penetrans*. *Nematology*, *3*(4), 335–341. https://doi.org/10.1163/156854101317020259

Oakley, A. J., Coggan, M., & Board, P. G. (2010). Identification and characterization of gamma-glutamylamine cyclotransferase, an enzyme responsible for gamma-glutamyl-epsilon-lysine catabolism. *The Journal of Biological Chemistry*, *285*(13), 9642–9648. https://doi.org/10.1074/jbc.M109.082099

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P., & Windham, E. (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences*, *105*(39), 14802–14807. https://doi.org/10.1073/pnas.0805946105

Paganini, J., Campan-Fournier, A., Da Rocha, M., Gouret, P., Pontarotti, P., Wajnberg, E., Abad, P., & Danchin, E. G. J. (2012). Contribution of lateral gene transfers to the genome composition and parasitic ability of root-knot nematodes. *PLoS ONE*, *7*(11), e50875. https://doi.org/10.1371/journal.pone.0050875

Rosso, M. N., Favery, B., Piotte, C., Arthaud, L., De Boer, J. M., Hussey, R. S., Bakker, J., Baum, T. J., & Abad, P. (1999). Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. *Molecular Plant-Microbe Interactions*, *12*(7), 585–591. https://doi.org/10.1094/MPMI.1999.12.7.585

Scholl, E. H., Thorne, J. L., McCarter, J. P., & Bird, D. M. (2003). Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biology*, *4*(6), R39. https://doi.org/10.1186/gb-2003-4-6-r39

Shukla, N., Yadav, R., Kaur, P., Rasmussen, S., Goel, S., Agarwal, M., Jagannath, A., Gupta, R., & Kumar, A. (2017). Transcriptome analysis of root-knot nematode (*Meloidogyne incognita*)-infected tomato (*Solanum lycopersicum*) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses. *Molecular Plant Pathology*, *19*(3), 615–633. https://doi.org/10.1111/mpp.12547

Sinha, S. C., & Smith, J. L. (2001). The PRT protein family. *Current Opinion in Structural Biology*, *11*(6), 733–739. https://doi.org/10.1016/S0959-440X(01)00274-3

Skovgaard, T., & Munch-Petersen, B. (2006). Purification and characterization of wild-type and mutant TK1 type kinases from *Caenorhabditis elegans*. *Nucleosides, Nucleotides & Nucleic Acids*, *25*(9–11), 1165–1169. https://doi.org/10.1080/15257770600894410

Smant, G., Stokkermans, J. P., Yan, Y., de Boer, J. M., Baum, T. J., Wang, X., Hussey, R. S., Gommers, F. J., Henrissat, B., Davis, E. L., Helder, J., Schots, A., & Bakker, J. (1998). Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(9), 4906–4911. https://doi.org/10.1073/pnas.95.9.4906

Thoden, J. B., Kim, J., Raushel, F. M., & Holden, H. M. (2003). The catalytic mechanism of galactose mutarotase. *Protein Science*, *12*(5), 1051–1059.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Wildermuth, M. C., Dewdney, J., Wu, G., & Ausubel, F. M. (2001). Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature*, *414*(6863), 562–565. https://doi.org/10.1038/35107108

Zhan, D., Bai, A., Yu, L., Han, W., & Feng, Y. (2014). Characterization of the PH1704 protease from *Pyrococcus horikoshii* OT3 and the critical functions of Tyr120. *PLoS ONE*, *9*(9), e103902. https://doi.org/10.1371/journal.pone.0103902

Zhang, J., Wang, F., Liang, F., Zhang, Y., Ma, L., Wang, H., & Liu, D. (2018). Functional analysis of a pathogenesis-related thaumatin-like protein gene TaLr35PR5 from wheat induced by leaf rust fungus. *BMC Plant Biology*, *18*. https://doi.org/10.1186/s12870-018-1297-2

CHAPTER IV

Genomic rearrangements are putative evolutionary traits in the facultative meitotic parthenogenetic nematode *Meloidogyne graminicola*

Summary in English

In Chapter III, we presented a quality assembly of the *M. graminicola* genome providing important genetic information, including gene structure and TE/HGT content. This reference genome also allowed us to perform a comparative analysis between *M. graminicola* isolates, presented in Chapter IV, in order to study intraspecific diversity and molecular mechanisms of genome evolution.

In order to answer the questions raised at the end of Chapter II, this Chapter IV focuses on the study of intraspecific diversity and the origin of the species. The genomic sequences of thirteen *M. graminicola* isolates from geographically distant rice fields in Southeast Asia and Brazil were compared to the reference genome to look for single nucleotide variants (SNVs) among the isolates. The reference genome that we realized is a haploid genomic sequence that consists mainly of hybrid "scaffolds" from the two divergent haplotypes (heterozygosity of 1.36% ± 0.78). The Hiseq sequencing data from the two divergent haplotypes will thus refer to the same scaffolds, resulting in SNVs that we have encoded ("0/1") for the heterozygous positions. These SNVs thus potentially represent a set comprising the two haplotypes of each genome as well as intra-species variants, without being distinguishable. Consequently, in our study, this type of position was excluded from the analysis of intraspecific variability. On the other hand, comparison of the genomes from the 12 isolates to haploid reference genome (except Mg-VN18 which is the reference genome) identified 44,527 homozygous inter-isolate SNVs. In other words, for each of these SNVs, both haplotypes have the same sequence within an isolate, but polymorphism exists between two or more isolates. Principal component analysis based on these SNVs showed an absence of population structure underlying the geographical origin. These SNVs identified in each isolate were then used to perform recombination tests ["four gametes" test and linkage disequilibrium (LD) between marker pairs]. An exponential and rapid increase in the proportion of marker pairs passing the "four gametes" test, combined with a decrease in the LD, suggested that abundant recombination (RE) events are found in the *M. graminicola* genome.

Visualization of all of these identified SNVs, among the 13 isolates on the reference genome scaffolds, revealed the existence of heterozygosity loss (LoH) in each of the 13 isolates. The regions of the genome affected by heterozygosity loss are different among the isolates and represent a cumulative length of up to 7.4% of the total length of the genome. Interestingly, 490 kb that were affected by these LoH (five "scaffolds", ~1.18% of genome size) were shared among nine of the isolates suggesting that certain regions were more susceptible to this heterozygosity loss. In addition, some potential effectors were affected by this LoH

phenomenon (single allele conservation) which could contribute to the adaptation of these isolates to their environment.

The 44,527 homozygous SNVs of the 12 isolates are for the most part linked to LoH regions and do not allow the study of polymorphism on the genome of all the isolates. Therefore, we selected a set of 40 SNVs located on regions other than those where LoHs were observed in order to compare the 12 isolates simultaneously. The PCA using these 40 SNPs offers a similar lack of structuring to that observed using the 44,527 SNVs. This lack of phylogenetic signal already observed in the study of the mitochondrial genome reinforces the hypothesis of a recent spread of this species.

The fused haploid reference genome has shown its limitations in detecting single nucleotide variants within the genome of different isolates. Therefore, two haplotypes of 18 genomic regions of the haploid genome (36 haplotypes in total, 354 kb) were constructed manually. Only two SNVs were detected among these 36 haplotypes between the 12 isolates (0.5×10^{-4} mutation per base). This is a very low rate compared to *M. incognita*. In addition, analysis of the coverage of these 36 haplotypes confirmed the existence of LoH events.

Variations in copy number (CNV) were also studied between the genomes of the 13 isolates. A total of 104 CNVs regions were found that may affect the expression of 266 genes. It is interesting to note that some isolates, which were distinguished on the basis of their SNVs, also exhibited gene losses and significant LoH events.

Unfortunately, details of the environments to which these 13 isolates were exposed before being collected from the rice field were not described. Moreover, no link between the occurrence of any of these molecular rearrangements and exposure to stress has been demonstrated. Therefore, the adaptive impact that SNV, LoH and CNV could have on the genome of the isolate remains to be studied. However, we can hypothesize that TE, SNV, LoH and CNV are molecular mechanisms that promote genomic plasticity and could be involved in the evolution of the *M. graminicola* genome. Lastly, the large proportion of LoHs and CNVs compared to the number of SNVs is intriguing because it is unexpected.

.

Résumé en Français

Dans le chapitre III, nous avons présenté un assemblage de qualité du génome de *M. graminicola* offrant des informations génétiques importantes, notamment sur la structure des gènes et le contenu en TE/HGT. Ce génome de référence nous a également permis de faire une analyse comparative entre les isolats de *M. graminicola,* présentée dans le chapitre IV, afin d'étudier la diversité intraspécifique et les mécanismes moléculaires de l'évolution du génome.

Pour répondre aux questions restées en suspens et évoquées à la fin du chapitre II, le présent chapitre IV se concentre sur l'étude de la diversité intraspécifique et l'origine de l'espèce. Les séquences génomiques de treize isolats de *M. graminicola*, provenant de rizières géographiquement éloignées de pays d'Asie du Sud-Est et du Brésil, ont été comparées au génome de référence afin de rechercher des variantes à un seul nucléotide (SNV) parmi les isolats. Le génome de référence que nous avons réalisé est une séquence génomique haploïde qui consiste principalement en des "scaffolds" hybrides provenant des deux haplotypes divergents (hétérozygotie de 1,36% ± 0,78). Les données de séquençage Hiseq provenant de deux haplotypes divergents se reporteront ainsi sur les mêmes "scaffolds", ce qui donnera des SNV que nous avons codé ("0/1") pour les positions hétérozygotes. Ces SNVs représentent donc potentiellement un ensemble comprenant les deux haplotypes de chaque génome ainsi que des variantes intra-espèces, sans que l'on puisse les distinguer. Par conséquent, dans notre étude, ce type de position a été exclue de l'analyse de la variabilité intraspécifique. En revanche la comparaison des génomes provenant des 12 isolats au génome haploïde de référence (sauf Mg-VN18 qui est le génome de référence) a permis d'identifier 44 527 SNVs homozygotes inter-isolat. En d'autres termes, pour chacun de ces SNVs, les deux haplotypes ont la même séquence au sein d'un isolat, mais le polymorphisme existe entre deux ou plusieurs isolats. L'analyse par composantes principales basée sur ces SNVs a montré une absence de structure des populations sous-jacente à l'origine géographique. Ces SNVs identifiés dans chaque isolat ont ensuite été utilisés pour réaliser des tests de recombinaison (test des "quatre gamètes" et déséquilibre de liaison (LD) entre les paires de marqueurs). Une augmentation exponentielle et rapide de la proportion de paires de marqueurs passant le test à "quatre gamètes", combinée à une diminution de la LD, ont permis de suggérer que des événements de recombinaison (RE) abondant se retrouvent dans le génome de *M. graminicola*.

La visualisation de l'ensemble de ces SNVs identifiés, parmi les 13 isolats sur les « scaffolds » du génome de référence, a révélé l'existence de perte d'hétérozygotie (LoH) dans chacun des 13 isolats. Les régions du génome affectées par la perte d'hétérozygotie sont différentes entres les isolats et représentent une longueur cumulée pouvant atteindre 7,4 % de

la longueur totale du génome. Il est intéressant de noter que 490 kb qui ont été affecté par ces LoH (cinq "scaffolds", ~1,18% de la taille du génome) étaient partagé entre neuf des isolats suggérant que certaines régions étaient plus propices à cette perte d'hétérozygotie. De plus, certains effecteurs potentiels ont été affectés par ce phénomène de LoH (conservation d'un seul allèle) ce qui pourrait participer à l'adaptation de ces isolats à leur environnement.

Les 44 527 SNVs homozygotes des 12 isolats se retrouvent pour la plupart liées aux régions de LoH et ne permettent pas d'étudier le polymorphisme sur le génome de l'ensemble des isolats. Par conséquent, nous avons sélectionné un ensemble de 40 SNVs situés sur des régions autres que celles où des LoHs ont été observés afin de pouvoir comparer simultanément les 12 isolats. L'analyse PCA utilisant ces 40 SNPs offre une absence de structuration similaire à celle observée en utilisant les 44 527 SNVs. Cette absence de signal phylogéographique déjà observé dans l'étude du génome mitochondrial, renforce l'hypothèse d'une propagation récente de cette espèce.

Le génome de référence haploïde fusionné nous a montré ses limites dans la détection de variantes de nucléotides uniques parmi le génome de différents isolats. Par conséquent, deux haplotypes de 18 régions génomiques du génome haploïde (36 haplotypes au total, 354 ko) ont été construits manuellement. Seuls deux SNVs ont été détectés parmi ces 36 haplotypes entre les 12 isolats ($0,5 \times 10^{-4}$ mutation par base). C'est un taux très faible comparé à *M. incognita*. D'autre part, l'analyse de la couverture de ces 36 haplotypes a confirmé l'existence d'événements de LoH.

Des variations du nombre de copies (CNV) ont également été étudiées entre le génome des 13 isolats. Au total, 104 régions CNVs ont été trouvées et peuvent affecter l'expression de 266 gènes. Il est intéressant de noter que certains isolats, qui ont été distingués sur la base de leurs SNV, ont également présenté des pertes de gènes et des événements LoH importants.

Malheureusement, le détail des environnements auxquels ont été exposés ces 13 isolats avant d'être prélevé en rizière n'a pas été décrit. De plus aucun lien entre l'apparition d'un de ces réarrangements moléculaires et l'exposition à un stress n'a été démontré. Par conséquent, l'impact adaptatif que pourraient jouer les SNV, LoH et CNV sur le génome de l'isolat reste à être étudié. Cependant, nous pouvons émettre l'hypothèse que les TE, SNV, LoH et CNV sont des mécanismes moléculaires qui favorisent la plasticité génomique et qui pourraient intervenir dans l'évolution du génome de *M. graminicola*. Enfin, la proportion importante de LoH et CNVs au regard du nombre de SNVs est intrigante car inattendue.

Genomic rearrangements are putative genomic evolutionary traits in the facultative meitotic parthenogenetic nematode *Meloidogyne graminicola*

Ngan Thi Phan[1], Etienne G.J. Danchin[2], Georgios D. Koutsovoulos[2], Marie-Liesse Vermeire[1], Guillaume Besnard[3,*], Stéphane Bellafiore[1,*]

[1]PHIM Plant Health Institute, Univ of Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

[2]Institut Sophia Agrobiotech, INRAE, Université Côte d'Azur, CNRS, 06903 Sophia Antipolis, France

[3]CNRS-UPS-IRD, UMR5174, EDB, 118 route de Narbonne, Université Paul Sabatier, 31062 Toulouse, France.

* co-senior authors for correspondence: guillaume.besnard@univ-tlse3.fr (G.B.); stephane.bellafiore@ird.fr (S.B.)

Abstract

Meloidogyne graminicola, commonly called the rice root-knot nematode, is a devastating pest causing serious damages at a global scale, especially on rice. The success of this parasite in a diversity of habitats and hosts raise questions about its adaption mechanisms, and particularly the ones responsible for an evolution of its virulence. This pathogen is supposedly mainly reproducing via meiotic parthenogenesis, with probable, but rare, sexual events. The impact of this reproductive mode on the genetic variation of *M. graminicola* and its adaptive response remains poorly understood. In this study, we used comparative genomics among 13 *M. graminicola* isolates sampled over a large geographical scale in order to test for genomic recombination, and to detect sequence polymorphisms including copy number variants (CNV). Loss of heterozygosity were frequently observed among isolates, which might be due to meiotic recombination, gene conversion, and/or translocation. The analysis of linkage disequilibrium and four-gamete test supported the presence of meiotic recombination. The accumulation of CNVs occurred independently from one isolate to another and seems to be associated with the accumulation of homozygous genomic blocks as well as single nucleotide mutations. Due to lack of reference sequence of two heterozygous haplotypes, single nucleotide polymorphisms on the entire genome of the 13 isolates was not addressed. A set of representative SNPs (40 SNPs) were identified among genome of 13 isolates. These polymorphisms did not reveal any clear phylogeographic structure, reinforcing the hypothesis of a recent and large-scale spread of the parasite especially in South-East Asia, and clonal adaptation to the environment.

Key words: comparative genomics, copy number variant (CNV), parthenogenetic, pathogen, root-knot nematode (RKN), sexual recombination, single nucleotide variant (SNV).

I. Introduction

More than 90% of the rice consumed in the world come from Asia, where around 700 million tons of paddy rice is annually produced (FAOSTAT, 2019). This crop is however increasingly threatened by the prevalence of *Meloidogyne graminicola*, the so-called rice Root-Knot Nematode (RKN). In Asia, RKNs is estimated to reduce rice yield by about 15% (reviewed in Mantelin et al., 2017). In particular, *M. graminicola* leads to yield reductions of 8-70% for upland rice in the Philippines, and 16-97% for lowland rice in India, Nepal, Bangladesh, and Indonesia (Bridge et al., 2009; Padgham et al., 2004). In the field, these losses may be exacerbated when combined with other biotic or abiotic stresses, such as drought. In addition, the damages caused by this pathogen are frequently underestimated due to incorrect disease diagnosis, since infected plants may present phenotypes similar to nutritional and water disorders. Not only causing damages in rice, *M. graminicola* is a polyphagous pathogen infecting more than 120 host plants, including both cultivated crops and weed species *(EPPO Global Database,* 2020). This devastating plant pathogen is therefore classified as a quarantine pest in several countries (e.g. Brazil, Madagascar, China; EPPO, 2017-2018) and was added recently to the EPPO Alert List in Europe after being introduced in northern Italy (Fanelli et al., 2017).

Recent surveys show that *M. graminicola* has an almost worldwide distribution. It was first found in the USA (Golden & Birchfield, 1965) and Laos (Golden & Birchfield, 1968) in the sixties, before being identified everywhere in Asia where rice is grown, and more recently, in Madagascar (Chapuis et al., 2016) and Italy (Fanelli et al., 2017). *Meloidogyne graminicola* has the ability to rapidly increase its population size and spread over a large area in a very short period of time. A freshly hatched juvenile can develop into an adult female laying 250 to 300 eggs after only 18-21 days (Bellafiore et al., 2015; Phan et al., 2018). For example, in northern Italy, where this pest has recently been detected, the total infected area has increased by about five folds in just one year (from 19 to 90 ha in 2016-2017; EPPO Reports, 2017). In Asia, modifications in agricultural practices in response to both climate and socioeconomic changes have also led to a dramatic *M. graminicola* increase (De Waele & Elsen, 2007). Despite the prevalence of this species on a large geographical scale, studies questioning the origin, spread, population structure and adaptation mechanisms of this pathogen are limited (Besnard et al., 2019). Morphological traits, esterase profiles or molecular data (rDNA or mitogenome) have been used to explore the intraspecific variability of *M. graminicola*, but did not reveal any geographical structure among isolates (Bellafiore et al., 2015; Besnard et al., 2019; Pokharel et al., 2010; Carneiro et al., 2000), suggesting a recent, wide expansion of this pathogen. However,

comparative and population genomics of isolates might confirm this hypothesis and reveal intramolecular variations responsible of species adaptation and parasitic success.

Reproduction mode plays a key role in the evolutionary and adaptive potential of pathogenic species (Gibson et al., 2017). Sexual reproduction allows combining the genetic material of two parental individuals to produce genetically-diverse offspring. New gene combinations are thus produced, offering opportunities to adapt to changing biotic interaction or environment (Burt, 2000). Besides, deleterious mutation are more efficiently eliminated in sexual species (Kondrashov, 1982; Lynch et al., 1993). An animal with clonal reproduction (e.g. parthenogenesis) has poorer adaptability because advantageous alleles from different individuals cannot be combined, and selection efficiency is impaired eventually leading to progressive accumulation of deleterious mutations (Glémin et al., 2019). Therefore, parthenogenesis was consider as evolutionary dead. However, asexual reproduction can be advantageous under favorable conditions, and allow a rapid increase of population. In this situation, the genotype adapted to the environment is more easily preserved than under sexual reproduction. Interestingly, several asexual species appeared to have amazing adaptability and comparative genomic has revealed several mechanisms of molecular evolution in their genome (Vakhrusheva et al., 2018, Archetti, 2004). The genome structure and sequence of asexual parasites can evolve via the insertion of transposable elements, recombination, loss of heterozygosity, and gene duplications/deletions (i.e. gene copy number variants - CNVs), that may help to bypass some selection pressures or breaking down plant's defense response (Archetti, 2004, Jaron et al. 2020). For instance, in the mitotic parthenogenetic *M. incognita*, convergent gene loss events have been observed in two isolates that broke down plant resistance (Castagnone-Sereno et al., 2019). *Meloidogyne graminicola* mainly reproduces through meiotic parthenogenesis. Sexual reproduction has been hypothesized to occur occasionally based on a few evidence of sperm pronuclei inside the eggs (~0.5%; Triantaphyllou, 1969). Pathogens with a mixed reproductive mode are thought to have an evolutionary advantage by combining clonal proliferation, when the genotype is adapted to the environment, and sexual reproduction with new allelic combinations in adverse conditions (McDonald & Linde, 2002). If this is verified for *M. graminicola*, it could explain the particularly strong aggressiveness of this pest towards rice in very different agrosystems. Unfortunately, these hypotheses on the *M. graminicola* reproduction mode are based only on cytogenetic observations (Triantaphylou, 1969), and, so far, have never been confirmed at the molecular level. The molecular evolution in *M. graminicola* genome remains unknown.

In this study, thanks to the recent publication of a high quality *M. graminicola* genome sequence (Phan et al., 2020), comparative genomics of isolates sampled on large area was conducted to (i) investigate the global genetic diversity at the genome level, and (ii) search for molecular markers to describe the evolution of the genome. To address these questions, we have sequenced the genome of 13 isolates mostly distributed in South-East Asia. Genomes of these isolates were compared to the reference genome to detect sequence variants such as single nucleotide variants (SNVs), loss of heterozyosity (LoH), and copy number variants (CNVs). SNVs identified among isolates were used to study the species phylogeography and to test for evidence of recombination. The results of our study revealed that *M. graminicola* populations exhibited evidences for recombination and numerous genomic reorganization with potential impact on their adaptive evolution. This pathogen also likely evolved a specific mechanism allowing it to adapt to changing environments through gains or losses of gene copies. The absence of clear phylogeographic signal based on genomic variants, similarly to that observed on the mitochondrial genome, reinforces the hypothesis of a recent spread of this species.

II. Materials & methods

2.1. **Nematode sampling, DNA extraction, sequencing and reads cleaning**

Thirteen isolates of *M. graminicola* originating from distant locations were collected in rice fields in South East Asia and Brazil (Table S1). These isolates were obtained after single nematode infection, as described in Bellafiore et al. (2015). Subsequently, each isolate was propagated for two months on the susceptible rice cultivar IR64 before DNA extraction according to the methods described in Besnard et al. (2019). Briefly, juveniles and eggs were extracted from roots using a hypochlorite extraction method and a blender, then treated for 15 min in 0.8% hypochlorite at room temperature, purified on a discontinuous sucrose gradient, and finally rinsed several times with sterile distilled water before being stored at -80 °C. The total genomic DNA (gDNA) was isolated after treatment with proteinase K and extraction with phenol-chloroform followed by ethanol precipitation, as described by Besnard et al. (2014). DNA quality and quantification were assessed with a NanoDrop spectrophotometer (Thermofisher, Waltham, MA, USA) and PicoGreen®dsDNA quantitation assay (Thermofisher). Finally, DNA integrity was checked by electrophoresis, loading 1 µL on a 1% agarose gel.

High-depth short-read sequencing was performed at the GeT-PlaGe core facility, INRAE Toulouse. For each accession, 300 ng of double-stranded DNA (dsDNA) was used for shotgun sequencing using Illumina technology (San Diego, CA, USA). DNAs were sonicated to get inserts of approximately 380 bp for adaptor ligation using the Illumina TruSeq DNA

Sample Prep v.2 kit. Libraries were multiplexed and inserts were then sequenced from both ends (2×100 bp, 2×125 bp, 2×150 bp) on HiSeq 2000 or 2500 (Illumina) (Table S1).

Illumina sequences were trimmed and cleaned. First, adapter sequences were removed with cutadapt (Martin, 2011). Then, overrepresented sequences and low quality bases ($q < 20$) were removed with Skewer (Jiang et al., 2014). Subsequently, the quality of sequences was assessed with fastQC (Andrews, 2010). Minor GC-rich peaks (ca. 70%) were revealed in five isolates (Mg-Bl, Mg-Bn, Mg-C21, Mg-C25, and Mg-P) indicating potential contaminants (e.g. from bacteria or fungi). Sequences of such contaminants were detected and removed based on the method proposed by Kumar et al. 2013. In short, the genome of each isolate was pre-assembled in contigs using Spades (Bankevich et al., 2012), that were blasted against the NCBI nt database to annotate taxa with high identity (>75%). Finally, the sequencing coverage, GC content, and length, as well as the annotated taxa were summarized and visualized on each pre-assembled contig using blobtools (Laetsch & Blaxter, 2017). Based on this information, contaminations of Proteobacteria and Firmicutes were detected only in these five isolates. Therefore, reads from these contaminants were removed using Bowtie2 (Langmead & Salzberg, 2012), resulting in the final set of cleaned reads. Five isolates (Mg-Bali, Mg-Borneo, Mg-Java2, Mg-L2, Mg-VN18) showed high genomic coverage (288 to 568× relative to the size of the reference haploid genome - 41.5 Mb), while other isolates had lower coverage ranging from 21 to 39×. Therefore, genomic sequences of those five isolates were down-sampled to 20% using the SEQTK tool (Li, 2020).

2.2 Calling genotype of nucleotide variants at entire genome of 13 isolates

To detect variants among the 13 genome sequences, Illumina reads of each isolate were mapped to the haploid reference genome of *M. graminicola* (isolate Mg-VN18) using the BWA-MEM software (Li & Durbin, 2009). SAMtools (Li et al., 2009) was used to filter alignments with MAPQ lower than 20, to sort the alignment file by reference position, and to remove multi-mapped alignments. GATK (McKenna et al., 2010) was used to mark and remove the duplicated reads (MarkDuplicate). FreeBayes (Garrison & Marth, 2012) was used to detect the variants (SNVs and Indels) using all alignments simultaneously to produce a variant call file (VCF). The VCF file was filtered using vcftools (Danecek et al., 2011), retaining only the positions that had more than 30 phred-scaled probability (minGQ), a minimum coverage depth (minDP) of ten and no missing data. The *M. graminicola* genome has an average heterozygosity of 1.36% ± 0.78 (Phan et al., 2020; Besnard et al., 2019). The final reference genome is a haploid genomic sequence that mainly consists of collapsed haplotypes (Phan et al., 2020).

Thus, after mapping of the Mg-VN18 reads to the Mg-VN18 haploid genomic reference sequence (Phan et al., 2020), the collapsed contigs obtained corresponded either to heterozygous regions [i.e., two haplotypes (or alleles) detected and coded as follow: genotype "0/1" where 0 = first haplotype sequence identical to the reference, and 1 = second haplotype sequence]; or to homozygous regions (identical to the reference sequence and coded "0/0"). When reads of other isolates were mapped to the reference genome, six possible genotypes were defined as follow: "0/0" = homozygous region with the same sequence as the reference, "0/1" = heterozygous region containing the reference haplotype sequence and the alternative haplotype of Mg-VN18, "1/1" = homozygous region bearing only the alternative haplotype of Mg-VN18, "0/2" = heterozygous region containing the reference haplotype sequence and a second alternative allele (distinct from this detected in Mg-VN18), "1/2" = heterozygous region bearing first and second alternative haplotypes, and "2/2" = homozygous region bearing only a second alternative allele (not detected in Mg-VN18). A matrix of the variant genotypes of the 13 isolates was constructed allowing us to score polymorphism for each position in the scaffolding. The distribution of variant genotypes for each isolate on the reference scaffolds was visualized using CIRCOS (http://circos.ca/) in order to reveal specificities of each isolate (i.e., polymorphism, genome organization). Note that homozygous "0/0" and heterozygous "0/1" genotypes among all isolates represent non-variable regions. Meanwhile, heterozygous "1/2" and "0/2" genotypes as well as homozygous "1/1" and "2/2" genotypes correspond to true variants between the reference Mg-VN18 and any other isolate. Therefore, the final matrix of single nucleotide variants (SNV) only includes loci which have at least one variant genotypes ("1/1", "2/2", "1/2", "0/2") among all isolates.

2.3 Principal component analysis using single nucleotide varants at entire genome of 13 isolates

To study the potential structuring of *M. graminicola* populations using the identified SNVs, a Principal Component Analysis (PCA) was conducted. The SNV matrix between the 13 isolates was used to perform a PCA using the SNPRelate package with default parameters (Zheng et al., 2012). We calculated fixation index (Fst) for the clusters and South East Asian population base on Weir and Cockerham's method 1984 using vcftools (Danecek et al., 2011).

2.4 Linkage disequilibrium and 4-gamete test single nucleotide variants at entire genome of 13 isolates

Homologous recombination in the *M. graminicola* genome can be revealed by studying the linkage disequilibrium between SNV markers as well as the proportion of pairs of markers

passing the 4-gamete test, as a function of their distance along the genome. The SNV matrix between isolates was thus used for applying the 4-gamete test and estimating the linkage disequilibrium (LD) between nucleotide variants, using the script developed by Koutsovoulos et al. (2019). Note that SNVs in each isolate were phased to two haplotypes using WhatsHap (Martin et al., 2016) before inputting to 4-gamete test and LD. If a recombination event occurred between two bi-allelic sites (or markers) during meiosis I, four gametes will be formed at the end of meiosis II, including two parental haplotypes and two recombinant haplotypes. Among the 13 isolates, the number of haplotypes formed by two markers (at diploid state) on the same scaffold was counted. Then, the number of marker pairs was grouped according to their physical distances. At each physical distance, the proportion of pairs of markers containing the four products of meiosis (haplotypes) was calculated. When the distance between two markers is large, recombination is likely to occur frequently, so if meiosis recombination occurs, the proportion of marker pairs that pass the 4-gamete test should increase with the distance between markers.

The LD (r^2) between the pairs of markers was calculated according to the following equation: $r^2 = (q_1q_2 – p_1p_2)^2/r_1r_2r_3r_4$, where q_1 and q_2 are the frequencies of two parental haplotypes (AB and ab); p_1 and p_2 are the frequencies of two recombinant haplotypes (Ab, aB); r_1, r_2, r_3 and r_4 are frequencies of four alleles (A, B, a, b) (Slatkin, 2008). When the two markers are close together, there is less chance of recombination, resulting in a low frequency of recombinant haplotypes and a high r^2 value. On the other hand, the greater distance between markers, the greater the possibility of recombination events leading to a decrease in the r^2 value. The value of r^2 was between 1 [no recombination, $p_1p_2 = 0$, $(q_1q_2)^2 = r_1r_2r_3r_4$] and 0 (when recombination always occurs, $q_1q_2 = p_1p_2$).

2.5 Coverage analysis and nucleotide variant calling using a set of nuclear divergent copies

The current assembly tools using a combination of both long and short reads were not able to assemble two distinct haplomes of *M. graminicola* genome. Aware of that, two-state genotypes were considered when selecting the nucleotide variants among isolates for principal component analysis and recombination test. To further support the reliability our results, principal component analysis and recombination test were repeated using a subset of curated haplotyped contigs in *M. graminicola* genome which were manually separated. As a result, the 16 pairs of low-copy contigs (165 kb) with sequence divergence ranging from 0.3 to 5.7% were extracted from the reference consensus genome of *M. graminicola*. For each contig, two divergent copies

were separately assembled using the methods described in Besnard et al. (2019). Briefly, the HiSeq sequences of the Mg-VN18 isolate were mapped to all 16 nuclear contigs using the BWA-MEM software (Li & Durbin, 2009). The mapping of paired-end reads on these contigs was also visualized and inspected with Geneious v.6 (Kearse et al., 2012). Then, based on the linkage between paired-end reads, reads were carefully phased and concatenated into separate haplotypes, resulting in the assembly of two divergent contig pairs (i.e. two haplotypes). The HiSeq reads of Mg-VN18 were again mapped to the assembled divergent copies of each contig to correct the sequence of the two copies. In addition, the two haplotypes from TAA6 and ACC6 contigs (Besnard et al., 2019) were also used in this study (Table S2). Finally, haplotypes of 18 contigs (178 kb) (including 16 news and two published contigs) were separated (in 36 sequences covering 354 kb) and used as a new reference.

The HiSeq reads of the 13 isolates were then mapped to these 36 regions using BWA-MEM. Each alignment was visualized on Geneious to manually verify that reads were clearly separated into two divergent copies with a correct read phasing. Then, the average read-coverage was calculated for each region. In addition, variants on the divergent copies were called among the isolates using the same method as described above. All detected variants were manually validated by a visual inspection of the alignment with Geneious. The quality control of the read mapping and variant search were also performed on five populations (Mg-Bali, Mg-Borneo, Mg-Java2, Mg-L2, and Mg-VN18) with very high sequencing coverage (288 to 568×).

2.6 Haplotype network, principal component analysis, and recombination test using a subset of single nucleotide mutations among genome of isolates

Visualization of positions and genotypes of nucleotide variants among genome of 13 isolates on scaffolds at section 2.2 show the clusters of homozygous SNVs (“1/1”) that represent genome reorganization events. Genomic regions of more than 3 kb showing suites of homozygous SNVs (“1/1”) separated by less than 100 bp were detected and categorized as genome reorganization events. Meanwhile, sporadically distributed SNVs (“1/1” and “2/2”) in the genome were classified as single nucleotide mutations. These single mutations were then manually checked on the alignment to ensure they are 1) highly supported (minimum coverage of 10×); 2) not belonging to duplicate regions; and 3) not located near mononucleotide stretches (e.g. poly-T or poly-A) that may cause sequencing errors. Finally, this set of nucleotide variants was also used as input into the recombination test (4-gamete test and LD) and the principal component analysis as described above. Finally, this SNVs set were aligned to construct a median joining network using PopArt v1.7 (Leigh & Bryant, 2015).

2.7 Detection of copy number variants

To identify putative copy number variants (CNVs) in the genome sequence of the 13 isolates, the cn.MOPS algorithm v.1.24.0 (Klambauer et al., 2012) was used. First, the alignment of Hiseq reads of each isolate against the haploid reference genome was used to calculate read coverage per 1-kb sliding window using Bedtools *multicov* (Quinlan & Hall, 2010). Then a matrix of read-coverage per 1-kb sliding windows along the scaffolds of all 13 isolates was used as input to cn.MOPS program. The cn.MOPS program was run using medium normalization mode, and default values for all other parameters. cn.MOPS is a multiple sample read depth method that applies a Bayesian approach to decompose read coverage variations across multiple samples at each genomic position into integer copy numbers. By using Poisson distributions, noise was detected and removed, thus, reducing false positives. Finally, low complexity regions of the genome were excluded from this analysis.

Protein-coding genes overlapping CNV regions (CNVRs) on at least 70% of their length were selected and considered involved in CNV. The classification of the proteins discovered in these CNVs was carried out by searching for functional domains (*InterProscan* domains) with the *InterProscan* tool (Mitchell et al., 2019).

III. Results

3.1 Detection of loss of heterozygosity among isolates

Variants calling allowed detecting 44,527 SNVs among 12 isolates (except Mg-VN18), including loci which have at least one variant genotypes ("1/1", "2/2", "1/2" , "0/2") among all isolates. Visualizing the positions of these variants on scaffolds revealed the variable genome structure between isolates (Figure S1). Indeed, variants are not uniformly distributed but instead enrichments (SNV clusters) are visible in certain genomic regions. In particular, at the same positions, some isolates show ranges of homozygous variants ("1/1" or "0/0"), while other isolates have a heterozygous status ("0/1"). For example, at position 20 to 55 kb in contig mg58 (Figure S1), Mg-Java2 and Mg-Borneo show a homozygous region with the reference haplotype ("0/0"), Mg-VN27 has a homozygous region with the alternative haplotype ("1/1"), while others were heterozygous ("0/1"). Genomic regions (> 3 kb) with such a contrasted pattern of homozygosity vs. heterozygosity among isolates were recorded (Table 1), showing that some isolates displayed more homozygous regions (loss of heterozygosity) than others. For instance, at least 20 loss of heterozygosity (LoH) regions (covering from 6 to 7.2% of the genome) were detected in Mg-C25, Mg-VN27, Mg-Bali, and Mg-P, whereas none were observed in Mg-VN11 and Mg-Brazil (Table 1). This phenomenon seems to have occurred

independently among isolates since there was no clear association with their geographic origin. For example, Vietnamese population Mg-VN27 showed the highest proportion of LoH (7.2%) while the opposite for others Vietnamese population (Mg-VN11, MgVN6, Mg-VN18) with less than 1.3% of LoH. Meanwhile, populations from distance continents such as Asia (Mg-VN11) and America (Mg-Brazil) share the same level of LoH (0%).

Table 1. Number and total length of genomic regions that have reached fixation (loss of heterozygosity) among 13 *M. graminicola* isolates

Isolate	# fragments (>3 kb)	Length (kb)	% genome size (haploid)
Mg-VN27	27	3 010	7.24
Mg-C25	26	2 675	6.44
Mg-Bali	21	2 532	6.09
Mg-P	21	2 480	5.97
Mg-L2	14	1 616	3.89
Mg-C21	18	1 463	3.52
Mg-Java2	13	1 303	3.14
Mg-Borneo	10	843	2.03
Mg-L1	8	544	1.31
Mg-VN6	6	520	1.25
Mg-VN18	6	469	1.13
Mg-VN11	0	0	0.00
Mg-Brazil	0	0	0.00

Notably, scaffolds mg78, mg87, mg145, mg157, and mg173, which cover 490 kb (~1.18% genome size), are fully homozygous in nine isolates (Mg-VN6, Mg-L1, Mg-Java2, Mg-Bali, Mg-Borneo, Mg-VN27, Mg-C25, Mg-C21, and Mg-P), while heterozygous in the four remaining genomes (Mg-VN18, Mg-VN11, Mg-Brazil, and Mg-L2; Figure 1). Among the nine isolates, Mg-P has fixed a different haplotype from all the others at scaffold mg78 (Figure 1). On these five scaffolds, 106 protein-coding genes were found (Table S3). The predictive functions of 63 genes were found on the basis of *pfam* domains present in the corresponding proteins (*InterProscan*) (Table S3). These genes are potentially involved in various molecular functions and biological processes including signal transduction, transcription control, molecular transportation, protein degradation, cellular components control, and others (Table S3). Interestingly, among 106 genes, 12 encode supposedly secreted proteins (11.7 % number of genes) (Petitot et al., 2015) (Table S3). Among the 29 galectin genes identified in the whole genome (Phan et al., 2020), eight were found on these five scaffolds (Table S3).

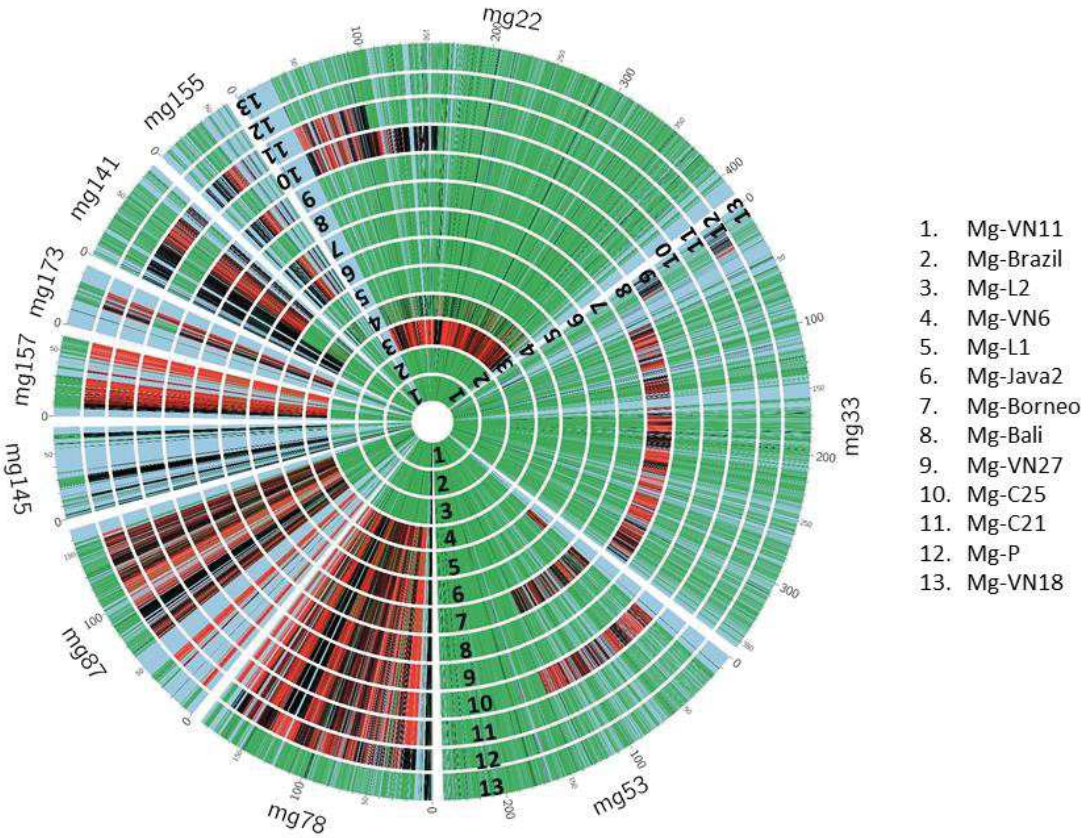


Figure 1. Presentation of loss of heterozygosity events on 13 isolates on ten scaffolds (mg78, mg87, mg145, mg157, mg173, mg141, mg155, mg22, mg33, and mg53) in which LoH were common among isolates. *The length of scaffolds was drawn in scale. Each circle represented genome sequence of each isolate. Each vertical line indicates the genotype and position of SNV on each scaffold with color code: green = heterozygous state (genotype "0/1"); black = homozygous state with reference haplotype (genotype "0/0"); red = homozygous state with the alternative haplotype (genotype "1/1"). Therefore, the "green" regions indicate heterozygous state while the "red and black" regions indicate loss of heterozygosity (LoH). The blue regions indicate the other regions (conserved homozygous regions and insertions/deletions between isolates).*

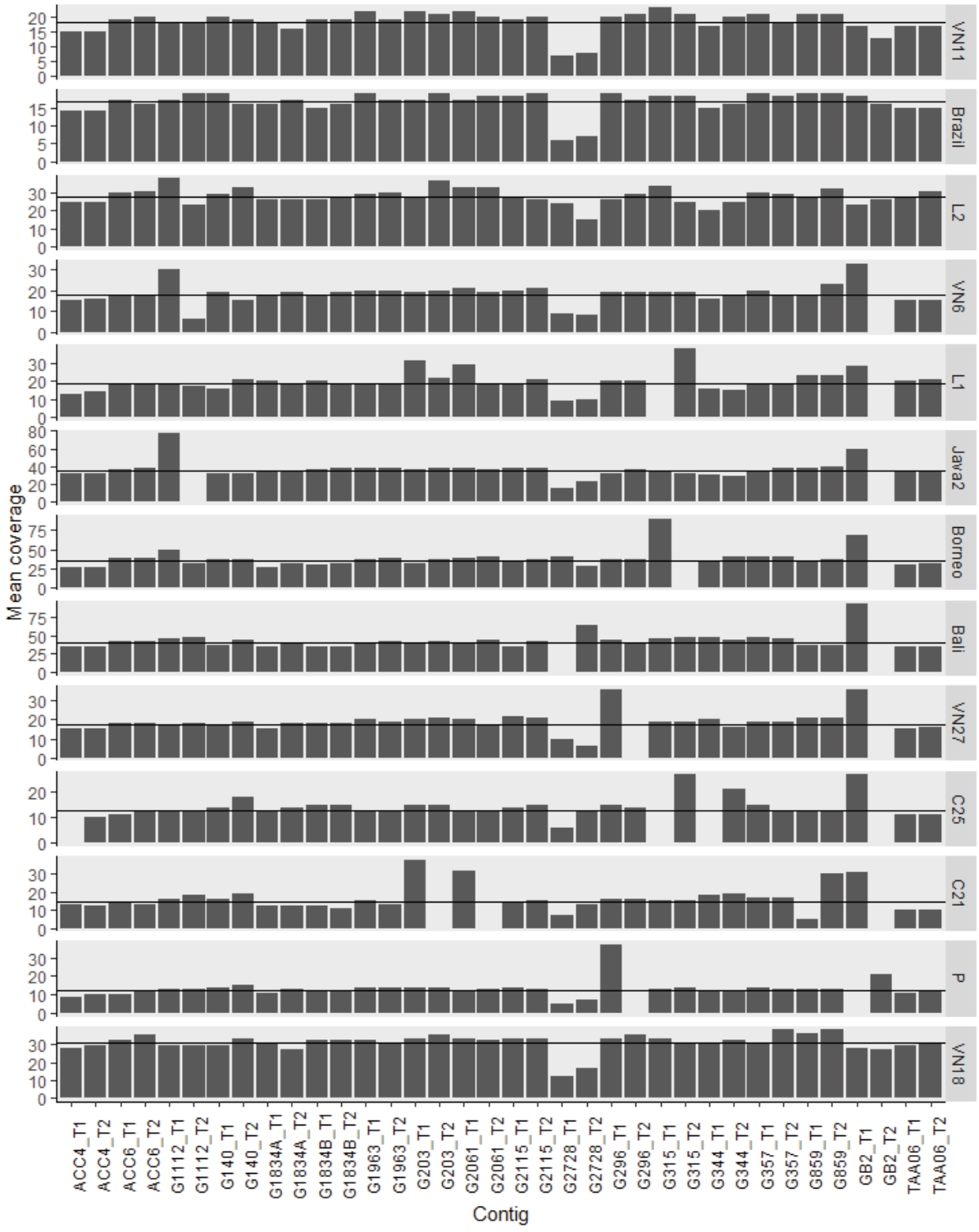


Figure 2. Average read-coverage on the two haplotypes of 18 contigs among 13 *M. graminicola* isolates. *The black horizontal line indicates mean value of read-coverage of all divergent copies in each isolate. Missing columns (read-coverage = 0×) indicates the absence of one haplotype*

Manually curated haplotypes (i.e. 18 pairs of contigs covering 354 kb; 0.43% total genome) were then used in order to specifically assess the level of heterozygosity in these

regions. Visual inspection of reads mapping on the divergent haplotypes revealed that some isolates have only one type (i.e. one haplotype is missing in some isolates). Note that reads assignment to each haplotype of all targeted regions was unambiguous, as shown on Figure S2. Therefore, for a given genomic contig, the average read-coverage on the two divergent haplotypes in each isolate showed their presence or absence. While all divergent haplotypes were present in four isolates (Mg-VN18, Mg-VN11, Mg-Brazil, and Mg-L2), some were absent in the remaining nine isolates (Figure 2). The GB2_T2 contig was absent in eight of the 13 isolates (Mg-C21, Mg-C25, Mg-Vn27, Mg-Bali, Mg-Borneo, Mg-Java2, Mg-L1 and Mg-VN6), contig G315_T1 was absent in two isolates (Mg-L1 and Mg-C25), and finally, nine other contigs were absent in only one isolate (Figure 2). Out of these 11 cases where a haplotype was missing, this was systematically accompanied by the doubling of sequencing coverage of its counterpart in ten cases. As an example, in Mg-C21, contig GB2_T2 was absent but its homologous GB2_T1 showed a 32× coverage, approximately twice the average sequencing depth of this isolate (15×; Figure 2). For these ten contigs, these observations suggest that one haplotype was present at the homozygous state (i.e. fixation). Only the contig ACC4_T1 and its counterpart ACC4_T2 did not appear to follow this rule in the Mg-C25 isolate. In this case, absence of ACC4_T1 was not accompanied by higher read coverage of its counterpart ACC4_T2, suggesting a single deletion of ACC4_T1 leading to a hemizygous region (Figure 2).

As the sequencing coverage was not high for some isolates (ca. 15-30×), the analysis was repeated using five isolates including Mg-Bali, Mg-Borneo, Mg-L2, Mg-Java2, and Mg-VN18, for which high sequencing coverage is available (288 – 568×) A similar pattern was observed with sequencing coverage doubling of the extant contig at the expense of its absent counterpart (Figure S3).

3.2 Evidence for meiotic homologous recombination events in *M. graminicola*

Using the genome-wide SNVs matrix, we performed a LD analysis as well as a 4-gamete test to look for evidence of meiotic recombination. LD analysis revealed a logarithmic-exponential downward trend in LD between markers as a function of inter-marker physical distance (Figure 3). The LD curve (r^2) started at high value (0.98), followed by a dramatic slope down close to 0 (r^2 = 0.125) at 20 kb and which will remain at this threshold beyond (Figure 3). In the 4-gamete test, we observed a rapid and exponential increase in the proportion of markers passing the 4-gamete test to reach a value of 0.99 as soon as markers are more than 20 kb away. By performing this same test but using only 40 single nucleotide mutations among isolates (See

method section 6) the same LD and 4-gamete profiles were observed (Figure S5). The drop of LD and the concomitant increasing proportion of bi-allelic markers passing the 4-gamete with inter-marker distance constitute strong signature for meiotic recombination.

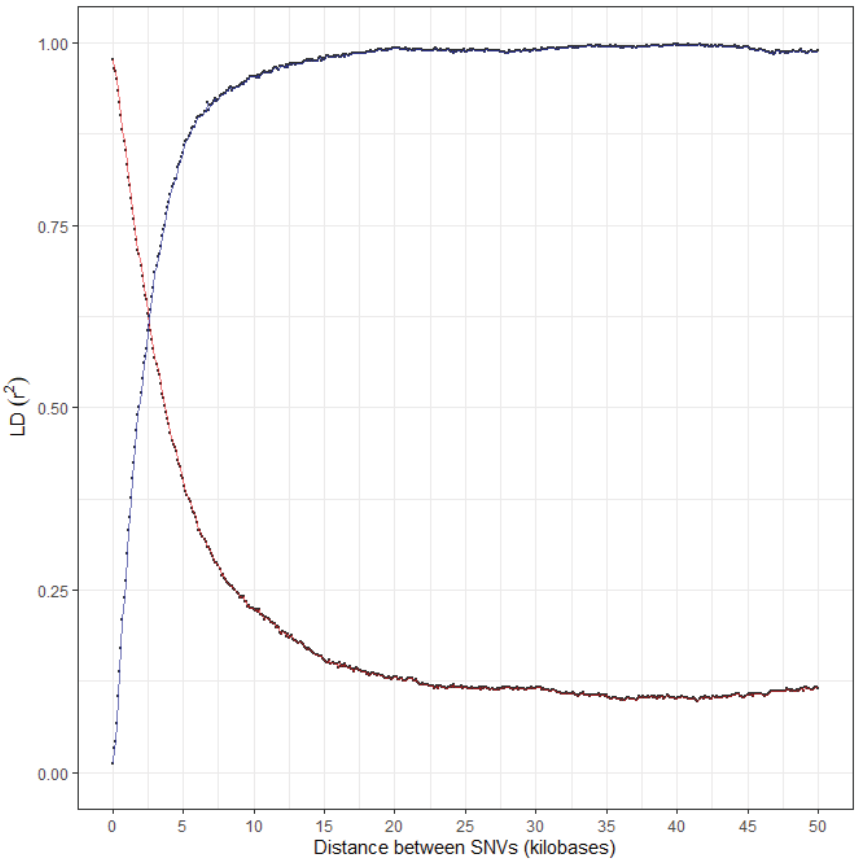


Figure 3. Linkage disequilibrium (LD) and 4-gamete test of *M. graminicola* isolates using the 44,527 SNVs matrix. *The red line indicated r^2 correlation between markers, indicating LD for each physical distance between DNA polymorphisms. The blue line indicated the proportion of pairs of bi-allelic markers that pass the 4-gamete test for each physical distance between the markers on a given scaffold*

3.3 Limited accumulation of point mutations and no clear geographical pattern of diversity

Careful visual inspection of reads mapping on the 36 manually curated haplotypes (354 kb) revealed only two SNVs on G859_T2 (one unique to Mg-C21 and another to Mg-P) and one InDel on G859_T1 (unique to Mg-C21). Surprisingly, only 40 single nucleotide mutations were found between 12 isolates compare to Mg-VN18 at whole genome after filtration and visual inspection (See method section 6). More point mutations were found in Mg-C25, Mg-P, Mg-C21, and Mg-VN27 isolates compare to Mg-VN18 genomic sequence (Figure S4).

The use of the 44,527 SNV did not allow the identification of population structure regarding geographical distribution as shown in the PCA (Figure 4). On the two first axes of this analysis (collectively accounting for approximately 45% of the variation), Mg-C25, Mg-P, Mg-VN27 and Mg-L2 were clearly distinguished from each other and from a relatively homogeneous group constituted by the other nine isolates. Similar pattern was observed using the matrix of 40 SNVs with a close relationship of eight isolates (Mg-Bali, Mg-Borneo, Mg-C21, Mg-Java2, Mg-P, Mg-VN11, Mg-VN6 and Mg-L1; Figure S6). It is interesting to note that although some isolates came from nearby geographic regions (e.g. Mg-C21 vs. Mg-C25; Mg-L1 vs. Mg-L2), they were distantly separated on the PCA. Conversely, Mg-Brazil, although isolated thousands of kilometers away from the others, presented a position close to our reference isolate Mg-VN18 from Vietnam (Figure S6). Therefore, the relationships between isolates did not relate to their geographical origin.

PCA plot do not allow us to cluster isolates. Therefore, fixation index (Fst) values was calculated to assess the genetic distance between populations from different countries belong to South-East Asia. Both mean and weighted Fst values between clusters were all less than 0.1, suggesting strongly genetic connections between the populations (Table S4).

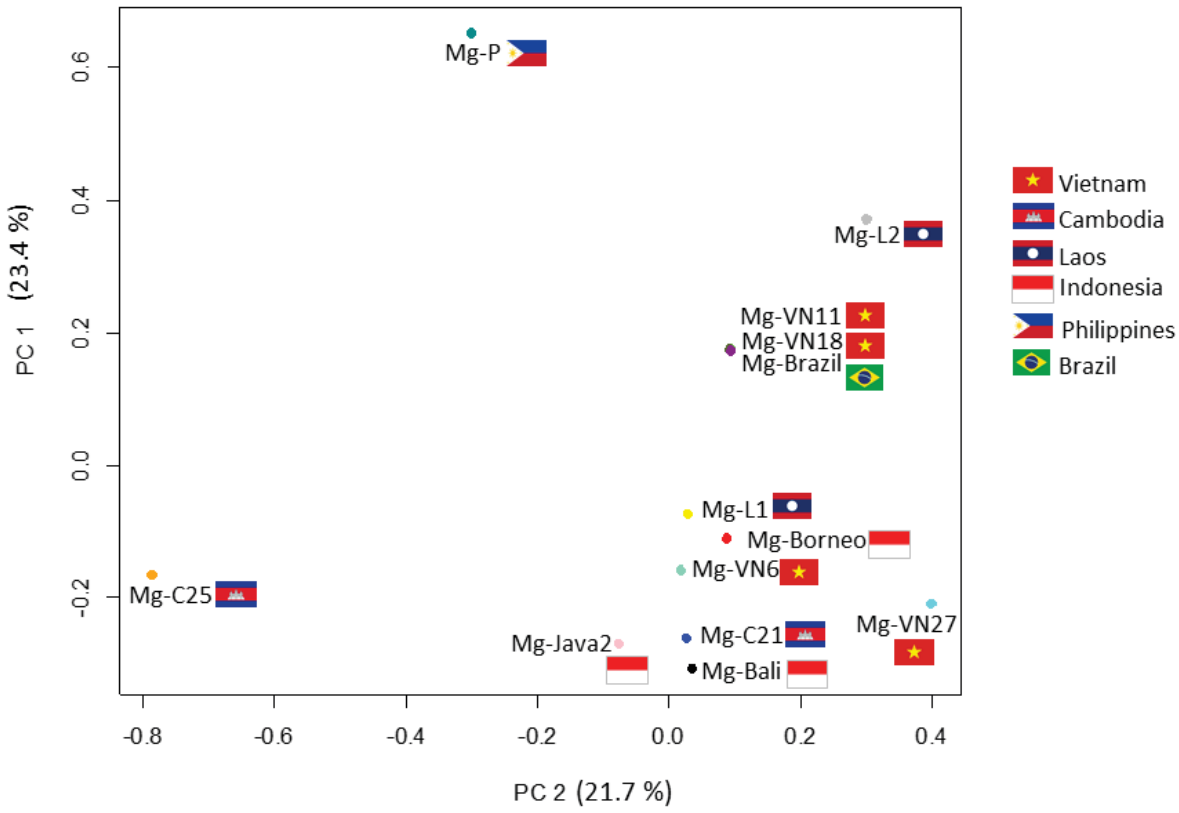


Figure 4. Principal component analysis of the 13 *M. graminicola* isolates using 44,527 SNVs showing no clear population structure underlying geographical origin.

3.4.Independent accumulation of copy number variants among isolates

Comparative genome analysis of the 13 isolates revealed 104 CNV regions in 39 scaffolds with CNVs ranging in size from 3 to 224 kb (Table S5, S6). While Mg-Bali, Mg-Java2, Mg-VN27, Mg-P, Mg-C21, and Mg-C25 showed more deletions than duplications, other isolates (Mg-VN18, Mg-VN6, Mg-VN11, Mg-Borneo, Mg-Brazil, Mg-L1 and Mg-L2) tended to accumulate many duplicated copies (Figure 5).

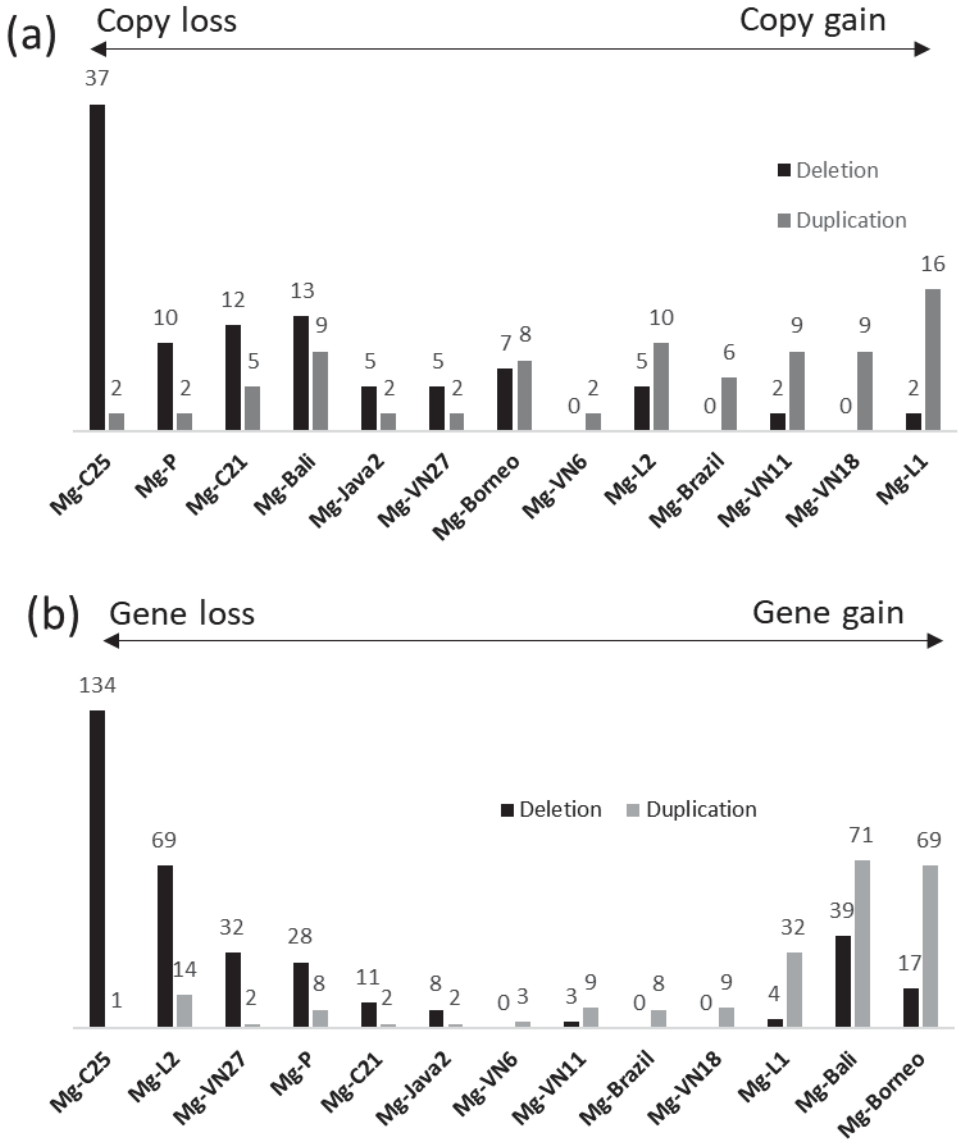


Figure 5. Copy number variants (CNVs) among 13 *M. graminicola* isolates: (a) Number of regions bearing CNVs among isolates, and (b) number of genes associated to regions experiencing gains or losses.

Interestingly, four of the five isolates (Mg-VN18, Mg-VN11, Mg-Brazil, and Mg-L2), which accumulated more duplications, were also those with the fewer LoH events as well as fewer point mutations (Figure 5, Table 1). At the same time, isolates that accumulate more deletions (Mg-P, Mg-C21, and Mg-C25) also showed more LoH events (Figure 5, Table 1) and higher number of point mutations (Figure 5, Figure S4). Furthermore, careful analysis of read coverage on the 18 pairs of homologous contigs suggested the ACC4 haplotype was lost in Mg-C25 (Figure S3)

Overall, the alignment with the reference genome (Phan et al., 2020) of the CNV regions (986.7 kb) revealed that 266 genes may be affected by CNVs. The Mg-C25 isolate is clearly distinguishable from the others with a large number of gene losses (134 genes) (Figure 5). Five isolates Mg-L2, Mg-C21, Mg-P, Mg-VN27, and Mg-Java2 had more deleted genes than duplicated genes. Four genes were found to be deleted on scaffolds mg87 in six isolates (Mg-Bali, Mg-C21, Mg-Java2, Mg-P, Mg-27, and Mg-25; Table S7). By investigating putative function of these genes (pfam; Table S7), we found that one gene encodes for a protein-tyrosine phosphatase-like (PTPLA), which is believed to play a role in signal transduction pathways and thus may affect the development, differentiation and maintenance of a number of tissue components. Function of the three remaining genes remains unknown (Table S7).

IV. Discussion

4.1 Recombination and loss of heterozygosity as evolutive mechanism in *M. graminicola*

This study indicates that genomic reorganizations have occurred in the nuclear genome of *M. graminicola* during its recent diversification. Using SNVs, we observed an exponential and rapid increase in the proportion of marker pairs passing the 4-gamete test combined with a decrease in the linkage disequilibrium (LD) as the physical distance between them increased. This 4-gamete test and LD pattern is different than the one of the apomictic *M. incognita,* wich reveal an absence of recombination (Koutsovoulos et al., 2019), but is similar than the one found in the meiotic amphimictic nematode *Globodera rostochiensis* (Koutsovoulos et al., 2019). In *G. rostochiensis*, the increase in the number of markers passing the 4-gamete test as well as the decrease in LD was observed at relatively short distance (< 1 kb; Koutsovoulos et al., 2019). A rapid LD decline over the genome had also been revealed in the Bdelloid rotifer *Adineta vaga* as resulting from homologous recombination (Vakhrusheva et al., 2018). These results suggests the occurrence of homologus recombination in *M. graminicola*. Although translocations and gene conversion can participate in breakage of LD (Lewin, 1990), this is not expected to follow a logarithmic-exponential distribution with inter-marker distance but would

rather create random noise around the distribution. Furthermore, translocations and gene conversions cannot explain that the proportion of pairs of bi-allelic markers passing the 4-gamete test increases exponentially with inter-marker distance as well while recombination does (Lewin, 1990).

Several large genomic regions have been fixed as homologous status in *M. graminicola* isolates suggests that a homogenization process occur, resulting in a loss of heterozygosity in some *M. graminicola* genome region (up to 7%) (Table 1). Loss of heterozygosity is expected in most forms of meiotic parthenogenesis (Schön et al. 2009; Engelstädter 2017). Loss of heterozygosity might unmask mutations that have little or no effect in heterozygous state, therefore, might be an important force for inducing variation among offspring. Different mechanisms can cause loss of heterozygosity in the genome of parthenogenetic animals such as recombination, gene conversion, and translocation. In facultative sexual *M. hapla*, recombination occurs preferentially as four-strand exchanges at similar locations between both pairs of non-sister chromatids in hybrid F1 female (Liu et al., 2007). The F2 inbred offspring segregation at 1:1 ratio suggesting sister chromosomes are rejoined at telophase II (Liu et al., 2007). Those mechanisms in facultative sexual parthenogenesis in *M. hapla* would be expected to result in rapid loss of heterozygosity (Liu et al., 2007). Besides, double-strand breaks (DSBs) repairing systems which repair damage DNA in the form of DSBs through homologous recombination (HR) was demonstrated to accumulate loss of heterozygosity. Following that, HR utilizes a sister chromatid, a homologous chromosome, or a homologous region of another chromosome as a template to heal the DNA break resulting in accumulation of homologous regions. Gene conversion, one form of HR repair, which results in the exchange of genetic material at flanking regions of sister chromatid can yield loss of heterozygosity (Birky 1996; Flot et al. 2013). Again, loss of heterozygosity emphasized the presence of meiotic recombination in *M. graminicola*. DSBs repairing (including gene conversion) were not investigated in this study, but might also occur in *M. graminicola* genome and should be worth to further intensive study. Bdelloid rotifers seem to survive extreme conditions (e.g. freezing, deep vacuum, UV and high doses of ionizing radiation) despite its asexual reproduction (evolutionary scandal) thanks to effective antioxidants and DNA repair mechanisms (Flot et al., 2013; Hespeels et al., 2020; Latta et al., 2019).

Meiotic recombination and LoH could provide interesting genomic plasticity for adaptive evolution in *M. graminicola*. Meiotic recombination might create chance to obtain new gene combination that favorable for parasitism. Meanwhile, expression of parasitic genes especially effector-coding genes might affected by LoH. The effectors secreted by the nematode are

known to promote parasitism by acting, for example, on the suppression of the defense mechanisms of the plant or the metabolism of the host cell (Haegeman et al., 2012; Hewezi & Baum, 2013; Mitchum et al., 2013). The alleles affected by LoH could be either harmful for the acquisition of virulence (suppression of the allele that compromises the compatible plant-pathogen interaction) or beneficial (dosage effect: fixation of a determining allele that could potentially have a higher level of expression). In *M. graminicola*, we observed convergent loss of heterozygous profiles of five scaffolds (~1.18 % genome length) fixed in nine isolates from very remote areas. We found 12% of protein-coding genes located in these regions encoding putative effectors according to Petitot et al. (2015) which is indeed significantly higher than the average percentage of putative effectors detected in *M. graminicola* (~2.8%) (Petitot et al., 2015). This suggests these loci could be the hotspot of genome recombination/reorganizations that regulating nematode parasitism. In addition, the impressive number of copies of galectin genes distributed in these hotspots is remarkable. Galectin appears to play an important role in host-parasite interaction (Young & Meeusen, 2002). Notably, the galectins produced by the parasite function as recognition and action factors that modulate the host's immune responses (Shi et al., 2018). As a result, some parasites can 'subvert' the recognition of the vector/host galectins to allow attachment or invasion of the parasite (Vasta et al., 2017). Moreover, in these hot spots we find genes that may play a role in the transduction of the signal in the plant (e.g. protein-tyrosine phosphatase-like, Romo1/Mgr2, and a Renin receptor-like) or the regulation of gene expression (e.g. TFIIH subunit TTDA/Tfb5, TSC-22 / Dip / Bun, and Zinc finger, GATA-type encoding transcription factors). Focusing on the study of these genes could help reveal the mechanism of adaptation of this pathogen to its hosts.

4.2 Recent spread of *Meloidogyne graminicola* and potential impact of clonal reproduction

In organisms with a short generation time and relatively small genomes, recombination defects could lead to the isolation of a large set of mutants (Alberts et al., 2002). Surprisingly, single nucleotide mutations are rare in *M. graminicola*. The search for such mutations is challenging on the whole genome, because of its heterozygous status and the fact that only the haploid reference is available. The 40 selected SNVs do not represent the total diversity of isolates, but only a subset of SNVs that were not linked to the genome rearrangement events among isolates in this study. While no sequence polymorphism was previously reported in nuclear regions covering 40.6 kb (Besnard et al., 2019), we detected only two SNVs in 354 kb, allowing us to estimate that the mutation rate of the *M. graminicola* genome is 2/354kb (0.5×10^{-4} mutation per base). This is even lower than mitotic *M. incognita* with 2×10^{-4} mutation per base (~0.02% nucleotide) (Koutsovoulos et al., 2019). Even though mitochondrial diversity is low, the

mutation rate is one hundred times higher than the one of the nuclear genome in *M. graminicola* (Besnard et al., 2019). This order of magnitude is consistent with that is observed in many species of the phylum Nematoda and especially in *Meloidogyne* species (Denver et al., 2004; Koutsovoulos et al., 2019; Neiman & Taylor, 2009). For example, in *C. elegans*, mutation rate estimates are 2.1×10^{-8} and 1.6×10^{-7} mutation per site per generation for the nuclear and mitochondrial genomes, respectively (Denver et al., 2004). The low level of polymorphisms between isolates might indicate SNVs might not be driving force of genome evolution in *M. graminicola*

PCAs based on SNVs revealed a lack of geographical structuring of isolates as already reported for mitochondrial genome (Besnard et al., 2019). Besides, other molecular diversity among isolates such as LoH and CNVs also showed no clear association with isolates' geographic origin. Furthermore, the small genetic distance between South East Asia populations suggest those populations shared a common ancestor. These observations reinforces the hypothesis of a recent expansion of *M. graminicola* isolates over Southeast Asia and probably globally. It is likely that these isolates were recently spread from a common and quickly accumulate molecular diversity (SNV, LoH, and CNVs). During adaptation process, some isolates (Mg-C25, Mg-C21, Mg-P, and Mg-VN27) have accumulated more mutations than others. Interestingly, haplotype networks using mutations in mitochondrial genome (11 SNPs, 5 InDels, 1 inversions) and nuclear genome (40 SNPs) among 13 isolates showed strongly correlation suggest common direction of molecular evolution of both mitogenome and nuclear genome toward changes in environments. Cytogenetic analysis also suggested that *M. graminicola* mainly reproduces through meiotic parthenogenesis (Triantaphyllou, 1969). Altogether, we could suggest *M. graminicola* isolates are currently spread from a common anscestor and potentially clonal evolution during from recent spread events over South East Asia of *M. graminicola*.

The *M. graminicola* genome has an average heterozygosity of 1.85% (at *k*-mer=21 bp) (Phan et al., 2020; Besnard et al., 2019), which is in the range observed in other meiotic and sexually reproducing animals (Leffler et al., 2012). This heterozygosity level is higher than in species with an intraspecific origin of parthenogenesis (0.03–0.83%) while belong to the range of heterozygosity level of the species with a known hybrid origin of parthenogenesis (1.73–8.5%) (Jaron et al. 2020). Two hypothesis might explain the observed heterozygosity level of *M. graminicola*. First, this species might have originated from a past homoploid hybridization with higher heterozygosity level (greater than 1.85%). Then, the heterozygosity level in *M. graminicola* genome was reduced owing to eventually genetic erosion (such as recombination,

gene conversion, translocations). Phylogenetic analysis on homologs of divergent low-copy nuclear regions also suggested *M. graminicola* might undergo potential hybrid origin with one common ancestor of *M. oryzae* (Besnard et al. 2019). Secondly, the maintenance of heterozygosity in the *M. graminicola* genome might be due to the occurrence of sexual reproduction between close relatives. However, the sexual reproduction mode of *M. graminicola* was proposed just based on a few evidence of sperm pronuclei inside the eggs (~ 0.5% among few hundred eggs) (Triantaphyllou, 1969) and should be further confirm at molecular level.

4.3 Genome structure variation as a potential adaptation mechanism

CNVs have been identified in all areas of life, from bacteria, to plants and animals and appear to play important functions in the evolution of genomes (Schughart et al., 1989). CNVs consists of DNA segments that are variable in copy number and dispersed throughout the genome (Alkan et al.2011). CNVs can either be inherited from the previous generation or appear de novo through duplication/deletion events, and their fixation by drift or selection may contribute to the creation of genetic novelty resulting in species adaptation to stressful or novel environments (Arlt et al., 2014; Bussotti et al., 2018).

Recent studies also detected CNVs in Cyst and RKN plant parasitic nematodes (Castagnone-Sereno et al., 2019; Cook et al., 2012; Patil et al., 2019). Some of these CNVs are associated with the ability of the parasite to bypass defense mechanisms and become virulent (Castagnone-Sereno et al., 2019; Cook et al., 2012; Patil et al., 2019). Interestingly, convergent gene loss events have been associated with virulent isolates of *M. incognita* (Castagnone-Sereno et al., 2019). The genetic variations generated by these CNVs could reflect an adaptive response to different environments and in particular to the host plant, in which this obligate parasite must establish a compatible interaction to be able to reproduce. In this study, we detected numerous CNVs in all *M. graminicola* isolates. These CNVs appeared independently among isolates without geographical relationship. It is likely that each isolate itself accumulates gene gain or loss to adapt to the environment and the host.

The identification of *M. graminicola* populations which have become virulent (i.e. able to overcome a resistance gene in a plant) would be particularly useful by making it possible to constitute a genetic material of choice to verify this hypothesis. Unfortunately this kind of genetic material is not currently available. However, it is clear that the isolates (Mg-C25, Mg-P, Mg-VN27, and Mg-L2), which were distinguished on the basis of their SNVs, exhibited substantial gene losses events. These isolates, excepted Mg-Bali which exhibited abundant gene

loss events, tend to have a profile with more homologous nuclear regions. The question of whether these CNVs participate in the evolution of the *M. graminicola* genome, notably via recombination events the signatures of which appear in large numbers, remains to be demonstrated.

V. **Conclusion**

This study reveal key points in the evolutionary history of *M. graminicola.* Meiotic recombination, loss of heterozygosity, and substantial amounts of CNV constitute genome plasticity factors that are probably actively involved in the adaptive evolution of this parasite. We also confirm a recent radiation and spread of *M. graminicola* in South-East Asia. Although additional experiments and analyzes will be needed to refine these, our study provides valuable genomic resources and new insights for further in-depth studies on how to understand the evolutionary history of this parasite.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). General recombination. *Molecular Biology of the Cell.* 4th Edition. https://www.ncbi.nlm.nih.gov/books/NBK26898/

Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, 12, 363–376. https ://doi.org/10.1038/nrg2958

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Archetti, M. (2004). Recombination and loss of complementation: A more than two-fold cost for parthenogenesis. *Journal of Evolutionary Biology*, *17*(5), 1084–1097. https://doi.org/10.1111/j.1420-9101.2004.00745.x

Arlt, M. F., Rajendran, S., Birkeland, S. R., Wilson, T. E., & Glover, T. W. (2014). Copy number variants are produced in response to low-dose ionizing radiation in cultured cells. *Environmental and Molecular Mutagenesis*, *55*(2), 103–113. https://doi.org/10.1002/em.21840

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bellafiore, S., Jougla, C., Chapuis, É.,Besnard, G., Suong, M., Vu, P. N., De Waele, D., Gantet, P., & Thi, X. N. (2015). Intraspecific variability of the facultative meiotic parthenogenetic root-knot nematode (*Meloidogyne graminicola*) from rice fields in Vietnam. *Comptes Rendus Biologies*, *338*(7), 471–483. https://doi.org/10.1016/j.crvi.2015.04.002

Besnard, G., Jühling, F., Chapuis, E., Zedane, L., Lhuillier, E., Mateille, T., & Bellafiore, S. (2014). Fast assembly of the mitochondrial genome of a plant parasitic nematode (*Meloidogyne graminicola*) using next generation sequencing. *Comptes Rendus – Biologies*, 337, 295-301.

Besnard, G., Thi-Phan, N., Ho-Bich, H., Dereeper, A., Trang Nguyen, H., Quénéhervé, P., Aribi, J., & Bellafiore, S. (2019). On the close relatedness of two rice-parasitic root-knot nematode species and the recent expansion of *Meloidogyne graminicola* in Southeast Asia. *Genes*, *10*(2). https://doi.org/10.3390/genes10020175

Birky C.W. Jr. (1996). Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics*. *144*(1):427–437.

Burt, A. (2000). Perspective: Sex, recombination, and the efficacy of selection--was Weismann right? *Evolution*, *54*(2), 337–351. https://doi.org/10.1111/j.0014-3820.2000.tb00038.x

Bussotti, G., Gouzelou, E., Boité, M. C., Kherachi, I., Harrat, Z., Eddaikra, N., Mottram, J. C., Antoniou, M., Christodoulou, V., Bali, A., Guerfali, F. Z., Laouini, D., Mukhtar, M., Dumetz, F., Dujardin, J.-C., Smirlis, D., Lechat, P., Pescher, P., Hamouchi, A. E., … Späth, G. F. (2018). Leishmania genome dynamics during environmental adaptation reveal strain-specific differences in gene copy number variation, karyotype instability, and telomeric amplification. *MBio*, *9*(6). https://doi.org/10.1128/mBio.01399-18

Bridge, J., Luc, M., & Plowright, R. A. (2009). Nematode parasites of rice. In Plant-parasitic nematodes in subtropical and tropical agriculture (pp. 69–108). CAB International.

Castagnone-Sereno, P., Mulet, K., Danchin, E. G. J., Koutsovoulos, G. D., Karaulic, M., Rocha, M. D., Bailly-Bechet, M., Pratx, L., Perfus-Barbeoch, L., & Abad, P. (2019). Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne i*ncognita. Molecular Ecology, 28(10), 2559–2572. https://doi.org/10.1111/mec.15095

Carneiro, R. M. D. G., Almeida, M. R., & Quénéhervé, P. (2000). Enzyme phenotypes of *Meloidogyne* spp. Populations. *Nematology*, 2, 645–654. https://www.cabi.org/isc/abstract/20013029216

Chapuis, E., Besnard, G., Andrianasetra, S., Rakotomalala, M., Nguyen, H. T., & Bellafiore, S. (2016). First report of the root-knot nematode *Meloidogyne graminicola* in Madagascar rice fields. *Australasian Plant Disease Notes*, *11*(1), 32. https://doi.org/10.1007/s13314-016-0222-5

Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., Wang, J., Hughes, T. J., Willis, D. K., Clemente, T. E., Diers, B. W., Jiang, J., Hudson, M. E., & Bent, A. F. (2012). Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science*, *338*(6111), 1206–1209. https://doi.org/10.1126/science.1228746

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Group, 1000 Genomes Project Analysis. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Waele, D., & Elsen, A. (2007). Challenges in tropical plant nematology. *Annual Review of Phytopathology*, *45*(1), 457–485. https://doi.org/10.1146/annurev.phyto.45.062806.094438

Denver, D. R., Morris, K., Lynch, M., & Thomas, W. K. (2004). High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature*, *430*(7000), 679–682. https://doi.org/10.1038/nature02697

Engelstädter J. 2017. Asexual but not clonal: evolutionary processes in automictic populations. *Genetics*. *206*:993–1009.

EPPO Global Database. (2020). https://gd.eppo.int/

FAOSTAT. (2019). http://www.fao.org/faostat/en/#data/QC

Fanelli, E., Cotroneo, A., Carisio, L., Troccoli, A., Grosso, S., Boero, C., Capriglia, F., & De Luca, F. (2017). Detection and molecular characterization of the rice root-knot nematode *Meloidogyne graminicola* in Italy. *European Journal of Plant Pathology*, *149*(2), 467–476. https://doi.org/10.1007/s10658-017-1196-7

Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E. G. J., Hejnol, A., Henrissat, B., Koszul, R., Aury, J.-M., Barbe, V., Barthélémy, R.-M., Bast, J., Bazykin, G. A., Chabrol, O., Couloux, A., Da Rocha, M., Da Silva, C., Gladyshev, E., … Van Doninck, K. (2013). Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adinetavaga*. *Nature*, *500*(7463), 453–457. https://doi.org/10.1038/nature12326

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv:1207.3907 [q-Bio]*. http://arxiv.org/abs/1207.3907

Gibson, A. K., Delph, L. F., & Lively, C. M. (2017). The two-fold cost of sex: Experimental evidence from a natural system. *Evolution Letters*, *1*(1), 6–15. https://doi.org/10.1002/evl3.1

Glémin, S., François, C. M., & Galtier, N. (2019). Genome evolution in outcrossing vs. selfing vs. asexual species. In M. Anisimova (Ed.), *Evolutionary Genomics: Statistical and Computational Methods* (pp. 331–369). Springer. https://doi.org/10.1007/978-1-4939-9074-0_11

Golden, A. M., & Birchfield, W. (1965). *Meloidogyne graminicola* (Heteroderidae) a new species of root-knot nematode from grass. *Proceedings of the Helminthological Society of Washington*, *32*(2), 228–231. https://www.cabdirect.org/cabdirect/abstract/19660801518

Golden, A. M., & Birchfield, W. (1968). Rice root-knot nematode (*Meloidogyne graminicola*) as a new pest of rice. *Plant Disease Reporter*, *52*(6), 243. https://www.cabdirect.org/cabdirect/abstract/19690806120

Haegeman, A., Mantelin, S., Jones, J. T., & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene*, *492*(1), 19–31. https://doi.org/10.1016/j.gene.2011.10.040

Hespeels, B., Penninckx, S., Cornet, V., Bruneau, L., Bopp, C., Baumlé, V., Redivo, B., Heuskin, A.-C., Moeller, R., Fujimori, A., Lucas, S., & Van Doninck, K. (2020). Iron ladies – How desiccated asexual rotifer *Adinetavaga* deal with X-rays and heavy ions? *Frontiers in Microbiology*, *11*, 1792. https://doi.org/10.3389/fmicb.2020.01792

Hewezi, T., & Baum, T. J. (2013). Manipulation of plant cells by cyst and root-knot nematode effectors. *Molecular Plant-Microbe Interactions*, *26*(1), 9–16. https://doi.org/10.1094/MPMI-05-12-0106-FI

Jaron, K.S., Bast, J., Nowell, R.W., Ranallo-Benavidez, T.R., Robinson-Rechavi, M., Schwander, T., 2020. Genomic Features of Parthenogenetic Animals. *J. Hered*. doi.org/10.1093/jhered/esaa031

Jiang, H., Lei, R., Ding, S.-W., & Zhu, S. (2014). Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, *15*(1), 182. https://doi.org/10.1186/1471-2105-15-182

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649.

Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.-A., Mitterecker, A., Bodenhofer, U., & Hochreiter, S. (2012). cn.MOPS: Mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Research*, *40*(9), e69. https://doi.org/10.1093/nar/gks003

Kondrashov, A. S. (1982). Selection against harmful mutations in large sexual and asexual populations. *Genetical Research*, *40*(3), 325–332. https://doi.org/10.1017/s0016672300019194

Koutsovoulos, G. D., Marques, E., Arguel, M.-J., Duret, L., Machado, A. C. Z., Carneiro, R. M. D. G., Kozlowski, D. K., Bailly-Bechet, M., Castagnone-Sereno, P., Albuquerque, E. V. S., & Danchin, E. G. J. (2019). Population genomics supports clonal reproduction and multiple gains and losses of parasitic abilities in the most devastating nematode plant pest. *Evolutionary Applications*, *13*(2), 442–457. https://doi.org/10.1111/eva.12881

Kumar, S., Jones M., Koutsovoulos G., Clarke M., & Blaxter M. (2013). Blobology: Exploring Raw Genome Data for Contaminants, Symbionts and Parasites Using Taxon-Annotated GC-Coverage Plots. *Frontiers in Genetics*. *4*(237), 1–12. https://doi.org/10.3389/fgene.2013.00237.

Laetsch, *D.* R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287. https://doi.org/10.12688/f1000research.12232.1

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Latta, L. C., Tucker, K. N., & Haney, R. A. (2019). The relationship between oxidative stress, reproduction, and survival in a bdelloid rotifer. *BMC Ecology*, *19*(1), 7. https://doi.org/10.1186/s12898-019-0223-2

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer W. K., Ségurel L., Venkat A., Andolfatto P., & Przeworski M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? *PLOS Biology*. *10*(9), e1001388. https://doi.org/10.1371/journal.pbio.1001388.

Lewin, B. (1990). *Genes* IV. Oxford University Press.

Leigh, J. W., & Bryant, D. (2015). popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Li, H. (2020). *lh3/seqtk: Toolkit for processing sequences in FASTA/Q formats* [C]. https://github.com/lh3/seqtk (Original work published 2012)

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, Q.L., Thomas, V.P., Williamson, V.M., 2007. Meiotic Parthenogenesis in a Root-Knot Nematode Results in Rapid Genomic Homozygosity. Genetics 176, 1483–1490. https://doi.org/10.1534/genetics.107.071134

Lynch, M., Bürger, R., Butcher, D., & Gabriel, W. (1993). The mutational meltdown in asexual populations. *The Journal of Heredity*, *84*(5), 339–344. https://doi.org/10.1093/oxfordjournals.jhered.a111354

Mantelin, S., Bellafiore, S., & Kyndt, T. (2017). *Meloidogyne graminicola*: A major threat to rice agriculture. *Molecular Plant Pathology*, *18*(1), 3–15. https://doi.org/10.1111/mpp.12394

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Martin, M., Patterson, M., Garg, S., Fischer, S. O., Pisanti, N., Klau, G. W., Schöenhuth, A., & Marschall, T. (2016). WhatsHap: Fast and accurate read-based phasing. *BioRxiv*, 085050. https://doi.org/10.1101/085050

Maynard Smith, J. (1986). Evolution: Contemplating life without sex. *Nature*, *324*(6095), 300–301. https://doi.org/10.1038/324300a0

McDonald, B. A., & Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, *40*(1), 349–379. https://doi.org/10.1146/annurev.phyto.40.120501.101443

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., … Finn, R. D. (2019). InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, *47*(D1), D351–D360. https://doi.org/10.1093/nar/gky1100

Mitchum, M. G., Hussey, R. S., Baum, T. J., Wang, X., Elling, A. A., Wubben, M., & Davis, E. L. (2013). Nematode effector proteins: An emerging paradigm of parasitism. *The New Phytologist*, *199*(4), 879–894. https://doi.org/10.1111/nph.12323

Neiman, M., & Taylor, D. R. (2009). The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1660), 1201–1209. https://doi.org/10.1098/rspb.2008.1758

Padgham, J. L., Duxbury, J. M., Mazid, A. M., Abawi, G. S., & Hossain, M. (2004). Yield loss caused by *Meloidogyne graminicola* on lowland rainfed rice in Bangladesh. *Journal of Nematology*, *36*(1), 42–48.

Patil, G. B., Lakhssassi, N., Wan, J., Song, L., Zhou, Z., Klepadlo, M., Vuong, T. D., Stec, A. O., Kahil, S. S., Colantonio, V., Valliyodan, B., Rice, J. H., Piya, S., Hewezi, T., Stupar, R. M., Meksem, K., & Nguyen, H. T. (2019). Whole-genome re-sequencing reveals the impact of the interaction of copy number variants of the *rhg1* and *Rhg4* genes on broad-based resistance to soybean cyst nematode. *Plant Biotechnology Journal*, *17*(8), 1595–1611. https://doi.org/10.1111/pbi.13086

Petitot, A.-S., Dereeper, A., Agbessi, M., Da Silva, C., Guy, J., Ardisson, M., & Fernandez, D. (2015). Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants. *Molecular Plant Pathology*, *17*(6), 860–874. https://doi.org/10.1111/mpp.12334

Phan, N. T., De Waele, D., Lorieux, M., Xiong, L., & Bellafiore, S. (2018). A hypersensitivity-like response to *Meloidogyne graminicola* in rice (*Oryza sativa*). *Phytopathology*, *108*(4), 521–528. https://doi.org/10.1094/PHYTO-07-17-0235-R

Phan, N. T., Orjuela, J., Danchin, E. G. J., Klopp, C., Perfus-Barbeoch, L., Kozlowski, D. K., Koutsovoulos, G. D., Lopez-Roques, C., Bouchez, O., Zahm, M., Besnard, G., & Bellafiore, S. (2020). Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*). *Ecology and Evolution*, *10*(20), 11006–11021. https://doi.org/10.1002/ece3.6680

Pokharel, R. R., Abawi, G. S., Duxbury, J. M., Smat, C. D., Wang, X., & Brito, J. A. (2010). Variability and the recognition of two races in *Meloidogyne graminicola*. *Australasian Plant Pathology*, *39*(4), 326–333. https://doi.org/10.1071/AP09100

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Schön I., Martens K., van Dijk P. 2009. Lost sex: the evolutionary biology of parthenogenesis. Berlin (Germany): *Springer*.

Schughart, K., Kappen, C., & Ruddle, F. H. (1989). Duplication of large genomic regions during the evolution of vertebrate homeobox genes. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(18), 7067–7071. https://doi.org/10.1073/pnas.86.18.7067

Shi, W., Xue, C., Su, X., & Lu, F. (2018). The roles of galectins in parasitic infections. *Acta Tropica*, *177*, 97–104. https://doi.org/10.1016/j.actatropica.2017.09.027

Slatkin, M., 2008. Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. *Nat Rev Genet 9*, 477–485. https://doi.org/10.1038/nrg2361

Triantaphyllou, A. C. (1969). Gametogenesis and the chromosomes of two root-knot nematodes, *Meloidogyne graminicola* and *M. naasi*. *Journal of Nematology*, *1*(1), 62–71. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2617796/

Vakhrusheva, O. A., Mnatsakanova, E. A., Galimov, Y. R., Neretina, T. V., Gerasimov, E. S., Ozerova, S. G., Zalevsky, A. O., Yushenova, I. A., Arkhipova, I. R., Penin, A. A., Logacheva, M. D., Bazykin, G. A., & Kondrashov, A. S. (2018). Recombination in a natural population of the bdelloid rotifer *Adinetavaga*. *BioRxiv*, 489393. https://doi.org/10.1101/489393

Vasta, G. R., Feng, C., González-Montalbán, N., Mancini, J., Yang, L., Abernathy, K., Frost, G., & Palm, C. (2017). Functions of galectins as 'self/non-self'-recognition and effector factors. *Pathogens and Disease*, *75*(5). https://doi.org/10.1093/femspd/ftx046

Weir B. S., Cockerham C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution, 38*(6): 1358–1370.

Young, A. R., & Meeusen, E. N. (2002). Galectins in parasite infection and allergic inflammation. *Glycoconjugate Journal*, *19*(7), 601–606. https://doi.org/10.1023/B:GLYC.0000014091.00844.0a

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Supporting information

Genomic rearrangements are putative genomic evolutionary traits in the facultative parthenogenic nematode *Meloidogyne graminicola*

Ngan Thi Phan[1], Etienne G.J. Danchin[2], Georgios D. Koutsovoulos[2], Marie-Liesse Vermeire[1], Guillaume Besnard[3,]*, Stéphane Bellafiore[1,]*

List of figures:

List of tables:

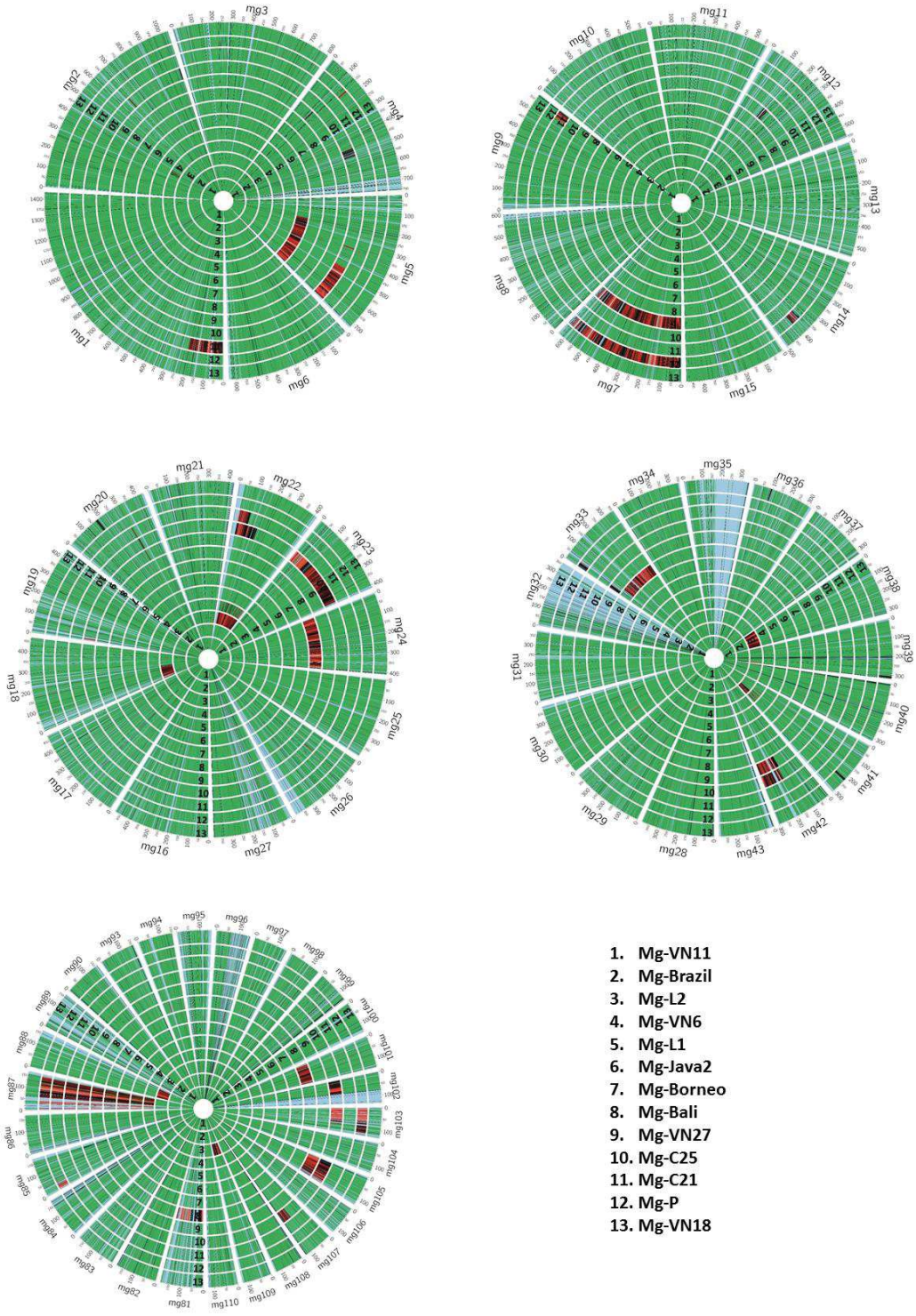


Figure S1. **Visualization of single nucleotide variants (SNV) called among 13 isolates on all 283 scaffolds of reference genome**. The length of scaffolds was drawn in scale. Each circle represented genome sequence of each isolate. Each vertical line indicates the genotype and position of SNV on each scaffold with color code: green = heterozygous state (genotype "0/1"); black = homozygous state with reference haplotype (genotype "0/0"); red = homozygous state with the alternative haplotype (genotype "1/1"). Therefore, the “green” regions indicate “heterozygous state” while the “red and black” regions indicated “homozygous state (LoH)”. The blue regions indicate the other regions (conserved homozygous regions and insertions/deletions between isolates).

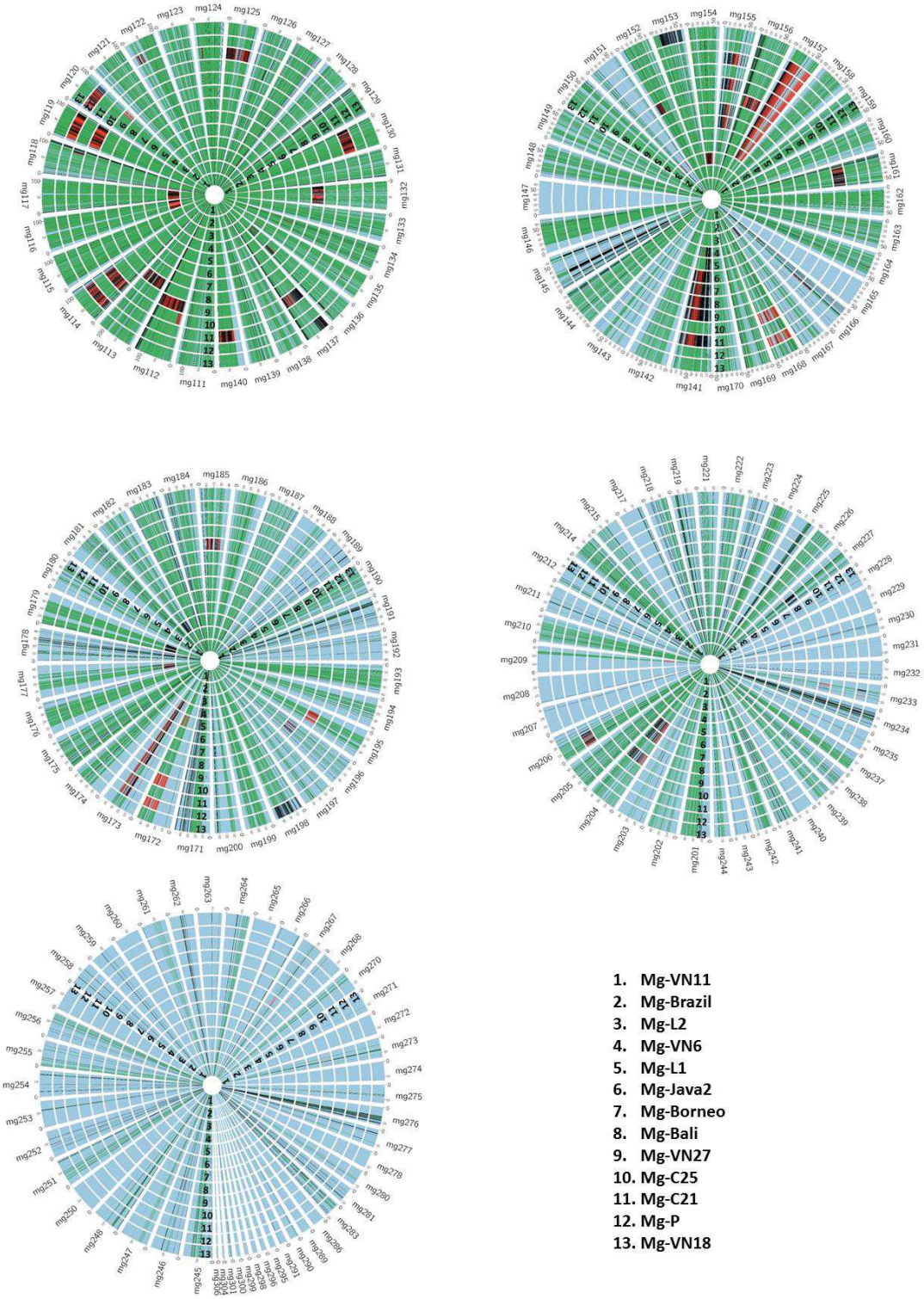


Figure S1. (Continue)

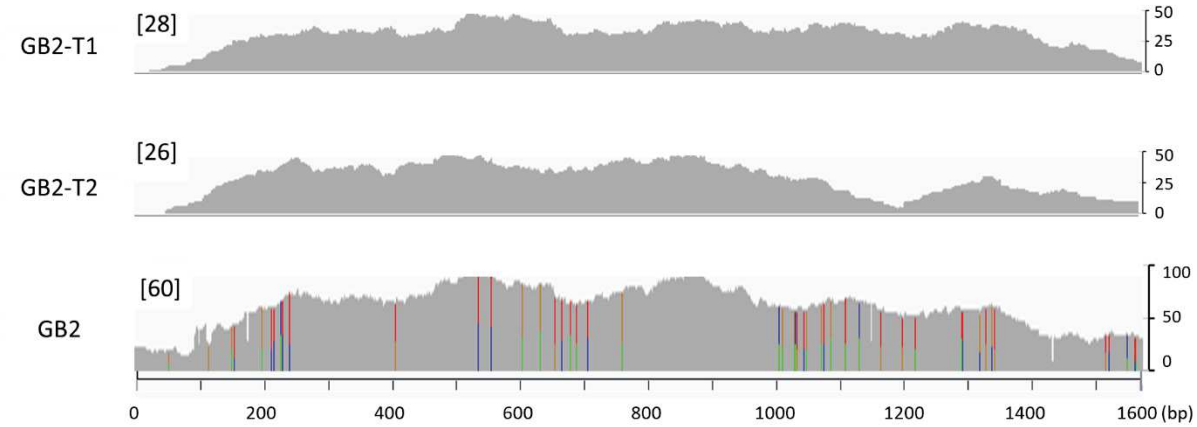


Figure S2. Per base coverage analysis of Hiseq data of *M. graminicola* isolates Mg-VN18 on consensus sequence of GB2 contig and two divergent copies (GB2_T1 and GB2_T2). The phased colors in the first track indicates the mapping of different reads belong to two divergent copies at the same position of consensus sequence. Meanwhile, reads were separated well when mapped onto two divergent copies. The number on the left side of each track indicates average coverage of reads on each fragment.

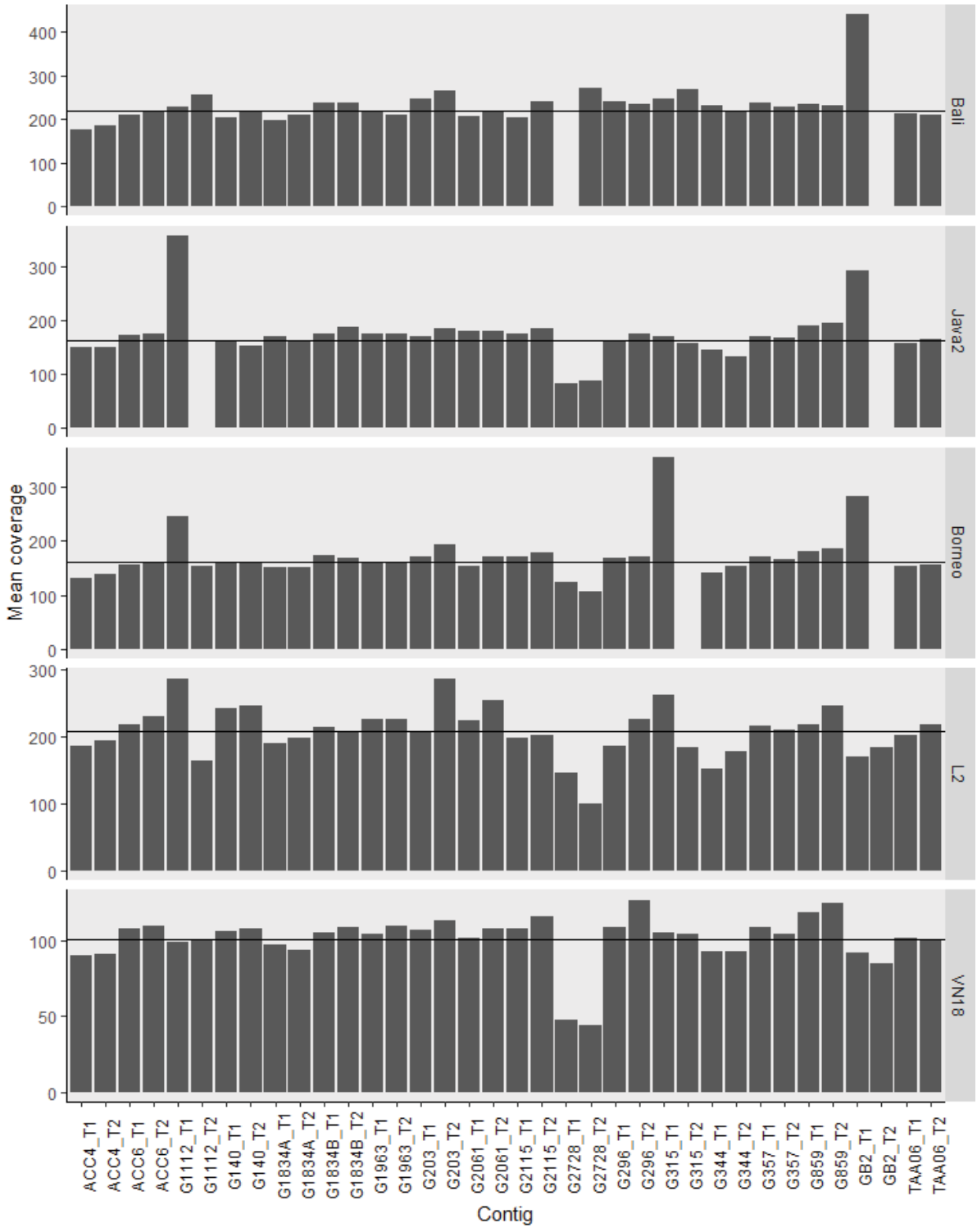


Figure S3. Average read-coverage on 36 divergent copies of 18 contigs in five isolates. The green horizontal line indicated mean value of read-coverage of all divergent copies in each isolate. The missing columns (average read-coverage = 0 ×) indicated the absence of the copy.

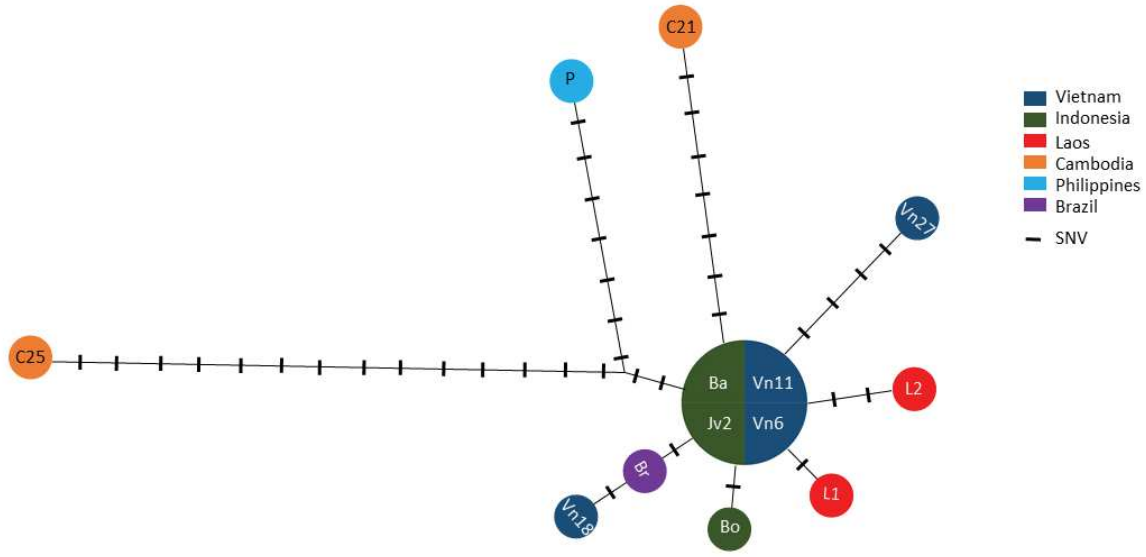


Figure S4. Haplotype networks among *M. graminicola* isolates using 40 SNVs. Code names of isolates are indicated in the circles and countries of origin are displayed by different colors. The number of SNV is shown on the branches with slashes.

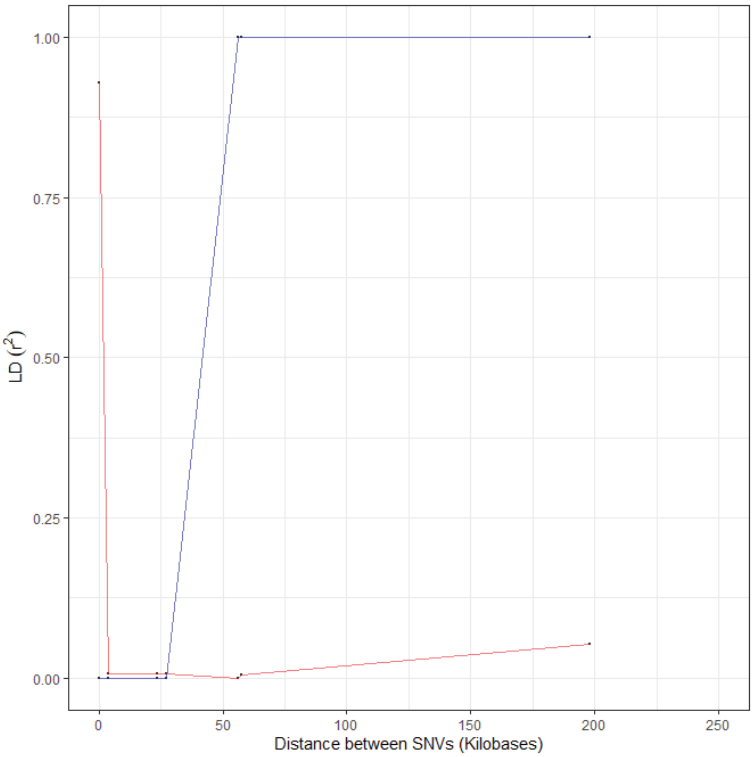


Figure S5 Linkage Disequilibrium and 4-gametes test of *M. graminicola* isolates using 40 fixed markers. The red line indicated r2 correlation between markers, indicating linkage disequilibrium (LD) for each physical distance between the SNV markers. The blue line indicated the proportion of pairs of two-state markers that pass the 4-gamete test for each physical distance between the markers on *M. graminicola* scaffold.

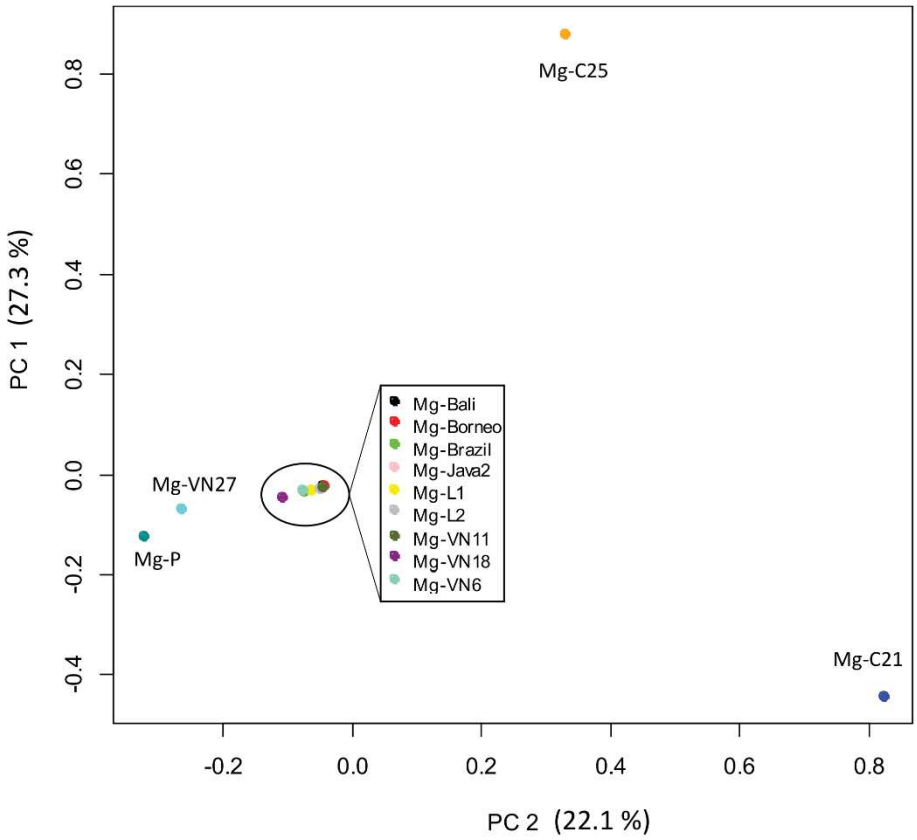


Figure S6. Principal component analysis of the 13 *M. graminicola* isolates using 41 fixed markers

Table S1. ***Meloidogyne graminicola*** **isolates used in this study and generated genomic data** (including number of Illumina reads, read length, GC content of the assembly, and mean coverage)

No	Population	Country	# raw PE reads	# cleaned PE reads	Read length	%GC	Coverage[2]
1	Mg-VN18[1]	Vietnam	122,890,657	86,991,142	150	24	288
2	Mg-Borneo	Indonesia	147,469,360	99,625,819	150	24	360
3	Mg-Bali	Indonesia	180,145,846	122,418,010	150	24	442
4	Mg-Java2	Indonesia	118,708,848	118,590,640	150	24	420
5	Mg-L2	Laos	157,229,694	156,871,340	150	24	568
6	Mg-L1	Laos	13,807,048	15,746,640	125	26	38
7	Mg-P	Philippines	15,524,276	9,867,172	100	24	24
8	Mg-Brazil	Brazil	16,880,430	16,339,898	100	26	39
9	Mg-C21	Cambodia	17,457,706	8,637,052	125	24	26
10	Mg-C25	Cambodia	26,347,326	7,045,100	125	25	21
11	Mg-VN6	Vietnam	16,374,636	15,182,532	100	24	37
12	Mg-VN11	Vietnam	17,227,778	15,746,640	100	25	38
13	Mg-VN27	Vietnam	16,298,908	15,137,918	100	24	36

[1] The genome of isolate Mg-VN18 was used to generate the haploid reference genome sequence in Phan et al. (2020). [2] Average read coverage considering assembly genome size of 41.5 Mb. Abbreviations: # = number; PE = paired-end reads; %GC = GC content.

Table S2. **List of two divergent copies of 18 contigs**

No	Contig name	Type	Divegent copy name	Length (bp)	Position on reference genome[1]	# Genes	Identity between 2 types
1	ACC4	Type 1 Type 2	ACC4_T1 ACC4_T2	11,247 11,214	mg4:505760-516936	2	94.21
2	G140	Type 1 Type 2	G140_T1 G140_T2	1,429 1,433	mg6:134432-135865	1	96.66
3	G296	Type 1 Type 2	G296_T1 G296_T2	12,282 12,323	mg7:6227-18522	2	97.67
4	ACC6[2]	Type 1 Type 2	ACC6_T1 ACC6_T2	6,310 6,151	mg15:201656-207657	1	94.73
5	G1963	Type 1 Type 2	G1963_T1 G1963_T2	8,687 8,690	mg25:79960-88655	5	96.56
6	G1834A	Type 1 Type 2	G1834A_T1 G1834A_T2	2,240 2,238	mg44:37041-39278	1	99.11
7	G1834B	Type 1 Type 2	G1834B_T1 G1834B_T2	3,110 3,111	mg44:39200-42308	1	99.71
8	G315	Type 1 Type 2	G315_T1 G315_T2	6,258 6,282	mg53:17740-24025	1	97.73
9	TAA06[2]	Type 1 Type 2	TAA06_T1 TAA06_T2	6,361 6,399	mg68:90615-96976	3	96.05
10	GB2	Type 1 Type 2	GB2_T1 GB2_T2	1,589 1,572	mg78:21353-22923	1	94.99
11	G2115	Type 1 Type 2	G2115_T1 G2115_T2	7,868 7,865	mg93:111858-119733	1	98.19
12	G2061	Type 1 Type 2	G2061_T1 G2061_T2	2,897 2,906	mg125:29921-32815	1	95.22
13	G203	Type 1 Type 2	G203_T1 G203_T2	17,866 17,172	mg125:68032-85564	3	98.50
14	G859	Type 1 Type 2	G859_T1 G859_T2	80,275 79,436	mg130:4438-84558	26	97.42
15	G1112	Type 1 Type 2	G1112_T1 G1112_T2	2,810 2,800	mg136:62364-65160	1	97.48
16	G2728	Type 1 Type 2	G2728_T1 G2728_T2	532 526	mg156:12687-13219	2	98.08
17	G344	Type 1 Type 2	G344_T1 G344_T2	2,065 2,061	mg161:11242-13302	1	97.83
18	G357	Type 1 Type 2	G357_T1 G357_T2	3,904 3,888	mg199:8629-12519	1	96.64

[1]scaffold name:start position-end position

[2]two contigs from Besnard et al. 2019

Table S3. ***InterProscan*** **and** ***pfam*** **domain of 106 genes located in five scaffolds** (mg78, mg87, mg145, mg157, and mg173) **bearing common LoH events in nine isolates**

InterProscan /pfam	domain/family	# genes	Putative secretomic protein in Petitot et al. 2015
IPR001079	Galectin, carbohydrate recognition domain	8	
IPR010678	Digestive organ expansion factor, predicted	2	
IPR001810	F-box domain	2	v
IPR001478	PDZ domain	2	v
IPR007482	Protein-tyrosine phosphatase-like, PTPLA	2	v
IPR001680	WD40 repeat	2	v
IPR003439	ABC transporter-like	1	v
IPR000582	Acyl-CoA-binding protein, ACBP	1	
IPR002300	Aminoacyl-tRNA synthetase, class Ia	1	
IPR007803	Aspartyl/asparaginy/proline hydroxylase	1	
IPR005811	ATP-citrate lyase/succinyl-CoA ligase	1	
IPR011611	Carbohydrate kinase PfkB	1	
IPR018484	Carbohydrate kinase, FGGY, N-terminal	1	
IPR002018	Carboxylesterase, type B	1	v
IPR022043	Chromatin assembly factor 1 subunit A	1	
IPR019187	Cyclin-dependent kinase 2-associated protein	1	
IPR001199	Cytochrome b5-like heme/steroid binding domain	1	
IPR000268	DNA-directed RNA polymerase, subunit N/Rpb10	1	
IPR006149	EB domain	1	v
IPR000742	EGF-like domain	1	v
IPR006076	FAD dependent oxidoreductase	1	v
IPR006838	FAR-17a/AIG1-like protein	1	
IPR014847	FERM adjacent (FA)	1	
IPR004045	Glutathione S-transferase, N-terminal	1	
IPR001296	Glycosyl transferase, family 1	1	
IPR002659	Glycosyl transferase, family 31	1	
IPR006121	Heavy metal-associated domain, HMA	1	
IPR024084	Isopropylmalate dehydrogenase-like domain	1	
IPR008657	Jumping translocation breakpoint	1	
IPR018108	Mitochondrial substrate/solute carrier	1	
IPR019392	Mitoguardin	1	

IPR007248	Mpv17/PMP22	1	
IPR006029	Neurotransmitter-gated ion-channel transmembrane domain	1	v
IPR004156	Organic anion transporter polypeptide	1	
IPR000819	Peptidase M17, leucyl aminopeptidase, C-terminal	1	
IPR005176	Potentiating neddylation domain	1	
IPR001313	Pumilio RNA-binding repeat	1	
IPR024135	Ragulator complex protein LAMTOR5	1	
IPR012493	Renin receptor-like	1	
IPR008932	Ribosomal protein L7/L12, oligomerisation	1	
IPR018450	Romo1/Mgr2	1	
IPR025749	Sphingomyelin synthase-like domain	1	
IPR009400	TFIIH subunit TTDA/Tfb5	1	v
IPR010339	TIP49, P-loop domain	1	
IPR000580	TSC-22 / Dip / Bun	1	
IPR015221	Ubiquitin-related modifier 1	1	
IPR000679	Zinc finger, GATA-type	1	
IPR024766	Zinc finger, RING-H2-type	1	v
IPR001841	Zinc finger, RING-type	1	v
IPR001293	Zinc finger, TRAF-type	1	
PF13087	AAA domain	1	
	Neuropeptide-like protein 36	1	
	tRNA pseudouridine synthase	1	
	Unknown	23	
	Uncharacterized protein similar to other nematode species	18	

Table S4. **Fixation index (FST) based on 44,527 SNVs among South East Asian populations** (upper part mean values, lower part weighted values).

	Indonesia	Cambodia	Vietnam	Laos	Philippines
Indonesia	x	0.004873	0.004250	0.004359	0.023668
Cambodia	0.005736	x	0.007243	0.001194	-0.011860
Vietnam	0.003957	0.008126	x	-0.000069	0.011437
Laos	0.004055	0.002622	0.001021	x	0.011670
Philippines	0.037149	0.005224	0.022120	0.022488	x

Table S5. **Number of regions bearing copy number variants (CNV) and number of genes overlap with CNV regions (at least 70% length) among 13 isolates**. (CN2 indicated normal diploid state; CN0 and CN1 indicated deletion of copy compare to normal diploid state; from CN3 to CN6 indicate duplication of copy number compare to the normal diploid state)

#copy	Number of CNV regions												
	Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-Java2	Mg-L1	Mg-L2	Mg-P	Mg-VN11	Mg-VN18	Mg-VN27	Mg-VN6
Deletion (CN0)	2	1	0	2	12	0	0	3	6	0	0	4	0
Half deletion (CN1)	11	6	0	10	25	5	2	2	4	2	0	1	0
Three copies (CN3)	8	7	6	3	1	2	12	9	0	8	8	2	2
Four copies (CN4)	0	1	0	2	1	0	3	0	1	0	0	0	0
Five copies (CN5)	1	0	0	0	0	0	0	0	0	0	0	0	0
Six copies (CN6)	0	0	0	0	0	0	1	1	1	1	1	0	0
#copy	Number of genes which have length overlap at least 70% with CNV regions												
Deletion (CN0)	29	0	0	5	16	0	0	68	16	0	0	31	0
Half deletion (CN1)	10	17	0	6	118	8	4	1	12	3	0	1	0
Three copies (CN3)	69	68	8	2	1	2	29	12	0	7	7	2	3
Four copies (CN4)	0	1	0	0	0	0	3	0	6	0	0	0	0
Five copies (CN5)	2	0	0	0	0	0	0	0	0	0	0	0	0
Six copies (CN6)	0	0	0	0	0	0	0	2	2	2	2	0	0

Table S6. Detail CNVs regions found on scaffolds among 13 isolates (CN2 indicated normal diploid state; CN0 and CN1 indicated deletion of copy compare to normal diploid state; from CN3 to CN6 indicate duplication of copy number compare to the normal diploid state)

No	Scaffolds	start	end	width	Number of copies												
					Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27
1	mg1	0	6000	6001	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
2	mg1	65000	72000	7001	CN2	CN2	CN2	CN0	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2
3	mg1	81000	86000	5001	CN2	CN2	CN2	CN1	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
4	mg1	145000	150000	5001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
5	mg1	168000	172000	4001	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
6	mg1	244000	247000	3001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
7	mg1	325000	330000	5001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
8	mg1	340000	344000	4001	CN2	CN2	CN2	CN1	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
9	mg1	382000	390000	8001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
10	mg1	406000	410000	4001	CN3	CN2	CN2	CN2	CN2	CN2	CN3	CN3	CN2	CN2	CN2	CN2	CN2
11	mg1	502000	505000	3001	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
12	mg1	534000	539000	5001	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
13	mg1	553000	557000	4001	CN3	CN3	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2
14	mg1	610000	615000	5001	CN3	CN3	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2
15	mg1	677000	680000	3001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
16	mg1	719000	722000	3001	CN3	CN3	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2
17	mg1	860000	863000	3001	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
18	mg1	1038000	1043000	5001	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN1
19	mg1	1261000	1264000	3001	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
20	mg1	1283000	1288000	5001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3
21	mg1	1338000	1342000	4001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3
22	mg1	1349000	1352000	3001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2
23	mg1	1358000	1362000	4001	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2
24	mg2	363000	366000	3001	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
25	mg2	377000	395552	18553	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2
26	mg2	380000	383000	3001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2
27	mg2	146000	150000	4001	CN1	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
28	mg2	364000	367000	3001	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
29	mg5	0	30000	30001	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
30	mg5	0	30000	30001	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
31	mg5	43000	59000	16001	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2
32	mg9	0	16893	16894	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
33	mg9	0	16301	16302	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
34	mg12	136000	140000	4001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
35	mg13	99000	102000	3001	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2

No	Scaffolds	start	end	width	Number of copies												
					Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27
36	mg13	222000	225000	3001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
37	mg14	356000	360000	4001	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
38	mg16	49000	52000	3001	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
39	mg16	61000	66000	5001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
40	mg16	188000	191000	3001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
41	mg16	193000	203000	10001	CN3	CN2	CN2	CN2	CN2	CN1	CN3	CN3	CN2	CN2	CN2	CN2	CN2
42	mg16	207000	213000	6001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
43	mg16	244000	248000	4001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
44	mg16	352000	361000	9001	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2
45	mg16	387000	390000	3001	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2
46	mg17	0	9000	9001	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2
47	mg18	109000	117000	8001	CN2	CN2	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2
48	mg19	405000	413000	8001	CN1	CN2	CN3	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN3	CN3	CN2
49	mg19	0	27000	27001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0
50	mg20	287000	290000	3001	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
51	mg21	41000	47000	6001	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
52	mg23	44000	49000	5001	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
53	mg28	0	6000	6001	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2
54	mg31	40000	44000	4001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
55	mg31	37000	43000	6001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
56	mg31	55000	59000	4001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
57	mg32	101000	105000	4001	CN1	CN1	CN2	CN4	CN2	CN2	CN1	CN1	CN2	CN2	CN2	CN2	CN2
58	mg32	185000	190000	5001	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
59	mg35	131000	134000	3001	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
60	mg35	174000	178000	4001	CN2	CN2	CN2	CN1	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
61	mg37	0	13000	13001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2
62	mg38	101000	105000	4001	CN3	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
63	mg38	69000	72000	3001	CN2	CN2	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2
64	mg38	0	224262	224263	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2
65	mg38	167000	172000	5001	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2
66	mg39	209000	214000	5001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN3	CN2
67	mg39	216000	219555	3556	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN3	CN2
68	mg40	165000	169000	4001	CN5	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2
69	mg40	129000	132000	3001	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
70	mg41	0	16000	16001	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2
71	mg45	147000	150000	3001	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
72	mg46	72000	75000	3001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
73	mg51	154000	162000	8001	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN1	CN2	CN2

No	Scaffolds	start	end	width	Number of copies												
					Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27
74	mg52	138000	163942	25943	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
75	mg53	128000	131000	3001	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
76	mg53	133000	139000	6001	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
77	mg53	7000	11000	4001	CN1	CN0	CN2	CN1	CN4	CN6	CN2	CN2	CN0	CN2	CN2	CN2	CN2
78	mg58	0	10000	10001	CN2	CN1	CN3	CN1	CN0	CN3	CN2	CN2	CN1	CN2	CN3	CN3	CN2
79	mg61	120000	127393	7394	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2
80	mg67	105000	110095	5096	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2
81	mg68	103000	107941	4942	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN3	CN2	CN2
82	mg71	10000	16000	6001	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2
83	mg71	28000	31000	3001	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2
84	mg71	0	20000	20001	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2
85	mg72	7000	11000	4001	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
86	mg75	31000	60957	29958	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0
87	mg76	43000	46000	3001	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
88	mg76	53000	59101	6102	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
89	mg77	42000	48000	6001	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2
90	mg77	0	52160	52161	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2
91	mg77	9000	23000	14001	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
92	mg78	0	11000	11001	CN2	CN2	CN3	CN2	CN2	CN2	CN6	CN2	CN6	CN2	CN6	CN6	CN2
93	mg79	0	11000	11001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2
94	mg85	23000	30000	7001	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2
95	mg85	12000	16000	4001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2
96	mg85	0	32492	32493	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2
97	mg86	22000	26000	4001	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2
98	mg86	26000	29728	3729	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2
99	mg87	0	8000	8001	CN1	CN2	CN3	CN2	CN2	CN2	CN3	CN1	CN2	CN2	CN3	CN3	CN2
100	mg87	0	6000	6001	CN1	CN2	CN3	CN2	CN2	CN2	CN3	CN1	CN2	CN2	CN3	CN3	CN2
101	mg87	0	24367	24368	CN1	CN2	CN3	CN0	CN0	CN2	CN3	CN1	CN0	CN2	CN3	CN3	CN0
102	mg92	0	13200	13201	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
103	mg92	0	12630	12631	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2
104	mg195	0	56000	56001	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0

Table S7. **List of protein-coding genes affected by CNVs among 13 isolates**. CN2 indicated normal diploid state; CN0 and CN1 indicated deletion of copy compare to normal diploid state; from CN3 to CN6 indicate duplication of copy number compare to the normal diploid state.

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00000001	mg1:1-3881	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR004859,PF03159,GO:0003676,GO:0004527
Mgra_00000016	mg1:64893-72272	CN2	CN2	CN2	CN0	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2	PF01163
Mgra_00000066	mg1:245104-245434	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00000092	mg1:328727-330489	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00000096	mg1:340210-344185	CN2	CN2	CN2	CN1	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000109,PF00854,GO:0005215,GO:0006810,GO:0016020
Mgra_00000143	mg1:533632-536135	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002486,IPR008160,PF01391,PF01484,GO:0042302
Mgra_00000134	mg1:502190-505406	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000719,PF00069,GO:0004672,GO:0005524,GO:0006468
Mgra_00000161	mg1:611218-611813	CN3	CN3	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00000300	mg1:1262289-1264175	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00000320	mg1:1349574-1351626	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	
Mgra_00000435	mg2:363326-364912	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR032436,PF16201
Mgra_00000436	mg2:365634-366569	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00000443	mg2:384011-387752	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	
Mgra_00000442	mg2:383542-383864	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	IPR001163,PF01423
Mgra_00001025	mg5:11282-24627	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001506,PF01400,GO:0004222,GO:0006508
Mgra_00001025	mg5:11282-24627	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001506,PF01400,GO:0004222,GO:0006508
Mgra_00001022	mg5:7-2138	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000242,PF00102,GO:0004725,GO:0006470
Mgra_00001022	mg5:7-2138	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000242,PF00102,GO:0004725,GO:0006470
Mgra_00001029	mg5:25845-29661	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR015868,IPR020683,PF04960,PF12796,GO:0004359,GO:0006541
Mgra_00001029	mg5:25845-29661	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR015868,IPR020683,PF04960,PF12796,GO:0004359,GO:0006541
Mgra_00001036	mg5:46974-48805	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	IPR019424,PF10320
Mgra_00001035	mg5:43656-44651	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00001024	mg5:6312-9611	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR012816,PF08719
Mgra_00001024	mg5:6312-9611	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR012816,PF08719
Mgra_00001037	mg5:49849-52907	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00001028	mg5:23988-24711	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR009071,PF09011

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00001028	mg5:23988-24711	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR009071,PF09011
Mgra_00001023	mg5:1950-4115	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR013057,PF01490
Mgra_00001023	mg5:1950-4115	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR013057,PF01490
Mgra_00001026	mg5:12195-12470	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00001026	mg5:12195-12470	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00001027	mg5:15126-18772	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR019382,PF10255,GO:0003743,GO:0005737,GO:0005852
Mgra_00001027	mg5:15126-18772	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR019382,PF10255,GO:0003743,GO:0005737,GO:0005852
Mgra_00001612	mg9:5367-8120	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR025750,PF14051
Mgra_00001612	mg9:5367-8120	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR025750,PF14051
Mgra_00001611	mg9:2295-5249	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00001611	mg9:2295-5249	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00001613	mg9:7975-9767	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR003613,PF04564,PF12895,GO:0004842,GO:0016567
Mgra_00001613	mg9:7975-9767	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR003613,PF04564,PF12895,GO:0004842,GO:0016567
Mgra_00001615	mg9:11530-17156	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002213,PF00201,GO:0008152,GO:0016758
Mgra_00001615	mg9:11530-17156	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002213,PF00201,GO:0008152,GO:0016758
Mgra_00001614	mg9:9968-10485	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00001614	mg9:9968-10485	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00002074	mg12:139321-140210	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00002208	mg13:99066-101014	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00002580	mg16:63677-64598	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000542,PF00755,GO:0016746
Mgra_00002577	mg16:50346-50750	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR008906,PF05699,GO:0046983
Mgra_00002612	mg16:196652-197616	CN3	CN2	CN2	CN2	CN2	CN1	CN3	CN3	CN2	CN2	CN2	CN2	CN2	
Mgra_00002616	mg16:208381-210390	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000276,PF00001,GO:0004930,GO:0007186,GO:0016021
Mgra_00002626	mg16:244722-246858	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000276,PF00001,GO:0004930,GO:0007186,GO:0016021
Mgra_00002613	mg16:199127-201659	CN3	CN2	CN2	CN2	CN2	CN1	CN3	CN3	CN2	CN2	CN2	CN2	CN2	IPR002524,IPR027470,PF01545,PF16916,GO:0006812,GO:0008324,GO:0016021,GO:0055085
Mgra_00002617	mg16:211047-212093	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00002659	mg16:352922-354313	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR006603,PF04193
Mgra_00002660	mg16:354546-361367	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR019791,PF03098

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00002698	mg17:6062-9501	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00002859	mg18:110139-110949	CN2	CN2	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00002984	mg19:2285-3687	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR019425,PF10321
Mgra_00002982	mg19:721-1451	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00002990	mg19:19936-25228	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR004898,PF03211,GO:0005576,GO:0030570
Mgra_00002989	mg19:13648-19746	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00002988	mg19:9703-12147	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR004898,PF03211,GO:0005576,GO:0030570
Mgra_00002983	mg19:1457-1811	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00002987	mg19:7451-8680	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00002991	mg19:21418-26169	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00002985	mg19:3499-5504	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR001873,PF00858,GO:0005272,GO:0006814,GO:0016020
Mgra_00002986	mg19:6048-7053	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00003216	mg21:43619-48011	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001356,PF00046,GO:0003677
Mgra_00003215	mg21:41859-42110	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002919,PF01826
Mgra_00004243	mg31:41351-41383	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004243	mg31:41351-41383	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004696	mg37:986-2941	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	IPR020070,PF01281
Mgra_00004697	mg37:3050-7263	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	
Mgra_00004812	mg38:60004-60900	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004798	mg38:4185-8494	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR005011,PF03343,GO:0000398
Mgra_00004808	mg38:37976-43018	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR002213,PF00201,GO:0008152,GO:0016758
Mgra_00004811	mg38:56713-58702	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004816	mg38:72258-74201	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR003347,PF02373
Mgra_00004806	mg38:33131-33511	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004807	mg38:34711-36889	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004801	mg38:16162-18085	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004809	mg38:43107-44137	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004805	mg38:31717-32536	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00004799	mg38:8137-10185	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004800	mg38:11029-12408	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR002048,PF13202,GO:0005509
Mgra_00004815	mg38:69109-71773	CN2	CN2	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004815	mg38:69109-71773	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004813	mg38:61342-67209	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001506,PF01400,GO:0004222,GO:0006508
Mgra_00004797	mg38:1274-3705	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR007588,PF04500
Mgra_00004804	mg38:24766-27334	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR004254,PF03006,GO:0016021
Mgra_00004810	mg38:45306-55067	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004803	mg38:21742-24280	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR000980,IPR001452,PF00017,PF00018,GO:0005515
Mgra_00004818	mg38:79114-81528	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001680,PF00400,GO:0005515
Mgra_00004817	mg38:75131-77439	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001019,PF00503,GO:0003924,GO:0004871,GO:0007186,GO:0019001,GO:0031683
Mgra_00004802	mg38:18567-21722	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001214,PF00856,GO:0005515
Mgra_00004814	mg38:67278-68385	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004796	mg38:258-875	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004839	mg38:136247-137650	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004824	mg38:94187-94483	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	PF00304
Mgra_00004827	mg38:106070-106545	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004842	mg38:146577-146962	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001841,PF13639,GO:0005515,GO:0008270
Mgra_00004823	mg38:91708-93551	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004833	mg38:117241-118895	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004851	mg38:179272-181204	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR006869,PF04784
Mgra_00004830	mg38:113594-116208	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001623,PF00226
Mgra_00004826	mg38:97675-99021	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001781,PF00412,GO:0008270
Mgra_00004837	mg38:124779-132324	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR003890,PF02854,GO:0003723,GO:0005515
Mgra_00004844	mg38:150180-152833	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR013099,PF07885
Mgra_00004841	mg38:140776-144264	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR004316,PF03083,GO:0016021
Mgra_00004847	mg38:159489-165100	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	PF00406
Mgra_00004828	mg38:107383-108466	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00004822	mg38:90059-91547	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004843	mg38:147529-149059	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001179,PF00254,GO:0006457
Mgra_00004821	mg38:86465-89166	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR030379,PF00735,GO:0005525
Mgra_00004820	mg38:83010-86186	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR002109,IPR004843,PF00149,PF00462,GO:0009055,GO:0015035,GO:0016787,GO:0045454
Mgra_00004819	mg38:81135-82931	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR000504,IPR000571,PF00076,PF00642,GO:0003676,GO:0046872
Mgra_00004845	mg38:153515-155588	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR003033,PF02036,
Mgra_00004836	mg38:121116-124554	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001806,IPR007143,PF00071,PF03997,GO:0005525,GO:0007264
Mgra_00004835	mg38:119807-120973	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR013907,PF08598,
Mgra_00004846	mg38:156807-158076	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR005002,PF03332,GO:0004615,GO:0005737,GO:0009298
Mgra_00004825	mg38:94654-95124	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001841,PF13639,GO:0005515,GO:0008270
Mgra_00004834	mg38:118901-119791	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004838	mg38:134697-135830	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004840	mg38:137996-139364	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004832	mg38:116246-117117	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004829	mg38:109440-113720	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	PF00304
Mgra_00004831	mg38:115661-116108	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR000781,PF01133,GO:0006221,GO:0007049,GO:0045747
Mgra_00004850	mg38:178508-179050	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00004859	mg38:207757-215303	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR000433,PF00569,GO:0008270
Mgra_00004858	mg38:201260-207472	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR000408,PF00415,PF13540
Mgra_00004860	mg38:215605-217322	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001451,IPR005835,PF00132,PF00483,GO:0009058,GO:0016779
Mgra_00004855	mg38:194092-195770	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR007125,PF00125,GO:0003677
Mgra_00004857	mg38:200565-201228	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR001515,PF01655,GO:0003735,GO:0005622,GO:0005840,GO:0006412
Mgra_00004856	mg38:195835-197734	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR006941,PF04857,GO:0005634
Mgra_00004849	mg38:173247-190101	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR000195,IPR000904,IPR001310,IPR001849,IPR003010,IPR004182,PF00566,PF00795, PF01230,PF01369,PF02893,PF16453,GO:0005086,GO:0006807,GO:0016810,GO:0032012
Mgra_00004848	mg38:167004-172736	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	IPR003131,PF02214,GO:0051260
Mgra_00004848	mg38:167004-172736	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR003131,PF02214,GO:0051260
Mgra_00004854	mg38:192704-194111	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR008855,PF05404,GO:0005783,GO:0016021
Mgra_00004853	mg38:191952-192506	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR009688,PF06916

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00004852	mg38:190199-191304	CN3	CN3	CN2	CN2	CN1	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR009688,PF06916
Mgra_00004985	mg40:166894-167718	CN5	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	IPR024445,PF12762
Mgra_00004984	mg40:166044-168857	CN5	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	IPR019425,PF10321
Mgra_00005020	mg41:7594-8417	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00005348	mg45:148111-149423	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR019425,PF10321
Mgra_00005413	mg46:72498-74490	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001503,PF00852,GO:0006486,GO:0008417,GO:0016020
Mgra_00005875	mg52:156112-163412	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001828,PF01094
Mgra_00005870	mg52:140931-143966	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000990,PF00876,GO:0005921
Mgra_00005874	mg52:148441-151080	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001506,PF01400,GO:0004222,GO:0006508
Mgra_00005872	mg52:145214-146987	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR016072,IPR016073,PF01466,PF03931,GO:0006511
Mgra_00005873	mg52:147006-148335	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00005871	mg52:144172-144793	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00005916	mg53:132825-135066	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00006341	mg61:120973-128269	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	IPR001223,IPR001752,IPR002557,PF00225,PF00704,PF01607,GO:0003777,GO:0004553,GO:0005524,GO:0005576,GO:0005975,GO:0006030,GO:0007018,GO:0008017,GO:0008061
Mgra_00006814	mg71:3443-3760	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006813	mg71:2547-3066	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006818	mg71:10805-11138	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006818	mg71:10805-11138	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006817	mg71:9465-10566	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006815	mg71:5987-8340	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006811	mg71:687-1331	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006812	mg71:1722-2342	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006816	mg71:8512-9280	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00006810	mg71:183-537	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00007032	mg75:44339-46588	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR016090,PF00068
Mgra_00007030	mg75:39406-41105	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00007033	mg75:48067-52967	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00007031	mg75:42227-42817	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR001534,PF01060,GO:0005615

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00007028	mg75:33047-34069	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00007034	mg75:58891-61343	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR001503,IPR031481,PF00852,PF17039,GO:0006486,GO:0008417,GO:0016020
Mgra_00007029	mg75:34845-37299	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00007079	mg76:42532-44355	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001506,PF01400,GO:0004222,GO:0006508
Mgra_00007112	mg77:15392-16314	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007112	mg77:15392-16314	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007118	mg77:21998-24792	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000980,IPR001245,PF00017,PF07714,GO:0004672,GO:0006468
Mgra_00007109	mg77:4477-6328	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002018,PF00135
Mgra_00007120	mg77:28057-35845	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001611,PF13855,GO:0005515
Mgra_00007123	mg77:41216-43093	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007129	mg77:49542-52562	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000198,IPR001251,PF00620,PF13716,GO:0007165
Mgra_00007124	mg77:43173-44278	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007124	mg77:43173-44278	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007119	mg77:25304-27856	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002130,PF00160,GO:0000413,GO:0003755,GO:0006457
Mgra_00007126	mg77:44413-45186	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001451,PF00132
Mgra_00007126	mg77:44413-45186	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001451,PF00132
Mgra_00007128	mg77:48981-49280	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007115	mg77:18493-20100	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007115	mg77:18493-20100	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007108	mg77:53-2153	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR006202,PF02931,GO:0005230,GO:0006810,GO:0016020
Mgra_00007122	mg77:38330-40987	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001623,IPR024586,PF00226,PF11875
Mgra_00007125	mg77:43718-47375	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR007109,IPR019330,PF04427,PF10185
Mgra_00007125	mg77:43718-47375	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR007109,IPR019330,PF04427,PF10185
Mgra_00007111	mg77:11160-13429	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007111	mg77:11160-13429	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007110	mg77:8113-10852	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001128,PF00067,GO:0005506,GO:0016705,GO:0020037,GO:0055114
Mgra_00007121	mg77:36581-37745	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007116	mg77:20828-21461	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR025954,PF14443

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00007116	mg77:20828-21461	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR025954,PF14443
Mgra_00007113	mg77:17016-18072	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002347,PF00106
Mgra_00007113	mg77:17016-18072	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002347,PF00106
Mgra_00007117	mg77:21548-21873	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007117	mg77:21548-21873	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007127	mg77:47798-48399	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007114	mg77:18213-18739	CN0	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007114	mg77:18213-18739	CN0	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007177	mg78:8294-8642	CN2	CN2	CN3	CN2	CN2	CN2	CN6	CN2	CN6	CN2	CN6	CN6	CN2	
Mgra_00007176	mg78:7047-9445	CN2	CN2	CN3	CN2	CN2	CN2	CN6	CN2	CN6	CN2	CN6	CN6	CN2	IPR006076,PF01266,GO:0016491,GO:0055114
Mgra_00007219	mg79:10006-11128	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	
Mgra_00007218	mg79:7522-9797	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	
Mgra_00007217	mg79:4926-6105	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	
Mgra_00007215	mg79:2429-3200	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	
Mgra_00007216	mg79:2562-3909	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	
Mgra_00007214	mg79:223-1924	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN4	CN2	CN2	CN2	CN2	
Mgra_00007470	mg85:19077-20968	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	
Mgra_00007464	mg85:1544-7554	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR000644,PF00571
Mgra_00007472	mg85:25319-29535	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR014782,IPR024571,PF01433,PF11838,GO:0008237,GO:0008270
Mgra_00007472	mg85:25319-29535	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR014782,IPR024571,PF01433,PF11838,GO:0008237,GO:0008270
Mgra_00007467	mg85:12125-13172	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	
Mgra_00007467	mg85:12125-13172	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	
Mgra_00007471	mg85:24189-25360	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR001247,PF01138
Mgra_00007471	mg85:24189-25360	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR001247,PF01138
Mgra_00007469	mg85:17639-23433	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR003689,PF02535,GO:0016020,GO:0030001,GO:0046873,GO:0055085
Mgra_00007465	mg85:7656-9110	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR010761,PF07062,GO:0016021
Mgra_00007466	mg85:9799-11146	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR010761,PF07062,GO:0016021
Mgra_00007468	mg85:13634-17324	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	IPR011442,PF07571,GO:0005634,GO:0051090

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00007533	mg87:7356-12037	CN1	CN2	CN3	CN0	CN0	CN2	CN3	CN1	CN0	CN2	CN3	CN3	CN0	
Mgra_00007535	mg87:21571-22792	CN1	CN2	CN3	CN0	CN0	CN2	CN3	CN1	CN0	CN2	CN3	CN3	CN0	IPR007482,PF04387
Mgra_00007532	mg87:809-5962	CN1	CN2	CN3	CN2	CN2	CN2	CN3	CN1	CN2	CN2	CN3	CN3	CN2	
Mgra_00007532	mg87:809-5962	CN1	CN2	CN3	CN2	CN2	CN2	CN3	CN1	CN2	CN2	CN3	CN3	CN2	
Mgra_00007532	mg87:809-5962	CN1	CN2	CN3	CN0	CN0	CN2	CN3	CN1	CN0	CN2	CN3	CN3	CN0	
Mgra_00007534	mg87:11649-20733	CN1	CN2	CN3	CN0	CN0	CN2	CN3	CN1	CN0	CN2	CN3	CN3	CN0	
Mgra_00007711	mg92:5250-6781	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002486,IPR008160,PF01391,PF01484,GO:0042302
Mgra_00007711	mg92:5250-6781	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002486,IPR008160,PF01391,PF01484,GO:0042302
Mgra_00007709	mg92:416-1561	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007709	mg92:416-1561	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007710	mg92:990-3868	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007710	mg92:990-3868	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007713	mg92:8357-10435	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007713	mg92:8357-10435	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007714	mg92:11496-13325	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007712	mg92:6632-7700	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00007712	mg92:6632-7700	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00009886	mg195:38644-39198	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00009878	mg195:10717-27294	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	PF14776
Mgra_00009877	mg195:6398-7108	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00009884	mg195:35246-36425	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00009885	mg195:36507-37520	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00009880	mg195:27665-29069	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR019425,PF10321
Mgra_00009883	mg195:32840-34072	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00009879	mg195:20666-23846	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR001841,PF17123,GO:0005515,GO:0008270
Mgra_00009881	mg195:28190-30886	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	IPR001873,PF00858,GO:0005272,GO:0006814,GO:0016020
Mgra_00009882	mg195:31200-32496	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN0	
Mgra_00000037	mg1:145000-150000	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR002073,PF00233,GO:0004114,GO:0007165

Gene name	Scaffold position	Number of copies													Interproscan/Pfam/GO Term
		Mg-Bali	Mg-Borneo	Mg-Brazil	Mg-C21	Mg-C25	Mg-L1	Mg-L2	Mg-Java2	Mg-P	Mg-VN6	Mg-VN11	Mg-VN18	Mg-VN27	
Mgra_00000106	mg1:382000-390000	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000699,IPR013662,IPR014821,IPR016093,PF01365,PF02815,PF08454,PF08709,GO:0005262,GO:0016020,GO:0070588
Mgra_00000175	mg1:677000-680000	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR028191,IPR028282,IPR028283,PF14744,PF14745,PF14746,
Mgra_00000212	mg1:860000-863000	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR003961,IPR013098,IPR026966,PF00041,PF07679,PF13882,PF13927,GO:0005515
Mgra_00000250	mg1:1038000-1043000	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN1	
Mgra_00000304	mg1:1283000-1288000	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	IPR000159,IPR000198,IPR001609,IPR002219,PF00063,PF00130,PF00620,PF00788,GO:0003774,GO:0005524,GO:0007165,GO:0016459,GO:0035556
Mgra_00000316	mg1:1338000-1342000	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	
Mgra_00000441	mg2:380000-383000	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	IPR001715,IPR004953,PF00307,PF03271,GO:0005515,GO:0008017
Mgra_00002611	mg16:188000-191000	CN2	CN2	CN2	CN2	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR000219,IPR000306,IPR001849,PF00169,PF00621,PF01363,GO:0005089,GO:0035023,GO:0046872
Mgra_00002860	mg18:109000-117000	CN2	CN2	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	CN2	CN2	IPR002049,IPR008211,PF00053,PF00055
Mgra_00003166	mg20:287000-290000	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR001623,IPR004179,IPR006993,PF00226,PF02889,PF04908
Mgra_00004974	mg40:129000-132000	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR013766,IPR017905,PF00085,PF04777,GO:0016972,GO:0045454,GO:0055114
Mgra_00005815	mg51:154000-162000	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN1	CN2	CN2	IPR019791,PF03098
Mgra_00005915	mg53:128000-131000	CN1	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	CN2	
Mgra_00006662	mg68:103000-107941	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN3	CN2	CN2	
Mgra_00006820	mg71:28000-31000	CN2	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN0	CN2	CN2	CN2	CN2	
Mgra_00007506	mg86:26000-29728	CN2	CN2	CN2	CN2	CN2	CN3	CN2	CN2	CN2	CN2	CN2	CN2	CN2	IPR011704,PF07728,GO:0005524,GO:0016887

CHAPTER V

Specifying genomic rearrangements in an emerging virulent isolate of rice root-knot nematode (*Meloidogyne graminicola*)

Summary in English

In Chapter IV, genomic comparison of population revealed mechanisms such as recombination, loss of heterozygosity, copy number variants and single nucleotide variants, which could potentially participate in the molecular evolution of the species. This chapter V shows that these mechanisms could thus be associated with the ability to bypass resistance in rice.

Meloidogyne graminicola can attack most cultivars of *Oryza sativa* and cause severe production losses. Several genotype screenings have found some cultivars resistant or tolerant to this pathogen. However, resistance in these cultivars can be bypassed by a virulent pathotype that has recently been isolated in Cambodia. The incompatible interaction resulting in a hypersensitive response (HR) in the plant usually responds to a gene-for-gene interaction. We thus hypothesized that the acquisition of virulence could have been obtained by the loss of HR-inducing effectors or by the acquisition of effectors to neutralize the induction of HR. In order to study the genomic modifications that may have targeted the potential virulent gene(s), the genome of the virulent isolate was sequenced and compared to the genomes of the 13 other avirulent isolates.

In total, 1.2Mb (~2.9% of the genome) have lost their heterozygosity in the virulent genome. Of these 1.2Mb, 547 kb are specific losses in the virulent genome. Of the 133 genes specifically affected in the virulent genome, 14 genes (~10% of the genes found) code for a potentially secreted protein. Thus, LoH could allow to suppress or increase (dose effect) the expression level of a gene encoding an effector or allow the phenotypic expression of a recessive allele.

Ten SNPs causing non-synonymous mutations and affecting seven genes have been found in the virulent genome. Among them, two genes carrying a signal peptide suggest that they are potential effectors.

The CNVs may have affected a total of 45 genes in the virulent genome. Twenty nine genes were absent from the virulent genome, four of which coded for potential effectors. The abundance of gene losses in the virulent genome reinforces the hypothesis that these losses could be linked to a genetic adaptation mechanism in response to a major stress such as having to reproduce in a resistant plant.

The function of genes encoding potential effectors that have been the target of these genomic variation events is discussed on the basis of the *InterProscan* domains found in their

deduced protein sequence. We hypothesize that among these genes are potentially genes encoding factors responsible for causing the hypersensitive response in *O. sativa*. However, it is also possible that these rearrangements modify regulatory sequences (Cis-regulatory elements) which in turn control the expression of the genes responsible for the incompatible interaction. Similarly, these genomic variations may have altered chromatin conformation, leading to changes in the expression of HR-inducing factors.

Résumé en Français

Dans le chapitre IV, la comparaison génomique des populations a révélé des mécanismes tels que la recombinaison, la perte d'hétérozygotie, les variantes du nombre de copies et les variantes d'un seul nucléotide, qui pourraient potentiellement participer à l'évolution moléculaire de l'espèce. Ce chapitre V montre que ces mécanismes pourraient ainsi être associés à la capacité de contourner la résistance du riz.

M. graminicola peut attaquer la plupart des cultivars d'*Oryza sativa* et provoquer de graves pertes de production. Plusieurs criblages de génotypes ont permis de trouver quelques cultivars résistants ou tolérants à cet agent pathogène. Cependant, la résistance chez ces cultivars peut être contourné par un pathotype virulent qui a récemment été isolé au Cambodge. L'interaction incompatible se traduisant par une réponse hypersensible (HR) chez la plante répond en générale à une interaction gène pour gène. Nous émettions ainsi l'hypothèse que l'acquisition de la virulence aurait pu être obtenu par la perte d'effecteurs induisant la HR ou bien par l'acquisition d'effecteurs permettant de neutraliser l'induction de la HR. Afin d'étudier les modifications génomiques qui pourraient avoir ciblé le(s) gène(s) de virulence potentiel, le génome de l'isolat virulent a été séquencé et comparé aux génomes des 13 autres isolats avirulents.

Au total, 1,2Mb (~2,9% du génome) ont perdu leur hétérozygotie dans le génome virulent. De ces 1,2Mb, 547 kb sont des pertes spécifiques au génome virulent. Parmi les 133 gènes affectés spécifiquement dans le génome virulent, 14 gènes (~10% des gènes trouvés) codent pour une protéine potentiellement sécrétée. Ainsi, les LoH pourrait permettre de supprimer ou augmenter (effet de dosage) le niveau d'expression d'un gène codant un effecteur ou permettre l'expression phénotypique d'un allèle récessif.

Dix SNPs provoquant des mutations non synonymes et affectant sept gènes ont été trouvés dans le génome virulent. Parmi eux, deux gènes portant un peptide signal suggère que ce sont des effecteurs potentiels.

Les CNVs peuvent avoir affecté au total 45 gènes du génome virulent. 29 gènes étaient absents du génome virulent dont quatre codant pour des effecteurs potentiels. L'abondance des pertes de gènes dans le génome virulent renforce l'hypothèse que ces pertes pourraient être lié à un mécanisme d'adaptation génétique en réponse à un stress important tel que de devoir se reproduire dans une plante résistante.

La fonction des gènes codant les effecteurs potentiels et qui ont été la cible de ces évènements de variation génomique est discuté sur la base des domaines *InterProscan* retrouvés dans leur séquence protéique déduite. Nous émettons l'hypothèse que parmi ces gènes se trouvent potentiellement des gènes codant pour les facteurs responsables du déclenchement de la réponse hypersensible chez *O. sativa*. Néanmoins, il est également possible que ces réarrangements modifient des séquences régulatrices (Cis-regulatory elements) qui a leur tour contrôlent l'expression des gènes responsables de l'interaction incompatible. De même, ces variations génomiques ont pu modifier la conformation de la chromatine ce qui a abouti à modifier l'expression des facteurs induisant la HR.

Specifying genomic rearrangements in an emerging virulent isolate of rice root-knot nematode (*Meloidogyne graminicola*)

Ngan Thi Phan[1], Hue Nguyen Thi[3], Guillaume Besnard[2*], Stephane Bellafiore[1]*

[1]PHIM Plant Health Institute, Univ of Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France.

[2]CNRS-UPS-IRD, UMR5174, EDB, 118 route de Narbonne, Université Paul Sabatier, 31062 Toulouse, France.

[3]LMI RICE, Agricultural Genetics Institute, University of Science and Technology of Hanoi, Hanoi, Viet Nam

* co-senior authors for correspondence: guillaume.besnard@univ-tlse3.fr (G.B.); stephane.bellafiore@ird.fr (S.B.)

Abstract

Meloidogyne graminicola causes extensive damage to rice production worldwide. Moreover, the recent emergence of a virulent pathotype against resistant varieties threatens the sustainability of the crop resistance. In this study, to reveal the potential genetic characteristics underlying the acquisition of virulence, a comparative genomic analysis was performed on two pathotypes (one virulent, and one avirulent). While the mitogenome of pathotypes is almost identical, loss of heterozygosity (LoH) on long genomic blocks, a variation in the number of gene copies (CNV), and single nucleotide variants (SNV) were detected in their nuclear genomes. The precise role of these genomic modifications remains unexplained but some of them may directly affect genes involved in plant-nematode interaction and parasitism. In particular, 28 genes present in the avirulent genome, including two genes coding for potential effectors, were not detected in the virulent genome. Besides, allelic variation of other putative effector-coding genes was eliminated in specific LoH regions of the virulent genome. Finally, non-synonymous SNPs were detected in seven nuclear genes in the virulent pathotype. This study confirms that different processes generate genomic variation in *M. graminicola* especially via LoHs and CNVs. These modifications may affect the expression of certain genes and alleles or potentially their level of expression (e.g. dosage effect), and some of them should affect the plant-pathogen interaction and the acquisition of virulence. The study of genes identified in this study in a context of adaptation of parasitism and acquisition of virulence should allow in the near future to better understanding the sophisticated dialogue that takes place between the parasite and its host leading to disease or resistance.

Key words: Comparative genomics, Copy number variation, Loss of heterozygosity, Plant-pathogen interaction, Single nucleotide variation, Virulence acquisition.

I. Introduction

Rice (*Oryza sativa*) is the staple food for more than half of the world population. Unfortunately, rice production is threatened by *Meloidogyne graminicola* (*Mg*), a rice root-knot nematode (RKN) in most of known agrosystems. The management of this pest remains a major challenge due to its sedentary endoparasitism and the lack of appropriate phytosanitary treatment following the ban on the use of most nematicides of chemical origin. The use of resistance varieties is a suitable control strategy due to its high efficiency and environmental friendliness. Resistance to *Mg* has been found in some wild rice species, including *Oryza longistaminata* (Soriano et al., 1999), *O. glumaepatula* (Mattos et al., 2019), and the African rice *O. glaberrima* (Cabasan et al., 2012). In Asian rice, *Oryza sativa,* only a few cultivars have been described as resistant and/or tolerant to *Mg* including Khao Pahk Maw, LD24 (Dimkpa et al., 2016), Shenliangyou 1, Cliangyou 4418 (Zhan et al., 2018), and Zhonghua 11 (Phan et al., 2018). In resistant rice cultivars, *Mg* juveniles can penetrate the roots but at most only a few individuals can infect and reproduce. Characterization of the resistant response indicated that resistant rice plants act against *Mg* by inducing hypersensitive-like reactions along migration sites and nematode feeding sites (Phan et al., 2018; Cabasan et al., 2014; Mattos et al., 2019). This suggested a gene-for-gene (Avr/R) relationship between rice and *Mg*. The hybrids resulting from the cross between 'Zhonghua 11' and the susceptible 'IR64' are all resistant indicating that dominant resistance gene(s) must be involved in this defense mechanism. These resistant cultivars are considered extremely valuable resources for future breeding programs aimed at producing resistance against *Mg* in Asian rice cultivars.

However, the durability of major rice resistance genes against *Mg* is threatened by the emergence of virulent pathotypes capable of overcoming plant resistance genes. Recently, a new virulent *Mg* pathotype isolated from a rice field in Cambodia was found to able to reproduce with a high reproduction factor on many *O. sativa* accessions such as 'Zhonghua 11', 'Khao Pahk Maw', and 'LD24' (Nguyen et al., unpublished data). Surprisingly, the histological analysis clearly illustrated the absence of phenolic compounds accumulation at the nematodes' neighboring root cells suggesting the disappearance of the hypersensitive-like reaction in 'Zhonghua 11'. This suggests that this virulent pathotype has developed sophisticated mechanisms to break down resistance mediated by potential R gene(s).

The co-evolution for the arms race between plant resistance and pathogen virulence has been investigated in plant-RKN interactions on other models. Molecular mechanism of interaction between plants and RKNs has been described and involved a multi-layered system

of plant defenses (Przybylska & Obrępalska-Stęplowska, 2020). At the first layer, plants recognize RKNs by pathogen-associated molecular patterns (PAMPs) on their surfaces as well as damage-associated molecular patterns (DAMPs) released by the disrupted host plant tissues. PAMPs and DAMPs are recognized by pattern recognition receptors (PRRs) which activate a level of basal defenses in the infected plant called PAMP-triggered immunity (PTI). PTI induces a range of responses including ROS production and strengthening to isolate and kill the invader (Jones & Dangl, 2006). Plant parasitic nematodes have also been shown to induce typical defense responses in plants such as callose production, cell wall thickening and ROS production (Ji et al., 2015; Waetzig et al., 1999). Interestingly, a family of nematode pheromones, called ascarosides, has been identified in several genera of plant parasitic nematodes. The detection of these ascarosides by plants induces characteristic defense responses, including the expression of genes associated with the immunity triggered by PAMPs (Manosalva et al., 2015). At the second layer, RKNs secrete effectors to notably suppress these defense responses and thus optimize the infection of the host. But these effectors can in turn be recognized by the plant and induce effector-triggered immunity (ETI). In the incompatible interaction (plant resistance against nematode), the ETI is accelerated and the immune response is amplified, resulting in a hypersensitive response (HR) at the site of infection. HR causes the death of the nematode's neighboring root cells, preventing the nematode from feeding, which will lead to its death. However, in the compatible interaction, RKNs could evolve to avoid and/or suppress ETI and contribute to making the plant susceptible [effector triggered susceptibility (ETS); (Jones & Dangl, 2006)].

Although the species mostly reproduce via asexual parthenogenesis, several mechanisms are reported to generate genetic diversity, such as the insertion of transposable elements (TE), recombination events, loss of heterozygosity (LoH) on long genomic blocks, or insertion/deletions producing copy number variations (CNV) (Phan et al 2020, in prep.). Such genetic diversity may promote the adaptation of the species to its environment, and, in particular, lead to the emergence of new virulent genotypes that could jeopardize varietal selection programs based on the introgression of resistance genes. Comparative genomic analysis between virulent and avirulent genotypes should make it possible to identify the molecular determinisms that govern adaptation/evolution and to understand how a virulent pathotype manages to bypass the plant's immune system. Applying this approach, we specified genome-specific molecular signatures of a virulent pathotype infecting a resistant rice variety to a wide panel of isolates. LoH and CNVs were detected, notably in several genes encoding

for potential effectors. CNVs observed in the virulent pathotype are essentially characterized by gene losses. An exhaustive list of candidate genes showing polymorphisms (i.e. presence/absence, point mutation, copy number) between closely related virulent and avirulent isolates was established in this study, opening new avenues for deciphering mechanisms that govern the circumvention of resistance in the plant-nematode interaction model.

II. Material & Methods

2.1 **Nematode DNA extraction and whole genome sequencing**

Virulent populations of *M. graminicola* were collected from 'Zhonghua 11' rice roots growing in a Cambodian rice field naturally infested by *Mg*. The most aggressive isolate (higher reproduction factor), called pathotype ZMgP1.8, was isolated and propagated on 'Zhonghua 11' from a single juvenile infection. The morphology and virulence of this pathotype against the resistant rice cultivar 'Zhonghua 11' were characterized (Nguyen et al., unpublished data). Subsequently, the virulent ZMgP1.8 and avirulent Mg-VN18 pathotypes were multiplied on 'Zhonghua 11' and 'IR64' rice varieties, respectively. Eggs and juveniles of these pathotypes were extracted from infected rice roots two months after inoculation (corresponding to two life cycles) using a hypochlorite extraction method and a blender as described previously (McClure et al., 1973; Bellafiore et al., 2015). After being extracted from roots, the mixture of eggs and juveniles was treated with a 0.8% sodium hypochlorite solution for 15 min at room temperature before being rinsed thoroughly with sterile ddH_2O. The eggs and juveniles were then purified using discontinuous sucrose gradient as described in Schaad & Walker (1975). Genomic DNA (gDNA) was extracted using a modified phenol–chloroform-based method as described in Phan et al. (2020). DNA quality and quantity were checked using electrophoresis, NanoDrop (Thermo Fisher Scientific), and the Qubit dsDNA HS Assay Kit (Life Technologies).

Pair-end read sequencing (2 x 150 bp) was performed for the two samples using Illumina MiSeq1-M00185 at the GeT-PlaGe core facility, INRA Toulouse (France). DNA-seq libraries with mean insert size of 417 bp for ZMgP1.8 and 444 bp for Mg-VN18 were used. Firstly, Illumina raw reads were trimmed from adapter using cutadapt (Martin, 2011), and read quality was controlled using FastQC (Andrews, 2010). Secondly, Skewer (Jiang et al., 2014) was used to trim pair-end reads considering a minimum quality score of 30 and a minimum read length of 51 bp. Sequence screening using BlobTools (Kumar et al., 2013) revealed no significant contamination by other organisms.

2.2 Calling of nucleotide variants and heterozygosity patterns across the genome

Three available reference sequences were used as templates to search for genetic variants namely: 1) the mitogenome (NC_024275.1); 2) a subset of diploid genomic sequences (354 kb) comprising pairs of haplotypes (i.e. alleles) of 18 nuclear regions (36 DNA fragments in total; Phan et al., unpublished); and 3) the haploid reference genome of the Mg-VN18 isolate (NCBI BioProjects: PRJNA615787; Phan et al., 2020). Illumina reads of both virulent and avirulent isolates were mapped to the above references separately using the BWA-MEM software (Li & Durbin, 2009). SAMtools (Li et al., 2009) was used to filter alignments with a MAPQ less than 20, to sort the alignment file by reference position, and to remove multi-mapped alignments. GATK (McKenna et al., 2010) was used to mark and remove the duplicated reads (MarkDuplicate). FreeBayes (Garrison & Marth, 2012) was used to detect the variants (SNVs and Indels) using all alignments simultaneously to produce a variant call file (VCF). The VCF file was filtered using vcftools (Danecek et al., 2011), retaining only the positions that had more than 30 phred-scaled probability (minGQ), a minimum coverage depth (minDP) of ten and no missing data.

Alignments of reads on the 36 nuclear fragments were visualized with Geneious v.9 (Kearse et al., 2012) to manually verify the correct phasing of paired-end reads assigned to each genomic region. Then, the average read-coverage was calculated for each region. The variants detected on the mitogenome and the 36 nuclear fragments were also visually inspected on the alignment with Geneious.

At the whole genome level, due to the collapse of the heterozygous haplotype sequences, the “0/1” genotype in the VCF file could represent two heterozygous alleles. Therefore, the two genotypes were called for the Mg-VN18 genome (when mapping to the haploid genome) including “0/0” and “0/1”. In the virulent genome, three genotypes, including “0/0”, "0/1" and "1/1", have been mainly called, while "2/2", "1/2", and “0/2” were rarely detected. To reveal variants in the genome organization (e.g. pattern of loss of heterozygosity on genomic blocks), all the genotypes called in the VCF files in both virulent/avirulent genomes were visualized along a reference scaffold using CIRCOS (http://circos.ca/). To study single nucleotide variants in the virulent population, only homologous ("0/0") regions found in the reference pathotype Mg-VN18 were considered. Genotypes found in the virulent pathotype and scored "1/1", "2/2", "1/2", and "0/2" were extracted. Genes presenting a polymorphism in the pathotype were identified and their putative function deduced by domain search (*Interproscan*) (Zdobnov & Apweiler, 2001) and GO annotation (Ashburner et al., 2000). Genes encoding putative effectors

were detected by checking their presence in the reference secretome published by Petitot et al. (2015) using Blastn.

2.3. Assembly of DNA regions from the virulent genome absent from references

Pair-end reads from the virulent genome that could not be mapped to the reference mitogenome (NC_024275.1) and the reference haploid genome (PRJNA615787) were finally extracted and assembled separately to construct additional contigs specific to the virulent pathotype. In brief, those reads were extracted using *samtools view* –F 0x4 and bedtools *bamtofastq*. Spades (Bankevich et al., 2012) was used to assembled unmapped pair-end reads into new contigs. The new constructed contigs were blasted to nucleotide database (nt NCBI) to identify genes.

2.4 Identification and function of genes with a variable copy number

CNVnator v.0.3.2 (Abyzov et al., 2011) and LUMPY v.0.4.13 (Layer et al., 2014) were used in parallel to identify putative copy number variations (CNVs) in the genome of the virulent isolate (ZMgP1.8). The CNVnator is a read depth method that uses a mean-shift-based approach to call CNVs based on the sequencing depth, while LUMPY is a probabilistic CNV discovery framework that integrates multiple detection signals, including split reads and paired-end mapping. The alignment of short reads on the haploid reference genome sequence was used as input to call CNVs using CNVnator. First, the window size (bin size) of 200 bp for read-depth calculation was selected based on author's optimizing method considering read coverage, read length, and data quality (Abyzov et al., 2011). Then, the software was run using a bin size of 200 bp, and all other parameters were set to default. CNVs detected with CNVnator were outputted in the VCF format. This result were then converted to the BEDPE format (using the script provided with LUMPY) to use them as input into LUMPY. Split reads or discordant reads (pair-end) were then extracted from the alignment with scripts provided with the software by Layer et al. (2014). Alignment of split reads, discordant pair-end reads and CNVnator's output were used to run LUMPY with default parameters. To eliminate artifacts, low complexity regions and high coverage regions (more than 500×) were excluded. Low complexity regions were detected using RepeatMasker v.4.0.7 (http://www.repeatmasker.org). High coverage regions were identified using python scripts implemented in LUMPY. Finally, genomic regions assigned to putative duplication, deletion, inversions and break-ends were outputted in VCF format. Furthermore, CNVs with weak read supports (less than 10×) or short length (less than 100 bp) were also removed.

In an attempt to reduce the number of false positives, CNVs were also called using the cn.MOPS algorithm v.1.24.0 (Klambauer et al., 2012). While CNVnator operates CNV regions based on the read count per bin size of 200 bp along the scaffolds of each isolate, cn.MOPS decomposes read count variants (in window length of 200 bp) across the two virulent/avirulent genomes. The cn.MOPS program was run using mean normalization mode and the default values for all other parameters.

CNVs identified by LUMPY that had at least 10% overlap with a CNV identified by cn.MOPS were retained for further analyses. Then, duplicated/deleted CNVs were further validated by visual inspection of the reading alignment (using CNVnator –view). Genes that overlapped by at least 70% of their length with the identified CNVs were extracted. The function of genes identified in these CNVs was approached by searching for conserved domains (*Interproscan*/*pfam*) and searching for the molecular functions, cellular locations, and processes gene products may carry out based on the GeneOntology (GO annotations) Finally, CNVs leading to gene deletion in the virulent pathotype were confirmed by PCR. List of primers and PCR conditions are provided in Table S5. Genes encoding putative effectors were detected by checking their presence in the reference secretome published by Petitot et al. (2015) using Blastn. In addition, for all genes with CNVs, we screened in their encoded proteins for the presence of secreted protein signal peptide (Sec/SPI) and the absence of transmembrane domain using SignalP-5.0 (Petersen et al., 2011).

III. Results

Sequencing of avirulent and virulent pathotypes generated 19.8 million and 16.0 million pair-end reads (2 x 150 bp), respectively. This corresponds to a sequencing depth of 72× and 58× compared to the reference haploid genome (PRJNA615787; 41.5 Mb). After cutting adapters and removing low-quality reads, the remaining cleaned pair-end reads for avirulent and virulent genomes had coverage of 71.6× and 57.7×, respectively (Table S1)

3.1 Patterns of heterozygosity loss on genomic blocks

Read coverage analysis on the 36 manually curated DNA regions (354 kb) revealed that haplotypes GB2_T2 and G296_T1 were absent from the virulent genome (Figure 1). In contrast, sequencing coverage of their homologs (i.e. GB2_T1 and G296_T2, respectively) was double, suggesting they were fixed (i.e. homozygous; Figure 1). Such a loss of heterozygosity (LoH) of a single haplotype was already observed among 12 additional isolates (Ngan et al. in prep.). In particular, GB2_T2 was also fixed in nine isolates (Figure S1A). On the other hand, G296_T2

was fixed only in the virulent isolate, while its homolog (G296_T1) was fixed in Mg-VN27 and Mg-C25. Two genes were identified in the G296 contig, one coding for a chemoreceptor (Srw) and the other for a COPII transport-vesicle coat protein (Sec23/Sec24).

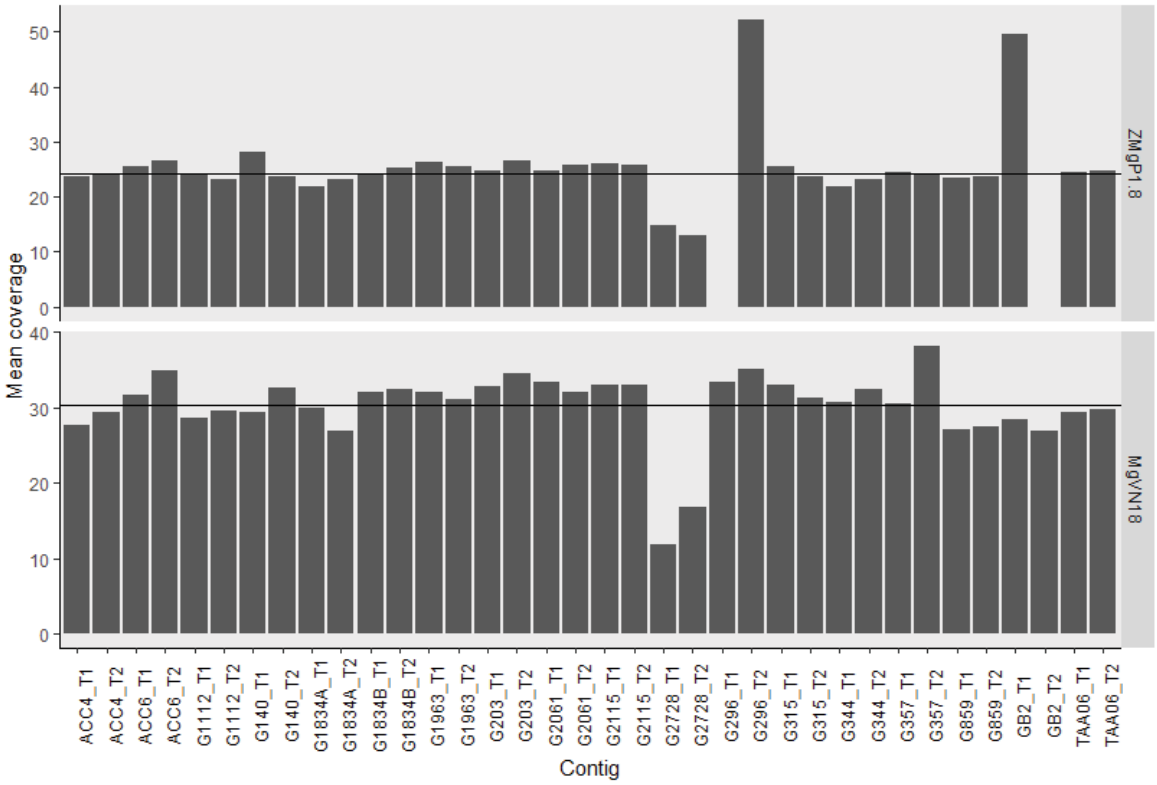


Figure 1. Average reading coverage for each haplotype in the two pathotypes on 36 genomic fragments (homozygous state) reconstituted from 18 nuclear contigs of the reference genome (heterozygous state). *The black horizontal line indicates the mean value of the sequencing coverage of all fragments in each isolate (Mg-VN18 and ZMg.P1.8). Missing columns (read-coverage = 0×) indicates the absence of one haplotype. T1 = haplotype 1, T2 = haplotype 2.*

Nucleotide variant calling detected 36,642 polymorphic sites between the two virulent and avirulent isolates. Visualization of their positions on scaffolds revealed a variable genome structure between these two isolates (Table 1, Figure 2). Twenty-two genomic regions (> 3 kb) covering ~2.9% of the genome (1.2 Mb) that harbor two haplotypes in Mg-VN18, appeared in a homozygous state in the virulent isolate (Table 1, Figure 2). Analysis of the 1-kb window sequencing coverage along these genomic regions shows a relatively homogeneous coverage (~60×) for 16 of these genomic regions (~1.15 Mb, ~2.21% of the genome) in ZMgP1.8 (Figure 2), indicating the fixation of one haplotype in the virulent genome (hereafter 'homozygous LoH'). In contrast, six others regions on scaffolds mg16, mg18, mg19, mg71, mg163, and mg255 (total length of 0.05 Mb, ~ 0.08% genome) showed half-read coverage (~30×)

suggesting hemizygous regions (hereafter 'hemizygous LoH') in the virulent genome (Figure 2). Nearly 50% (609 kb) of homozygous LoH regions that were present on eight scaffolds (mg78, mg87, mg105, mg141, mg145, mg157, mg166, and mg173) presented also a homozygous pattern in other isolates (Table 1). The remaining eight homozygous LoH regions (547 kb) were specific to the virulent genome (Table 1). A total of 112 genes and 21 genes located respectively on these homologous LoH and hemizygous LoH regions were found (Tables S2 and S3). Among them, 14 genes (~10%) encode a putative secreted protein. No enrichment in GO annotation was found for these regions affected by LoH (Figure S1).

Table 1. List of genomic regions that have experienced a loss of heterozygosity (LoH) in the virulent genome compare to other 13 avirulent isolates.

No.	Scaffold	Start (bp)	Stop (bp)	Length (bp)	Note
1	mg3	299433	308424	8991	Homozygous LoH regions specific to virulent genome
2	mg6	280300	283359	3059	
3	mg7	898	177575	176677	
4	mg42	202	296323	296121	
5	mg57	204518	207799	3281	
6	mg127	185	50356	50171	
7	mg268	2210	7497	5287	
8	mg274	598	4029	3431	
9	mg78	0	184797	184797	LoH found in Mg-VN6, Mg-L1, Mg-Java2, Mg-Bali, Mg-Borneo, Mg-VN27, Mg-C25, Mg-C21, and Mg-P
10	mg87	0	153709	153709	
11	mg145	19667	64849	45182	
12	mg157	6105	60561	54456	
13	mg173	19587	34093	14506	
14	mg141	1320	44120	42800	LoH found in Mg-VN6, Mg-L1, Mg-Java2
15	mg105	7906	117956	110050	LoH found in Mg-C25 and Mg-VN27
16	mg166	16336	23292	6956	
17	mg16	192654	203138	10484	Hemizygous LoH regions specific to virulent genome
18	mg18	182017	204006	21989	
19	mg19	52077	57417	5340	
20	mg71	3064	9532	6468	
21	mg163	3279	8574	5295	
22	mg255	110	5662	5552	

A similar phenomenon of heterozygosity depletion was observed in Mg-VN18: homologous LoH on scaffolds mg20 and mg39, and hemizygous LoH on scaffolds mg20, mg36, mg41, mg70, mg153, and mg198. These regions, which represent 1.13% of the genome, were already identified by Phan et al. (in prep.) and confirmed in this study (Figure 2)

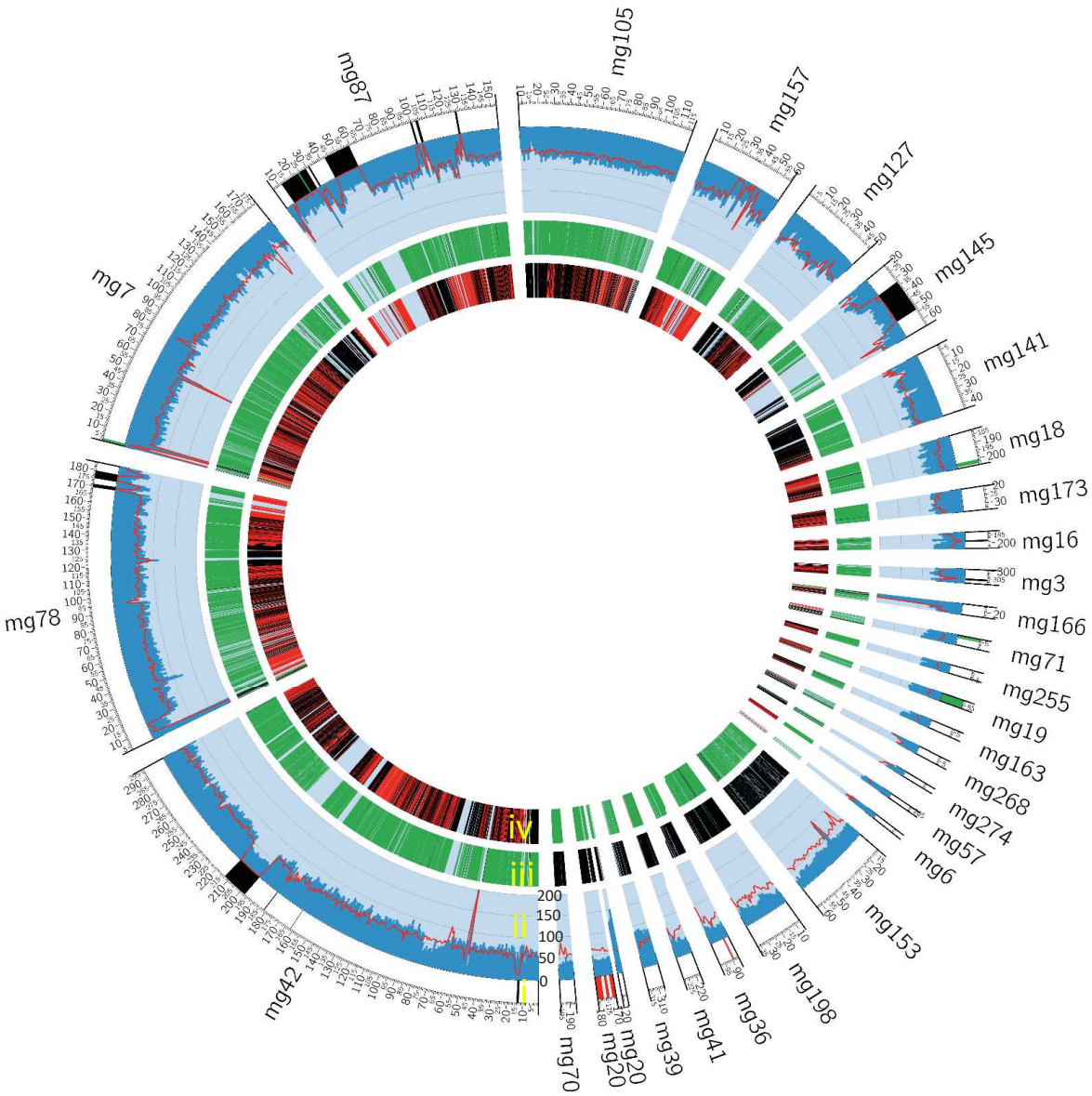


Figure 2. Graphical comparison of genome characteristics (LoH, CNV) and sequencing depth of 29 scaffolds for avirulent (Mg-VN18) and virulent (ZMgP1.8) pathotypes. *The size of the genomic regions were drawn to scale and sorted by length in a clockwise direction from longest to shortest. Circle shows three layers: i) Copy number variation regions (CNVR) with following coding colors: black = deleted regions, green = copy loss regions, and red = copy gain regions; ii) short-read depth per 1-kb sliding window along the scaffold for the avirulent Mg-VN18 (histogram) and virulent ZMgP1.8 (red line) genomes; iii) and iv) Variants of loci observed along scaffolds in the genome of Mg-VN18 (layer iii) and ZMgP1.8 (layer iv), respectively. For layers iii and iv, red = homozygous variants (genotype 1/1); green = heterozygous variants (genotype 0/1); black = homozygous reference (genotype 0/0). The total size of these 29 scaffolds for the reference genome Mg-VN18 corresponds to 1.35 Mb.*

3.2 Few single nucleotide polymorphisms and no new gene detected in the genome of the virulent isolate

Comparison of the virulent mitogenome with the avirulent mitogenome excluding the minisatellite 111R region (Besnard et al., 2014) revealed one non-synonymous SNP located in *nad5*. Two heteroplasmic sites were also detected: one in *nad4* and the other on the ribosomal RNA gene *rrnL* (Table S4). These three variants have never been reported in other *Mg* mitogenomes (Besnard et al., 2019) and may thus be specific to the virulent isolate. At the nuclear level, among the 36 regions analyzed, only one synonymous SNP was found on contig G315_T2. Based on these observations, the mutation rate of the mitogenome (~17 kb without the 111R region) is estimated to be about 60 times higher than that observed on the nuclear genome (~354 kb for the 36 DNA fragments).

At the whole genome level, we first considered variants from homozygous regions (genotype "0/0") in the reference genome (Mg-VN18) in order to get rid of the complexity related to *Mg* heterozygosity. A total of 42 SNPs and 19 InDels were detected. No InDels but 29 SNPs were found and located in 20 genes (Table 2). Of these 29 SNPs, ten were non-synonymous and located in two genes coding for galectin (two genes), one gene for Spliceosome-Associated Protein (SAP), and four genes encoding protein of unknown function (Table 2). Among them, the two genes encoding galectin have peptid signal Sec/PSI indicating putative secreted effectors (Table 2).

As our reference may not contain all the genomic information among *Mg* isolates, we then assembled new contigs for the virulent genome. We thus obtained 1,611 additional contigs (size from 200 to 7,218 bp) using pair-end reads that unmapped to the reference genome. However, the average read coverage of all contigs was low (inferior to 7×) indicating those contigs are likely artifacts (or contaminants). Besides, none of them showed significant similarity to gene of species belonging to the nematoda phylum. Therefore, no new gene was detected in the virulent genome.

Table 2. List of genes presenting SNPs in the virulent genome.

Scaffolds	SNP position	Gene name	Start gene	Stop gene	Putative function base on *InterProscan* domain	intron/(non) synonymous
mg4	730498	Mgra_00001016	728470	735008	Galectin	intron
mg21	243955	Mgra_00003271	243941	244896	Uncharacterized protein	intron
mg32	116748	Mgra_00004345	116617	116891	Unknown function	nonsynonymous
mg37	303011	**Mgra_00004790**[12]	300969	303886	Uncharacterized protein	intron
mg59	213793	Mgra_00006242	212306	226842	Sec63 domain;DEAD/DEAH box helicase domain	synonymous
mg67	76106	Mgra_00006603	67898	76361	SAP domain-containing protein	nonsynonymous
mg78	184246	**Mgra_00007212**[2]	179449	184528	Galectin	nonsynonymous
mg78	184254					nonsynonymous
mg78	184339					intron
mg78	33345	Mgra_00007180	20267	34811	Ribosomal protein L7/L12, oligomerisation/ C-terminal	intron
mg84	48200	Mgra_00007431	45973	49694	Galactose oxidase, central domain	intron
mg87	9294	**Mgra_00007533**[2]	7356	12037	Unknown function	intron
mg87	34498	Mgra_00007536	26637	36720	Unknown function	intron
mg87	40836	Mgra_00007538	37879	40840	Unknown function	nonsynonymous
mg87	99410	**Mgra_00007554**[2]	97939	108594	Galectin	nonsynonymous
mg87	99484					nonsynonymous
mg87	100732					intron
mg100	83510	**Mgra_00008027**[1]	79678	91822	Uncharacterized protein	intron
mg116	56592	**Mgra_00008575**[1]	55666	61957	Uncharacterized protein	intron
mg143	68207	Mgra_00009205	67836	70011	Uncharacterized protein	nonsynonymous
mg145	32294	**Mgra_00009230**[2]	23636	35602	Unknown function	intron
mg145	34204					intron
mg145	35111					synonymous
mg145	35501					nonsynonymous
mg145	35538					nonsynonymous
mg157	39900	Mgra_00009411	39485	40323	Unknown function	intron
mg157	39912					intron
mg209	18753	Mgra_00009997	16718	20376	Histidine phosphatase superfamily, clade-1	intron
mg250	13882	**Mgra_00010225**[2]	11598	14587	ser1: Chymotrypsin-like serine proteinase	intron

[1] *Genes were annotated as putative effectors by Petitot et al. 2015;* [2]*Genes encoding proteins with a secreted standard Sec/SPI signal peptide*

3.3 Detection of copy number variations in the virulent genome

A total of 1,520 CNVs were detected in the virulent genome using a combination of CNVnator and Lumpy tools. Among them, there were 495 deletions, 101 duplications, 14 inversions, and 910 break-ends. The use of cn.MOPs then detected 135 deletions and 29 duplications. Finally, 35 deletions (for a total of 130 kb) and 10 duplications (30 kb) were confirmed using both CNVnator and Lumpy tools. By a visual inspection of the reading alignment of each CNV, we were able to confirm the presence of 31 deletions (127 kb) and six duplications (10 kb) (Figure 3, Table 3). Of these deletions, 13 regions involved partial losses with one or two copies absent in the virulent genome compared to the reference. Deletions in 18 other regions of the virulent genome resulted in the complete loss of genomic blocks (Table 3, Figure 3).

Forty-five genes were found in these 31 deletions and 6 duplications (Table 4). Among them, 29 genes were found only in the avirulent genome. Most of the lost genes (24/29) were located in regions of LoH (Figure 1). A polymerase chain reaction (PCR) was performed on four of these genes to confirm their absence in the virulent genome (Figure S2). Two of these four genes code for proteins containing the galectin domain, one of them having a standard Sec/SPI secretion signal peptide (Mgra_00007558; Figure S2). .

Table 3. Genomic regions carrying copy number variations (CNVs) between virulent (ZMgP1.8) and avirulent (Mg-VN18) genomes.

CNV type	Scaffolds	CNV start (bp)	CNV end (bp)	CNV length (bp)	Mg-VN18 (# copy)	ZMgP1.8 (# copy)
Deleted regions	mg3	305400	306400	1000	CN2*	CN0
	mg16	197997	198817	820	CN2	CN0
	mg42	12600	13800	1200	CN2	CN0
	mg42	161522	161656	134	CN2	CN0
	mg42	181404	181769	365	CN2	CN0
	mg42	196800	213200	16400	CN2	CN0
	mg78	169200	171200	2000	CN2	CN0
	mg78	174800	179400	4600	CN2	CN0
	mg80	150000	158000	8000	CN2	CN0
	mg87	3514	5644	2130	CN2	CN0
	mg87	11800	29200	17400	CN2	CN0
	mg87	31200	32000	800	CN2	CN0
	mg87	42800	59400	16600	CN2	CN0
	mg87	99600	100600	1000	CN2	CN0
	mg87	103000	104400	1400	CN2	CN0
	mg87	127800	128800	1000	CN2	CN0
	mg145	32400	53000	20600	CN2	CN0
	mg194	0	7200	7200	CN2	CN0
Copy loss regions	mg1	506000	506800	801	CN2	CN1
	mg6	2000	2600	601	CN3	CN2
	mg18	200800	203000	2201	CN2	CN1
	mg19	51800	60000	8201	CN2	CN1
	mg20	105600	106800	1201	CN3	CN2
	mg20	112000	112800	801	CN3	CN2
	mg20	114400	116600	2201	CN2	CN1
	mg20	126400	127000	601	CN3	CN2
	mg20	129200	133400	4201	CN3	CN2
	mg20	135200	136000	801	CN2	CN1
	mg71	3600	4200	601	CN2	CN1
	mg71	4400	5000	601	CN2	CN1
	mg6	0	1200	1201	CN3	CN1
Copy gain regions	mg19	184000	184800	801	CN1	CN2
	mg20	170000	172000	2001	CN2	CN3
	mg20	431600	432800	1201	CN1	CN2
	mg36	89000	89800	801	CN2	CN3
	mg20	173800	175400	1601	CN1	CN3
	mg20	177400	181000	3601	CN1	CN3

**CN2 indicated normal diploid state; CN0 and CN1 indicated loss of copy compare to normal diploid state; CN3 indicated gain of copy number compare to the normal diploid state*

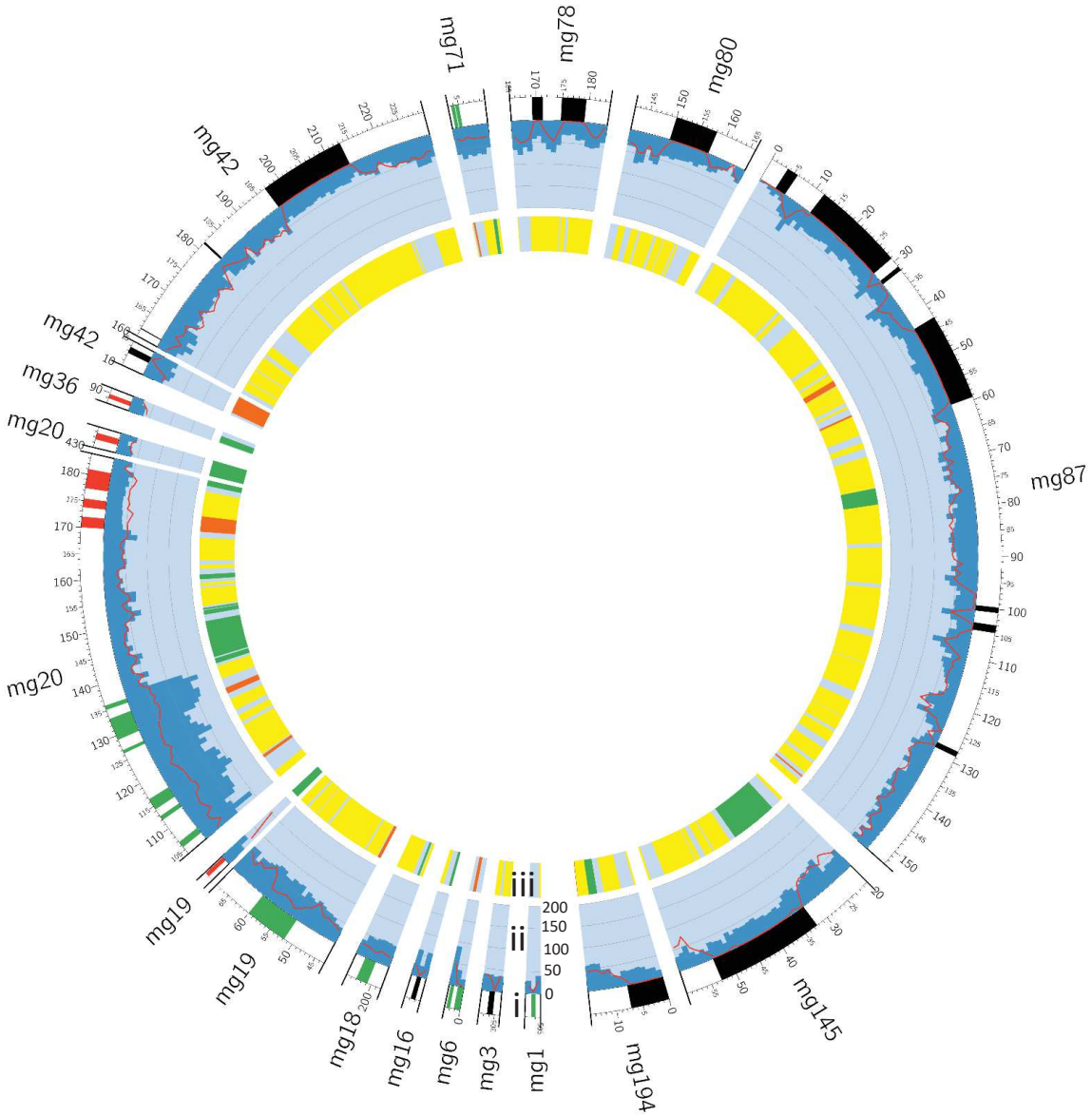


Figure 3. Visualization of genomic regions carrying copy number variations (CNVs) between virulent (ZMgP1.8) and avirulent (Mg-VN18) genomes. *The position of the genomic regions on scaffolding of the reference genome Mg-VN18 (accession number PRJNA615787) was indicated and the size of regions was drawn to scale. Circle shows three layers: i) Copy number variation regions (CNVR) found in virulent genome with following coding colors: black = deleted regions, green = copy loss regions, and red = copy gain regions; ii) short reading depth per 1-kb sliding window along the scaffold for the avirulent Mg-VN18 (histogram) and the virulent ZMgP1.8 (red line) genomes; iii) Genes distribution on genomic regions, each gene was displayed by a color representing its GC content: green = GC < 20%; yellow = 20 ≤ GC < 30; orange = 30 ≤ GC < 40.*

Among the 29 genes lost in the virulent genome, two genes have been previously predicted to code for putative effectors (Petitot et al., 2015), including an M12 protease (Mgra_00005142)

and a protein with unknown function (Mgra_00005101). In addition, two genes coding for a galectin (Mgra_00007558) and a protein with unknown function (Mgra_00005132) exhibited a standard Sec/SPI secretion signal peptide which is one of the criteria for the identification of nematode effectors. Other genes absent from the virulent genome would code for proteins with diverse functions such as ubiquitin, zinc ion binding, collagen, regulation of apoptotic process (Fas apoptotic inhibitory and LAMTOR5 ragulator), transporters (cation efflux protein, organic anion transporter, and amino acid transporter), chromosome assembly and modification (chromatin assembly factor 1 and jumping translocation breakpoint), biosynthesis (carbohydrate kinase PfkB and Isopropylmalate dehydrogenase), and unknown function. Nine genes, which appeared to undergo the loss of some of their copies, are expected to be involved in molecular functions such as chromatin remodeling, protein kinase activity, transmembrane receptor, DNA replication, recombination and repair, and other unknown functions. At the same time, seven genes related in nucleotidyltransferases activity, GTPase activity, DNA strand breakage and rejoining, and protein degradation have more copies in the virulent genome than in the avirulent Mg-VN18 genome (Table 4). Seven genes coding for a polymerase, a lon protease, a neurotransmitter, a GTPase, a DNA topoisomerase, and a proteasomal ubiquitin receptor appeared to have more copies in the virulent genome than in avirulent genome (Table 4). Four genes were also affected by CNV in six avirulent isolates (Mg-C25, Mg-C21, Mg-P, Mg-VN27, Mg-Java2 and Mg-Bali; Figure 4, Table 4). Of these, one gene would code for a protein-tyrosine phosphatase and the other three genes had unknown function.

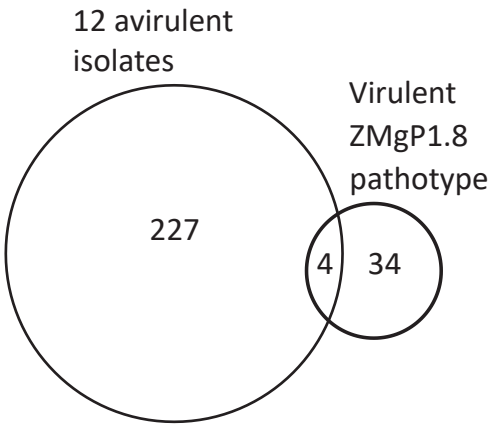


Figure 4. Venn diagram illustrating the distribution of gene (copy) loss detected between the virulent isolate and the 12 avirulent isolates.

Table 4. Genes affected by CNVs (gene loss, copy loss and copy gain) in the virulent genome ZMgP1.8

Type	Gene name	Gene position	Gene length (bp)	*InterProscan* domain/family	#genes in family	Putative function
Gene loss	Mgra_00000700	mg3:304839-306097	1258	IPR000626: Ubiquitin-like domain IPR001841: Zinc finger, RING-type	16	Ubiquitination-proteasome
	Mgra_00005142[1]	mg42:198638-218317	19679	IPR001506: Peptidase M12A IPR010695: Fas apoptotic inhibitory molecule 1	28	
	Mgra_00007541	mg87:44941-45622	681	IPR024766: Zinc finger, RING-H2-type	2	
	Mgra_00007535[2]	mg87:21571-22792	1221	IPR007482: Protein-tyrosine phosphatase-like, PTPLA	3	Role in the development, differentiation, and maintenance of a number of tissue types
	Mgra_00007543	mg87:48046-49608	1562	IPR007482: Protein-tyrosine phosphatase-like, PTPLA		
	Mgra_00009864	mg194:3929-8301	4372	IPR008160: Collagen triple helix repeat	62	Cuticule component
	Mgra_00007558[3*]	mg87:127454-131899	4445	Galectin domain-containing protein	29	Potentially manipulates host immunity
	Mgra_00009234[3]	mg145:38181-41876	3695	Galectin domain-containing protein		
	Mgra_00009236	mg145:45853-47004	1151	IPR024135: Ragulator complex protein LAMTOR5	2	Apoptotic cell death regulation
	Mgra_00007546[3]	mg87:53570-58614	5044	IPR004156: Organic anion transporter polypeptide	5	Transporter
	Mgra_00009863	mg194:2-378	376	IPR013057: Amino acid transporter, transmembrane domain	12	
	Mgra_00007302	mg80:149904-153273	3369	IPR002524, IPR027470: Cation efflux protein	6	
	Mgra_00007209	mg78:168218-175579	7361	IPR022043: Chromatin assembly factor 1 subunit A	1	Chromosome assembly and modification

Type	Gene name	Gene position	Gene length (bp)	*InterProscan* domain/family	#genes in family	Putative function
Gene loss	Mgra_00007539	mg87:41458-44685	3227	IPR008657: Jumping translocation breakpoint	1	
	Mgra_00007542	mg87:45753-47889	2136	IPR024084: Isopropylmalate dehydrogenase-like domain	4	Biosynthesis of leucin
	Mgra_00007211	mg78:177941-180487	2546	IPR011611: Carbohydrate kinase PfkB	5	Biosynthesis of thiamine pyrophosphate
	Mgra_00005101[1]	mg42:10686-16751	6065		2	Unknown function
	Mgra_00005136	mg42:179584-182486	2902		3	
	Mgra_00007210	mg78:175954-177128	1174		2	
	Mgra_00007304	mg80:156211-158863	2652		2	
	Mgra_00007540	mg87:43299-44732	1433		3	
	Mgra_00007536	mg87:97939-108594	10655		4	
	Mgra_00005132*	mg42:161088-163244	2156		1	
	Mgra_00007532[2]	mg87:809-5962	5153		3	
	Mgra_00007534[2]	mg87:11649-20733	9084		3	
	Mgra_00007545	mg87:53104-53685	581		6	
	Mgra_00007544	mg87:50632-51242	610		4	
	Mgra_00009235[3]	mg145:42310-44365	2055		5	

Type	Gene name	Gene position	Gene length (bp)	*InterProscan* domain/family	#genes in family	Putative function
Gene copy loss	Mgra_00002998	mg19:59056-62463	3407	IPR001005,IPR007526,IPR032448,IPR032451: SANT/Myb; SWIRM; SMARCC	7	Nuclear receptor co-repressors/chromatin-remodeling
	Mgra_00003123	mg20:111424-120002	8578	IPR001245: Serine-threonine/tyrosine-protein kinase IPR006085,IPR006086: DNA endonuclease (XPG-I) IPR020683: Ankyrin repeat-containing domain	42	Protein kinase activity
	Mgra_00003126	mg20:124297-127348	3051	IPR006029: Neurotransmitter-gated ion-channel transmembrane domain IPR006202: Neurotransmitter-gated ion-channel ligand-binding domain	26	Transmembrane receptor
	Mgra_00003127	mg20:128952-130420	1468	IPR000424: Primosome PriB/single-strand DNA-binding	1	DNA replication, recombination and repair.
	Mgra_00002891[1]	mg18:200633-205110	4477	UDENN domain-containing protein	5	Unknown function
	Mgra_00002997	mg19:50113-58376	8263	IPR019409,IPR019441,IPR019443: FMP27 (DUF2405; GFWDK; C-terminal)	1	
	Mgra_00001182	mg6:1371-2095	724		-	
	Mgra_00003128	mg20:132957-136323	3366		-	
	Mgra_00006814[2]	mg71:3443-3760	317		-	
	Mgra_00003020	mg19:182372-187216	4844	IPR002934: Polymerase, nucleotidyl transferase domain IPR007012: Poly(A) polymerase, central domain	5	Nucleotidyl transferases activity

Type	Gene name	Gene position	Gene length (bp)	*InterProscan* domain/family	#genes in family	Putative function
Gene copy gain	Mgra_00003141	mg20:170857-174769	3912	IPR003111: Lon protease	2	Protein degradation
	Mgra_00003204	mg20:424558-434994	10436	IPR006029: Neurotransmitter-gated ion-channel transmembrane domain IPR006202: Neurotransmitter-gated ion-channel ligand-binding domain	26	Transmembrane receptor
	Mgra_00004630	mg36:88657-89818	1161	IPR001806: Small GTPase	42	GTPase activity
	Mgra_00003142	mg20:174777-178615	3838	IPR006171,IPR013497: DNA topoisomerase	3	DNA strand breakage and rejoining
	Mgra_00003143	mg20:178458-181071	2613	IPR006773: Proteasomal ubiquitin receptor Rpn13/ADRM1 IPR032368: UCH-binding domain	1	Protein-degradation
	Mgra_00004629	mg36:88215-89821	1606		-	Unknown function

[1] Genes were annotated as putative effectors by Petitot et al. 2015; [2] The loss of these genes was also detected in six isolates (Mg-C25, Mg-C21, Mg-P, Mg-VN27, Mg-Java2 and Mg-Bali); [3] deletion of genes were confirmed by Polymerase chain reaction (PCR); *Genes encoding proteins with a secreted standard Sec/SPI signal peptide.

IV. Discussion

In this study, we used a comparative genomics approach to study the genomic signatures distinguishing a virulent pathotype from an avirulent pathotype in the nematode *Meloidogyne graminicola*. The aim of this study was not to identify the avirulence factor(s) but to determine whether the adaptation of this species, which reproduces by facultative meiotic parthenogenesis, is accompanied by significant modifications in the genome. In this, and while little variability in SNPs was observed between isolates of *M. graminicola* from distant countries (Besnard et al., 2019), we found that profound modifications such as loss of heterozygosity (LoH) on long genomic blocks and copy number variants (CNVs) distinguish the two pathotypes. Interestingly, these modifications affect important genes, including effectors whose role in host-pathogen interaction has yet to be determined.

4.1 Gene loss affect potential effectors

Comparative genomic analysis revealed genomic features that could potentially explain the acquisition of virulence in *M. graminicola* in response to a resistant rice cultivar, including loss of heterozygosity (LoH) on long genomic blocks, a few single nucleotide variants (SNV), and copy number variations (CNVs). CNVs have been identified as an important mechanism of genomic variation that may be involved in the evolution of many plant and animal organisms (Clop et al., 2012; Zhang et al., 2009; Żmieńko et al., 2014). CNVs appear as a result of duplication/deletion events, are heritable traits and can contribute to the adaptation of species to their environments (Kondrashov, 2012). For example, CNVs have been shown to cause insecticide resistance in the *Aedes aegypti* mosquito (Faucon et al., 2015) and copper resistance in yeast (Hull et al., 2017).

Genomic studies in the nematode *Caenorhabditis elegans*, revealed that the rate of CNV per gene and per generation is twice as high as the point mutation rate observed per nucleotide within a gene (Lipinski et al., 2011). It is interesting to note that CNVs produce significant changes in the *Meloidogyne incognita* genome when experimenting with virulent pathotypes that can multiply on resistant tomato plants. These genetic variations could reflect an adaptive response of the parthenogenetic mitotic nematode to host resistance (Castagnone-Sereno et al., 2019). Furthermore, the loss of one copy of the small *Cg-1* gene family in a virulent pathotype of *Meloidogyne javanica* was shown to be associated with its adaptability against the *Mi-1* tomato resistance gene (Gleason et al., 2008). The abnormally enriched proportion of genes coding for putative effectors that are targeted by these CNVs suggests that the deletion of certain

effectors could participate in the evolution of pathotypes to become virulent. In our study, the proportion of genes coding for putative effectors of *M. graminicola* was about 7% of all genes affected by CNVs. This percentage is indeed significantly higher than the average percentage of putative effectors detected in *M. graminicola* (~2.8%) (Petitot et al., 2015). Although our study does not demonstrate that genes that have been affected by these profound rearrangements are the key to adaptation to a resistant plant, it is particularly intriguing to note that many of the genes targeted by CNVs involve effectors. Among these potential effectors, some could play important functions in the host-pathogen interaction.

The virulent genome showed a loss of two genes of among 29 genes of the galectin family as well as a non-synonymous substitution on two other genes. This suggests that the galectin family was particularly affected by genomic rearrangements. Galectins were not detected as potential effectors in the analysis of Petitot et al. (2015), however, one of the galectins identified in our analysis has a peptide signal suggesting that it is potentially secreted. Galectins present in both the host and the pathogen play an important role in host-parasite interaction (Young & Meeusen, 2002). Host-produced galectins can act as pattern recognition receptors (PRRs) detecting conserved pathogen-associated molecular patterns (PAMPs), while parasite-produced galectins can modulate host immune responses (Ayona et al., 2020; Shi et al., 2018). For example, in the animal-parasitic nematode *Haemonchus contortus*, the galectin produced by the nematode alters the host's cellular activities, including cytokine production, to evade host immunity (Lu et al., 2017). In addition, galectins from the excretory/secretory products of parasitic helminths have been identified as mediators of the host immune response (Young & Meeusen, 2002). Interestingly, in the plant-parasitic nematode *M. incognita* and *Globodera rostochiensis*, galectin was previously described as putative secreted protein from the oesophageal secretory gland region (Dubreuil et al., 2007; Price et al., 2020). Galectin from *G. rostochiensis* has also been identified at the surface layer of the nematode, where it may interact with degradation products of the host cell wall to prevent detection of PAMP by the host (Price et al., 2020).

Effectors involved in the ubiquitination-proteasome pathway secreted by plant parasitic nematodes are thought to interfere with the signaling and regulatory process of plant cells, thereby contributing to the attenuation of host defenses to promote parasitism (Davis et al., 2004). Three genes belong to ubiquitination-proteasome pathways, including ubiquitin-like, Ring-H2-type Zinc finger, and peptidase M12A are missing in the virulent genome. This observation is interesting considering that the absence of genes coding for peptidase has also

been observed in the *M. incognita* virulent genome (Castagnone-Sereno et al., 2019). Putative secreted protein in ubiquitination-proteasome pathway could be involved in the disruption of SA-mediated defense signaling (Haegeman et al., 2012).

Other genes, which were not previously recognized as effectors (Petitot et al. 2015) and which do not present a secretion signal peptide, have also been the target of rearrangements in the virulent genome. However, these genes could also be important for the acquisition of virulent pathotypes. Those genes are involved in growth process, regulation process, transporter, chromosome assembly and modification, and biosynthesis. Notably, two genes of three protein-tyrosine phosphatase-like (PTPLA) coding genes annotated in the *M. graminicola* genome were missing in the virulent genome. PTPLA proteins are thought to play a role in the development, differentiation, and maintenance of a number of tissue types (Uwanogho et al., 1999). The genes coding for collagen molecules with triple helix repetition belong to a large family of genes in RKNs, as for example with the 176 genes in *Meloidogyne hapla* (Opperman et al., 2008) and 61 genes in *M. graminicola* (Phan et al., 2020). Collagen forms an important part of the cuticle of the nematode in *M. incognita* and participates in the maintenance of the general shape of the animal and its protection against the external environment (Banerjee et al., 2017). The genes coding for collagen are down-regulated during the incompatible interaction between *M. graminicola* and *O. glaberrima* (Petitot et al., 2020), and *Pratylenchus penetrans* and a resistant alfalfa cultivar (Vieira et al., 2019). In this study, a gene coding for collagen was absent in the virulent isolate. This suggests that a decrease in the amount of collagen present on the nematode would be accompanied by the process of adaptation to the interaction with a resistant plant. Other processes are also potentially affected with the loss of one copy of a gene. For example, a gene coding for a predicted effector serine-threonin/tyrosine protein kinase has suffered copy loss in the virulent genome. Serine-threonine kinases (STK) play a role in a multitude of cellular processes, including division, proliferation, apoptosis, and differentiation (Manning et al., 2002; McCubrey et al., 2000). STK is an important factor in host-pathogen interactions. In the plant defense pathway, STK relays signal from the receptor containing leucine-rich repeat (LRR) to activate the MAPK pathway (Afzal et al., 2008). On the other hand, STKs have already demonstrated a role as a pathogenicity factor in bacteria (Canova & Molle, 2014; Theeya et al., 2015) and in the eukaryotic pathogen *Toxopmasma gondi* (Taylor et al., 2006).

Finally, a few genes involved in the epigenetic functions of the nematode have been affected by CNVs. Epigenetic mechanisms are thought to be involved in transcriptional

reprogramming observed in RKN-plant interactions (Ji et al., 2013; Portillo et al., 2013). Indeed, genes involved in chromatin remodeling, DNA methylation, small RNA formation and histone modifications are all highly expressed within *Mg*-induced giant cells in rice (Ji et al., 2013), as well as in *M. incognita*-induced giant cells in tomato (Portillo et al., 2013). In the virulent pathotype, the genes encoding a chromatin assembly factor and a jumping translocation breakpoint were absent. Similarly, a gene related in chromatin-remodeling suffered copy loss in the virulent genome. In addition, a non-synonymous substitution was detected in a gene coding for a spliceosome-associated protein (SAP) in the virulent genome.

4.2 Genes located on LoH regions could be associated with the adaptation of the virulence pathotype to the resistant plant

Loss of heterozygosity occurs over time in most organisms that reproduce by meiotic parthenogenesis (Jaron et al., 2020). LoH are the consequence of homology-based DSB repair during homologous recombination and gene conversion (Engelstädter, 2017). It has been shown that LoH affects fitness and leads to the adaptation of hybrid yeast to its environment (Heil et al., 2017). The LoH, also known as "Loss of complementation—LOC hypothesis" (Archetti, 2004), associated with recombination processes during asexual reproduction leads to the unmasking of recessive mutations which have little or no effect in the heterozygous state but which would be necessary to allow the species to adapt to its environment (Dukić et al., 2019).

In a previous study of 13 avirulent *M. graminicola* isolates, LoH was detected and affect 0 to 7.29% of their genome (Phan et al. in preparation). Comparison of the LoH regions identified in the virulent genome with the 13 avirulent pathotypes reveals that these isolates share some LoH but without a clear structure in their distribution among the pathotypes. Nevertheless, during the adaptive process, the virulent isolate has accumulated specific LoH regions. Indeed, about ~10% of the genes identified in the LoH regions were presumptive effectors, which is nearly five times higher than the average number of effectors annotated in the transcriptome (according to Petitot et al. 2015). For example, among the genes coding for putative effectors affected by homozygous LoH, 12 encode proteins involved in detoxification of ROS production (carboxylesterase, UDP glycosyltransferases), signaling pathways (protein kinase, kinase-like receptor), protein degradation (Peptidase C1A, Peptidase M12A, Ubiquitin), chemosensory (G protein-coupled receptor, serpentine receptor), nematode neurobiology (CUB domain, Neurotransmitter), and transcription factor (Myc).

Some genes located in the hemizygous LoH regions of the virulent genome are also potentially crucial for nematode parasitism. For example, cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1 (CAP) superfamily are known, or proposed, to play roles in parasite development and reproduction, and in modulating host immune attack and infection processes (Stroehlein et al., 2016). Among these genes, two code for venom allergen-like proteins (VAP), belonging to the CAP superfamily, that are induced in *M. graminicola* during the incompatible interaction with *O. glaberrima* (Petitot et al., 2020). It also contains genes coding for acyltransferase, which could be involved in the detoxification process (Chen, 2020) .

V. Conclusions and perspectives

To understand how the virulent strain *M. graminicola* ZMgP1.8 could bypass resistance in the rice cultivar 'Zhonghua 11', we performed a whole-genome comparative analysis between the virulent ZMgP1.8 and 12 other avirulent pathotypes. The results pointed out specific signatures in the genome of ZMgP1.8, including CNVs, LoH, and SNVs. Loss of heterozygosity (LoH), which could be a powerful mechanism in genomic evolution, is a common feature of parthenogenetic animals. The abundance of genes loss detected in the ZMgP1.8 genome supports the idea that gene losses could be a mechanism of genetic adaptation in response to a resistant host in plant parasitic nematodes. It is particularly intriguing to observe the abundance of CNVs and the rarity of SNPs in an organism that reproduces by parthenogenesis and that these genomic modifications seem to lead to the adaptation of the organism to its environment. Unfortunately, the role of *M. graminicola*'s mode of reproduction (facultative parthenogenetic meiosis) in this accumulation of CNVs cannot be addressed in this study. However, the genus *Meloidogyne* offers a genetic material of choice since all modes of reproduction are observed, from amphimixis to mitotic parthenogenesis, and abundant CNVs have also been observed in the mitotic parthenogenetic *M. incognita* to bypass plant resistance mechanism. The higher proportion of putative effectors affected by CNVs, LoH, and SNVs relative to the whole genome suggests that these modifications could contribute to the acquisition of virulence in ZMgP1.8. Hence, it roses the hypothesis that one of these lost/modified genes in ZMgP1.8 could potentially be the avirulent factor which is recognized in the 'Zhonghua 11' variety by a resistance gene and whose activation triggers the hypersensitive response.

References

Abyzov, A., Urban, A. E., Snyder, M., & Gerstein, M. (2011). CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research*, *21*(6), 974–984. https://doi.org/10.1101/gr.114876.110

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Afzal, A. J., Wood, A. J., & Lightfoot, D. A. (2008). Plant receptor-like serine threonine kinases: Roles in signaling and plant defense. *Molecular Plant-Microbe Interactions*, *21*(5), 507–517. https://doi.org/10.1094/MPMI-21-5-0507

Archetti, M. (2004). Recombination and loss of complementation: A more than two-fold cost for parthenogenesis. *Journal of Evolutionary Biology*, *17*(5), 1084–1097. https://doi.org/10.1111/j.1420-9101.2004.00745.x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Ayona, D., Fournier, P.-E., Henrissat, B., & Desnues, B. (2020). Utilization of galectins by pathogens for infection. *Frontiers in Immunology*, *11*, 1877. https://doi.org/10.3389/fimmu.2020.01877

Banerjee, S., Gill, S. S., Jain, P. K., & Sirohi, A. (2017). Isolation, cloning, and characterization of a cuticle collagen gene, Mi-col-5, in *Meloidogyne incognita*. *3 Biotech*, *7*(1), 64. https://doi.org/10.1007/s13205-017-0665-1

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bellafiore, S., Jougla, C., Chapuis, É.,Besnard, G., Suong, M., Vu, P. N., De Waele, D., Gantet, P., & Thi, X. N. (2015). Intraspecific variability of the facultative meiotic parthenogenetic root-knot nematode (*Meloidogyne graminicola*) from rice fields in Vietnam. *Comptes Rendus Biologies*, *338*(7), 471–483. https://doi.org/10.1016/j.crvi.2015.04.002

Besnard, G., Jühling, F., Chapuis, É., Zedane, L., Lhuillier, É., Mateille, T., & Bellafiore, S. (2014). Fast assembly of the mitochondrial genome of a plant parasitic nematode *Meloidogyne graminicola* using next generation sequencing. *Comptes Rendus Biologies*, *337*(5), 295–301. https://doi.org/10.1016/j.crvi.2014.03.003

Besnard, G., Thi-Phan, N., Ho-Bich, H., Dereeper, A., Trang Nguyen, H., Quénéhervé, P., Aribi, J., & Bellafiore, S. (2019). On the close relatedness of two rice-parasitic root-knot nematode species and the recent expansion of *Meloidogyne graminicola* in Southeast Asia. *Genes*, *10*(2), 175. https://doi.org/10.3390/genes10020175

Cabasan, M. T. N., Kumar, A., Bellafiore, S., & Waele, D. D. (2014). Histopathology of the rice root-knot nematode, *Meloidogyne graminicola*, on *Oryza sativa* and *O. glaberrima*. *Nematology*, *16*(1), 73–81. https://doi.org/10.1163/15685411-00002746

Cabasan, M. T. N., Kumar, A., & Waele, D. D. (2012). Comparison of migration, penetration, development and reproduction of *Meloidogyne graminicola* on susceptible and resistant rice genotypes. *Nematology*, *14*(4), 405–415. https://doi.org/10.1163/156854111X602613

Canova, M. J., & Molle, V. (2014). Bacterial serine/threonine protein kinases in host-pathogen interactions. *The Journal of Biological Chemistry*, *289*(14), 9473–9479. https://doi.org/10.1074/jbc.R113.529917

Castagnone-Sereno, P., Mulet, K., Danchin, E. G. J., Koutsovoulos, G. D., Karaulic, M., Rocha, M. D., Bailly-Bechet, M., Pratx, L., Perfus-Barbeoch, L., & Abad, P. (2019). Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita* . *Molecular Ecology*, *28*(10), 2559–2572. https://doi.org/10.1111/mec.15095

Chen, C.-H. (2020). Detoxifying metabolism: detoxification enzymes. In C.-H. Chen (Ed.), *Xenobiotic Metabolic Enzymes: Bioactivation and Antioxidant Defense* (pp. 71–81). Springer International Publishing. https://doi.org/10.1007/978-3-030-41679-9_7

Clop, A., Vidal, O., & Amills, M. (2012). Copy number variation in the genomes of domestic animals. *Animal Genetics*, *43*(5), 503–517. https://doi.org/10.1111/j.1365-2052.2012.02317.x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Group, 1000 Genomes Project Analysis. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davis, E. L., Hussey, R. S., & Baum, T. J. (2004). Getting to the roots of parasitism by nematodes. *Trends in Parasitology*, *20*(3), 134–141. https://doi.org/10.1016/j.pt.2004.01.005

Dimkpa, S. O. N., Lahari, Z., Shrestha, R., Douglas, A., Gheysen, G., & Price, A. H. (2016). A genome-wide association study of a global rice panel reveals resistance in *Oryza sativa* to root-knot nematodes. *Journal of Experimental Botany*, *67*(4), 1191–1200. https://doi.org/10.1093/jxb/erv470

Dubreuil, G., Magliano, M., Deleury, E., Abad, P., & Rosso, M. N. (2007). Transcriptome analysis of root-knot nematode functions induced in the early stages of parasitism. *New Phytologist*, *176*(2), 426–436. https://doi.org/10.1111/j.1469-8137.2007.02181.x

Dukić, M., Berner, D., Haag, C. R., & Ebert, D. (2019). How clonal are clones? A quest for loss of heterozygosity during asexual reproduction in *Daphnia magna*. *Journal of Evolutionary Biology*, *32*(6), 619–628. https://doi.org/10.1111/jeb.13443

Engelstädter, J. (2017). Asexual but not clonal: evolutionary processes in automictic populations. *Genetics*, *206*(2), 993–1009. https://doi.org/10.1534/genetics.116.196873

Faucon, F., Dusfour, I., Gaude, T., Navratil, V., Boyer, F., Chandre, F., Sirisopa, P., Thanispong, K., Juntarajumnong, W., Poupardin, R., Chareonviriyaphap, T., Girod, R., Corbel, V., Reynaud, S., & David, J.-P. (2015). Identifying genomic changes associated with insecticide resistance in the dengue mosquito *Aedes aegypti* by deep targeted sequencing. *Genome Research*, *25*(9), 1347–1359. https://doi.org/10.1101/gr.189225.115

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv:1207.3907 [q-Bio]*. http://arxiv.org/abs/1207.3907

Gleason, C. A., Liu, Q. L., & Williamson, V. M. (2008). Silencing a candidate nematode effector gene corresponding to the tomato resistance gene *Mi-1* leads to acquisition of virulence. *Molecular Plant-Microbe Interactions*, *21*(5), 576–585. https://doi.org/10.1094/MPMI-21-5-0576

Haegeman, A., Mantelin, S., Jones, J. T., & Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. *Gene*, *492*(1), 19–31. https://doi.org/10.1016/j.gene.2011.10.040

Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., & Dunham, M. J. (2017). Loss of heterozygosity drives adaptation in hybrid yeast. *Molecular Biology and Evolution*, *34*(7), 1596–1612. https://doi.org/10.1093/molbev/msx098

Hull, R. M., Cruz, C., Jack, C. V., & Houseley, J. (2017). Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biology*, *15*(6), e2001333. https://doi.org/10.1371/journal.pbio.2001333

Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M., & Schwander, T. (2020). Genomic features of parthenogenetic animals. *BioRxiv*, 497495. https://doi.org/10.1101/497495

Ji, H., Gheysen, G., Denil, S., Lindsey, K., Topping, J. F., Nahar, K., Haegeman, A., De Vos, W. H., Trooskens, G., Van Criekinge, W., De Meyer, T., & Kyndt, T. (2013). Transcriptional analysis through RNA sequencing of giant cells induced by *Meloidogyne graminicola* in rice roots. *Journal of Experimental Botany*, *64*(12), 3885–3898. https://doi.org/10.1093/jxb/ert219

Ji, H., Kyndt, T., He, W., Vanholme, B., & Gheysen, G. (2015). β-aminobutyric acid–induced resistance against root-knot nematodes in rice is based on increased basal defense. *Molecular Plant-Microbe Interactions*, *28*(5), 519–533. https://doi.org/10.1094/MPMI-09-14-0260-R

Jiang, H., Lei, R., Ding, S.-W., & Zhu, S. (2014). Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, *15*(1), 182. https://doi.org/10.1186/1471-2105-15-182

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323–329. https://doi.org/10.1038/nature05286

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.-A., Mitterecker, A., Bodenhofer, U., & Hochreiter, S. (2012). cn.MOPS: Mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Research*, *40*(9), e69. https://doi.org/10.1093/nar/gks003

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 5048–5057. https://doi.org/10.1098/rspb.2012.1108

Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., & Blaxter, M. (2013). Blobology: Exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Frontiers in Genetics*, *4*, 237. https://doi.org/10.3389/fgene.2013.00237

Layer, R. M., Chiang, C., Quinlan, A. R., & Hall, I. M. (2014). LUMPY: A probabilistic framework for structural variant discovery. *Genome Biology*, *15*(6), R84. https://doi.org/10.1186/gb-2014-15-6-r84

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lipinski, K. J., Farslow, J. C., Fitzpatrick, K. A., Lynch, M., Katju, V., & Bergthorsson, U. (2011). High spontaneous rate of gene duplication in *Caenorhabditis elegans*. *Current Biology*, *21*(4), 306–310. https://doi.org/10.1016/j.cub.2011.01.026

Lu, M., Tian, X., Yang, X., Yuan, C., Ehsan, M., Liu, X., Yan, R., Xu, L., Song, X., & Li, X. (2017). The N- and C-terminal carbohydrate recognition domains of *Haemonchus contortus* galectin bind to distinct receptors of goat PBMC and contribute differently to its immunomodulatory functions in host-parasite interactions. *Parasites & Vectors*, *10*(1), 409. https://doi.org/10.1186/s13071-017-2353-8

Manning, G., Plowman, G. D., Hunter, T., & Sudarsanam, S. (2002). Evolution of protein kinase signaling from yeast to man. *Trends in Biochemical Sciences*, *27*(10), 514–520. https://doi.org/10.1016/s0968-0004(02)02179-5

Manosalva, P., Manohar, M., von Reuss, S. H., Chen, S., Koch, A., Kaplan, F., Choe, A., Micikas, R. J., Wang, X., Kogel, K.-H., Sternberg, P. W., Williamson, V. M., Schroeder, F. C., & Klessig, D. F. (2015). Conserved nematode signalling molecules elicit plant defenses and pathogen resistance. *Nature Communications*, *6*(1), 7795. https://doi.org/10.1038/ncomms8795

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Mattos, V. S., Leite, R. R., Cares, J. E., Gomes, A. C. M. M., Moita, A. W., Lobo, V. L. S., & Carneiro, R. M. D. G. (2019). *Oryza glumaepatula*, a new source of resistance to *Meloidogyne graminicola* and histological characterization of its defense mechanisms. *Phytopathology*, *109*(11), 1941–1948. https://doi.org/10.1094/PHYTO-02-19-0044-R

McClure, M. A., Kruk, T. H., & Misaghi, I. (1973). A method for obtaining quantities of clean *Meloidogyne* eggs. *Journal of Nematology*, *5*(3), 230.

McCubrey, J. A., May, W. S., Duronio, V., & Mufson, A. (2000). Serine/threonine phosphorylation in cytokine signal transduction. *Leukemia*, *14*(1), 9–21. https://doi.org/10.1038/sj.leu.2401657

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer,

R., Scholl, E., Sosinski, B. R., Thomas, V. P., & Windham, E. (2008). Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(39), 14802–14807. https://doi.org/10.1073/pnas.0805946105

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Petitot, A.-S., Dereeper, A., Agbessi, M., Da Silva, C., Guy, J., Ardisson, M., & Fernandez, D. (2015). Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants. *Molecular Plant Pathology*, *17*(6), 860–874. https://doi.org/10.1111/mpp.12334

Petitot, A.-S., Dereeper, A., Silva, C. D., Guy, J., & Fernandez, D. (2020). Analyses of the root-knot nematode (*Meloidogyne graminicola*) transcriptome during host infection highlight specific gene expression profiling in resistant rice plants. *Pathogens*, *9*(8). https://doi.org/10.3390/pathogens9080644

Phan, N. T., De Waele, D., Lorieux, M., Xiong, L., & Bellafiore, S. (2018). A hypersensitivity-like response to *Meloidogyne graminicola* in rice (*Oryza sativa*). *Phytopathology*, *108*(4), 521–528. https://doi.org/10.1094/PHYTO-07-17-0235-R

Phan, N. T., Orjuela, J., Danchin, E. G. J., Klopp, C., Perfus-Barbeoch, L., Kozlowski, D. K., Koutsovoulos, G. D., Lopez-Roques, C., Bouchez, O., Zahm, M., Besnard, G., & Bellafiore, S. (2020). Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*). *Ecology and Evolution*, *10*(20), 11006–11021. https://doi.org/10.1002/ece3.6680

Portillo, M., Cabrera, J., Lindsey, K., Topping, J., Andrés, M. F., Emiliozzi, M., Oliveros, J. C., García-Casado, G., Solano, R., Koltai, H., Resnick, N., Fenoll, C., & Escobar, C. (2013). Distinct and conserved transcriptomic changes during nematode-induced giant cell development in tomato compared with *Arabidopsis*: A functional role for gene repression. *The New Phytologist*, *197*(4), 1276–1290. https://doi.org/10.1111/nph.12121

Price, J. A., Smith, T. K., & Jones, J. T. (2020). Surface coat proteins of the potato cyst nematode, *Globodera rostochiensis*. *Nematology*, *23*(1), 113–123. https://doi.org/10.1163/15685411-bja10035

Przybylska, A., & Obrępalska-Stęplowska, A. (2020). Plant defense responses in monocotyledonous and dicotyledonous host plants during root-knot nematode infection. *Plant and Soil*, *451*(1), 239–260. https://doi.org/10.1007/s11104-020-04533-0

Schaad, N. W., & Walker, J. T. (1975). The use of density-gradient centrifugation for the purification of eggs of *Meloidogyne* spp. *Journal of Nematology*, *7*(2), 203–204.

Shi, W., Xue, C., Su, X., & Lu, F. (2018). The roles of galectins in parasitic infections. *Acta Tropica*, *177*, 97–104. https://doi.org/10.1016/j.actatropica.2017.09.027

Soriano, I. R., Schmit, V., Brar, D. S., Prot, J.-C., & Reversat, G. (1999). Resistance to rice root-knot nematode *Meloidogyne graminicola* identified in *Oryza longistaminata* and *O. glaberrima*. *Nematology*, *1*(4), 395–398. https://doi.org/10.1163/156854199508397

Stroehlein, A. J., Young, N. D., Hall, R. S., Korhonen, P. K., Hofmann, A., Sternberg, P. W., Jabbar, A., & Gasser, R. B. (2016). CAP protein superfamily members in *Toxocara canis*. *Parasites & Vectors*, *9*, 360. https://doi.org/10.1186/s13071-016-1642-y

Taylor, S., Barragan, A., Su, C., Fux, B., Fentress, S. J., Tang, K., Beatty, W. L., Hajj, H. E., Jerome, M., Behnke, M. S., White, M., Wootton, J. C., & Sibley, L. D. (2006). A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. *Science*, *314*(5806), 1776–1780. https://doi.org/10.1126/science.1133643

Theeya, N., Ta, A., Das, S., Mandal, R. S., Chakrabarti, O., Chakrabarti, S., Ghosh, A. N., & Das, S. (2015). An inducible and secreted eukaryote-like serine/threonine kinase of *Salmonella enterica* serovar Typhi promotes intracellular survival and pathogenesis. *Infection and Immunity*, *83*(2), 522–533. https://doi.org/10.1128/IAI.02521-14

Uwanogho, D. A., Hardcastle, Z., Balogh, P., Mirza, G., Thornburg, K. L., Ragoussis, J., & Sharpe, P. T. (1999). Molecular cloning, chromosomal mapping, and developmental expression of a novel protein tyrosine phosphatase-like gene. *Genomics*, *62*(3), 406–416. https://doi.org/10.1006/geno.1999.5950

Vieira, P., Mowery, J., Eisenback, J. D., Shao, J., & Nemchinov, L. G. (2019). Cellular and Transcriptional responses of resistant and susceptible cultivars of alfalfa to the root lesion nematode, *Pratylenchus penetrans*. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00971

Waetzig, G., Sobczak, M., & Grundler, F. (1999). Localization of hydrogen peroxide during the defence response of *Arabidopsis thaliana* against the plant-parasitic nematode *Heterodera glycines*. *Nematology*, *1*(7), 681–686. https://doi.org/10.1163/156854199508702

Young, A. R., & Meeusen, E. N. (2002). Galectins in parasite infection and allergic inflammation. *Glycoconjugate Journal*, *19*(7), 601–606. https://doi.org/10.1023/B:GLYC.0000014091.00844.0a

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*(9), 847–848. https://doi.org/10.1093/bioinformatics/17.9.847

Zhan, L., Ding, Z., Peng, D., Peng, H., Kong, L., Liu, S., Liu, Y., Li, Z., & Huag, W. (2018). Evaluation of Chinese rice varieties resistant to the root-knot nematode *Meloidogyne graminicola*. *Journal of Integrative Agriculture*, *17*(3), 621–630. https://doi.org/10.1016/S2095-3119(17)61802-1

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics*, *10*, 451–481. https://doi.org/10.1146/annurev.genom.9.081307.164217

Żmieńko, A., Samelak, A., Kozłowski, P., & Figlerowicz, M. (2014). Copy number polymorphism in plant genomes. *Theoretical and Applied Genetics*, *127*(1), 1–18. https://doi.org/10.1007/s00122-013-2177-7

Supporting Informations

Specifying genomic rearrangements in an emerging virulent isolate of rice root-knot nematode (*Meloidogyne graminicola*

Ngan Thi Phan[1], Hue Thi Nguyen[3], Guillaume Besnard[2*], Stephane Bellafiore[1]*

This supporting information contains Tables and Figures as listed below:

List of tables:

List of figures:

Table S1. Genomic data generated by Illumina Miseq for avirulent/virulent isolates.

No	Population	Country	# raw PE reads	# Cleaned PE reads	Read length (bp)	%GC	Read coverage *	Note
1	Mg-VN18	Vietnam	19,812,842	19,810,452	150	24	71.6 x	avirulent
2	ZMgP1.8	Cambodia	15,975,742	15,974,774	150	24	57.7 x	virulent

*Read coverage compare to haploid reference genome size (41.5 Mb)

Table S2. List of genes located on specific homozygous LoH regions in the virulent genome.

Scaff-olds	Start region (bp)	Stop region (bp)	Gene name	Start gene (bp)	Stop gene (bp)	*Interproscan*/*Pfam* domain
mg3	299433	308424	Mgra_00000698	300701	303397	IPR005522: Inositol polyphosphate kinase
			Mgra_00000699	303602	304665	IPR000626,IPR001841: Ubiquitin-like domain
			Mgra_00000700	304839	306097	IPR000626,IPR001841: Ubiquitin-like domain
mg6	280300	283359	Mgra_00001244	280976	283359	
mg7	898	177575	**Mgra_00001347**[2]	103651	105342	
			Mgra_00001348	107710	110912	IPR000922: D-galactoside/L-rhamnose binding SUEL lectin
			Mgra_00001349	110984	120189	IPR014928: Serine rich protein interaction domain
			Mgra_00001350[1]	120970	129718	IPR000873: AMP-dependent synthetase/ligase
			Mgra_00001351[1]	130901	135230	
			Mgra_00001352	141961	142930	IPR000782: FAS1 domain
			Mgra_00001353[12]	144847	148976	IPR000782: FAS1 domain
			Mgra_00001354	148957	150172	
			Mgra_00001355	150050	152708	
			Mgra_00001356	152713	153689	IPR018614: Keratinocyte-associated protein 2
			Mgra_00001320	16154	20678	IPR006895,IPR006896: Zinc finger, Sec23/Sec24-type
			Mgra_00001357	173430	174518	
			Mgra_00001321	21513	28774	IPR013721: STAG
			Mgra_00001322[2]	29076	30441	IPR033438: Modulator of levamisole receptor-1
			Mgra_00001323	31394	32895	
			Mgra_00001324	31394	36835	IPR000668: Peptidase C1A, papain C-terminal
			Mgra_00001318	3555	4075	
			Mgra_00001325	36897	41765	IPR003316,IPR032198: E2F/DP family
			Mgra_00001326	42248	42968	IPR001873: Epithelial sodium channel
			Mgra_00001327	43054	43984	

Scaff-olds	Start region (bp)	Stop region (bp)	Gene name	Start gene (bp)	Stop gene (bp)	*Interproscan*/*Pfam* domain
			Mgra_00001328[2]	44572	46703	IPR002213: UDP-glucuronosyl/UDP-glucosyltransferase
			Mgra_00001329	47270	48212	
			Mgra_00001330	49512	52138	IPR002018: Carboxylesterase, type B
			Mgra_00001331	52173	53007	
			Mgra_00001332	53966	57116	
			Mgra_00001319	5532	20678	IPR019427: 7TM GPCR, serpentine receptor class w (Srw)
			Mgra_00001333	59086	63155	IPR019774: Aromatic amino acid hydroxylase, C-terminal
			Mgra_00001334	61818	62844	IPR002048: EF-hand domain
			Mgra_00001335	63628	68247	IPR001382: Glycoside hydrolase family 47
			Mgra_00001336	65035	66720	IPR005135: Endonuclease/exonuclease/phosphatase
			Mgra_00001337	68876	70455	IPR001506: Peptidase M12A
			Mgra_00001338	71818	73815	IPR006029,IPR006202: Neurotransmitter-gated ion-channel
			Mgra_00001339	73819	75373	IPR006029: Neurotransmitter-gated ion-channel
			Mgra_00001340	74432	75282	IPR016072: SKP1 component, dimerisation
			Mgra_00001341[2]	75509	83565	IPR013032: EGF-like, conserved site
			Mgra_00001342	83782	92064	IPR000719: Protein kinase domain
			Mgra_00001343	85046	85129	
			Mgra_00001344	92632	93026	
			Mgra_00001345	94340	95583	
			Mgra_00001346[1]	97537	101828	
mg42	202	296323	Mgra_00005122	105356	107467	IPR007822: Lanthionine synthetase C-like
			Mgra_00005101[1]	10686	16751	
			Mgra_00005123[2]	107529	110330	IPR006202: Neurotransmitter-gated ion-channel
			Mgra_00005124	112614	114881	
			Mgra_00005125	115506	118235	IPR001873: Epithelial sodium channel
			Mgra_00005126	118291	121383	IPR023631: Amidase signature domain

Scaff -olds	Start region (bp)	Stop region (bp)	Gene name	Start gene (bp)	Stop gene (bp)	*Interproscan*/*Pfam* domain
			Mgra_00005127	122188	130018	IPR000859,IPR001752: CUB domain
			Mgra_00005128	130384	134003	IPR001164: Arf GTPase activating protein
			Mgra_00005129	134231	136692	IPR000276: G protein-coupled receptor, rhodopsin-like
			Mgra_00005130	137201	138813	
			Mgra_00005131	143948	157141	IPR000219: Dbl homology (DH) domain
			Mgra_00005132	161088	163244	
			Mgra_00005102	16170	21383	IPR001680: WD40 repeat
			Mgra_00005133	163572	166805	
			Mgra_00005134	167158	170663	
			Mgra_00005135	171672	174026	
			Mgra_00005136	179584	182486	
			Mgra_00005137	182277	188064	
			Mgra_00005138	188708	191522	
			Mgra_00005139	191881	194437	
			Mgra_00005140	194896	196876	
			Mgra_00005141	197024	197897	
			Mgra_00005142[1]	198638	218317	IPR001506,IPR010695: Peptidase M12A
			Mgra_00005103	21441	28390	
			Mgra_00005143	218890	219169	
			Mgra_00005144	224795	242158	IPR011641,IPR024731: Tyrosine-protein kinase ephrin type A/B receptor-like
			Mgra_00005145	242351	243832	IPR003582: ShKT domain
			Mgra_00005146[1]	244139	246424	IPR009071: High mobility group box domain
			Mgra_00005147	247445	249088	IPR001878: Zinc finger, CCHC-type
			Mgra_00005148[12]	249850	252237	IPR000917: Sulfatase, N-terminal
			Mgra_00005149[1]	252138	253488	

Scaff -olds	Start region (bp)	Stop region (bp)	Gene name	Start gene (bp)	Stop gene (bp)	*Interproscan*/*Pfam* domain
			Mgra_00005099	2541	6980	IPR000276: G protein-coupled receptor, rhodopsin-like
			Mgra_00005150[2]	254270	257859	
			Mgra_00005151	258319	266102	IPR000008: C2 domain
			Mgra_00005152	266715	271755	IPR007651,IPR013209: Lipin, N-terminal
			Mgra_00005104	26946	28274	IPR013078: Histidine phosphatase superfamily, clade-1
			Mgra_00005153	273860	274457	IPR001478: PDZ domain
			Mgra_00005154	275047	282243	IPR001478: PDZ domain
			Mgra_00005155	282368	285733	
			Mgra_00005156	283327	285863	
			Mgra_00005105	28409	35643	IPR021773: Trafficking protein particle complex subunit 11
			Mgra_00005157	286751	294115	IPR005817: Wnt
			Mgra_00005158	286837	294335	
			Mgra_00005106	42898	46063	IPR001839: Transforming growth factor-beta, C-terminal
			Mgra_00005107[2]	49697	54095	IPR002213: UDP-glucuronosyl/UDP-glucosyltransferase
			Mgra_00005108[2]	54711	55989	
			Mgra_00005109[2]	56410	59361	
			Mgra_00005110[1]	58890	65052	
			Mgra_00005111[1]	65975	76601	IPR001650,IPR011545: Helicase, C-terminal
			Mgra_00005100	6924	9701	
			Mgra_00005112	70872	72181	
			Mgra_00005113[2]	77376	79041	
			Mgra_00005114[2]	79653	81134	
			Mgra_00005115	81155	85772	IPR000195: Rab-GTPase-TBC domain
			Mgra_00005116	85784	88886	
			Mgra_00005117[1]	89003	89597	
			Mgra_00005118	93431	93868	IPR011598: Myc-type, basic helix-loop-helix (bHLH) domain

Scaff-olds	Start region (bp)	Stop region (bp)	Gene name	Start gene (bp)	Stop gene (bp)	*Interproscan*/*Pfam* domain
			Mgra_00005119	95187	98493	
			Mgra_00005120	97317	98314	
			Mgra_00005121	98508	105162	
mg57	204518	207799	Mgra_00006142	204519	207799	IPR001245,IPR001627,IPR012877: Serine-threonine/tyrosine-protein kinase, catalytic domain
			Mgra_00008847	17424	23884	IPR000719: Protein kinase domain
			Mgra_00008845	186	974	
			Mgra_00008848	27533	28045	
mg127	185	50356	Mgra_00008849	28904	30450	
			Mgra_00008846	3198	6992	
			Mgra_00008850	35532	37462	IPR000609: 7TM GPCR, serpentine receptor class g (Srg)
			Mgra_00008851[2]	37760	40851	IPR023631: Amidase signature domain

[1] Genes were annotated as putative effectors by Petitot et al. 2015; [2] these genes encode for proteins bearing signal peptide Sec/SPI of standard secreted proteins

Table S3. List of genes located on specific hemizygous LoH regions in the virulent genome.

Scaff-olds	Start region (bp)	Stop region (bp)	Gene name	Start gene (bp)	Stop gene (bp)	*Interproscan* domain
mg16	192654	203138	Mgra_00002612	196652	197616	
			Mgra_00002613	199127	201659	IPR002524,IPR027470: Cation efflux protein
mg18	182017	204006	Mgra_00002881	182442	183308	
			Mgra_00002884	190303	191498	IPR000535: Major sperm protein (MSP) domain
			Mgra_00002886[1]	194552	195543	IPR014044: CAP domain
			Mgra_00002891[1]	200633	204006	UDENN domain-containing protein
			Mgra_00002889	196740	198605	IPR002656: Acyltransferase 3
			Mgra_00002882	186114	187405	
			Mgra_00002885	192751	193931	
			Mgra_00002888	196017	196502	
			Mgra_00002890	198910	199396	
			Mgra_00002883	189409	190544	IPR000535: Major sperm protein (MSP) domain
			Mgra_00002887	195642	196831	
mg19	52077	57417	Mgra_00002997	52078	57417	IPR019409,IPR019441,IPR019443 FMP27, domain of unknown function DUF2405
mg71	3064	9532	Mgra_00006814	3443	3760	
			Mgra_00006815	5987	8340	
			Mgra_00006816	8512	9280	
mg163	3279	8574	Mgra_00009490	6539	8574	
			Mgra_00009488	3280	5083	IPR016090: Phospholipase A2 domain
mg255	110	5662	Mgra_00010245	3742	5662	IPR013057: Amino acid transporter, transmembrane domain
			Mgra_00010244	260	2363	IPR000743: Glycoside hydrolase, family 28

[1] Genes were annotated as putative effectors by Petitot et al. 2015

Table S4. Detection of variants on the mitochondrial genome of the virulent isolate by comparison with the reference sequence NC_024275.

Position on NC_024275.1	Coverage (x)	Reference sequence	SNP*/ heteroplasmic site	Gene	Amino acid change
6892	6328	G (82%)	T (17%)	rrnL (rRNA)	
9764	6723	A (73%)	G (26%)	ND4	Asn -> Asp
19297	5719	T (1%)	C* (99%)	ND5	Phe -> Leu

Table S5. List of primers used in PCR reaction for validation lost genes in virulent genome

Gene name	Primers forward/reverse	Primer sequence	Tm	Product size (bp)
Mgra_00005142	mg42_D3_GP1_FW	GCTCCTTTTTCAGCGTTTCC	58	506
	mg42_D3_GP1_RV	ATTGCCCTAAACCCTGAAGC	58	
Mgra_00005142	mg42_D3_GP2_FW	GGATTTATGGGCATGTCGGT	58	563
	mg42_D3_GP2_RV	AATGGCGAAGGGTTGAAGTT	58	
Mgra_00007209	mg78_D1_GP1_FW	TACGACGTCGAACAGTGTCC	60	524
	mg78_D1_GP1_RV	ATGAACCTGGCTTCTGGTGG	60	
Mgra_00007302	mg80_D1_G1P1_FW	AGACTTCCTCCACCACAAAGT	59	557
	mg80_D1_G1P1_RV	CGATGGATGGTTAGGTGGTCA	60	
Mgra_00007540	mg87_D3_G1_FW	CAGCCGCAACCTGTAATACT	58	570
	mg87_D3_G1_RV	GTACGAGTGCCAAATGTTGC	58	
Mgra_00007542	mg87_D3_G3_FW	CACTCATTAACGATCACCGG	56	539
	mg87_D3_G3_RV	AGGCACATCTCTGACTGGAA	58	
Mgra_00007543	mg87_D3_G4_FW	AACTGTGCGTGGTTTAACCT	58	537
	mg87_D3_G4_RV	GCGAAATTACTCGAAGCTGC	58	
Mgra_00007544	mg87_D3_G5_FW	AAGGGTCATATTAAGGAGCACC	58	571
	mg87_D3_G5_RV	GGTTAGCGGATTGAAGCACT	58	
Mgra_00007545	mg87_D3_G6_FW	CCATATACTAAAGCGACCCCG	58	593
	mg87_D3_G6_RV	GAGAACCTCGACCTCGACTA	58	
Mgra_00007546	mg87_D3_G7_FW	CGTCAAACACGTCCACAAAG	58	555
	mg87_D3_G7_RV	ATACGCCAAAGCATCTCGTT	58	
Mgra_00007558	mg87_D5_GP1_FW	GGTTCTGTACTCTGGCATGGA	59	504
	mg87_D5_GP1_RV	GCCCGGCATATCTGTGAAGA	60	
Mgra_00009234	mg145_D_G1_P1_FW	AGAAAAGGCCATTTCAGCGC	60	560
	mg145_D_G1_P1_RV	TGTGCGGCAATGAAAAGGAC	60	
Mgra_00009236	mg145_D_G3_FW	AACCAGACAAGGAAACGGCT	60	585
	mg145_D_G3_RV	TGCGGATTACTATTTGCGGAC	59	
Mgra_00009235	mg145_D_G2_FW	ACCTGCACAATCCTCTATGCC	60	529
	mg145_D_G2_RV	GCCTTAGAACCTTAGGCGCC	60	

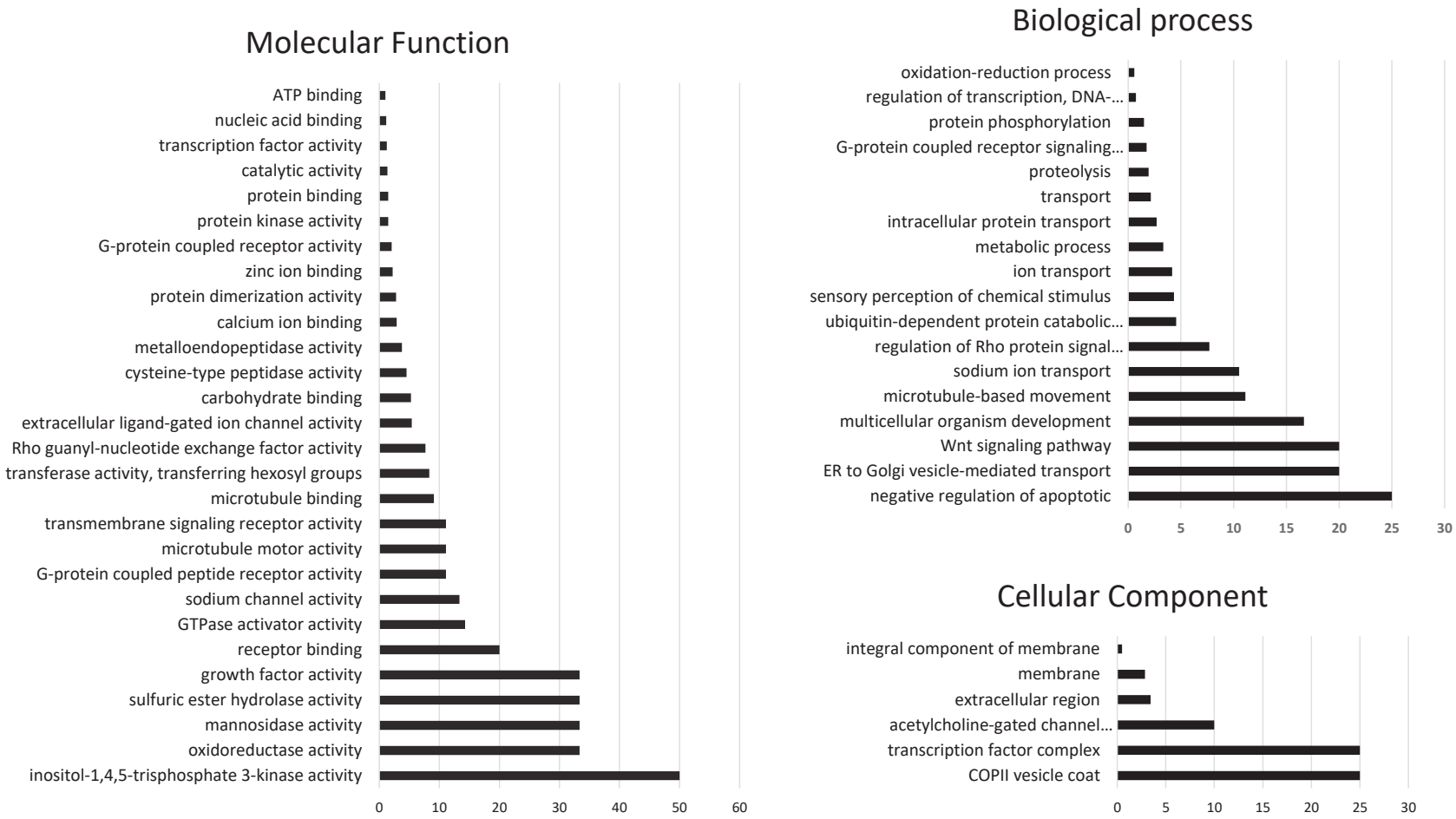


Figure S1. Gene Ontology annotation of genes located on the regions where a specific LoH has been observed in the genome of the virulent pathotype. The proportion indicates the number of genes affected by these LoHs for each biological class (the molecular functions, cellular locations, and processes gene products) in relation to the total number of genes have same GO in whole genome.

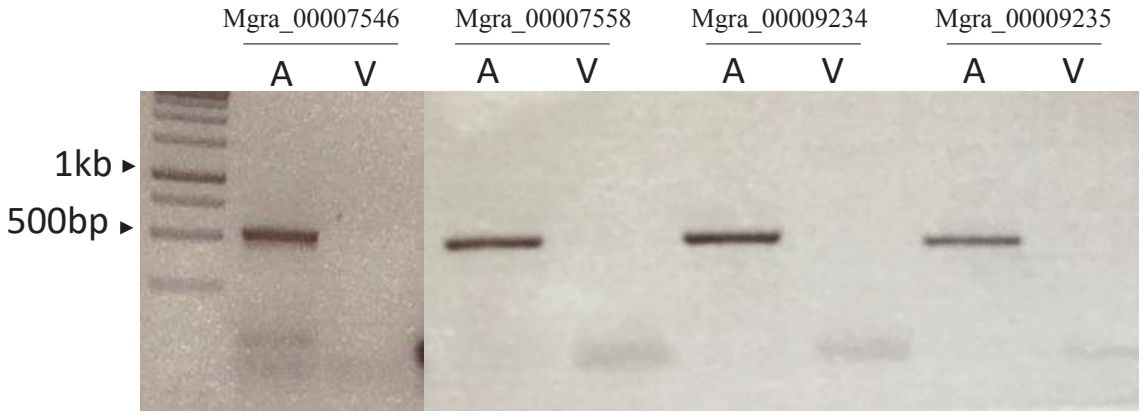


Figure S2. Gel electrophoresis of PCR reaction (product of ~500bp) confirmed the presence of four genes in the genome of the avirulent pathotype Mg-VN18 (A) and their absences in the genome of the pathotype virulent ZMgP1.8 (V).

CHAPTER VI

General discussions, conclusions, and perspectives

Knowledge of the spatial distribution and genetic diversity within and between populations of rice nematode pests is a crucial step in establishing crop management strategies to ensure food security. *Meloidogyne graminicola*, also known as the rice Root Knot Nematode, is a particularly aggressive pest for rice, significantly affecting yields, especially in Asia where more than 90% of the world's rice is produced. Despite its importance, this parasite is one of the little studied organisms at the genomic level and little is known about the phylogeography and genetic structure of this pest. The general objective of the thesis was to understand the evolutionary history of the parthenogenetic nematode *Meloidogyne graminicola* using genomic approaches. The complete sequencing of the genomes of different isolates allowed to understand the molecular diversity and phylogenetic relationships at the intra- and interspecific level and to describe the mechanisms that appear to be involved in the evolution of the species. We were able to suggest that *M. graminicola* could have a hybrid origin like its related sister species of *Meloidogyne*, but which reproduce by mitotic parthenogenesis whereas *M. graminicola* reproduces after meiotic parthenogenesis. The genome of *M. graminicola* is heterozygous, probably diploid and relatively small for a *Meloidogyne*. Phylogenetic study of *M. graminicola* isolates collected mainly from Southeast Asian countries show a low level of intraspecific diversity and an absence of phylogenetic structure that could be related to recent migratory events and its clonal reproduction pattern. Nevertheless, the *M. graminicola* genome shows abundant evolutionary markers and in particular the different isolates are distinguished from each other by different CNVs, LoHs and SNVs. Surprisingly, putative effectors seem to be targeted by these CNVs, LoH and SNVs, suggesting that the parasite could use these evolutionary mechanisms to adapt to constraining environments such as growing on resistant plants.

In each chapter, we have discussed in detail the relevance of our findings and how they relate to current knowledge on the corresponding topic. Here we present a summary of the main findings and provide some key discussions and perspectives.

I. *Meloidogyne graminicola* has a small, heterozygous, and likely diploid genome

M. graminicola was discovered nearly 60 years ago and is a major threat to global rice production. Despite its importance for phytosanitary reasons, almost no study has addressed genomic aspects because of the difficulties (fragility of DNA, heterozygosity) of reconstructing genomic scaffolds. By optimizing DNA extraction methods and using the advantages of long reading sequencing, we have built a very complete and contiguous genome compared to the previously published version (Somvanshi et al., 2018). The assembly of the haploid genomes *of M. graminicola* reveals that it is the smallest among the RKN genomes available to date (41.5 Mb). The total DNA content measured experimentally (81.5-83.8 Mb) is twice the size of that obtained in the assembly, suggesting that the *M. graminicola* genome is probably diploid. Its DNA is fragile certainly due to its GC content (23.5%) which is the lowest observed among the other RKNs described to date (up to 30% in mitotic RKNs). *M. graminicola* has a moderate level of heterozygosity (~1.85% at *k*-mer=21) compared to the genomes of other RKNs [ranging from 0.3% (*M. exigua* unpublished data) to 8.3%] (*M. javanica*; Blanc-Mathieu et al., 2017)]. A relatively small number of protein-encoding genes (10,284) were predicted for the small *M. graminicola* genome. This is consistent with the size of its genome, as the genome of *M. incognita* (triploid, 183 Mb) has 45,351 coding genes. Nevertheless, the use of MAKER to predict genes should be complemented by the use of other automated annotation pipelines such as EuGene which has demonstrated its efficiency in gene prediction in nematodes (http://eugene.toulouse.inra.fr/). The high-quality assembly of *M. graminicola* genome has allowed us to implement a series of comparative genomic analyses to study the evolution of *M. graminicola* genome. However, the haploid genome assembly consists mainly of collapsed haplotypes. The absence of haplotype sequences has caused several drawbacks for downstream analyses (eg. detection of single nucleotide polymorphisms at whole genome level). We have used other assemblers such as Platanus-allee (Kajitani et al., 2014) and HapCut2 (Edge et al., 2017) which should be effective on low heterozygous genome. Both methods produced two distinct haplotypes with a total genome assembly size of approximately 84 Mb, which corresponds to the total DNA content of the *M. graminicola* genome (81.5-83.8 Mb). However,

the 36 haplotigs assembled and checked manually (see Chapter IV) are not perfectly identical to the haplotypes obtained by this automatic assembly, which indicates that there are still errors during haplotyping. This indicates that in a context with ~1.85% heterozygosity, nanopore sequencing technology alone is still not sufficient to separate divergent haplotypes within the *Mg* genome. Several techniques such as PacBio, mate-pair Illumina, and high-throughput chromosome conformation capture (Hi-C) could be used to increase the length of sequencing reads and assembly quality. Among these, Hi-C is a new technique that targets chromatin interactions and allows the mapping of dynamic conformations throughout the genome (Belton et al., 2012). It is the most promising method today for producing heterozygous and complex genome assemblages that allow us to know all the haplotypes. For example, the *Adineta vaga* genome with a heterozygosity level of 1.7% (similar to the rate observed in *M. graminicola*) was fully assembled by combining short readings, long readings and Hi-C (Simion et al., 2020). In the future, this technique should be considered to resolve all divergent haplotypes in the genome of *M. graminicola* or other nematodes.

II. *M. graminicola* is a nematode that reproduces by meiotic parthenogenesis and seems to have promoted specific evolutionary mechanisms to adapt its parasitic lifestyle

For a long time we thought that parthenogenetic animals represented evolutionary "dead ends". That they should most certainly disappear due to the accumulation of deleterious mutations and a limited evolutionary capacity, while conversely sexual species seemed to be more evolved and able to adapt more easily to changing environments (Maynard-Smith, 1978). This seemed to correspond in particular to the low number of asexual animals in nature (Rice, 2002). However, contradicting this widespread idea, several observations have shown that asexual animals (e.g. bdelloid rotifer, the nematodes Panagrolaimus and *Meloidogyne incognita*) were considered an “evolutionary scandal” due to their outstanding ability to adapt to environments and hosts (Blanc-Mathieu et al., 2017; Flot et al., 2013; Schiffer et al., 2019). Parthenogenetic animals have acquired various means to promote genomic plasticity, which could enhance their adaptive capacity. These include the acquisition of genes by horizontal gene transfer (Schiffer et al., 2019), load of transposable element (Kozlowski et al., 2020),

recombination (Archetti, 2004), copy number variants (Castagnone-Sereno et al., 2019), and loss of heterozygosity (Dukić et al., 2019). In addition to parthenogenetic meiotic reproduction, sexual reproduction was supposed to occur occasionally in *M. graminicola* based on a few evidence of sperm pronuclei inside the eggs (~0.5%; Triantaphyllou, 1969). The evidence accumulated in our study indicates that *M. graminicola* possesses those genetic features of meiotic parthenogenetic animals that could lead to extreme adaptability and parasitic success of this pathogen.

M. graminicola may have acquired 67 genes (~0.74% of the total number of protein-encoding genes) through HGT during its evolution (see Chapter III). Most of these genes (~92%) are of bacterial origin and probably contributed to the success of its parasitism. These HGT genes would potentially be involved in the degradation of the plant cell wall (which would promote the migration of pests into the root tissue), biosynthesis and transformation of nutrients, detoxification and hijacking of the host plant's defenses. Several HGT genes are present in the form of multigenic families, suggesting that positive selective pressure would have favored the selection of individuals harboring multiple copies of these genes. Such a phenomenon has been described in other species of RKN (*M. incognita*, *M. hapla*) and PPN (e.g. *G. pallida*, *H. glycines*) (Akker et al., 2016; Opperman et al., 2008; Scholl et al., 2003). Further phylogenetic analysis of these families of HGT genes presumably originating from bacterial/fungi could reveal important points regarding the acquisition of these families of HGT genes by PPNs and in particular by *M. graminicola*.

Transposable elements (TEs), by their repetitive and mobile nature, can have both a passive and active impact on the plasticity of the genome. TEs may be involved in illegitimate genomic rearrangements resulting in the loss of genomic portions or the expansion of gene copy numbers. In addition, TEs can insert themselves into coding or regulatory regions and have an impact on gene expression or on the structure/function of the genes themselves. TEs represent 2.61% of the *M. graminicola* genome and their impact can be significant (see Chapter 3). It is likely that the abundance of TEs in mitotic RKNs is greater in number and diversity than in meiotic/sexual RKNs. For example, the number of TEs in mitotic RKNs such as *M. incognita*

appears to be slightly higher (4.67%) than in facultative meiotic RKNs such as *M. graminicola* using the same TE annotation pipeline (Kozlowski et al., 2020). In addition, mitotic RKNs (*M. incognita*, *M. javanica* and *M. arenaria*) also have twice the TE load as the facultative sexual *M. hapla* (Blanc-Mathieu et al., 2017). The more abundance DNA transposons combining with the high identity of TE copies in *M. incognita* genome suggests they have been recently active (Kozlowski et al., 2020). Insertions of TEs within genic regions and in the upstream regulatory regions of *M. incognita* genome suggests a functional impact (Kozlowski et al., 2020). This suggests contribution of TEs on genomic dynamics and plasticity, which may contribute to *M. incognita*'s adaptability to his environment (Kozlowski et al., 2020). As an example, the deletion of the *Cg-1* gene is associated with virulent pathotypes of *M. javanica*, and this gene has been identified in the Tm1 transposon (Gross and Williamson, 2011). In *M. graminicola*, two suspected HGTs, coding for DNA integrates whose potential function is to integrate foreign DNA into the genome, are linked to a TE, suggesting that some TEs may have been transferred laterally from the bacteria to the *M. graminicola* genome and/or that the TEs may promote diversity of acquired HGTs by replicating these genes. It remains to be determined whether TEs activity could represent a mechanism supporting genome plasticity thus contributing to the adaptive evolution of the species. In order to study the impact of TEs on the adaptability of *M. graminicola*, other studies could focus on the diversity of TEs underlying the parasite's phenotypes and in particular the acquisition of its virulence. It could also be considered whether the insertion of certain TEs in gene coding or regulatory regions leads to the acquisition of new adaptive functions.

Our study revealed that recombination events occurred in the *M. graminicola* genome (see Chapter IV). They were revealed by the presence of four-gamete products that must have been set up during meiosis and by the linkage disequilibrium between pairs of loci as a function of their physical distance. Since homologous recombination increases the chances of creating a new gene combination, it could provide an interesting genomic plasticity for adaptive evolution in *M. graminicola*. In *Meloidogyne* species, recombination is associated with species reproducing sexually and/or by meiotic parthenogenesis (e.g. *M. hapla* and *M. floridensis*) but

not with mitotic RKN (Koutsovoulos et al., 2019; Liu et al., 2007; Szitenberg et al., 2017). One of the consequences of recombination is the loss of heterozygosity, which led to the fixation of 0-7.24% of the genome in the studied *M. graminicola* isolates. Genes affected by LoH have only one allele, which can lead to important changes such as suppression of the expression of a deleterious allele, increase in the expression of an allele (dosage effect) or the expression of the phenotype of a recessive allele. A five-fold higher proportion of genes coding for putative effectors was found in these LoH regions of the genome of the virulent pathotype of *M. graminicola*. This suggests the LoH could participate in the adaptive evolution of *M. graminicola* by promoting the selection of phenotypic traits. A model that takes into account the evolutionary role of LoH during asexual reproduction is the "Loss of complementation—LOC hypothesis" proposed by Archeti 2004. The basic logic of this loss of complementation hypothesis is that recombination processes causing LoH during asexual reproduction will lead to the unmasking of deleterious recessive mutations. The consequences of LoCs will depend on the number of deleterious recessive mutations (lethal equivalents) accumulated and the proportion of the genome that becomes homozygous with each asexual generation (Archetti, 2010, 2004). As demonstrated by Archetti, LoC can result in the cost of asexual reproduction being more than twice the cost of sexual reproduction, but only with certain combinations of parameters, including a sufficiently high LoH rate (Archetti, 2010, 2004). LoH is also a consequence of gene conversion. The LoH, by gene conversion, can either expose new mutations to selection by making them homozygous, or suppress them as they occur thanks to their complementary allelic version, thus slowing down Muller's ratchet (i.e. the irreversible accumulation of harmful mutations in asexual populations of finite size). Gene conversion has been reported to play an important role in the evolution of the genome of RKN clade I (Szitenberg et al., 2017) and bdelloid rotifer (Flot et al., 2013). With the presence of LoH, gene conversion should also occur in the *M. graminicola* genome. It may be of interest to explore evidence of gene conversions and their evolutionary consequences in the *M. graminicola* genome. Sequences in palindromes could also limit the accumulation of deleterious mutations in parthenogenetic genomes (Trombetta and Cruciani, 2017). Palindromes, which are regions

duplicated on a single locus and in reverse orientation, are also present in *M. graminicola* genome (data not shown). The identification of palindromic sequences and genes important for parasitism (e.g. effector genes) affected by these palindromes could also be important to study the adaptive evolution of *M. graminicola*.

Copy number variants have been identified as an important mechanism involved in the evolution of genomes in most model organisms in the plant and animal kingdoms (Clop et al., 2012; Zhang et al., 2009; Żmieńko et al., 2014). CNVs have been identified in mitotic RKNs where they appear to be associated with the ability of the pest to overcome host plant resistance. This genetic variation could reflect an adaptive response to host resistance in this parthenogenetic species (Castagnone-Sereno et al., 2019). In addition, the loss of one or more copies of the *Cg-1* gene in the virulent pathotype of *M. javanica* appears to be associated with the gain in virulence against the resistant tomato genotype expressing *Mi-1.2* (Gleason et al., 2008). Our study also identified CNVs in different isolates of *M. graminicola* collected from different parts of the world (see Chapter IV). Although the living conditions under which these isolates were collected have not been described (e.g. edaphic factors, biotic and abiotic stress), variations in copy number among their genomes suggest a potential adaptation mechanism to these unknown different environments. In particular, CNVs are potentially associated in *M. graminicola* with adaptation to rice resistance by the virulent pathotype (see Chapter V). A significant number of genes have been lost in the virulent genome, among them some genes coding for putative effectors have been targeted, suggesting that CNVs may be directly involved in the acquisition of the virulence trait. The genes targeted by the CNVs will need to be further characterized, particularly with regard to their functions in host-plant interaction, in order to confirm the evolutionary role that CNVs seem to play in *M. graminicola*.

Single nucleotide variants (SNVs) have targeted some putative effectors in the virulent pathotype (see Chapter V). In addition, the accumulation of SNVs appears to be associated with LoH and gene loss events (see Chapter IV). These elements suggest that SNVs may contribute to the adaptation process of *M. graminicola*. However, due to the lack of a reference sequence including all haplotypes, SNVs at the whole genome level have not been studied. In the future,

an estimate of the substitution rate at the nuclear genome level should provide valuable information on whether SNVs contribute to the genomic plasticity of the *M. graminicola* genome. In addition, the study of non-synonymous mutations carried by these SNVs on parasitism genes (e.g. effectors) could help determine the evolutionary impact of these SNVs on the *M. graminicola* genome.

III. Recent expansion of *M. graminicola* and the potential impact of clonal reproduction

The study of the intraspecific diversity between the different genotypes (mainly isolated in Southeast Asia) and based on genomic variants (SNV, CNV and LoH) has confirmed a lack of clear phylogenetic signals both at the level of the mitochondrial and nuclear genome (see chapter II and chapter IV). These observations suggest that *M. graminicola* has spread recently and on a large scale from a likely single site. The natural spread of *M. graminicola* is in theory very limited because it is an obligate parasite and only pre-parasitic juveniles (pre-J2) can actively move in the soil, but only for a few days and over a limited distance (a few tens of centimeters). It is therefore believed that the large-scale expansion of this parasite is associated with the migratory activity of humans, animals and/or runoff water. However, the hypothesis presented here of recent expansion is based on the study of isolates mainly from South-East Asian countries (with the exception of one isolate from South America) and a larger study including isolates on a global scale would be valuable. Additional isolates from other regions such as the Americas, Italy and Madagascar would provide valuable information on population origin and dynamics. Since *M. graminicola* was first discovered in USA (Golden and Birchfield, 1965) before it was discovered in many countries in the Americas, several populations representative of the Americas should be included. On the other hand, it would be interesting to determine precisely the point mutation rate per generation from experiments conducted in the laboratory (under stress or not) in order to determine the rate of accumulation of mutations.

Haplotype networks based on SNPs from the mitochondrial and nuclear genomes have shown a strong correlation (see chapter II and chapter IV). In addition, some isolates show more genetic variation (such as CNVs, SNPs and LoHs) than others without a relationship being found based on their geographic origin. It is likely that each isolate has accumulated specific genomic changes (CNVs, SNPs and LoHs) to adapt to its environment. The narrow genetic distance (Fst < 0.1) between these isolates could suggest the presence of a common recent ancestor but also raises the question of whether sexual reproduction is active or not. Indeed, the meiotic parthenogenetic mode of reproduction is supposed to homogenize the genome and possibly produce a weak heterozygosity (Castagnone-Sereno et al., 2013; Triantaphyllou, 1985). From this perspective, the relatively high heterozygosity in *M. graminicola* is unexpected. This may suggest a hybrid origin between closely related species [(the hybrid origin is discussed in the next part (section IV, chapter VI)] or the outcrossing events that contribute to the observed heterozygosity gains. Reproductive mode is one of the main key to tackle genome diversity and evolution during the adaptation process. Therefore, to study the evolution of the *M. graminicola* genome, it is important to clarify its mode of reproduction. To study the mode of reproduction of *M. graminicola*, we propose here some experiments. First of all, a nematode crossing method was implemented by Liu et al., 2007 for *M. hapla* and a similar approach could be developed for *M. graminicola*. Males and females from two different isolates could be co-inoculated into the same plant. Then, after one generation, egg masses could be collected and tested for the presence of genetic markers on these offspring using molecular markers specific to each isolate. Secondly, to study the genetic exchanges potentially taking place between individuals living together in the field, several isolates collected individually can be sequenced. Then, the calculation of the genetic distance (Fst) based on the genetic markers could indicate whether genetic exchange (crossing) between these isolates has taken place.

IV. *M. graminicola* has a potential hybrid origin

Phylogenetic analysis of the divergent haplotypes (copies) of some low copy nuclear contigs showed a haplotype shared between the genome of *M. oryzae* and that of *M. graminicola* (see Chapter II). In addition, the mitochondrial genomes of the two species are very identical. This

pattern suggests that *M. graminicola* and *M. oryzae* could be the result of a hybridization event between close ancestors, or that *M. graminicola* (36 chromosomes) could be one of the parents for the genesis of *M. oryzae* (54 chromosomes). In addition, hybridization between the parental sexual species should generate a very heterozygous genome. The level of heterozygosity of the *M. graminicola* genome (1.85% at k-mer=21) is higher than that of parthenogenetic animals of intraspecific origin (0.03-0.83%) (Jaron et al., 2020). At the same time, this rate corresponds to that observed in parthenogenetic animals of hybrid origin (1.73-8.5%) (Jaron et al., 2020). As LoH always seems to take place in the genome of *M. graminicola*, the level of heterozygosity at the time of species formation should be higher than the current state. Therefore, we hypothesize that *M. graminicola* could have a hybrid origin. According to cytogenetic study, the original number of chromosome of *Meloidogyne* species may have been n=9 as represented by majority of *Heterodera* species (Triantaphyllou, 1985). *M. graminicola* might have an alloteteraploid genome (4n=36) with possibly two distinct genomes (AABB) due to the genome fusion of two closely related ancestors or hybridization followed by duplication of the whole genome (alloploidization event).

In order to test the ploidy level and structure of heterozygosity using a bioinformatics approach, a new technique called Smudgeplot has recently been developed and has successfully profiled the ploidy level of mitotic *Meloidogyne* species (Ranallo-Benavidez et al., 2020). This k-mer-based method should be easily applicable to check the ploidy level of the *M. graminicola* genome. On the other hand, using fluorescence *in situ* hybridization (FISH) it should be possible to confirm experimentally whether the observed divergent copies have a similar allele (AB) or belong to two distinct genomes (AABB). Unfortunately, this FISH approach is made difficult by the extremely small size of the chromosomes in *M. graminicola* and the holocentric nature of the chromosomes in the genus *Meloidogyne*.

In order to demonstrate the hypothesis of a hybrid origin, strong evidence must be provided at the genomic level. Phylogenetic analysis using copies/haplotypes of homologous genomic building blocks from closely related species should definitively answer this question. At the beginning of this thesis, to address the question of the hybrid origin of *M. graminicola*, *M.*

exigua was selected as a potential ancestor because of their close relationship both geographically and phylogenetically. Consequently, the genome sequence of *M. exigua* was sequenced and assembled. However, phylogenetic analysis using homologous nuclear contigs between *M. graminicola*, *M. oryzae* and *M. exigua* indicates that the *M. exigua* genome does not share any haplotype with *M. graminicola* (data not presented). This indicates that *M. exigua* is most certainly not a direct ancestor of *M. graminicola*. We suggest that other species related to *M. graminicola* and *M. oryzae*, such as *M. trifoliophila* and *M. naasi*, should in the future be used to study the hybrid origin of *M. graminicola*.

V. **Molecular mechanisms involved in the adaptation of *M. graminicola* to plant resistance**

The co-evolution of the arms race is intense between the host and the pathogen. To date, only a handful of *Oryza sativa* cultivars present the ability to resist to *M. graminicola* (Dimkpa et al., 2016; Zhan et al., 2018). However, the resistance of those cultivars has been compromised since the emergence of a new virulent pathotype (Nguyen Thi Hue's unpublished data). Comparative analysis between the virulent and avirulent genotypes in our study revealed several genetic variantions, including CNVs, LoH and SNVs (see chapter V). Genes encoding putative effectors that are targeted by these genetic modifications may be involved in the acquisition of virulence. A hypersensitive-like reaction is present in the resistant *Oryza sativa* cultivar (Zhonghua 11) – avirulent (Mg-VN18) interaction (Phan et al., 2018) but absent in the Zhonghua 11 – virulent (ZMgP1.8) interaction (Nguyen Thi Hue's unpublished data). If the incompatible interaction between RKN and the host is considered to follow a gene-for-gene relationship, it is likely that the pathogen-derived avirulence (Avr) gene product recognized by Zhonghua 11 and inducing HR is absent from the virulent pathotype. Therefore, we hypothesize that among these lost/modified genes in the virulent isolate is potentially the avirulence factor that triggers the hypersensitive response in the Zhonghua 11 variety. To test this hypothesis, we propose here some experimental procedures to identify the avirulence gene(s) in *M. graminicola*. Theoretically, the avirulence factor should be specifically and significantly expressed in a secretory organ, then the corresponding protein should be secreted into the root

and thus be recognized by a receptor protein (R-protein) of the host which will trigger the hypersensitive response. A first possibility could therefore consist in cloning from an avirulent population (e.g. Mg-VN18) all the potential effectors that are absent or whose coding sequence is modified in the virulent pathotype. If we discover in the meantime the resistance gene in these resistant varieties, it would be possible to use a heterologous expression system such as tobacco to transiently express the resistance gene and the potential avirulence factors (Ma et al., 2012). A screen based on induction of cell death would reveal the Avr/R pair that appears to control this incompatible interaction. A second possibility would be to directly express the candidate Avr genes transiently in the resistant rice variety. This could be done using protoplasts (Page et al., 2019) or in rice seedlings (Burman et al., 2020). A third possibility, which would not involve knowledge of the R gene, would be to repress or suppress the Avr factor in the avirulent population of *M. graminicola* and to phenotype the plant-pathogen interaction as a result of these modifications. The protocol for silencing PPNs gene using RNAi technique has been established in both *in vitro* (Gleason et al., 2008; Huang et al., 2006; Rosso et al., 2005), and *in planta* (Arguel et al., 2012; Gheysen and Vanholme, 2007; Iqbal et al., 2020; Nguyen et al., 2018). A protocol for gene inactivation with the CRISPR/Cas9 machinery has been established in *C. elegans* (Paix et al., 2015), and in the human-parasitic nematode *Strongyloides stercoralis* (Gang et al., 2017). However, no similar protocol is currently available for PPN genome editing.

V. Final conclusion

In conclusion, my Ph.D work explored part of the evolutionary history of plant parasitic nematode *Meloidogyne graminicola*. For the first time, the evolutionary history of *M. graminicola* was studied at the level of genomics and population genomics. High quality genome assembly represents a valuable molecular resource for future phylogenomic studies of *Meloidogyne* species. This work suggests that a recent expansion event has led to the spread of the species in Southeast Asian countries. The accumulation of evidence suggesting horizontal gene transfers, transposable elements, recombinations, loss of heterozygosity and copy number variants are major evolutionary markers that appear to have shaped the *M. graminicola* genome

and may explain the extreme adaptive capacities of this parasite. In particular, the effectors targeted by these evolutionary mechanisms could lead to the emergence of virulent pathotypes that can overcome plant resistance. The results of my thesis also open several important hypotheses necessary for future studies to better understand the evolutionary history of *Meloidogyne graminicola* and its closely related species.

References

Akker, S.E. den, Laetsch, D.R., Thorpe, P., Lilley, C.J., Danchin, E.G.J., Rocha, M.D., Rancurel, C., Holroyd, N.E., Cotton, J.A., Szitenberg, A., Grenier, E., Montarry, J., Mimee, B., Duceppe, M.-O., Boyes, I., Marvin, J.M.C., Jones, L.M., Yusup, H.B., Lafond-Lapalme, J., Esquibet, M., Sabeh, M., Rott, M., Overmars, H., Finkers-Tomczak, A., Smant, G., Koutsovoulos, G., Blok, V., Mantelin, S., Cock, P.J.A., Phillips, W., Henrissat, B., Urwin, P.E., Blaxter, M., Jones, J.T., 2016. The genome of the yellow potato cyst nematode, *Globodera rostochiensis* , reveals insights into the basis of parasitism and virulence. Genome Biol 17, 1–23. https://doi.org/10.1186/s13059-016-0985-1

Archetti, M., 2010. Complementation, genetic conflict, and the evolution of sex and recombination. J Hered 101 Suppl 1, S21-33. https://doi.org/10.1093/jhered/esq009

Archetti, M., 2004. Recombination and loss of complementation: a more than two-fold cost for parthenogenesis. Journal of Evolutionary Biology 17, 1084–1097. https://doi.org/10.1111/j.1420-9101.2004.00745.x

Arguel, M.-J., Jaouannet, M., Magliano, M., Abad, P., Rosso, M.-N., 2012. siRNAs Trigger Efficient Silencing of a Parasitism Gene in Plant Parasitic Root-Knot Nematodes. Genes 3, 391–408. https://doi.org/10.3390/genes3030391

Belton, J.-M., McCord, R.P., Gibcus, J., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi-C: A comprehensive technique to capture the conformation of genomes. Methods 58. https://doi.org/10.1016/j.ymeth.2012.05.001

Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J.-M., Rocha, M.D., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J.-F., Kozlowski, D.K., Cazareth, J., Couloux, A., Silva, C.D., Guy, J., Kim-Jo, Y.-J., Rancurel, C., Schiex, T., Abad, P., Wincker, P., Danchin, E.G.J., 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. PLOS Genet. 13, e1006777. https://doi.org/10.1371/journal.pgen.1006777

Burman, N., Chandran, D., Khurana, J.P., 2020. A Rapid and Highly Efficient Method for Transient Gene Expression in Rice Plants. Front. Plant Sci. 11. https://doi.org/10.3389/fpls.2020.584011

Castagnone-Sereno, P., Danchin, E.G.J., Perfus-Barbeoch, L., Abad, P., 2013. Diversity and evolution of root-knot nematodes, genus *Meloidogyne*: new insights from the genomic era. Annu. Rev. Phytopathol. 51, 203–220. https://doi.org/10.1146/annurev-phyto-082712-102300

Castagnone-Sereno, P., Mulet, K., Danchin, E.G.J., Koutsovoulos, G.D., Karaulic, M., Rocha, M.D., Bailly-Bechet, M., Pratx, L., Perfus-Barbeoch, L., Abad, P., 2019. Gene copy number variations

as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita* . Molecular Ecology 28, 2559–2572. https://doi.org/10.1111/mec.15095

Clop, A., Vidal, O., Amills, M., 2012. Copy number variation in the genomes of domestic animals. Anim Genet 43, 503–517. https://doi.org/10.1111/j.1365-2052.2012.02317.x

Dimkpa, S.O.N., Lahari, Z., Shrestha, R., Douglas, A., Gheysen, G., Price, A.H., 2016. A genome-wide association study of a global rice panel reveals resistance in *Oryza sativa* to root-knot nematodes. J Exp Bot 67, 1191–1200. https://doi.org/10.1093/jxb/erv470

Dukić, M., Berner, D., Haag, C.R., Ebert, D., 2019. How clonal are clones? A quest for loss of heterozygosity during asexual reproduction in *Daphnia magna*. J Evol Biol 32, 619–628. https://doi.org/10.1111/jeb.13443

Edge, P., Bafna, V., Bansal, V., 2017. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res 27, 801–812. https://doi.org/10.1101/gr.213462.116

Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E.G.J., Hejnol, A., Henrissat, B., Koszul, R., Aury, J.-M., Barbe, V., Barthélémy, R.-M., Bast, J., Bazykin, G.A., Chabrol, O., Couloux, A., Da Rocha, M., Da Silva, C., Gladyshev, E., Gouret, P., Hallatschek, O., Hecox-Lea, B., Labadie, K., Lejeune, B., Piskurek, O., Poulain, J., Rodriguez, F., Ryan, J.F., Vakhrusheva, O.A., Wajnberg, E., Wirth, B., Yushenova, I., Kellis, M., Kondrashov, A.S., Mark Welch, D.B., Pontarotti, P., Weissenbach, J., Wincker, P., Jaillon, O., Van Doninck, K., 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. Nature 500, 453–457. https://doi.org/10.1038/nature12326

Gang, S.S., Castelletto, M.L., Bryant, A.S., Yang, E., Mancuso, N., Lopez, J.B., Pellegrini, M., Hallem, E.A., 2017. Targeted mutagenesis in a human-parasitic nematode. PLOS Pathogens 13, e1006675. https://doi.org/10.1371/journal.ppat.1006675

Gheysen, G., Vanholme, B., 2007. RNAi from plants to nematodes. Trends in Biotechnology 25, 89–92. https://doi.org/10.1016/j.tibtech.2007.01.007

Gleason, C.A., Liu, Q.L., Williamson, V.M., 2008. Silencing a candidate nematode effector gene corresponding to the tomato resistance gene *Mi-1* leads to acquisition of virulence. Mol Plant Microbe Interact 21, 576–585. https://doi.org/10.1094/MPMI-21-5-0576

Golden, A.M., Birchfield, W., 1965. *Meloidogyne graminicola* (Heteroderidae) a new species of root-knot nematode from grass. Proc. Helminthol. Soc. Wash. 32, 228–231.

Gross, S.M., Williamson, V.M., 2011. Tm1: A *mutator*/*foldback* transposable element family in root-knot nematodes. PLOS ONE 6, e24534. https://doi.org/10.1371/journal.pone.0024534

Huang, G., Allen, R., Davis, E.L., Baum, T.J., Hussey, R.S., 2006. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. PNAS 103, 14302–14306. https://doi.org/10.1073/pnas.0604698103

Iqbal, S., Fosu-Nyarko, J., Jones, M.G.K., 2020. Attempt to Silence Genes of the RNAi Pathways of the Root-Knot Nematode, *Meloidogyne incognita* Results in Diverse Responses Including Increase and No Change in Expression of Some Genes. Front Plant Sci 11. https://doi.org/10.3389/fpls.2020.00328

Jaron, K.S., Bast, J., Nowell, R.W., Ranallo-Benavidez, T.R., Robinson-Rechavi, M., Schwander, T., 2020. Genomic Features of Parthenogenetic Animals. Journal of Heredity. https://doi.org/10.1093/jhered/esaa031

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., Kohara, Y., Fujiyama, A., Hayashi, T., Itoh, T., 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 24, 1384–1395. https://doi.org/10.1101/gr.170720.113

Koutsovoulos, G.D., Marques, E., Arguel, M.-J., Duret, L., Machado, A.C.Z., Carneiro, R.M.D.G., Kozlowski, D.K., Bailly-Bechet, M., Castagnone-Sereno, P., Albuquerque, E.V.S., Danchin, E.G.J., 2019. Population genomics supports clonal reproduction and multiple independent gains and losses of parasitic abilities in the most devastating nematode pest. Evolutionary Applications 13, 442–457. https://doi.org/10.1111/eva.12881

Kozlowski, D.K., Hassanaly-Goulamhoussen, R., Rocha, M.D., Koutsovoulos, G.D., Bailly-Bechet, M., Danchin, E.G., 2020. Transposable Elements are an evolutionary force shaping genomic plasticity in the parthenogenetic root-knot nematode *Meloidogyne incognita*. bioRxiv 2020.04.30.069948. https://doi.org/10.1101/2020.04.30.069948

Liu, Q.L., Thomas, V.P., Williamson, V.M., 2007. Meiotic Parthenogenesis in a Root-Knot Nematode Results in Rapid Genomic Homozygosity. Genetics 176, 1483–1490. https://doi.org/10.1534/genetics.107.071134

Ma, L., Lukasik, E., Gawehns, F., Takken, F.L.W., 2012. The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves. Methods Mol Biol 835, 61–74. https://doi.org/10.1007/978-1-61779-501-5_4

Maynard-Smith, J., 1978. The Evolution of Sex. Cambridge University Press, Cambridge.

Nguyen, C.-N., Perfus-Barbeoch, L., Quentin, M., Zhao, J., Magliano, M., Marteu, N., Da Rocha, M., Nottet, N., Abad, P., Favery, B., 2018. A root-knot nematode small glycine and cysteine-rich secreted effector, MiSGCR1, is involved in plant parasitism. New Phytol 217, 687–699. https://doi.org/10.1111/nph.14837

Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T.D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B.R., Thomas, V.P., Windham, E., 2008. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. Proc. Natl. Acad. Sci. U.S.A. 105, 14802–14807. https://doi.org/10.1073/pnas.0805946105

Page, M.T., Parry, M.A.J., Carmo-Silva, E., 2019. A high-throughput transient expression system for rice. Plant Cell Environ 42, 2057–2064. https://doi.org/10.1111/pce.13542

Paix, A., Folkmann, A., Rasoloson, D., Seydoux, G., 2015. High Efficiency, Homology-Directed Genome Editing in *Caenorhabditis elegans* Using CRISPR-Cas9 Ribonucleoprotein Complexes. Genetics 201, 47–54. https://doi.org/10.1534/genetics.115.179382

Phan, N.T., De Waele, D., Lorieux, M., Xiong, L., Bellafiore, S., 2018. A Hypersensitivity-Like Response to *Meloidogyne graminicola* in Rice (*Oryza sativa*). Phytopathology 108, 521–528. https://doi.org/10.1094/PHYTO-07-17-0235-R

Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nature Communications 11, 1432. https://doi.org/10.1038/s41467-020-14998-3

Rice, W.R., 2002. Experimental tests of the adaptive significance of sexual recombination. Nat Rev Genet 3, 241–251. https://doi.org/10.1038/nrg760

Rosso, M.N., Dubrana, M.P., Cimbolini, N., Jaubert, S., Abad, P., 2005. Application of RNA interference to root-knot nematode genes encoding esophageal gland proteins. Mol Plant Microbe Interact 18, 615–620. https://doi.org/10.1094/MPMI-18-0615

Schiffer, P.H., Danchin, E.G.J., Burnell, A.M., Creevey, C.J., Wong, S., Dix, I., O'Mahony, G., Culleton, B.A., Rancurel, C., Stier, G., Martínez-Salazar, E.A., Marconi, A., Trivedi, U., Kroiher, M., Thorne, M.A.S., Schierenberg, E., Wiehe, T., Blaxter, M., 2019. Signatures of the Evolution of Parthenogenesis and Cryptobiosis in the Genomes of *Panagrolaimid* Nematodes. iScience 21, 587–602. https://doi.org/10.1016/j.isci.2019.10.039

Scholl, E.H., Thorne, J.L., McCarter, J.P., Bird, D.M., 2003. Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. Genome Biol 4, R39. https://doi.org/10.1186/gb-2003-4-6-r39

Somvanshi, V.S.;Tathode, M.; Shukla, R.N.; Rao, U.Nematode genome announcement: A draft genome for rice root-knot nematode, *Meloidogyne graminicola*. *J. Nematol*. **2018**, *50*, 111–116. doi:10.21307/jofnem-2018-018.

Simion, P., Narayan, J., Houtain, A., Derzelle, A., Baudry, L., Nicolas, E., Cariou, M., Guiglielmoni, N., Kozlowski, D.K., Gaudray, F.R., Terwagne, M., Virgo, J., Noel, B., Wincker, P., Danchin, E.G., Marbouty, M., Hallet, B., Koszul, R., Limasset, A., Flot, J.-F., Doninck, K.V., 2020. Homologous chromosomes in asexual rotifer *Adineta vaga* suggest automixis. bioRxiv 2020.06.16.155473. https://doi.org/10.1101/2020.06.16.155473

Szitenberg, A., Salazar-Jaramillo, L., Blok, V.C., Laetsch, D.R., Joseph, S., Williamson, V.M., Blaxter, M.L., Lunt, D.H., 2017. Comparative genomics of apomictic root-knot nematodes: Hybridization, ploidy, and dynamic genome change. Genome Biol. Evol. 9, 2844–2861. https://doi.org/10.1093/gbe/evx201

Triantaphyllou, A.C., 1985. Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes, in: Sasser, J.N., Carter, C.C. (Eds.), An Advanced Treatise on *Meloidogyne*. Biology and Control. Raleigh, North Carolina State University, pp. 113–126.

Triantaphyllou, A.C., 1969. Gametogenesis and the chromosomes of two root-knot nematodes, *Meloidogyne graminicola* and *M. naasi*. J. Nematol. 1, 62–71.

Trombetta, B., Cruciani, F., 2017. Y chromosome palindromes and gene conversion. Hum Genet 136, 605–619. https://doi.org/10.1007/s00439-017-1777-8

Zhan, L., Ding, Z., Peng, D., Peng, H., Kong, L., Liu, S., Liu, Y., Li, Z., Huag, W., 2018. Evaluation of Chinese rice varieties resistant to the root-knot nematode *Meloidogyne graminicola*. Journal of Integrative Agriculture 17, 621–630. https://doi.org/10.1016/S2095-3119(17)61802-1

Zhang, F., Gu, W., Hurles, M.E., Lupski, J.R., 2009. Copy Number Variation in Human Health, Disease, and Evolution. Annu Rev Genomics Hum Genet 10, 451–481. https://doi.org/10.1146/annurev.genom.9.081307.164217

Żmieńko, A., Samelak, A., Kozłowski, P., Figlerowicz, M., 2014. Copy number polymorphism in plant genomes. Theor Appl Genet 127, 1–18. https://doi.org/10.1007/s00122-013-2177-7

ANNEXES

PUBLICATIONS

Publication in PhD thesis

1. **Phan** et al. Comparative genomic reveals genetic signatures of a new virulent *Meloidogyne graminicola* isolate against resistant rice *Oryza sativa* cultivar Zhonghua 11. In preparation.
2. **Phan** et al. Genomic rearrangements are putative genomic evolutionary traits in the facultative parthenogenic nematode *Meloidogyne graminicola.* In preparation.
3. **Phan** et al. 2021. High-quality genome sequence of the coffee Root-Knot Nematode *Meloidogyne exigua*. In preparation.
4. **Phan**, N. T., Orjuela, J., Danchin, E. G. J., Klopp, C., Perfus-Barbeoch, L., Kozlowski, D. K., Koutsovoulos, G. D., Lopez-Roques, C., Bouchez, O., Zahm, M., Besnard, G., & Bellafiore, S. (2020). Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*). *Ecology and Evolution*, *10*(20), 11006–11021. https://doi.org/10.1002/ece3.6680
5. Besnard[1], G., **Phan**[1], N. T., Ho-Bich, H., Dereeper, A., Trang Nguyen, H., Quénéhervé, P., Aribi, J., & Bellafiore*, S. (2019). On the close relatedness of two rice-parasitic root-knot nematode species and the recent expansion of *Meloidogyne graminicola* in Southeast Asia. *Genes*, *10*(2). https://doi.org/10.3390/genes10020175

Previous publications

6. **Phan**, N.T., De Waele, D., Lorieux, M., Xiong, L., Bellafiore, S., 2018. A Hypersensitivity-Like Response to *Meloidogyne graminicola* in Rice (*Oryza sativa*). Phytopathology 108, 521–528. https://doi.org/10.1094/PHYTO-07-17-0235-R
7. **Phan**, N.T., Trinh, L.T., Rho, M.-Y., Park, T.-S., Kim, O.-R., Zhao, J., Kim, H.-M., Sim, S.-C., 2019. Identification of loci associated with fruit traits using genome-wide single nucleotide polymorphisms in a core collection of tomato (*Solanum lycopersicum* L.). Scientia Horticulturae 243, 567–574. https://doi.org/10.1016/j.scienta.2018.09.003
8. **Phan**, N.T., Sim, S.C., 2017. Genomic Tools and Their Implications for Vegetable Breeding 149–164. https://doi.org/10.12972/kjhst.20170018
9. **Phan**, N.T., Kim, M.-K., Sim, S.-C., 2016. Genetic variations of F1 tomato cultivars revealed by a core set of SSR and InDel markers. Scientia Horticulturae 212, 155–161. https://doi.org/10.1016/j.scienta.2016.09.043
10. Tran, T.T., Nga, N.V., **Phan**, N.T., Hong, N.T., Szurek, B., Koebnik, R., Ham, L.H., Cuong, H.V., Cunnac, S., 2015. Confirmation of Bacterial Leaf Streak of Rice Caused by *Xanthomonas oryzae* pv. *oryzicola* in Vietnam. Plant Disease 99, 1853–1853. https://doi.org/10.1094/PDIS-03-15-0289-PDN

SCIENTIFIC COMMUNICATION

1. [Poster presentation] Ngan Thi Phan, Guillaume Besnard, Hai Ho-Bich, Alexis Dereeper, Hieu Trang Nguyen, Patrick Quénéhervé, Stéphane Bellafiore. Reduced genetic diversity among *Meloidogyne graminicola* isolates suggests a recent worldwide expansion. **European Society of Nematologist (ESN) conference** 2018. Ghent, Belgium
2. [Oral presentation] Ngan Thi Phan. Evolutionary history of plant-root parasite *Meloidogyne graminicola*. **PhD days** 2018 and 2019. IRD Montpellier, France

TRAININGS/FORMATIONS

Catégorie : Enseignement à distance : 15h

☛ MOOC Research integrity in scientific professions (07 novembre 2018 - 2 janvier 2019)

Catégorie : Formations scientifiques : 37h

☛ Guide de survie à Linux IRD Montpellier, France (15 heures enregistrées par GAIA)

☛ Initiation aux analyses de données Metagenomiques IRD Montpellier, France (15 heures enregistrées par GAIA)

☛ Initiation aux analyses de données transcriptomiques IRD Montpellier, France (7 heures enregistrées par GAIA)

Catégorie : Langues vivantes : 76h

☛ FLE - Français Langue Étrangère (session janvier 2018) (29 janvier 2018 - 26 avril 2018) Université de Montpellier, Département des Langues Bâtiment 5 (30 heures, Note : A1)

☛ FLE - Français Langue Étrangère (session Septembre 2018) (01 octobre 2018) Université de Montpellier, Département des Langues Bâtiment 5 (22 heures, Note : A2)

☛ FLE - Français Langue Étrangère (session septembre 2019) (30 septembre 2019) Université de Montpellier, Département des Langues Bâtiment 5 (24 heures, Note : A1+)

Catégorie : Outils et méthodes : 44h

☛ Dealing with a scientific litterature (16 avril 2018 - 18 avril 2018) Université de Montpellier (20 heures)

☛ Start with R (04 septembre 2019) Université de Montpellier (14 heures)

☛ Writing and defense a PhD (07 janvier 2020) salle de conférence Institut Européen des Membranes (10 heures)

Total participation : 176 heures / 11 modules

Article

Genome sequence of the coffee root-knot nematode *Meloidogyne exigua*

Ngan Thi Phan[1], Guillaume Besnard[2], Rania Ouazahrou[3], William Solano Sánchez[4], Lisa Gil[5], Sophie Manzi[2], Stéphane Bellafiore[1]

Received for publication:

[1]PHIM Plant Health Institute, Univ. Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

[2]CNRS-UPS-IRD, UMR5174, EDB, 118 route de Narbonne, Université Paul Sabatier, 31062 Toulouse, France

[3]University of Rennes 1, UFR SVE, 35065 Rennes, France

[4]CATIE - Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica

[5]US 1426, GeT-PlaGe, Genotoul, INRAE, Castanet-Tolosan, France

Acknowledgments

This research was funded by the Consultative Group for International Agricultural Research Program on rice-agrifood systems (CRP-RICE, 2017–2022). The sequencing was performed at the GeT-PlaGe core facility, INRAE Toulouse, France. Ngan Thi Phan was supported by a PhD fellowship from French Embassy in Vietnam. Guillaume Besnard is member of the EDB lab that is supported by the LABEX TULIP (ANR-10-LABX-0041) and CEBA (ANR-10-LABX-25-01). The authors are grateful to J. Aribi (PHIM) for his technical assistance.

Email stephane.bellafiore@ird.fr

This paper was edited by

Running Head: *Meloidogyne exigua* genome: *Phan et al.*

Abstract: Root-knot nematodes (*Meloidogyne* spp.) cause serious damages on most crops. Here, we report a high-quality genome sequence of *M. exigua* (population Mex1, Costa Rica), a major pathogen of coffee. Its mitogenome (20,974 bp) was first assembled and annotated. The nuclear genome was then constructed consisting of 206 contigs, with an N50 length of 1.89 Mb and a total assembly length of 42.1 Mb.

Keywords: Genomics, Illumina, Mitogenome, Nanopore sequencing, Nuclear genome, Root-knot nematode.

Root-knot nematodes (RKN) parasitize a wide range of host plants and have a global distribution. They are considered the most important group of plant-parasitic nematodes (Jones et al., 2013). Several *Meloidogyne* species can attack coffee plants, but only *Meloidogyne exigua* Goeldi (Goeldi, 1892) has a significant impact on coffee production. This pathogen is the most widely distributed nematode in the coffee production areas in Central and South America (Campos and Villain, 2005), with estimated yield losses of up to 45% in the Rio de Janeiro State (Barbosa et al., 2004) and between 15-20% in Central America as a whole (Anzueto et al., 1995). Despite these serious impacts on coffee production, diversity and adaptation of *M. exigua* has been poorly documented, and so far, the only published study on the species was based on isozyme profiles and random amplified polymorphic DNA (RAPD) markers (Muniz et al., 2008). With the advent of high throughput sequencing methods, the analysis of its genome has become possible and may open new avenues for studying its evolutionary history.

Comparative genomics of RKN species has revealed a striking diversity in genome structure (e.g chromosome counts, ploidy level, duplicated regions, heterozygosity) that might be linked to their different reproductive modes and species origin (Blanc-Mathieu et al., 2017; Castagnone-Sereno et al., 2013; Jaron et al., 2020; Triantaphyllou, 1985). Interestingly, despite prominent asexual reproduction in several RKN species, various mechanisms can generate genomic variability and may play a major role in their adaptability against different environments and hosts. These include, in particular, horizontal gene transfers (Danchin et al., 2016; Opperman et al., 2008; Phan et al., 2020), insertion of transposable elements (Kozlowski et al., 2020), and gene duplications/deletions (i.e. gene copy number variants; Castagnone-Sereno et al., 2019). *Meloidogyne exigua* is a successful pathogen on coffee with a parthenogenesis reproduction mode (Triantaphyllou, 1985), and as demonstrated in other RKNs, its adaptation to various conditions may be also favored by above mentioned mechanisms. Here, we report a high-quality genome assembly of the genome of *M. exigua* population 'Mex1'. The assembly represents a valuable molecular resource for future studies of phylogenomics on *Meloidogyne* species. In particular, this will foster comparative genomics to

investigate and understand the evolutionary history of this nematode, the results of which may help in the development of new strategies for its management.

We used long-read Oxford Nanopore Technology (ONT) and short-read Illumina HiSeq sequencing data to generate the genome assembly. The population 'Mex1' was isolated from coffee roots collected in Hacienda Aquiares located in Turrialba, Cartago, Costa Rica (9°56'18.09"N, 83°43'43.86"W). A single juvenile was inoculated and multiplied on tomato (*Solanum lycopersicum* 'Money maker'). The procedures for sequencing of *M. exigua* genome including nematode extraction and purification, genomic DNA extraction and purification, library preparation, and sequencing processes for the ONT and Illumina platforms were as described by Phan et al. (2020). For ONT sequencing, six micrograms of purified DNA were used to produce 3,150,177 raw reads with a total length of 15.16 Gb (N50 length = 13.9 kb; ca. 150-fold genome coverage). The ONT reads were trimmed to remove adapters using Porechop v.0.2.3 (Wick, 2019). Then, sequence was filtered for quality (Q-score ≥ 9) and length ($L \geq$ 500 bp) using NanoFilt v.1.1.0 (De Coster et al., 2018). Finally, 13.75 Gb of trimmed long reads (coverage of 137×) were selected for further analysis. Reads from the Illumina Technology were obtained from 3 µg of gDNA using the HiSeq3000 platform as described by Phan et al. (2020). Paired-end reads of 150 bp were generated (mean insert size = 452 bp), yielding 43.08 million reads (64.6 Gb; ca. 153-fold genome coverage). The quality of Illumina raw reads were assessed using FastQC (Andrews, 2010). Spades v.3.14.1 (Bankevich et al., 2012) and Blobtools v.2.1 (Kumar et al., 2013) were used to identify possible contamination; however, no potential contamination was detected. The Skewer v.0.2.2 software (Jiang et al., 2014) was used to trim reads using a minimum quality score of 30 and a minimum read length of 51 bp. Finally, the reads were error-corrected using Musket v.1.1 (Liu et al., 2013). Finally, 43.01 million trimmed pair-end reads (64.4 Mb, coverage of 152×) were used for the genome assembly.

Table 1: Statistics of the genome assembly for *Meloidogyne exigua* obtained in our study (with Canu; Koren et al., 2017)

Assembly features	*M. exigua* genome
Total #scaffolds	260
Total length (bp)	42,101,073
Largest contig (bp)	3,958,915
N50 (bp)	1,882,513
N90 (bp	1,045,864
L50(# scaffolds)	10
L90 (#scaffolds)	18
GC (%)	25.55
Mismatches	0
Gaps	0
CEGMA completeness[1] (n:248)	C:95.97% (C+P : 97.18%)
BUSCO completeness[2] (n:303)	C:89.4%[S:89.1%,D:0.3%]

[1]C: Complete; C+P: Complete + Partial.
[2]C: Complete; S: Complete and single-copy; D: Complete and duplicated

The mitochondrial genome (mitogenome) of *M. exigua* was *de novo* assembled using short reads following the experimental procedure described by Besnard et al. (2014). Long reads were used to resolve the repeated sequences. A mitogenome sequence of 20,974 bp was constructed with an average coverage of 18,698×. Protein-coding genes were annotated using the protein prediction pipeline of Mitos (Bernt et al., 2013) with the invertebrate mitochondrial code. Transfer RNAs (tRNAs) were annotated using Blastn search against the mitogenomes of *M. graminicola* (NC_024275.1) and *M. chitwoodii* (KJ476150), and were checked manually for the start/stop position. Fourteen protein coding genes (*atp6*, *nad5*, *cox1*, *nad1*, *nad2*, *cox3*, *nad6*, *nad4L*, *cox2*, *rrnL*, *nad3*, *cob*, and *nad4*), two ribosomal RNA (rRNA) genes (*rrnS* and *rrnL*), two repeated regions (102R and 313R), and 22 transfer ribonucleic acid (tRNA) genes (*trnM*, *trnW*, *trnS*$_2$, *trnQ*,*trnA*, *trnR*,*trnV*, *trnE*,*trnS*$_1$, *trnT*,*trnY*, *trnL*$_2$,*trnI*, *trnN*, *trnF*, *trnG*, *trnK*, *trnC*, *trnH*, *trnL*$_1$, *trnP*, and *trnD*) were finally annotated from the mitogenome sequence. The mitogenome structure was visualized using the CIRCOS software (http://circos.ca/) (Fig. 1).

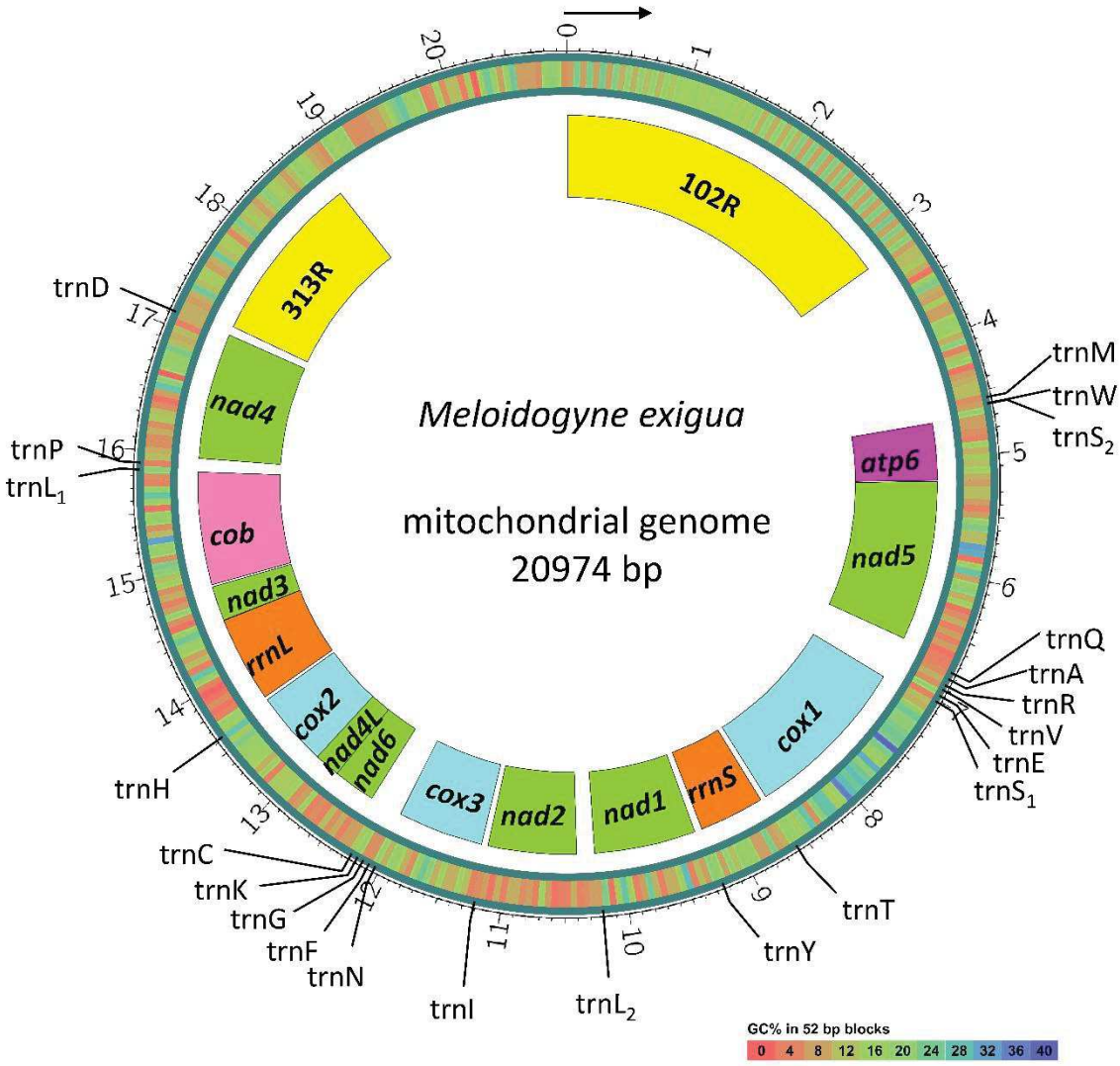


Figure 1. Circular gene map of the complete mitochondrial genome of *Meloidogyne exigua*. Protein-coding genes, rRNA genes, and repeated regions are represented as boxes. Position of tRNAs are indicated by black lines. The direction of transcription of all genes is the same and indicated by the arrow. Abbreviations of protein coding and rRNA genes are: *nadi* = subunit i of NADH dehydrogenase; *coxi* = subunit i of cytochrome c oxydase; *cob* = cytochrome b; *atp6* = subunit 6 of ATP-synthase; *rrnS* = small subunit ribosomal RNA (12S); *rrnL* = large subunit ribosomal RNA (16S). tRNA genes (trnX) are named with a single-letter amino acid abbreviation (X) except for those coding for leucine and serine, which are named as L_1 (anticodon uag), L_2 (uaa), S_1 (ucu), and S_2 (uga). Two minisatellite regions, namely 102R and 313R, are composed of 102-bp and 313-bp repeats.

The reads that mapped to the mitogenome (with 100% identity; CIGAR = 100M) were removed from the cleaned long and short reads datasets and the remaining sequences used for

assembly of nuclear genome. The Canu v.1.8 software (Koren et al., 2017) was first used for the assembly. Subsequently, Racon v.1.4.3 (Vaser et al., 2017) and Pilon v.1.23 (Walker et al., 2014) were used to correct bases and homopolymer lengths. Contigs that had low read-coverage (< 10×) were eliminated from the assembly to avoid artifacts and possible contamination. Finally, the assembled genome consisted of 206 contigs with a maximum contig length of 3,958 Kb and N50 of 1,882 Kb (Table 1). The total length of the assembly is 42.10 Mb, which matches the estimated haploid genome length of 43.2 Mb based on *k*-mer analyses (at $k = 21$) using Jellyfish v.1.0 (Marçais and Kingsford, 2011) and GenomeScope v.2.0 (Vurture et al., 2017) (Table 1, Fig. 2A). Smudgeplot v.0.1.3 (Ranallo-Benavidez et al., 2020) and KMC v.3.0.0 softwares (Kokot et al., 2017) were used to estimate genome ploidy based on the *k*-mers counting ($k = 21$) of the short-read data. The genome is estimated to be diploid (AB) with heterozygosity of 0.03% (Fig. 2B). Blobtools (Laetsch and Blaxter, 2017) was used to assess contaminant DNA presence on the final genome assembly (Fig. 2C). Most of the genome assembly belong to Nematoda phylum (93.1%; Fig. 2C). One scaffold (1.18 Mb) was, however, assigned to the Arthopoda phylum (Fig. 2C). However, sequencing coverage and GC content of this scaffold were similar to other contigs of the genome assembly, and should thus be part of the nematode genome (Fig. 2C). The GC content of the assembled genome was 25.5% (Table 1). The Core Eukaryotic Genes Mapping Approach (CEGMA v.2.5) analysis (Parra et al., 2007) revealed that genome assembly contain 95.75% among 248 Eukaryotic Orthologs. The average number of orthologs per core gene at 1.09 indicated a haploid genome assembly. Besides, the genome assembly was 89.4% complete based on the eukaryote set (n = 303) of Benchmarking Universal Single-Copy Orthologs (BUSCO v.3.0.2) (Simão et al., 2015). Among available *Meloidogyne* genomes, this new assembly yields the second highest BUSCO completeness (after *Meloidogyne javanica*, summarized in Koutsovoulos et al., 2020) and the second largest N50 length (after *Meloidogyne chitwoodi*; Bali et al., 2021). This reference will assist a range of genetic, genomic, and phylogenetic studies to uncover diversity and evolution of *M. exigua* and other related RKNs.

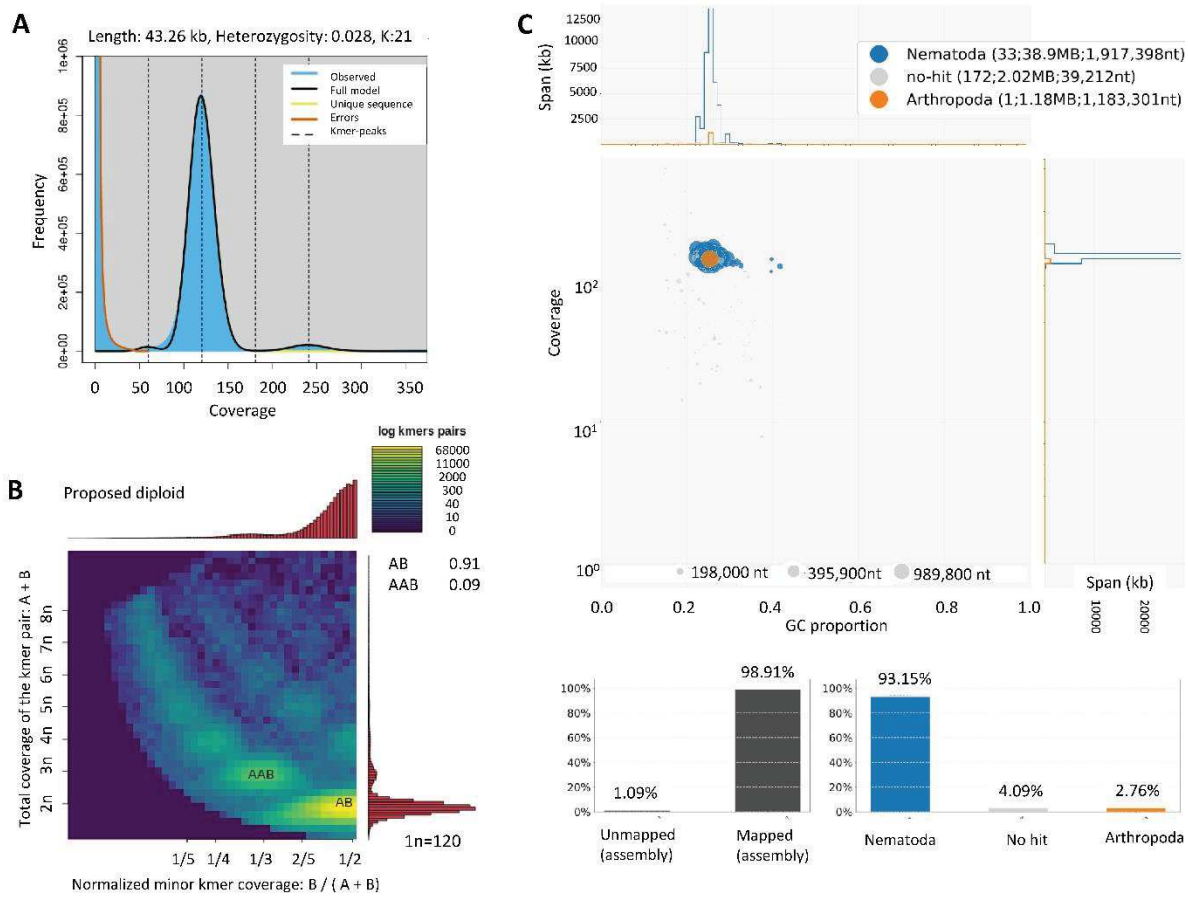


Figure 2. Haploid genome length, genome ploidy estimation and contaminant analysis of the *Meloidogyne exigua* genome assembly. (A) GenomeScope profile showing estimated genome length of 43.26 Mb and heterozygosity of 0.028% at *k*-mer =21. (B) Smudge plots showing the coverage and distribution of *k*-mer pairs that fit to diploid genome model. (C) Blobplot showing the lack of contamination in the final assembly by foreign (non-Nematoda) genetic material

Data availability and accession number(s)

Procedural information concerning the genome assembly and analysis presented in this paper can be found at the GitHub repository at https://github.com/PhanNgan/genome_assembly_mex. The mitogenome and nuclear genome sequences have been deposited in DDBJ/ENA/GenBank under the accession numbers xx and SAMN18905899, respectively.

Literature Cited

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Babraham Bioinformatics, The Babraham Institute, Cambridge, available at: www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed March 15, 2020).

Anzueto, F., Bertrand, B. and Dufour, M. 1995. "Nemaya", desarrollo de una variedad porta-injerto resistente a los principales nemátodos de América Central. Programa Cooperativo para el Desarrollo Tecnológico y la Modernización de la Caficultura (PROMECAFE). Boletín PROCAFE-IICA-CATIE. N° 66-67. http://repositorio.iica.int/handle/11324/7910

Bali, S., Hu, S., Vining, K., Brown, C.R., Majtahedi, H., Zhang, L., Gleason, C., Sathuvalli, V. 2021. Nematode Genome Announcement: draft genome of *Meloidogyne chitwoodi,* an economically important pest of potato in the Pacific Northwest. IS-MPMI Congress 0:0-0.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. and Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19:455–477.

Barbosa, D. H. S. G., Duarte, H., Souza, R., Viana, A. P. and Silva, C. P. 2004. Field estimates of coffee yield losses and damage threshold by *Meloidogyne exigua*. Nematologia Brasileira 28:49–54.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M. and Stadler, P. F. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69:313–319.

Besnard, G., Jühling, F., Chapuis, É., Zedane, L., Lhuillier, É., Mateille, T. and Bellafiore, S. 2014. Fast assembly of the mitochondrial genome of a plant parasitic nematode *Meloidogyne graminicola* using next generation sequencing. Comptes Rendus Biologies 337:295–301.

Campos, V. P. and Villain, L. 2005. Nematode parasites of coffee and cocoa. Pp.529–579 *in* M. Luc, R.A. Sikora, J. Bridge, ed. Plant Parasitic Nematodes in Subtropical and Tropical Agriculture. Wallingford: CABI.

Castagnone-Sereno, P., Mulet, K., Danchin, E. G. J., Koutsovoulos, G. D., Karaulic, M., Rocha, M. D., Bailly-Bechet, M., Pratx, L., Perfus-Barbeoch, L. and Abad, P. 2019. Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita*. Molecular Ecology 28:2559–2572.

Danchin, G. J. E., Guzeeva, A. E., Mantelin, S., Berepiki, A. and Jones, T. J. 2016. Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. Molecular Biology and Evolution 33:1571–1579.

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. and Van Broeckhoven, C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669.

Goeldi, E. A. 1892. Relatoria sobre a molestia do cafeeiro na provincia do Rio de Janeiro. Archivos do Museu Nacional do Rio de Janeiro 8:1–121.

Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M., & Schwander, T. 2020. Genomic features of parthenogenetic animals. Journal of Heredity 1:19–33.

Jiang, H., Lei, R., Ding, S.-W. and Zhu, S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15:182.

Jones, J. T., Haegeman, A., Danchin, E. G. J., Gaur, H. S., Helder, J., Jones, M. G. K., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J. E., Wesemael, W. M. L. and Perry, R. N. 2013. Top 10 plant-parasitic nematodes in molecular plant pathology. Molecular Plant Pathology 14:946–961.

Kokot, M., Długosz, M. and Deorowicz, S. 2017. KMC 3: counting and manipulating *k*-mer statistics. Bioinformatics 33:2759–2761.

Koutsovoulos, G. D., Poullet, M., Ashry, A. E., Kozlowski, D. K., Sallet, E., Rocha, M. D., Perfus-Barbeoch L, Martin-Jimenez C., Frey J. E., Ahrens C. H., Kiewnick S. Danchin, E. G. J. 2020. The polyploid genome of the mitotic parthenogenetic root-knot nematode *Meloidogyne enterolobii*. Scientific Data 7:324.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Research 27:722–736.

Kozlowski, D. K., Hassanaly-Goulamhoussen, R., Rocha, M. D., Koutsovoulos, G. D., Bailly-Bechet, M. and Danchin, E. G. J. 2020. Transposable Elements are an evolutionary force shaping genomic plasticity in the parthenogenetic root-knot nematode *Meloidogyne incognita*. bioRxiv 2020.04.30.069948.

Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M. and Blaxter, M. 2013. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Frontiers in Genetics 4:237.

Laetsch, D. R. and Blaxter, M. L. 2017. BlobTools: Interrogation of genome assemblies. F1000Research 6:1287.

Liu, Y., Schröder, J. and Schmidt, B. 2013. Musket: a multistage *k*-mer spectrum-based error corrector for Illumina sequence data. Bioinformatics 29:308–315.

Marçais, G. and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics 27:764–770.

Muniz, M. de F., Carneiro, R., Almeida, M. R., Campos, V. P., Castagnone-Sereno, P. and Castro, J. M. C. 2008. Diversity of *Meloidogyne exigua* (Tylenchida: Meloidogynidae) populations from coffee and rubber tree. Nematology 10:897–910.

Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P. and Windham, E. 2008. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. Proceedings of the National Academy of Sciences of the United States of America 105:14802–14807.

Parra, G., Bradnam, K. and Korf, I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067.

Phan, N. T., Orjuela, J., Danchin, E. G. J., Klopp, C., Perfus-Barbeoch, L., Kozlowski, D. K., Koutsovoulos, G. D., Lopez-Roques, C., Bouchez, O., Zahm, M., Besnard, G. and Bellafiore, S.

2020. Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*). Ecology and Evolution 10:11006–11021.

Ranallo-Benavidez, T. R., Jaron, K. S. and Schatz, M. C. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nature Communications 11:1432.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Triantaphyllou, A. C. 1985. Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes. Pp. 113–126 *in*: J.N. Sasser, C.C. Carter, ed. An Advanced Treatise on *Meloidogyne*. Biology and Control. Raleigh, North Carolina State University.

Vaser, R., Sovic, I., Nagarajan, N. and Sikic, M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Research 27:737–746.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J. and Schatz, M. C. 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33:2202–2204.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S.K. and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9: e112963.

Wick, R. 2019. Porechop [C++]. Retrieved from https://github.com/rrwick/Porechop (Original work published 2017).