# Genome evolution and adaptations to plant parasitism in nematodes

Etienne Danchin

# Genome evolution and adaptations to plant parasitism in nematodes

Etienne G.J. Danchin

**Habilitation à Diriger des Recherches**

**Ecole Doctorale des Sciences de la Vie et de la Santé (ED85)**

**Université de Nice - Sophia Antipolis**

**Soutenue le 11 Mars 2014, devant le jury constitué de:**

| | | |
|---|---|---|
| **Dr. Jean-Christophe Simon** | Institut de Génétique, Environnement et Protection des Plantes, UMR INRA, Agrocampus Ouest, Université de Rennes I, Rennes, France | **Rapporteur** |
| **Prof. Isa Schön** | Royal Belgian Institute of Natural Science, Freshwater Biology, Bruxelles, Belgique | **Rapporteur** |
| **Dr. Richard Cordaux** | Ecologie et Biologie des Interactions, UMR: CNRS, Université de Poitiers, Poitiers, France | **Rapporteur** |
| **Dr. Pascal Barbry** | Institut de Pharmacologie Moléculaire et Cellulaire, UMR: CNRS, Université de Nice - Sophia Antipolis, Sophia Antipolis, France | **Examinateur** |
| **Prof. Marylène Poirié** | Institut Sophia Agrobiotech, UMR: INRA, Université de Nice - Sophia Antipolis, CNRS, Sophia Antipolis, France | **Examinateur** |
| **Dr. Pierre Abad** | Institut Sophia Agrobiotech, UMR: INRA, Université de Nice - Sophia Antipolis, CNRS, Sophia Antipolis, France | **Examinateur** |

Institut Sophia Agrobiotech – Santé des plantes - Environnement

INRA Science & Impact

Université Nice Sophia Antipolis

CNRS

# Table of Contents

# Index of illustrations



# Index of tables

# Résumé : évolution génomique et adaptations au parasitisme des plantes chez les nématodes

Les nématodes sont des vers ronds non segmentés pour la plupart translucides et mesurant moins d'un mm. Bien que peu visibles, ils constituent l'un des groupes d'animaux les plus riches en espèces et en individus. La majorité des nématodes sont des espèces libres se nourrissant de bactéries ou de champignons mais les espèces les plus connues de l'homme sont les parasites d'animaux et de plantes. Les nématodes parasites de plantes causent des dégâts considérables à l'agriculture mondiale. Ils sont responsables de pertes économiques avoisinant une centaine de milliards d'euros par an. Le parasitisme des plantes est apparu au moins quatre fois indépendamment au cours de l'histoire évolutive des nématodes. Des similitudes morphologiques ont émergé de manière convergente chez ces parasites et il est possible que des singularités caractéristiques se retrouvent également dans leurs génomes.

Le premier génome d'un nématode parasite de plantes, celui du nématode à galles *Meloidogyne incognita* a été séquencé et annoté en 2008 dans le cadre d'un consortium international coordonné par notre équipe. Dès lors, la comparaison avec les génomes d'autres nématodes a permis de révéler des singularités pouvant êtres liées à une adaptation au parasitisme des plantes.

Une partie importante de ma recherche est consacrée à l'identification de singularités dans les génomes de nématodes, en lien avec le parasitisme des plantes. J'ai pu montrer que les nématodes avaient acquis, par transferts horizontaux, des gènes d'origines bactérienne et fongique impliqués dans le succès parasitaire (effecteurs). Ces événements évolutifs, considérés comme très rares chez les animaux, auraient joué un rôle important dans l'adaptation au parasitisme des plantes. La comparaison des génomes de parasites de plantes à ceux d'autres eukaryotes nous a également permis de mettre en évidence un jeu de gènes absents de toutes les autres espèces. Cette spécificité aux phyto-parasites, suggère que ces gènes seraient impliqués dans le parasitisme des plantes. Nos travaux ont montré que l'inactivation de ces gènes atténuait considérablement le succès parasitaire des nématodes. Etant absents des génomes d'autres espèces, ces gènes constituent des cibles particulièrement intéressants pour le développement de nouvelles méthode de luttes plus spécifiques.

Une autre singularité du génome de *Meloidogyne incognita*, concerne sa structure. En effet, le génome est majoritairement présent sous la forme de deux copies divergentes au niveau de leur séquence. Cette caractéristique est peut être liée au mode de reproduction particulier de ce nématode. *M. incognita* se reproduit de manière asexuée et sans méiose. On suppose qu'en absence de méiose, l'équivalent des chromosomes homologues des espèces à reproduction sexuée sont libres de diverger considérablement. Une partie de mes recherches est consacrée à l'étude de l'origine et des conséquences fonctionnelles de cette architecture génomique particulière. La reproduction asexuée est considérée comme une impasse évolutive chez les animaux. Pourtant, *M. incognita* est capable d'attaquer plus de plantes que ses cousins sexués et occupe une aire géographique plus grande. De plus, il est capable de contourner la résistance des plantes et donc de s'adapter. Il est possible que la présence d'une partie des gènes en copies divergentes joue un rôle dans cette plasticité. Analyser les capacités évolutives chez des animaux à reproduction asexuée, en particulier des ravageurs de culture est une thématique que je souhaité développer dans les prochaines années.

# Abstract: genome evolution and adaptations to plant parasitism in nematodes

Nematodes are non-segmented roundworms, mostly transparent and less than 1 mm long. Although nearly invisible, they are one of the most species-rich groups of animals. The majority of nematodes are free-living and feed on bacteria and fungi, but the most notorious species are animal and plant parasites. Plant-parasitic nematodes cause approximately 100 billion Euros in damages to the world agriculture every year. Plant parasitism has appeared at least 4 times independently during the evolutionary history of nematodes. Morphological similarities have emerged in a convergent manner in these parasites and specific genomic singularities might be associated as well.

The first genome for a plant-parasitic nematode, the root-knot nematode *Meloidogyne incognita* was sequenced and annotated in 2008 as part of an international consortium coordinated by our team. Comparisons with the genomes of other nematodes were made possible and revealed idiosyncrasies that might be linked to plant parasitism.

A major part of my research consists in identifying such idiosyncrasies in nematode genomes that could be related to adaptations to plant parasitism. I have shown that nematodes acquired, via lateral gene transfers, a plethora of genes of bacterial and fungal origins involved in successful parasitic interactions with plants (effectors). These evolutionary events, considered as rare in animals, have apparently played an important role in adaptation to plant parasitism. Large scale comparisons of the genomes of plant parasites with those of other eukaryotes allowed us to identify a set of genes absent from non parasitic species. This specificity to parasites suggests that the genes could be involved in plant parasitism. We have shown that, indeed, silencing of these genes significantly reduced the parasitic success of root-knot nematodes. Because they are absent from other species, these genes constitute interesting targets for the development of new and more specific control methods against phyto-nematodes.

Another singularity of the *Meloidogyne incognita* genome resides in its structure itself. Indeed, the genome is mainly present in pairs of copies similar yet divergent at the nucleotide level. This singular structure might be related to the *M. incognita* peculiar reproductive mode. This root-knot nematode reproduces asexually and without meiosis. We suppose that in the absence of sexual reproduction, the equivalents of homologous chromosomes found in sexual species are free to diverge substantially. Sexual reproduction is considered as an evolutionary dead end in animals. Nevertheless, *M. incognita* is able to infest more plants than its sexual cousins and possesses a larger geographic distribution. Furthermore, it is able to overcome plant resistance and thus adapt to changes. Presence of a proportion of genes in divergent copies might constitute a pool for plasticity. Evolution capabilities in animals lacking sexual reproduction, in particular plant pests, is a topic that I will try to develop in the next few years.

# 1 Scientific and agronomic context

Since my arrival, by the end of 2007 as an INRA researcher in Sophia-Antipolis, I have been working in the team "Plants - Nematodes Interactions" (IPN), one of the ten research teams that constitute the laboratory "Institut Sophia Agrobiotech" (ISA). My main research topic has been so far the evolutionary and comparative genomics of plant-parasitic nematodes.

## 1.1 Nematodes

Nematodes or roundworms are non-segmented worm-like creatures. Because most of them are smaller than 1 mm and generally colorless, they are largely ignored by the wider public (most people have never seen any nematode). However, some rare species can reach exceptional sizes and, to date the biggest reported nematode species is *Placentonema gigantissima*. These nematodes parasitize the placenta of sperm whales and the females can grow longer than 8 m and be 2.5 cm thick (Gubanov, 1951). The nematode species which is probably the most famous to life scientists is *Caenorhabditis elegans*. This free-living nematode has become a recognized model for development and molecular biology (Blaxter, 2011). First described in 1900 by Emile Maupas, a French biologist (Maupas, 1901), *C. elegans* has become a model of significant impact to molecular biology, development and ageing because it had been chosen as an easy to rear laboratory model by Sydney Brenner in Cambridge and the subject of intense and detailed research (Plasterk, 2011). Collectively, *C. elegans* scientists have been awarded a total of three Noble Prizes to date. The nematode *C. elegans* is also the first animal to have had its genome sequenced and this is the only animal genome so far to be considered as complete (The C. elegans Genome Sequencing Consortium, 1998). More generally and despite being ignored by a majority of the public, Nematodes are extremely abundant, species-rich and dominate, in number, many ecosystems. Although no more than 23,000 species have been described in the phylum Nematoda, it is estimated that the total number of species may range between 0.5 to more than 10 million (Blaxter, 2011; Lambshead, 1993; Lambshead and Boucher, 2003). Besides free-living species that feed on bacteria (e.g. *C. elegans*), fungi, or are predators of other nematodes, the phylum Nematoda also comprises parasites of vertebrates, invertebrates and plants. These parasites, although also generally small and colorless are notorious by the damages they cause to animal and human health as well as to agriculture.

Nematodes are thought to have emerged in the Paleozoic era from a marine habitat during the Cambrian explosion (550-600 million years ago), but due to the extremely scarce fossil record and poor taxonomic keys to identify them, this remains speculative (van Megen et al., 2009). Although older nematode-like fossils from the Precambrian (Proterozoic era) have been reported (Poinar, 2011), the earliest clear nematode fossil identified to date is from the Paleozoic era. More precisely, this nematode, named *Palaeonema phyticum* has been dated via radiometry to be ca. 396 million years old (Poinar et al., 2008) and thus belong to the Devonian period (360-410 million years ago). Interestingly, this nematode has been found in stomatal chambers of a fossilized early land plant, *Aglaophyton major*, and may thus represent an early plant-associated lineage.

Phylogenetically, nematodes form a group of bilaterian animals named Ecdysozoa, together with arthropods, priapulids and relatives (Aguinaldo et al., 1997). All ecdysozoan have a cuticle that undergoes at least one molt during their life cycle. Because only a few morphological characters,

generally very difficult to distinguish and that can be the result of convergent evolution, exist in nematodes, systematics in the phylum nematoda, based on morphology has been so far unable to resolve deep-branches in phylogenies. Availability of more and more molecular markers, such as small subunit ribosomal DNA (SSU rDNA) has allowed significant advances in assessing the phylogeny of nematodes. In 1998, a phylogeny, based on SSU rDNA from 53 different nematode species provided a first general representation of the phylum nematoda and confirmed that convergent morphological characters emerged during their evolutionary history (Blaxter et al., 1998). Since this reference phylogeny has been published, several enrichments in the number of included nematode species have been released and, in 2009, a phylogeny including SSU rDNA from more than 1,200 nematode species has been published (van Megen et al., 2009). This phylogeny, represents, to date, the most complete view of the nematoda tree of life and has defined 12 different clades within the phylum. This classification will be used as a reference during all the rest of this manuscript and a schematic representation is provided (Figure 1).

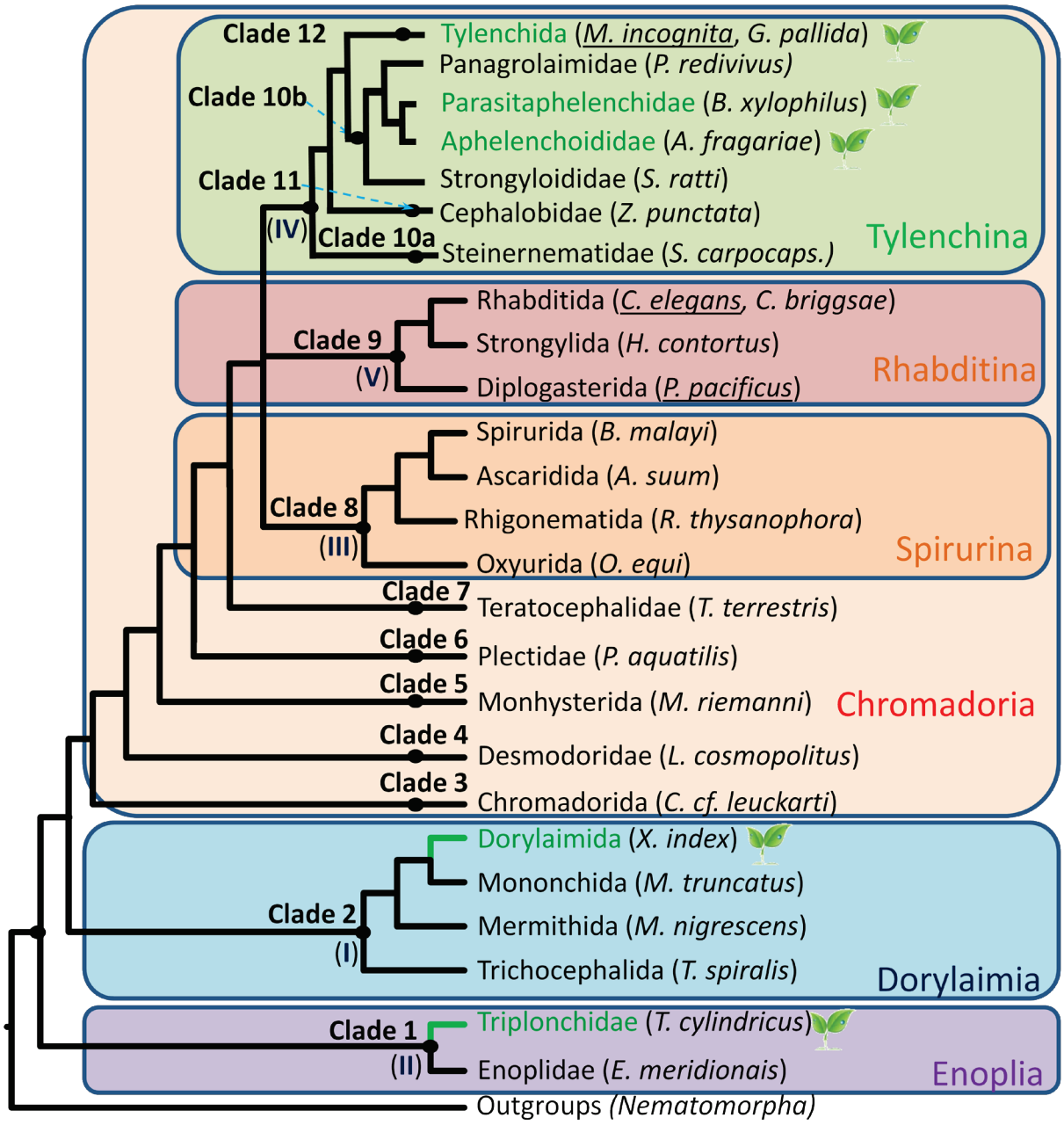



**Figure 1 Schematic phylogeny of Nematoda**

This simplified tree topology is modified from (van Megen et al., 2009) based on SSU rDNA. Clades 1 - 12 are according to the classification proposed by (van Megen et al., 2009). Roman numbers I - V correspond to clades that had been defined in (Blaxter et al., 1998). The three major Nematode lineages Enoplia, Dorylaimia and Chromadoria as described in (De Ley, 2006) are represented by colored rectangles. The Chromadoria lineage is further sub-divided in Spirurina, Rhabditina and Tylenchina. Taxonomic groups in which plant-parasitic species are found are colored in green and highlighted by a leaf symbol. Nematomorpha, a group mainly constituted of parasites of arthropods is the closest outgroup to nematodes.

## 1.2 Evolution of plant parasitism in nematodes

One of the most evident observations that can be made, examining Figure 1, is that plant-parasitic nematodes do not form a single monophyletic clade in the Nematoda phylogeny. Instead, they are spread across the phylogeny and, for instance, they are present within 3 of the 5 clades defined by (Blaxter et al., 1998) and within at least 4 of the 12 clades defined by (van Megen et al., 2009). Furthermore, clades comprising plant-parasitic nematodes are very distant one another. For example the last common ancestor of plant-parasitic nematodes from clade 12 (Tylenchida) and from clade 1 (Triplonchidae) is the last common ancestor of all nematodes. Because plant-parasitic nematode have a patchy distribution in the phylogeny and are interspersed by clades comprising free-living species, predators and parasites of animals, the most likely hypothesis is that plant-parasitism has emerged several times independently during the course of nematode evolution. Currently, it appears that plant parasitism arose at least 4 times independently (i.e. within Clade 1 Triplonchidae, within Clade 2 Dorylaimia, within Clade 10b Aphelenchoidea and within Clade 12 Tylenchida). To make things more complicated, plant-parasitic species generally do not even form a monophyletic group within a given clade. For instance, within the Tylenchida, although most species are plant-parasitic, at the base of this clade lies Hexatylina, a suborder constituted by parasites of insects. Things are more complex in other clades such as Aphelenchoidea. Within this clade, the Apelenchoides genus is mainly constituted by fungivorous nematodes but plant-parasitic (foliar) nematodes like *Aphelenchoides fragariae* or *Aphelenchoides besseyi* are interspersed between fungivorous species (Rybarczyk-Mydłowska et al., 2012). Hence, it is unclear whether plant-parasitism was the ancestral state at the basis of each of these clades or whether this lifestyle has emerged independently several times within the clades themselves. Interestingly, as stated above, the oldest nematode fossil identified so far, *Palaeonema phyticum* was found associated with stomatal chambers from a Devonian plant (Poinar et al., 2008). Hence, it can be hypothesized that this nematode was from a plant-parasitic species and that this lifestyle is at least as old as 396 million of years. Although this early nematode could not be assigned to an exact modern genus, it has been tentatively classified into the Enoplia lineage (Clade 1, Figure 1). This lineage contains known plant parasites (e.g. Triplonchidae) and holds the most basal position in the current Nematoda tree of life, which makes sense considering the age of the early putative plant-parasitic nematode from the Devonian.

Besides their common feeding style, which apparently evolved multiple times independently, plant-parasitic nematodes have at least one common morphological characteristic, the presence of a syringe-like stylet used to inject secretion and pump out nutrient from plant cells. All plant-parasitic nematodes examined at the morphological level so far bear a stylet. However from one clade to another, the exact morphology, ontogeny and structure differs substantially. For instance, at least three types of stylets have been defined: the stomatostyle in Tylenchida (clade 12) and Aphelenchoidea (Clade 10b), the odontostlye in Dorylaimida (clade 2) and the onchiostyle in Triplonchidae (clade 1) (Baldwin et al., 2004). All types of stylets perform similar basic functions: they allow puncturing plant cells and soaking up nutrients. However, fine structural and functional differences exist. For instance, while the stomatostyle of Tylenchida and Aphelenchoidea has a lumen that allows nutrients and secretion to pass through, this is not the case of the onchiostlye. Indeed, in the case of ochionstlyes, nutrients pass directly from the puncture to the stoma along the stylet and not through (Baldwin et al., 2004). This structural difference reinforces the idea of a distinct and independent origin of stylets despite their apparent functional similarities. Furthermore, while bearing a stylet seems to be a prerequisite for plant parasitism, several stylet-bearing

nematodes are not plant parasites. For example, predators of other nematodes like *Labronema ferox* (Clade 2), blood-sucking animal parasites like *Haemonchus spp* (Clade 9) or fungivorous nematodes like *Aphelenchoides spp* (Clade 10b), all bear a stylet used to assist their feeding. Although it has been hypothesized that plant-parasitic Tylenchida (clade 12) evolved from stylet-bearing fungivorous ancestors, due to close outgroup phylogenetic positions, this is not the case for other clades of plant-parasitic nematodes (clade 1 and clade 2) that have no close fungivorous relatives (Bert et al., 2011). Altogether, the scattered phylogenetic positions of plant-parasitic nematodes, the non-homology (in the evolutionary sense) of their stylets and the presence of a stylet in species not associated with plants supports the idea that emergence of this organ as well as of plant-parasitism itself is most probably the result of multiple convergent evolutions.

The mode of plant-parasitism itself is different from one clade to another. Although ecto-parasites, that stay outside plant tissue are found in the four clades (1, 2, 10b and 12), endoparasites are only found within Tylenchida (clade 12). Endo-parasitism of plants includes migratory species like the lesion nematodes (*Pratylenchus spp*) or sedentary species that induce the development of a feeding structure in plant tissues (e.g. root-knot and cyst nematodes).

## 1.3 Economic impact of plant-parasitic nematodes

Plant-parasitic nematodes have a worldwide distribution and are virtually able to infest any human-cultivated plant. Estimating a global economic impact is not straightforward because not only the symptoms of infection are generally not easy to diagnosis but economic data lacks for many countries. According to a book published by George N. Agrios in 2005, the total annual production for all agricultural crops worldwide in 2002 reached about 1,200 billion € (Agrios, 2005). It has been estimated that as much as 430 billion € of production are lost annually because of diseases caused by pathogens, parasites (including nematodes) and weeds. Without crop protection practices used currently, an additional 355 billion € would be lost annually. According to same author, annual losses caused by nematodes on life-sustaining crops, (including all grains and legumes, banana, cassava, coconut, potato, sugar beet, sugarcane, sweet potato, and yam) are estimated to be about 11%. For the rest of economically important crops (vegetables, fruits and non-edible field crops) the loss exceeds 14%. Given these percentages, plant-parasitic nematodes are suspected to cause approximately 65 billion € of production loss annually, worldwide. Another frequently cited reference concerning the economic impact results from an international survey launched in 1987 that collected opinions from 371 nematologists (Sasser and Freckman, 1987). The results of this survey lead to an estimated annual range of economic losses of 62-98 billion €. An extrapolation of this estimate by James McCarter to year 2001 on 40 of the most important crops production volume reached an amount of 92 billion € or 11% of the production (Table 1). It is important to note that damages to non-food plants such as ornamental plants or forest trees were not included, so the total impact might be even higher. Overall, and keeping in mind the uncertainties of these estimates into account, the range of annual worldwide damage to agriculture can reasonably be considered to exceed 90 billion €.

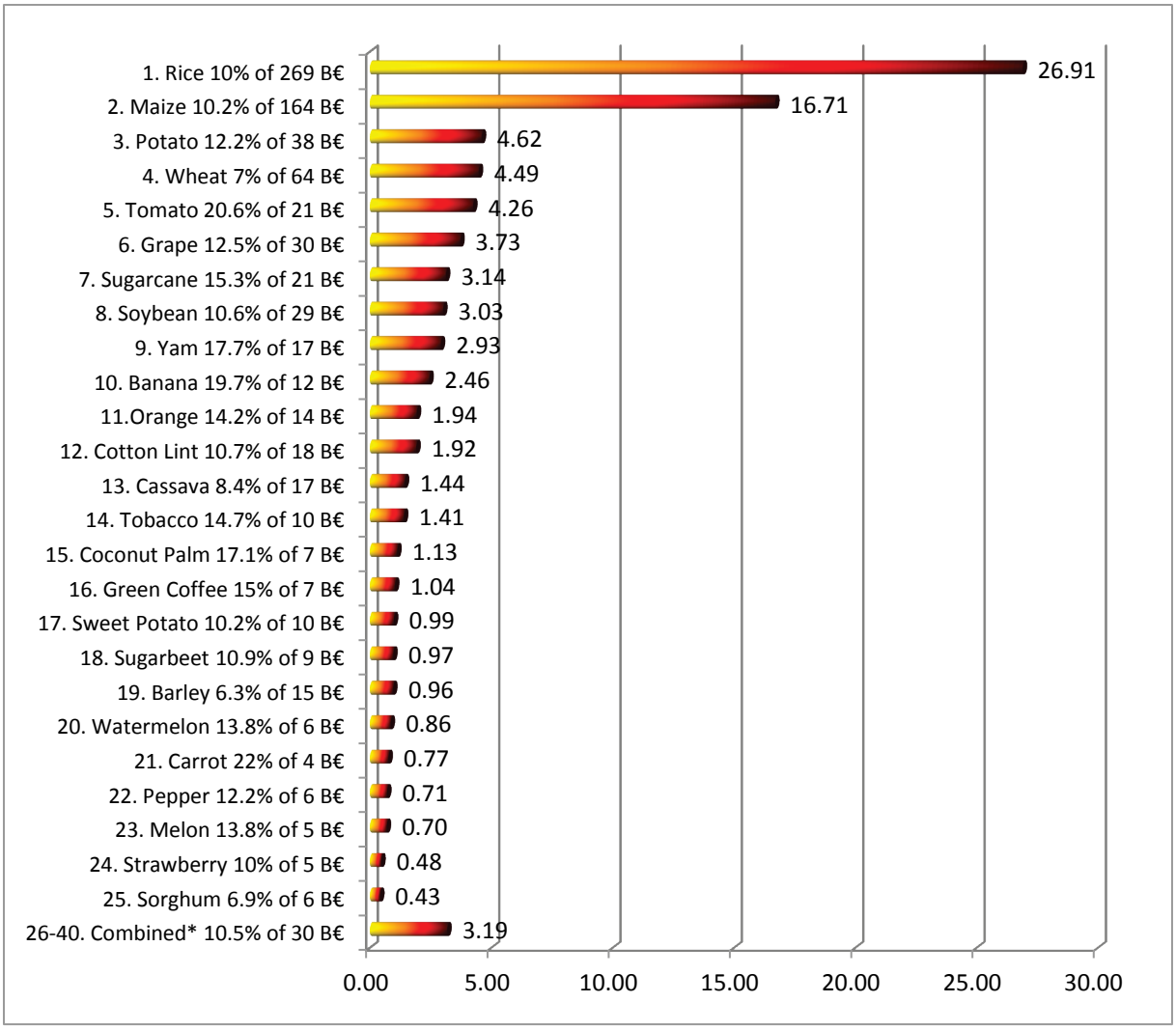



**Table 1 Damages caused by plant-parasitic nematodes to crops annually in billion €**

Estimated 2001 potential global nematode damage in billion € to 40 crops based on the 1987 survey by (Sasser and Freckman, 1987). The top 25 crops are ranked by decreasing total damage. Crops of ranks 26-40 are combined and comprise: millet, cow peas, lemons, chick peas, pineapple, broad beans (green), cocoa beans, tea leaves, eggplant, papaya, oats, pigeon pea, grapefruit, broad beans (dry), and rye. Percent loss due to nematodes range from >20% for Carrot and Tomatoes to <7% for Sorghum and Barley. This table is adapted from (McCarter, 2009).

## 1.4 Root-knot nematodes

Root-knot nematodes (*Meloidogyne* genus) are present worldwide in temperate, tropical sub-tropical and Mediterranean regions as well as everywhere in greenhouses. However, their precise geographical origin is unknown so far. Hence, whether their current distribution is a consequence of recent extensive spread due to human activities and agriculture, or ancient colonization of an old species is still unclear (Trudgill, 1995; Trudgill et al., 1996). These nematodes attack more than 2,000 plant species and are virtually able to infect any cultivated plant of economic interest (Agrios, 2005). Due to their large geographical distribution and huge host spectrum, root-knot nematodes are generally considered as the most damaging plant-parasitic nematodes (Sasser and Freckman, 1987; Trudgill and Blok, 2001)(Jones et al., 2013). Within this nematode genus, it is generally admitted that the following species are of major economic impact: *Meloidogyne incognita*, *Meloidogyne javanica*, *Meloidogyne arenaria*, *Meloidogyne hapla*, *Meloidogyne enterolobii* and *Meloidogyne chitwoodi* (Agrios, 2005). Besides their economic impact, root-knot nematodes are also characterized by their variety of reproductive modes, ranging from 'classical' obligate sexual reproduction to strict parthenogenesis with species are able to alternate between sexual and parthenogenetic reproduction (Castagnone-Sereno et al., 2013).

Nematodes from the *Meloidogyne* genus are, together with the cyst nematodes, the only plant-parasitic nematodes that are fully sedentary endoparasites. They are obligate plant parasites and cannot survive without parasitizing plant root system. Because they depend on live plant material for their development, these nematodes are said to be biotrophs. Biotrophy implies that the nematodes have to overcome plant detection and defense systems for the whole duration of their endoparasitic phase. The root-knot nematodes are able to induce the development of a feeding structure in plant roots from which they feed and which is necessary for the completion of their infective life cycle (Figure 2).

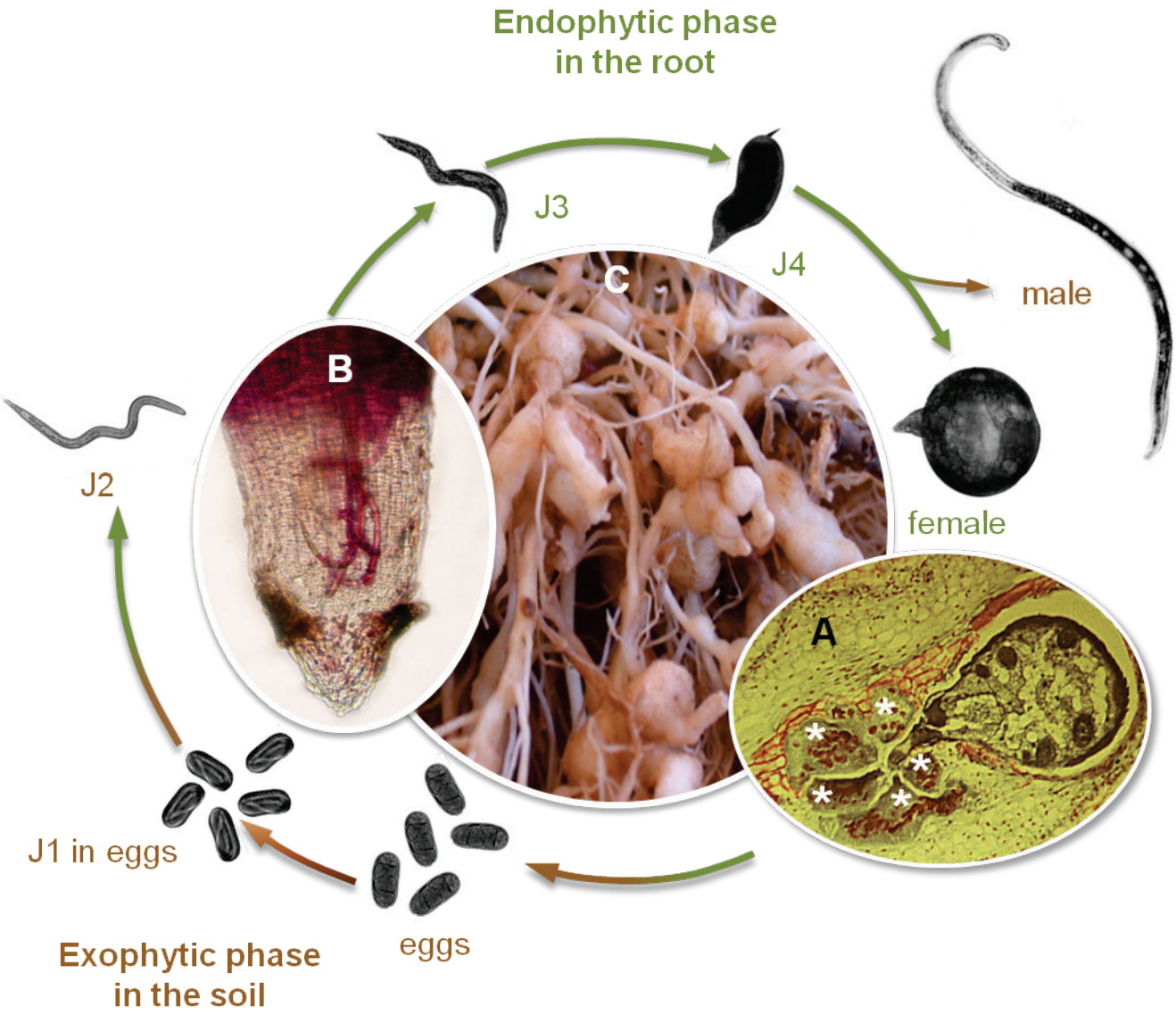



**Figure 2 Life cycle of parthenogenetic root-knot nematodes**

Eggs are laid across the root, embedded in a gelatinous matrix by pear-shaped females (A). A first molt takes place in the egg yielding to stage 1 juveniles (J1). Stage 2 juveniles (J2) hatch from the egg and constitute the infective larvae. These J2s penetrate root tissue near the root tips and migrate towards the apex where they accomplish a U-turn (B). The J2s then navigate in the vascular cylinder and use their needle-like stylet to inject secretions in plant cells. Following this injection, plant cells become multinucleate and giant and will serve as feeding site for the nematode. These giant cells, combined with proliferation of surrounding cells lead to external root galls that form the 'visible' symptoms of an infection (C). The nematode then remains sedentary and undergoes additional molts (J3, J4) until it reaches the adult stage. Adult females then keep inflating and become pear-shaped and filled by hundreds to thousands of eggs. The eggs are extruded from the root under the form of an egg mass and allow repeating the infective cycle. Adult males become vermiform again like J2s and escape the root tissue. Unlike root-knot nematodes with sexual reproduction, in strict parthenogenetic *Meloidogyne*, male genetic material does not contribute to the offspring. This figure is adapted from (Castagnone-Sereno et al., 2013).

## 1.5 Genomes and transcriptomes of plant-parasitic nematodes

Although the genome sequence of the free-living nematode *C. elegans* has been released in 1998 and constituted the first genome for an animal (The C. elegans Genome Sequencing Consortium, 1998), the genomic era reached plant-parasitic nematodes only ten years after. Our team has coordinated genome sequencing and analysis of the root-knot nematode *Meloidogyne incognita*, which has been published and publicly released in 2008 (Abad et al., 2008). This genome constituted several firsts. This was not only the first genome for a plant-parasitic nematode but also the first for an animal able to parasitize plants. Besides plant parasitism aspects, this genome was also the first for an animal with strict asexual reproduction and without meiosis. This singularity will constitute another important topic of my research. Soon after the *M. incognita* genome was released, another root-knot nematode genome was published in 2008. Genome sequencing and analysis of *Meloidogyne hapla*, able to alternate between sexual reproduction and meiotic asexual reproduction has been coordinated and published by our colleagues from North Carolina State University in the USA (Opperman et al., 2008). This second root-knot nematode genome offered the first opportunity for comparative genomics among plant-parasitic representatives of nematodes. To date, only one other plant-parasitic nematode genome sequence has been published in the literature and this is from *Bursaphelenchus xylophilus* (Kikuchi et al., 2011). This nematode is the agent responsible for pine wilt disease and has a complex lifecycle that includes beetles as vector and that comprise a fungivorous phase. It is only distantly related to cyst and root-knot nematodes and probably represents an independent recent evolution of plant-parasitism. No other publication describing the whole genome for an additional truly plant-parasitic nematode has yet been released. However, a draft, partially assembled and un-annotated genome for the soybean cyst nematode *Heterodera glycines* has been sequenced as a collaborative effort between two companies (Monsanto and Divergence) and a US patent has been deposited (Boukharov et al., 2007). The genome of the potato cyst nematode *Globodera pallida* has recently been sequenced at the Wellcome Trust Sanger Institute and a manuscript describing its content is under preparation (Jones JT, personal communication). Overall, the genomic era for plant-parasitic nematodes is thus very recent. However, many other genome projects are ongoing, including in our laboratory (e.g. *Meloidogyne javanica* and *Meloidogyne arenaria*) and we can expect a significant enrichment of available genome data in the next few years as costs for sequencing continually decrease (Kumar et al., 2012).

Before the venue of the genomic era itself, previous large-scale sequence datasets were already available for plant-parasitic nematodes through classical EST analysis. These EST efforts represented partial yet valuable views of the transcriptomes of these nematodes. EST sequences from plant-parasitic nematodes have been released in Genbank's dbEST library since 1997 (Koltai et al., 1997) and as of 2013, there are around 210,000 available sequences, though with lot of redundancy and partial sequences. Our laboratory has collectively produced a total of 63,816 ESTs sequences for *M. incognita* that have been assembled in 22,350 distinct unisequences (Jaouannet et al., 2012). Efforts have been made to collect, prune, organize and assemble EST data and make the results available to the wider community. For instance, NEMBASE, developed in the laboratory of Mark Blaxter in the University of Edinburgh proposes a constantly updated portal to these resources (Elsworth et al., 2011; Parkinson et al., 2004). Similar resources are available in the US through the nematode.net portal (Martin et al., 2011). A shift from classical Sanger sequencing to second generation sequencing (i.e. 454 and illumina) technologies have boosted the throughput of produced transcriptomic data. Consequently, as a complement to the few whole genome sequences already available, some

transcriptomes for plant-parasitic nematodes can also be accessed. For example, the transcriptomes of two root lesion nematodes *Pratylenchus coffeae* and *Pratylenchus thornei* have been published in 2011 (Haegeman et al., 2011a) and 2012 (Nicol et al., 2012), respectively. Concerning the root-knot nematodes, besides the two whole genomes available, the transcriptome of *Meloidogyne graminicola*, a relative able to infect cereals has recently been made available (Haegeman et al., 2013). These transcriptomes were all obtained via 454 technology at lower costs and higher throughputs than classical Sanger EST projects. No whole transcriptome using illumina technology has been published so far, but several projects are ongoing and with the additional coverage and throughput offered by this method, we can expect an acceleration of the process. As part of a collaboration with Genoscope, our group has sequenced the transcriptomes of 4 plant-parasitic nematodes, *P. coffeae*, *Ditylenchus dipsaci*, *B. xylophilus* and *X. index* using illumina technology. These transcriptomes are under analysis in our lab and it can be expected that publications describing the obtained results will be published in the next few years.

# 2 Research Topics

Since my recruitment at INRA in Sophia-Antipolis, most of my research has been focused on the genomes of plant-parasitic nematodes and more intensively on that of the root-knot nematode *Meloidogyne incognita*. This nematode has been the main subject of research of the team Plant-Nematode Interaction (IPN) not only on molecular aspects of the parasitic interactions but also in terms of durability of plant resistance and population studies. When I arrived in this team, by the end of 2007, the genome of *M. incognita* had been assembled and gene predictions were just finished. The challenge was then to constitute a consortium for the functional annotation of this genome, assemble the results and try to write a story reporting the most interesting findings (if any). During my postdoc in the AFMB laboratory in Marseilles (see section 5), just before my recruitment, I had been involved in annotation consortia for a total of 7 fungal genomes, including some for species involved in pathogenic or symbiotic interactions with plants (Amselem et al., 2011; Coleman et al., 2009; Espagne et al., 2008; Ma et al., 2010; Martin et al., 2008; Martinez et al., 2008; Pel et al., 2007). This previous experience surely was beneficial to my involvement in setting up the annotation consortium for *Meloidogyne incognita* and for my contribution to writing the associated genome paper (Abad et al., 2008). Analyzing the genome and assembling the results obviously occupied most of my early years at INRA. Since the publication of the genome paper, my main activity consisted in using comparative and evolutionary genomics to try to identify singularities in the genomes of root-knot nematodes that could be linked to an adaptation to plant-parasitism. Besides representing the first genome for a plant-parasitic animal, the genome of *M. incognita* was also the first for an animal that reproduces exclusively without sexual reproduction and without meiosis (mitotic parthenogenesis). Despite its asexual mode of reproduction, *M. incognita* has a wider host spectrum than its sexual cousins and is able to overcome plant resistance (Castagnone-Sereno et al., 2013). How an animal species can evolve and adapt in the absence of sexual reproduction and inter-individual genetic exchange is a question that occupied a progressively more and important part of my recent research activity.

## 2.1 Genomic signatures of adaptation to plant parasitism in nematodes

With the genome sequence of the root-knot nematode *M. incognita* unveiled (Abad et al., 2008), the possibility to compare the genome of a plant parasite to those of animal parasites and free-living species was open. Analysis of the *M. incognita* genome in a comparative framework allowed identifying a series of singularities. Some of these singularities might just be a consequence of having sequenced a nematode from an as yet unexplored lineage, but, others might reflect adaptation to plant-parasitism or other aspects of the biology of root-knot nematodes. We are of course mainly interested in those singularities linked to biological properties but distinguishing between the two is not always evident.

The initial annotation and comparative analysis of the *M. incognita* genome (Abad et al., 2008) allowed identifying the following singularities:

- **A peculiar genome structure.**

The majority of the *M. incognita* genome is present in two copies presenting substantial divergence at the nucleotide level. Whether this is linked to the mitotic asexual reproduction of this species will be discussed in the section "2.2.1 A genome mainly composed of pairs of regions re-arranged along scaffolds. " and not further detailed here.

- **An unprecedented repertoire of enzymes for the degradation of the plant cell wall.**

A total of 19,212 protein-coding genes were predicted in the genome of *M. incognita*. During the functional annotation, we identified a set of genes coding for enzymes potentially involved in the degradation of the plant cell wall (PCW) and that have higher similarity to microbial genes (Table 2). The abundance of PCW-degrading enzymes was unprecedented in animals and high similarity to bacterial genes suggested an acquisition via lateral gene transfers (LGT). Investigating the role of LGT in the emergence of plant-parasitism has constituted an important part of my research that will be detailed in the section "2.1.1 Lateral gene transfers".

- **A set of "orphan" genes (lacking evident homology to other genomes).**

In an initial comparative analysis of the *M. incognita* gene set against those of 7 other species (including the free living *C. elegans* and the animal parasite *B. malayi*), we found that an important part of *M. incognita* predicted genes had no evident homologs in the other species. Some of these genes might represent true orphans (specific to root-knot nematodes). We can hypothesize that among genes specific to root-knot nematodes; some are involved in functions supporting plant-parasitism. Distinguishing those genes of interest from the rest of orphan genes has also constituted an important part of my research which will be further detailed in the section "2.1.2 Parasitism genes specific to nematodes".

| Substrate: | **Cellulose / Xylan** | | **Pectin / Pectate** | | **Arabinose** | |
|---|---|---|---|---|---|---|
| Species / Enzyme families | GH5_2 | GH30 | GH28 | PL3 | GH43 | Total |
| ***Meloidogyne incognita*** | **21** | **6** | **2** | **30** | **2** | **61** |
| ***Meloidogyne hapla*** | **6** | **1** | **2** | **22** | **2-3** | **33-34** |
| *Pristionchus pacificus* | 0 (7)* | 0 | 0 | 0 | 0 | 0 (7)* |
| *Caenorhabditis elegans* | 0 | 0 | 0 | 0 | 0 | 0 |
| *Brugia malayi* | 0 | 0 | 0 | 0 | 0 | 0 |
| *Drosophila melanogaster* | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 2 Putative plant cell wall-degrading enzymes in nematode genomes**

Plant parasitic nematodes are represented in bold and the other nematodes are free-living or animal parasites. The fruit fly *D. melanogaster* was used as an outgroup. *Cellulases from CAZy family GH5 have been found in *P. pacificus*. The cellulase activity has been experimentally confirmed. However, these *P. pacificus* cellulases are from a subfamily distinct from that of root-knot nematodes and more resemble protists cellulases than bacterial cellulases (Dieterich et al., 2008; Mayer et al., 2011; Rodelsperger and Sommer, 2011).

- **A high proportion of repeats and mobile elements.**

Annotation of repeats and transposable elements in *M. incognita* revealed that they cover 36% of the genome. This proportion is substantially higher than observed in *C. elegans* (Stein et al., 2003) and whether this might be related to the asexual mode of reproduction of *M. incognita* will be further discussed in the section "2.2.2 A high proportion of mobile elements and repeats. ".

- **A reduced gene set for detoxification, immunity and defense against bacteria and fungi.**

Comparison of the repertoire of genes putatively involved in defense / detoxification and immunity in *M. incognita* and *C. elegans* revealed marked differences. While orthologs to almost all genes involved in *C. elegans* innate immunity pathways were found in *M. incognita*, entire categories of immune effectors or detoxification genes are depleted. For instance, considering detoxification, while 44 Glutathione S-transferases (GST ) belonging to three classes (Omega, Sigma and Zeta) have been identified in *C. elegans* (Lindblom and Dodd, 2006), only five GSTs were found in *M. incognita* and all belonged to the sigma class (Table 3). Similarly to the reduced set of detoxification genes, we observed that, compared to other nematodes, *M. incognita* had a very low number of genes encoding chitinases and chitin-binding proteins (Table 4). Chitinases can be used by nematodes as antifungal enzymes (the fungal cell wall is chitin-rich). Anti-bacterial peptides and lysozymes were also found to be depleted in *M. incognita* compared to *C. elegans*. Overall, this reduced arsenal of genes putatively involved in detoxification or defense in *M. incognita* may represent an adaptation to the parasite's habitat. Indeed, life within the plant root tissue can be seen as an environment preserved from usual nematode parasites and pathogens.

| Function | Mode of action / Family name | Number of genes in *M. incognita* | Number of genes in *C. elegans* as referred to public databases |
|---|---|---|---|
| **Antioxidant** | catalase | 3 | 3 |
| | Peroxiredoxin | 7 | 3 |
| | Superoxide dismutase | 3 | 5 |
| | Copper chaperonin | 2 | 1 |
| | Glutathione peroxidase | 2 | 6 |
| | Glutathione synthetase | 4 | 1 |
| **glutathione-S-transferase** | GST class sigma and sigma-like* | 5 | 21 |
| **cytochrome P450** | CYP13 | 6 | 14 |
| | CYP23 | 1 | 1 |
| | CYP25-like | 1 | 6 |
| | CYP31 | 2 | 4 |
| | CYP32 | 3 | 1 |
| | CYP33-like | 11 | 17 |
| | CYP42 | 2 | 1 |
| | partial CYP | 1 | NA |

**Table 3 Putative detoxification genes in *M. incognita* and *C. elegans***

*Besides GST from sigma class, *C. elegans* also possesses GST from Omega and Zeta classes and totals 44 predicted GSTs.

| | Degradation | | | | Binding | | |
|---|---|---|---|---|---|---|---|
| Family / Species | GH18 | GH19 | GH20 | Total | CBM14 | CBM18 | Total |
| ***M. incognita*** | **2** | **2** | ***4*** | **8** | **7** | **0** | **7** |
| ***M. hapla*** | **2** | **1** | **1** | **4** | **3** | **0** | **3** |
| *B. malayi* | 6 | 0 | *3* | 9 | 5 | 0 | 5 |
| *C. elegans* | 38 | 5 | *5* | 48 | 41 | 7 | 48 |
| *P. pacificus* | 14 | 3 | *5* | 22 | 19 | 4 | 23 |
| *D. melanogaster* | 12 | 0 | *4* | 16 | 214 | 0 | 214 |

**Table 4 Putative chitinases and chitin-binding proteins in nematode genomes**

Plant parasitic nematodes are represented in bold and the other nematodes are free-living or animal parasites. The fruit fly *D. melanogaster* was used as an outgroup.

### 2.1.1 Lateral gene transfers

Lateral gene transfer (LGT) is the transmission of genes by means other than direct (vertical) inheritance from the parental generation to the offspring. This phenomenon has been largely documented in prokaryotes, and for instance, resistance to antibiotics as well as pathogenicity factors are easily exchanged horizontally between bacteria regardless whether the gene is present in the bacterial genome itself or plasmid-borne (Boucher et al., 2003; Gogarten and Townsend, 2005). It has long been considered that LGT occurs very rarely in eukaryotes and particularly in animals. This phenomenon has neither been considered as having significantly contributed to the making of an animal genome nor to its biology. However, numerous cases of LGT from prokaryotes to eukaryotes have been reported (Andersson, 2005), including from bacteria to animals (Dunning Hotopp, 2011), suggesting that this phenomenon is more prevalent than thought so far. Because, many reported cases of LGT either led to genes that are not expressed in the receiver animal or that perform no evident function, it has been hypothesizes that they were probably biologically insignificant (Blaxter, 2007). However, a survey of the literature allowed revealing some cases of LGT from micro-organisms to animals with clear link between the transferred genes and a biological function in the receiver organism. As discussed in (Danchin, 2011; Danchin and Rosso, 2012), LGT to nematodes obviously belong to this category and estimating their importance in the emergence of plant-parasitism has been the subject of an important part of my research.

The history of LGT and plant-parasitic nematodes really started with the first discovery of cellulase genes in animal in 1998. Indeed, functional cellulases, secreted by cyst nematodes were identified and their absence from animal genomes available at that time combined with highest similarity to bacterial cellulases suggested that they have been acquired via LGT (Smant et al., 1998). Far from being negligible in these plant-parasitic nematodes, the cellulases were shown to be involved in the degradation of the cellulose composing the protective plant cell wall. The plant cell wall, mainly composed of various oligo and polysaccharides, acts as a physical barrier against aggressions by parasites and pathogens and plant-parasitic nematodes need to break or soften this barrier at several occasions during their life cycle. Following identification of cellulases in cyst nematodes, cellulases were first identified in root-knot nematodes as a result of research efforts led by Marie-Noëlle Rosso in our team (Ledger et al., 2006; Rosso et al., 1999). Besides cellulases, a series of other plant cell wall-degrading enzymes were progressively identified in plant-parasitic nematodes, including xylanases (Mitreva-Dautova et al., 2006), pectate lyases (Doyle and Lambert, 2002) and polygalacturonases (Jaubert et al., 2002). Completing this arsenal of enzymes, expansin-like proteins that putatively act on non-covalent bonds in the plant cell wall were also identified (Qin et al., 2004). However, in the absence of an available whole genome sequence for a plant-parasitic nematode at that time, the total arsenal of proteins involved in the degradation / softening of the plant cell wall was impossible to assess.

All the enzymes involved in degradation of plant cell wall oligo and polysaccharides belong to the enclosing category of Carbohydrate-Active enZymes (or CAZymes). CAZymes are classified in classes, families and sub-families in the CAZy database (http://www.cazy.org), according to sequence similarity (Cantarel et al., 2009). Assignment of new proteins, based on their sequence similarities to families and families allows prediction of the putative function. During the annotation of the set of predicted proteins in *M. incognita*, I have been in charge of the detection and annotation of CAZymes in collaboration with the CAZy-team in Marseilles.

**- The repertoire of enzymes active on plant oligo and poly-saccharides in the *M. incognita* genome**

Our annotation of CAZymes in the *M. incognita* genome allowed assessing the full repertoire of plant cell wall-degrading enzymes in a plant-parasitic nematodes. We found that the root-knot nematode *M. incognita* possesses as many as 61 putative PCW-degrading enzymes (Abad et al., 2008). This repertoire broke all the records in animals and encompassed 5 different families of CAZymes (Table 2). Besides the cellulases, xylanases, pectate lyases and polygalacturonases, we also found candidate arabinanases (GH43 family, possibly involved in degradation of PCW sugars) and candidate invertases (GH32 family). Invertases are not PCW-degrading enzymes, but they can be involved in the degradation of sucrose, the major sugar form circulating in plants, into glucose and fructose, that can be readily metabolized by nematodes. The presence of a bacterial-like invertase in *Bombyx mori*, a lepidopteran insect that feeds on plants, further reinforce this possibility (Daimon et al., 2008). Besides these enzymes, we also identified in the *M. incognita* predicted proteins, a total of 20 expansin-like proteins that can assist degradation / softening of the PCW by enzymes. A simple BLAST (Altschul et al., 1997) analysis of these different proteins active on plant sugars showed that their highest similarities were to bacterial and fungal proteins (Table 5). Higher similarity to microbial proteins suggested that the corresponding genes had been acquired via LGT.

| Predicted activity | CAZy family | Abundance in *M. incognita* | Best hits in CAZy |
|---|---|---|---|
| Cellulase | GH5 | 21 | Tylenchina, Proteobacteria, Cytophaga, Firmicutes, Coleoptera Insecta |
| Xylanase | GH30 | 6 | *M. incognita*, Firmicutes and Gamma Proteobacteria |
| Arabinanase | GH43 | 2 | Actynomycetales, Fungi, Gamma Proteobacteria |
| Polygalacturonase | GH28 | 2 | Tylenchida ,Gamma and Beta Proteobacteria |
| Pectate Lyase | PL3 | 30 | Tylenchida, Actynomycetales, Fungi |
| Expansin | EXPN | 20 | Tylenchida, Actynomycetales, Fungi, Delta proteobacteria |
| Invertase | GH32 | 2 | Rhizobium Proteobacteria |

**Table 5 Highest similarities to *M. incognita* proteins active on plant sugars.**

Animal species appear in red, bacteria in blue and fungi in green. In most cases the only significant hits in animals are against *M. incognita* itself or other plant-parasitic nematodes from Clade 12 (Tylenchida).

**- Origin and evolutionary history of plant cell wall-degrading enzymes in *M. incognita***

Although BLAST-based similarity results suggested a possible acquisition via LGT for the genes encoding PCW-degrading enzymes, only a precise phylogenetic analysis would be able to decipher the evolutionary history of these proteins. Determining whether the genes were acquired via LGT was an important point because, if confirmed, this would suggest that LGT have played an important role in the evolution of plant-parasitism in nematodes. We thus decided to perform a systematic phylogenetic analysis for each of the proteins and protein families presented in Table 5. This analysis has been the occasion to apply phylogenetic techniques and evolutionary concepts that I had learned during my PhD in the laboratory of Pierre Pontarotti (5 Curriculum Vitae). We first started to constitute a collection of homologs to the different protein and protein families in public databases and in the CAZy database. Systematic phylogenetic analysis, using maximum likelihood and Bayesian approaches allowed to draw a series of conclusions (Danchin et al., 2010):

- *Confirmation of closest homology to bacterial and fungal proteins*

Xylanases, pectate lyases, polyglacturonases, candidate arabinases and expansin-like proteins all showed closest homology to bacterial or fungal proteins. The topologies indicating closest homology to microbial proteins were highly supported and their likelihoods were significantly higher than those of alternative topologies. The only exception was for cellulases. Here, the tree topology showed closest homology to cellulases of two insects that feed on plant material then immediately after, to bacteria. One possibility is that cellulase genes from bacteria not yet sampled in sequence database have been transferred twice. One transfer happened in plant-parasitic nematodes, and another in phytophagous insects. The otherwise absence of cellulase in the many non-phytophagous insects sampled in sequence databases argues in favor of a transfer in insects. However whether these cellulases are ancestral in insects or have been acquired via LGT is still the subject of debate (Watanabe and Tokuda, 2010).

- *Elimination of the probability of a bacterial contamination*

Because the identified proteins active on the PCW in *M. incognita* generally present highest similarity to bacterial proteins, it was important to rule out the hypothesis of a contamination. Availability of the genome sequence allowed extracting solid arguments against the possibility of contamination. First, the corresponding genes contain from one to numerous spliceosomal introns, a feature distinctive from bacterial genes. Second, these genes have been assembled in the *M. incognita* genome on contigs where they are interspersed with true evolutionarily conserved nematode genes. Thus it is clear that these proteins are well present in the endogenous *M. incognita* genome and not encoded by symbiotic or associated bacteria.

- *Adaptation to the M. incognita genome characteristics*

Besides the gain of introns, genes of microbial origin have adopted some features of the *M. incognita* genome. For instance, it appears that genes acquired via LGT cannot be differentiated from the rest of *M. incognita* protein-coding genes based on their GC content and codon usage. It seems that the genes have been domesticated by the nematode genome.

- *Age of acquisition in plant-parasitic nematodes*

Although, at that time, some genes were identified only in root-knot nematodes (e.g. GH43, GH28 and GH32), others were shared by different lineages of plant-parasitic nematodes in clade 12. For instance, pectate lyases, cellulases and expansins are also present in cyst nematodes and many of these genes can also be found in Radopholus and Pratylenchus. Monophyly of PCW-degrading proteins from root-knot nematodes, cyst-nematodes, Pratylenchus and Radopholus strongly suggests that they have been acquired early in a common ancestor of all these nematodes. For cellulases, an ancestral acquisition in clade 12 plant-parasitic nematodes was recently confirmed (Rybarczyk-Mydlowska et al., 2012).

- *Some genes have undergone massive duplications since their acquisition*

Cellulases, pectate lyases and expansin-like proteins are present in 21, 30 and 20 copies, respectively, in *M. incognita*, suggesting that the corresponding genes have undergone massive duplications and now form multigene families. Analysis of the tree topologies confirms that duplications have started early, at least in an ancestor of clade 12 plant-parasitic nematodes and have since continued to duplicate in the different descending species. We hypothesize that gene duplications have been subject to selection and that nematodes harboring more copies of these genes had a selective advantage, regardless their mode of reproduction.

- *Multiple independent transfers from different sources probably occurred.*

Analysis of the distribution of genes encoding plant cell wall-degrading enzymes along the genome revealed an absence of genomic clusters grouping genes from different CAZy families in a same genomic location (a same region in a same scaffold). In contrast, genes for degradation of the plant cell wall appeared scattered around different scaffolds and surrounded by evolutionarily conserved nematode genes not predicted to be involved in degradation of plant sugars. Furthermore, comparing the tree topologies obtained for the different CAZyme families showed that microbial proteins most closely related to those of nematodes were neither from the same species nor from a same lineage. These two features argue for multiple independent acquisitions from different species of micro-organisms.

Altogether, this ensemble of results suggests that nematodes have borrowed genes from micro-organisms that now play important function regarding plant-parasitism (e.g. degradation of the plant cell wall). LGT has thus probably played an important role in the emergence of a plant-parasitic lifestyle. Publication of our article in PNAS (Danchin et al., 2010) received interesting feedback in the Press, and, for instance, nematodes have been consequently considered as "gene stealers" (Figure 3).

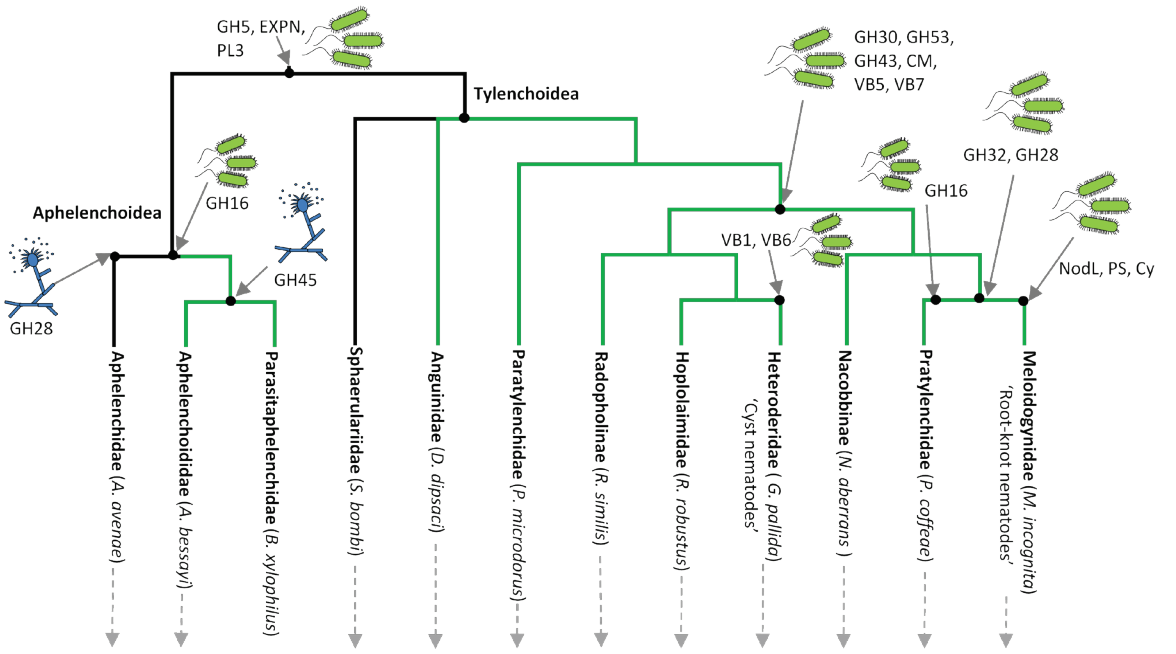

**Figure 3 Plant-parasitic nematodes viewed as gene-stealers**

This illustration from a French magazine (Dubon, 2011), was inspired by our results published in PNAS in 2010 describing lateral gene transfers in plant-parasitic nematodes (Danchin et al., 2010). Here, two nematodes are represented as burglars that have stolen genes from bacteria to support plant-parasitism. Detective Columbo is disappointed because no law punishing gene stealers exists.

**- The importance of lateral gene transfers in the plant-parasitic ability of nematodes**

The plant-cell wall-degrading enzymes constituted a clear example of acquisitions via LGT in the genome of a nematode with evident roles in plant parasitism. However, the total contribution of foreign genes to plant-parasitic capacity in nematodes might not be restricted to the degradation of the plant cell wall. Together with Annelies Haegeman and John T. Jones, two nematologist colleagues also interested in LGT and evolution of plant parasitism, we performed a scan of the literature to gather all the reported cases of LGT in plant-parasitic nematodes and tried to decipher whether processes other than degradation of the plant cell wall could be associated to these transfers. Our results published in the journal MPMI (Haegeman et al., 2011b) showed that besides plant cell wall degradation, LGT have also contributed to other important processes supporting parasitism. We have been able to identify three additional processes in which genes acquires via LGT putatively play roles: (i) suppression of host defenses (e.g. Cyanate Lyases and Chrorismate Mutases), (ii) nutrient processing (e.g. candidate invertase and enzymes from the salvage pathway of vitamins B1, B5, B6 and B7) or (iii) establishment of a feeding structure (e.g. NodL like).

Reporting the presence / absence pattern of the corresponding genes on a simplified phylogeny of plant-parasitic nematodes as well as the putative donor, according to the literature, allowed us to reconstruct a possible scenario of acquisition (Figure 4). The majority of genes acquired via LGT are most similar to bacterial genes, suggesting they have been acquired from bacteria. Genes of possible fungal origins are found in Aphelenchoidea (e.g. GH45 cellulases and GH28 polygalacturonases). Interestingly, fungivorous nematodes are frequent in Aphelenchoidea. Whenever a gene of candidate bacterial or fungal origin was present in multiple different plant-parasitic nematode species, we assumed it was present in the last common ancestor of these species. We thus plotted the putative origin of the LGT on the simplified nematode phylogeny (Figure 4). Some genes like GH5 cellulases, PL3 pectate lyases or expansin like proteins appear to be shared by many plant-parasitic nematodes across the phylogeny and they may have been acquired very ancestrally. Other genes, like GH32 candidate invertases, which had only been found in root-knot nematodes so far appear more specific to a lineage and may have been acquired more recently. It should be noted that the scenario reported on Figure 4, was according to the state of the literature in 2011 and it can change as new cases of LGT are reported (particularly from upcoming genomes). The picture has already been slightly evolving since 2011 and, for instance, GH5 cellulases have been identified in *Aphelenchoides fragariae* (Fu et al., 2012) while candidate GH32 invertases, so far considered as specific to root-knot nematodes have recently been identified in the genome of the cyst nematode *G. pallida* (J.T. Jones, personal communication).




| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GH5 | Y | .(Y) | . | . | Y | . | Y | . | Y | . | Y | Y | PCW |
| GH45 | . | Y | Y | . | . | . | . | . | N | . | . | N | |
| GH30 | . | . | . | . | . | . | Y | . | N | . | Y | Y | |
| EXPN | Y | . | Y | . | Y | . | . | . | Y | . | Y | Y | |
| GH28 | Y | . | . | . | ? | . | . | . | N | . | Y | Y | |
| GH53 | . | . | . | . | . | . | . | . | Y | . | Y | N | |
| GH43 | . | . | . | . | . | . | . | . | Y | . | . | Y | |
| PL3 | Y | . | Y | . | . | . | . | . | Y | . | Y | Y | |
| GH16 | Y | Y? | Y | . | . | . | . | . | N | . | Y | N | FCW |
| CM | . | . | . | . | . | . | . | . | Y | . | Y | Y | Defense sup. |
| CyanL | . | . | . | . | . | . | . | . | N | . | . | Y | |
| GH32 | . | . | . | . | . | . | . | . | N | . | Y | Y | Nutrient process. |
| VB1 | . | . | . | . | . | . | . | . | Y | . | . | N | |
| VB5 | . | . | . | . | . | . | . | . | Y | . | . | Y | |
| VB6 | . | . | . | . | . | . | . | . | Y | . | . | N | |
| VB7 | . | . | . | . | . | . | . | . | Y | . | . | Y | |
| PolS | . | . | . | . | . | . | . | . | N | . | . | Y | |
| NodL | . | . | . | . | . | . | . | . | ? | . | . | Y | Feed. str. |

**Figure 4 Roles and origin of genes acquired via LGT in plant-parasitic nematodes**

A 'Y' in the table indicates that the gene has been reported, a '.' indicates that the gene has not yet been reported and an 'N' indicates that the gene has been searched in genome sequences but not found. Gene family names are as follows: GH5 and GH45: cellulases; GH30: xylanases; EXPN: expansin-like proteins; GH28: polygalacturonases; GH53: candidate arabinogalactan endo-1,4-beta-galactosidases; GH43: candidate arabinanases; PL3: pectate lyases; GH16: beta-1,3-glucanases; CM: chorismate mutases; CyanL: candidate cyanate lyases; GH32: candidate invertase; VB1, VB5, VB6 and VB7: B1, B5, B6 and B7 vitamins biosynthetic and salvage pathways genes; PolS: polyglutamate synthase; NodL: similar to rhizobial nodulation factor NodL. Putative processes in which genes are involved are indicated in the last column. PCW and FCW stand for plant and fungal cell wall degradation, respectively. Species names are as follows from left to right: *Aphelenchus avenae, Aphelenchoides bessayi, Bursaphelenchus xylophilus, Sphaerularia bombi, Ditylenchus dipsaci, Paratylenchus microdorus, Radopholus similis, Rotylenchus robustus, Globodera pallida, Nacobbus aberrans, and Pratylenchus coffeae.*

Our survey of the literature further highlighted the potential importance of LGT in the emergence of plant-parasitism in nematodes. It appears that genes from putative bacterial and fungal origin have been inserted in the genomes of plant-parasitic nematodes and that they have been functionally integrated. Most of these genes have been originally identified in EST sequences, including from nematode secretory glands or identified in proteomics analysis of secretions, indicating that they are correctly transcribed, translated and secreted by the nematodes. Furthermore, the function of the gene products, either experimentally characterized or predicted by sequence similarity, allowed to identify several processes that appear crucial to plant-parasitism, besides degradation of the plant cell wall. We thus hypothesized that LGT has been a catalyst for plant-parasitism, providing the nematodes with new abilities that certainly allowed them to access a new ecological niche. Because these LGT events were probably associated with a selective advantage, I hypothesized that they have been themselves under positive selection. Individuals harboring these laterally-acquired genes probably had a significant advantage over their relatives to access new resources. Hence their offspring, harboring the foreign genes, probably invaded the population generation after generation. This strong positive selection associated to the transfers probably was responsible for successful fixation of the foreign gene at the level of populations then of the species itself (Danchin, 2011). We think that with the ensemble of examples of LGT with clear roles in plant-parasitism processes, we have established that this phenomenon, although usually considered as insignificant in animal evolution, has been of crucial importance in nematodes for the emergence of phytoparasitism.

**- The contribution of lateral gene transfers to the genome composition of root-knot nematodes**

Although, it now appears clear that LGT have played a significant role in the biology of plant-parasitic nematodes, including in the root-knot nematodes, their total contribution to the set of protein-coding genes in a nematode genome was unknown. Indeed, most reported cases in the literature so far have been "by-products" of efforts to characterize the set of parasitism genes and do not result from a systematic search for LGT in a genome. The prevalence of LGT and their contribution to bacterial genomes is so high that the concept of a Darwinian tree of life to represent the evolution of bacteria itself has been considered as challenged (Raoult, 2010). In an opinion article we discussed this concept and whether it could apply to animals too (Danchin and Rosso, 2012). We used plant-parasitic nematodes as an example because we knew a variety of biologically significant examples of LGT were reported. We concluded that despite evidences for several LGT in plant-parasitic nematodes, it was not yet possible to assess to what extent these events might challenge a tree-like representation in animal too. A systematic scan for LGT event was much needed to estimate the proportion of genes of foreign origin in a plant-parasitic nematode and whether this can reach ratios comparable to those observed in bacteria.

Starting from this observation, we decided to systematically scan the two available whole root-knot nematode genomes (*M. incognita* and *M. hapla*) to identify genes of putative non-metazoan origin (Paganini et al., 2012). Using a combination of OrthoMCL and BLAST analyzes we identified 609 non-redundant root-knot nematode genes that were absent from model metazoan species and presented higher similarity to non-metazoan genes. Overall, 385 cases were from *M. incognita* and considering redundancy (i.e. species-specific duplications), this represents a total of 680 protein-coding genes or 3.34 % of the total predicted gene set. A significant proportion of the genes in a root-knot nematode

species are thus of potential non-metazoan origin and may have been acquired via LGT. To further confirm with higher support, potential acquisition via LGT, we performed an automated phylogenetic analysis of cases identified via OrthoMCL and BLAST. For this, in collaboration with the laboratory of Pierre Pontarotti in Marseilles, we used the FIGENIX platform that we had developed during my PhD (Gouret et al., 2005). To detect topologies indicative of an LGT event from the obtained phylogenies, we used the program PhyloPattern (Gouret et al., 2009). We considered the general topology represented in (Figure 5) as indicative on a LGT event. We found that 122 phylogenies, containing a total of 513 *M. incognita* genes supported an acquisition via LGT of non-metazoan origin, which represents 2.52% of predicted genes. Hence, overall between 2.52 and 3.34% of protein-coding genes in *M. incognita* have been potentially acquired via LGT. The reason why there were much more genes than phylogenies is explained by the prevalence of gene duplications. A given phylogeny usually contained many *M. incognita* genes within a same monophyletic group together with genes from *M. hapla* and other plant-parasitic nematodes. We observed that 80% of LGT genes present both in *M. incognita* and *M. hapla* have underwent duplications. The large majority of these duplications started in a common ancestor of the two species and continued independently in the two root-knot nematodes after their separation. This tendency for duplication after acquisition via LGT confirmed, at a larger and more general scale, what we had already observed for plant cell wall-degrading enzymes (Danchin et al., 2010).

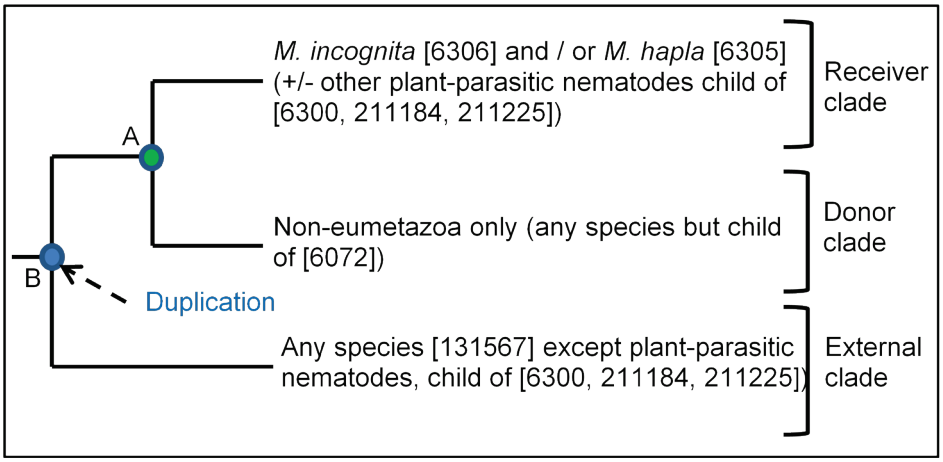



**Figure 5 Topology searched to identify LGT in root-knot nematode genomes**

The topology searched is composed of three main clades. In every clade, species or taxonomic division authorized or forbidden as well as their NCBI's taxonomy identifiers are indicated. The "receiver clade" must contain at least one sequence from *M. incognita* or from *M. hapla* and possibly from other plant-parasitic nematodes. The "donor clade" can contain any species but eumetazoan (e.g. bacteria, fungi, plant,...). The external clade can contain any species but plant-parasitic nematodes. Presence of a node "A" connecting the receiver clade and the donor clade to the exclusion of the external clade is required and constitutes a minimal phylogenetic support for LGT. Stronger support for LGT was assigned when, additionally, a node "B", defined as follows was found. This node "B" must connect node "A" to the external clade and this node must be detected as a duplication node due to presence of at least one non-metazoan species in the external clade.

As an important validation, our method allowed retrieving all previously published cases of LGT in root-knot nematodes, including the plant cell wall-degrading enzymes. However, the majority of LGT cases we identified were novel and we had no a priori idea of their functions. We thus lunched a prediction of function based on Pfam domains and Gene Ontology mapping on the whole predicted proteomes of *M. incognita* and *M. hapla*. We then compared the relative abundance of the different predicted functional categories with those observed in the set of genes predicted as acquired via LGT. We observed that while the *M. incognita* and *M. hapla* whole proteomes had very similar relative abundance of functional categories, the distribution was totally different in the proteins predicted to be of non-metazoan origin. Functional categories related to metabolism, catalytic activity, protein and carbohydrate modification or degradation were substantially more abundant in the set of LGT proteins than in the whole root-knot nematode proteomes. This observation is consistent with previous reports of LGT that, in their majority concerned enzymes. It is also reasonable to imagine that a gene product performing a simple basic activity such as degrading a peptidic or glycosidic bond can more easily perform successfully the same operation in a distant receiver species than a gene product involved in regulatory or gene expression pathways.

Concerning the origin of the genes putatively transferred, we remarked that the majority was from bacteria and fungi, consistent with previous reports of LGT, but we also identified genes of possible archaeal, plant or protist origin. Among candidate donors, a number of plant-pathogens, plant-symbionts and plant-parasitic bacteria and fungi were found. Interestingly, several genes of candidate bacterial origin, including those coding for experimentally-characterized enzymes such as GH30 xylanases or Chorismate mutases were plasmid-borne in bacteria. This suggests that these mobile elements that easily move from one bacterium to another may have played a role in the transfer to the nematode. We also remarked that the density of transposable elements was significantly higher around genes acquired via LGT than around the rest of protein-coding genes. Reinforcing this trend, an abundance of transposable elements around genes acquired via lateral transfer has also been observed in the necromenic nematode *P. pacificus* (Rodelsperger and Sommer, 2011) and in the bdelloid rotifer *Adineta vaga* (Flot et al., 2013). Whether these genomic regions correspond to hotspots of integration of genetic elements or whether transposable elements themselves are involved in the transfer would be interesting to determine. Supporting such possibility, transposable elements can be horizontally transferred between more or less distant phylogenetic groups (Schaack et al., 2010). Although most transfers of TE described so far are between eukaryotes or between prokaryotes, cases of TE transfer from prokaryote to eukaryote have been described (Gilbert and Cordaux, 2013). Recently, compelling evidence for a role of viruses in the horizontal transfer of TE between eukaryotes has been reported (Gilbert et al., 2014). While mechanisms of lateral transfers between prokaryotes are widely documented, evidences for a role of viruses constitute one of the first clearly documented cases of transfer to eukaryotes.

Similarly to what we had already observed for PCW-degradation genes (Danchin et al., 2010), we noticed that the GC content and codon usages of genes potentially acquired via LGT could not be differentiated from codon usage and GC contents of the rest of protein-coding genes. This suggests either an old acquisition followed by a progressive domestication of the transferred genes by the nematode genome, or that the only genes that can be successfully transferred are those that already present the same characteristics than prototypic nematode genes. Presence of multiple spliceosomal introns in the majority of these genes suggests that, in any case, they have undergone sequence / structure modifications since their acquisition.

**- General conclusions on lateral gene transfers**

Our ensemble of studies relative to LGT has contributed to support the idea that this evolutionary phenomenon has played an important role in the biology of plant-parasitic species. We have shown that besides their potential role in supporting a plant-parasitic life style, LGT events have significantly contributed to the genome composition itself, at least in root-knot nematodes. Although the proportion of genes putatively acquired via LGT in the plant-parasitic nematodes does not reach levels comparable to those observed in bacteria, they are still significant and we expect similar importance in upcoming nematode genomes.

### 2.1.2 Parasitism genes specific to nematodes

Parasitism genes are defined here as genes that support plant-parasitism ability in nematodes. It is reasonable to imagine that a proportion of genes specific to root-knot nematodes (i.e. without homologs in other species) is involved in plant parasitism processes. Some of these genes might also represent lineage-specific features with no clear link to plant-parasitism itself. My main objective here was to mine the genomes of root-knot nematodes to identify novel potential parasitism genes. However, differentiating between lineage-specific singularities and those linked to adaptation to plant parasitism without *a priori* is not evident.

During the initial annotation of protein-coding genes in the *M. incognita* genome, we compared the predicted proteome against those of seven other species: three other nematodes (*Caenorhabditis elegans*, *Caenorhabditis briggsae* and *Brugia malayi*); one insect (*Drosophila melanogaster*) and three fungi (*Magnaporthe grisea*, *Giberella zea* and *Neurospora crassa*). We used OrthoMCL (Li et al., 2003) to perform all against all comparisons of the eight different proteomes and group orthologs and species-specific paralogs (in-paralogs) into clusters. We found that out of the 19,212 protein models in *M. incognita*, 9,960 (51.8%) had no predicted orthologs in any other species and were apparently specific to this species (Abad et al., 2008). However, this possibility has to be taken with caution because this results from a comparison with only a limited number of other species. It is possible that some of the apparently *M. incognita*-specific proteins are in fact conserved in other plant parasites. But, because the genome of *M. incognita* was the first for a plant-parasitic animal, the absence of other genomes at that time precluded a more thorough determination of the conservation pattern. Another possibility that has to be taken into account is that some predicted *M. incognita* proteins may be the result of over-prediction from genome annotation software. This hypothesis can be refuted for at least 2,154 predicted proteins whose corresponding gene models received transcriptional support from ESTs (Table 6). These apparently *M. incognita*-specific genes are thus effectively transcribed and among the corresponding proteins, some might be involved in processes linked to plant parasitism.

One classical category of parasitism genes that has been extensively studied in nematodes is effectors. There are almost as many definitions of what is an effector than laboratories working on plant-parasitic nematodes (Haegeman et al., 2012, 2013; Jones et al., 2009; Mitchum et al., 2013; Rosso et al., 2012). The most restrictive definitions state that effectors are proteins secreted by the nematode *in planta* that manipulate plant functions (such as defense or metabolism) to the benefit of the parasite. More pervasive definitions state that an effector is a protein secreted by the nematode and that assist parasitism (in general). This latter definition encompasses plant cell wall-

degrading enzymes and other categories that do not directly manipulate plant function yet participate in successful parasitism. We will adopt here this more general definition of effectors. One characteristic of nematode effectors is that they are secreted *in planta* and, thus, the proteins frequently bear a signal peptide for secretion. We thus searched for predicted signal peptides in the set of *M. incognita* predicted proteins. Interestingly, 2,758 apparently *M. incognita*-specific proteins had a predicted signal peptide (Table 6). Focusing on genes that were supported by EST transcriptomic data, we remarked that 338 encoded proteins were apparently *M. incognita*-specific, have a signal peptide for secretion and lack known protein domain. This set of protein thus represents an interesting pool to identify candidate new effectors specific to root-knot nematodes.

| **19,212 *M. incognita* proteins** | **10,748 with IPR (54.2 %)** | **4,250 with SP (22.1 %)** | **6,858 with ESTs (35.7%)** |
|---|---|---|---|
| **Mi specific 9,960** (51.8%) | with IPR **3,128** (31.4 %) | SP **759** | **202** |
| | | no SP 2,369 | **684** |
| Mi-restricted groups: 6,522<br>Mi not in groups: 3,438 | without IPR 6,832 (68.6 %) | SP **1,819** | **338** |
| | | no SP 5,013 | **930** |
| | | **2,578 with SP (25.9 %)** | **2,154 with ESTs (21.6 %)** |
| **Mi shared 9,252** (48.2%) | with IPR **7,290** (78.8 %) | SP **1,258** | **661** |
| | | no SP 6,032 | **3,330** |
| | without IPR 1,962 (21.2 %) | SP **414** | **137** |
| | | no SP 1,548 | **576** |
| | | **1,672 with SP (18.1 %)** | **4,704 with ESTs (50.8 %)** |

**Table 6 "Orphan" genes in *M. incognita* potentially contain novel parasitism genes**

The top row indicates results of Interpro domain search (IPR), presence of a predicted signal peptide for secretion (SP) and transcriptional support form EST data on the 19,212 protein-coding genes *in M. incognita*. "Mi specific" represents *M. incognita* genes that did not cluster with genes from other species in the OrthoMCL analysis while "MI shared" represents genes that were clustered with at least one gene from another species.

Mining the *M. incognita* genome may allow identifying new parasitism genes. For instance, genes that resemble known effectors and that present no significant similarity to other species. Thus, we decided to investigate this point into more details and launched a more comprehensive analysis, taking advantage of the latest genomes released since the publication of the *M. incognita* genome. The main goal behind this new genome mining project was to identify protein-coding genes in root-knot nematodes that could represent promising new targets for the development of more specific and efficient control methods. One of the main pitfalls of control methods used so far against plant-parasitic nematodes is their poor specificity. Many of the chemical compounds that represented the most efficient nematicides also impact other species, including those that are beneficial to agriculture or have potential consequences on human health and environment. This toxicity and poor

specificity have led to banning of most of these chemicals. To circumvent the problem of poor specificity, one important criterion that we have strictly applied during our genome mining strategy was to identify root-knot nematode genes lacking homolog in non-target species. The basic aim of our project was to identify genes conserved in root-knot nematodes and possibly other plant-damaging species but absent from the genomes of non target species such as those of chordates, mollusks, insect pollinators, annelids or plants. The rationale behind this strategy is that the more a gene is conserved in plant-damaging species but absent from the genomes of non-target species, the more it is likely to be involved in plant-parasitism. Genes potentially involved in parasitism and specific to plant pests constitute target of particular interest for the development of efficient and specific control methods. Indeed, methods or chemicals, targeting these genes would potentially affect parasitism and are supposed not to have any effect on non-target species.

We wrote a grant proposal around this project and obtained a funding from the "Agence Nationale de la Recherche" (ANR) to complete our analysis. The project, that we named NEMATARGETS, included a scan of whole root-knot nematode genomes and a comparative genomics analysis to identify genes absent from non-target species but present in plant-damaging species. Besides bioinformatics aspects, this project also included an important experimental part that consisted in two main tasks: (i) producing additional transcriptomic data to support predicted genes as "true", transcriptionally active nematode genes (ii) testing the effect of inactivation of the genes identified through the bioinformatics mining on the success of parasitism. In this project, I have been mainly involved in supervising the bioinformatics mining process and elaboration of the filtering strategy to identify novel candidate parasitism genes of interest.

Our starting material consisted in the whole sets of predicted proteins in *M. incognita* (19,212) an *M. hapla* (14,421), which collectively represented 34,780 protein models. Our first immediate intention was to eliminate from this set of proteins, those that had predicted orthologs in non-target species. To achieve this objective, we combined an OrthoMCL analysis with the whole proteomes of 23 other species and a BLAST analysis against the NCBI's nr library.

For the OrthoMCL analysis, we selected species with a whole genome available and with a predicted set of proteins of good quality and considered as relatively complete. Our selection of species included 4 other nematodes, 5 insects, 9 vertebrates (including 7 mammals), 4 fungi and one plant (Figure 6). Besides the two root-knot nematodes, we included in our selection, 4 other plant-damaging species (the phytophagous lepidopteran *Bombyx mori*, the pea aphid *Acyrtosiphon pisum*, and two plant-pathogenic fungi, *Fusarium graminearum* and *Magnaporthe grisea*). All the other species were considered as non-targets and every root-knot nematode protein that had a predicted ortholog in any of these non-targets was systematically eliminated. Overall, this step allowed the elimination of 15,181 root-knot nematode proteins.

The rest of RKN proteins (19,599) lacked predicted ortholog in the non-target species and passed this filter. Among remaining proteins, we then eliminated redundancy between *M. incognita* and *M. hapla*. Because subsequent biological assays will be performed in *M. incognita*, we kept as representative the *M. incognita* versions whenever the gene was present in the two RKN. At the end of this OrthoMCL filtering step, a total of 17,153 root-knot nematode proteins were kept for further analysis.

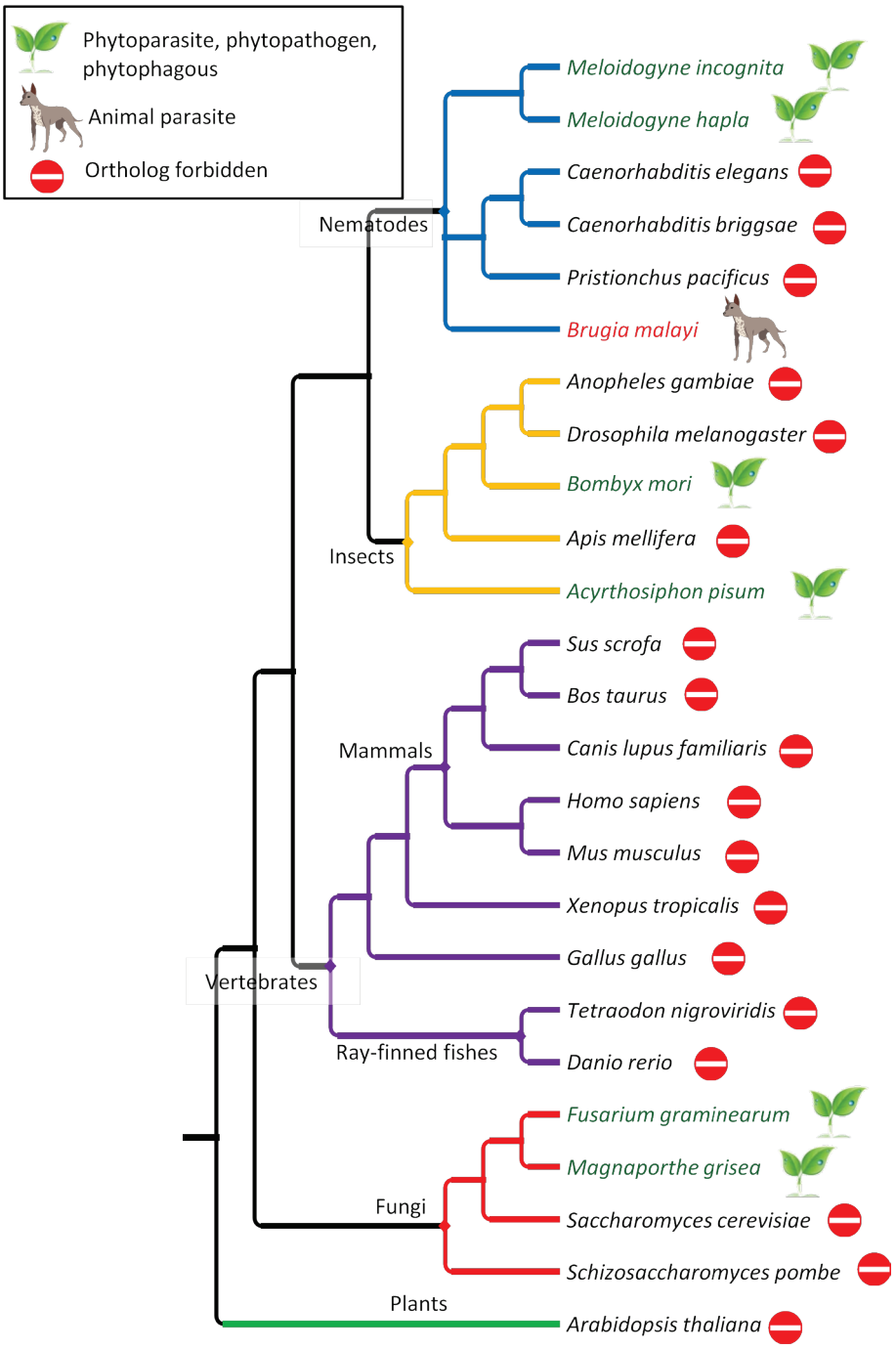



**Figure 6 Species included in the NEMATARGETS OrthoMCL filtering step**

Simplified taxonomy of the 25 species included in the OrthoMCL comparison of whole proteomes. Plant-damaging species are highlighted in green and with a plant symbol. The animal-parasitic nematode *B. malayi* is highlighted in red and with a dog symbol. All the other species, in black, with a "wrong way" sign, are blacklisted.

This OrthoMCL comparative analysis, including a total of 25 species, encompassed more than 500,000 proteins and allowed eliminating ~20,000 root-knot nematode proteins that had predicted orthologs in non-target species. However, many species not represented in this OrthoMCL analysis have at least some associated partial sequence data in public database. To take this data into account, we compared the 17,153 RKN proteins that passed the OrthoMCL filter against the NCBI's nr

library using BLAST. Predicted orthology was determined using % similarity and % query coverage thresholds deduced from one-to-one orthologs obtained during the OrthoMCL analysis. Retrieving taxonomic information from predicted orthologs, we systematically eliminated RKN proteins having putative orthologs in non-target species. Because there is no comprehensive database indicating the lifestyles of the plethora of species with a sequence in the nr library, we generated a list of non-target taxa. In total, our list included 170,258 species covering 4 whole clades (annelida, chordata, mollusca and viridiplantae). Overall, a total of 10,105 RKN proteins did not return any significant BLASTp hit in nr and more than half of these proteins (5,536) also lacked predicted ortholog in the OrthoMCL analysis. These proteins are thus considered as potentially orphan or restricted to RKN. In contrast, 1,201 RKN proteins returned BLASTp hits indicative of potential orthology in at least one non-target species and were thus discarded.

Overall, a total of 15,952 RKN proteins were predicted to be absent from non-target species. In order to state whether some of these proteins were conserved in other plant-damaging species, we filtered OrthoMCL and BLAST results. At this occasion, we built up a list of potentially plant-damaging species from the literature, discussion with plant-pathologist colleagues and existing partial databases (Hamilton et al., 2011; Winnenburg et al., 2008). We identified a total of 5,297 root-knot nematode proteins absent from non-target species but present in at least two different plant-damaging species (including RKN).

Automatic annotation of RKN proteins absent from non-target species revealed a significant over-abundance of putative transcription factors. In total, we identified 190 putative transcription factors in this set of proteins. If these proteins actually function as transcription factors, they represent particularly interesting targets. Indeed, they may be involved in the regulation of expression of various parasitism genes and their inactivation may lead to severely reduced parasitism efficiency. Because they are absent from non-target species, use of chemicals or other inactivation methods against these genes is supposed to be specific and thus safe to other species.

Because effectors are classical parasitism genes whose inactivation / targeting is supposed to impair success of parasitism, we also searched in the ensemble of root-knot nematode proteins, those that share the same characteristics than effectors. As previously introduced, canonical effectors bear a signal peptide for secretion and have no transmembrane region. As part of a collaborative work with Celine Vens from K.U. Leuven (Belgium) we have developed a software, named MERCI (for (Motif—EmeRging and with Classes—Identification) to identify protein motifs specific from a set of sequences of interest (Vens et al., 2011). Using MERCI, we identified protein motifs that were frequent in known nematode effectors but absent from evolutionarily conserved housekeeping genes. In addition to signal peptide and absence of transmembrane region, we thus also searched for presence of these MERCI motifs. Starting from the 17,153 RKN proteins that lacked ortholog in non-target species, we identified 993 candidate effectors that cumulated these 3 characteristics.

Having identified a series of genes of interest for the development of more specific control methods, our main objective was to verify that some of these genes actually play important role in the success of plant-nematode parasitic interaction. The idea was then to extract the most promising of these genes and experimentally check the effect of inactivation of each of these genes, using siRNA on the success of parasitism. We thus started by pruning the set of candidate proteins of interest to identify those that were the most amenable to inactivation experiments. Because, these proteins have been

software-predicted during the initial annotation of the *M. incognita* and *M. hapla* genomes, some may simply be computational predictions that have no biological existence. Thus, we first checked whether experimental transcription data supported the existence of parts of these genes as true genes. Because the rest of experimental procedures will be applied to *M. incognita*, we focused our analysis on *M. incognita* predicted genes. To assess transcriptional support, we made use of 63,816 ESTs sequences that were assembled in 22,350 distinct unisequences in our lab (Jaouannet et al., 2012). Although informative, EST sequences are necessarily incomplete and redundant (due to splice variants) and some genes expressed at a particular life stage may not be represented in EST data. To circumvent this limitation, we generated high throughput transcriptomic data, using illumina sequencing on different developmental life stages of *M. incognita* (Figure 7).

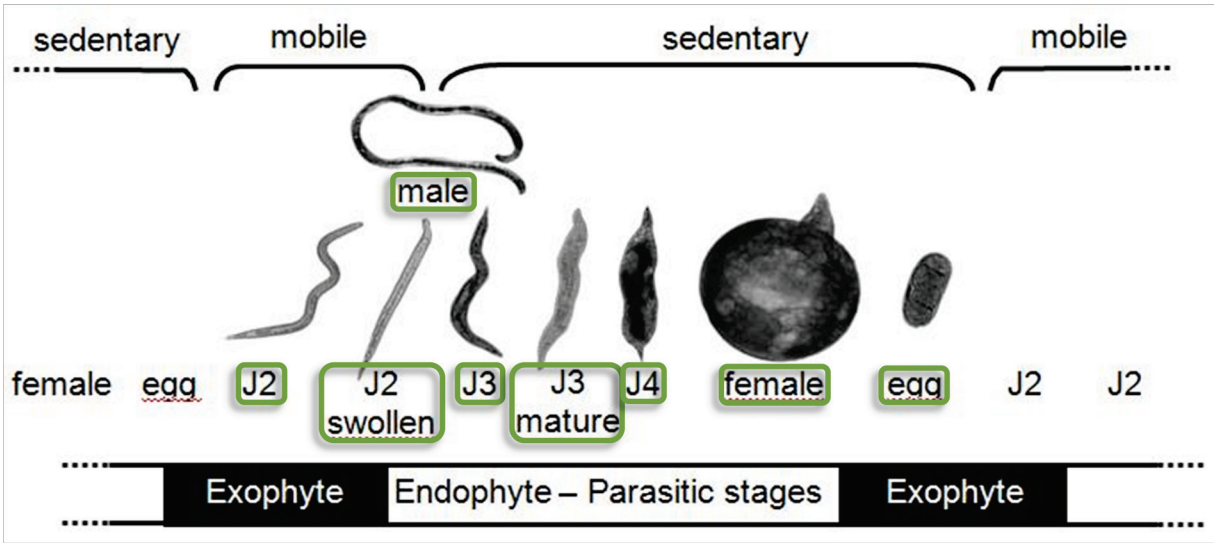



**Figure 7 *M. incognita* developmental life stages used for illumina sequencing.**

Different developmental life stages were isolated (J2 infective juveniles, parasitic "swollen" J2 larvae, J3 larvae, mature J3 larvae, J4 larvae, adult males, adult females and eggs). RNA was extracted and prepared for RNA-seq illumina sequencing.

Overall, RNA-seq illumina sequencing generated more than 190 million reads. These reads were assembled in 137,733 contigs using velvet / OASES assembly software. Together with the EST data previously generated, we had, at our disposal, a solid source for transcriptional validation of *M. incognita* predicted genes.

A total of 12 putative transcription factors out of the 190 identified were supported by transcriptional evidence and present in at least two plant-damaging species. Similarly, out of the 993 identified effector-like proteins, 232 were supported by transcriptional data and present in at least two plant-damaging species. Interestingly, we found 42 previously known effectors among this set of 232 effector-like proteins, an important confirmation for the validity of our screening procedure. Further pruning of the genes of interest included elimination of genes present as multigene families (to avoid potential compensation of function between paralogs) and elimination of previously published effectors (because our objective was to find novel interesting candidates). An important step was to experimentally confirm these predicted proteins as effectors by observing the effect of inactivation of the corresponding genes via siRNA. To achieve this goal, two other mandatory criteria

were added. The gene must be compatible with: (i) the design of gene-specific small interfering RNAs (ii) the design of gene-specific primers to monitor variation in expression level via qPCR following siRNA treatment.

Overall, 16 genes satisfied all these criteria (one putative transcription factor and 15 effector-like proteins). We tested the effect of inactivation of each of these genes, individually, via siRNA on the efficiency of plant parasitism. We put *M. incognita* J2 infective juveniles on a water solution containing the siRNA and then inoculated these treated J2s on tomato plant. As a control, we used a siRNA that does not match any sequence from *M. incognita*. To estimate the effect on the infection, we compared the number of egg masses and galls on tomato roots infested by each of the 16 samples of siRNA-treated J2s against tomato root infested by control J2s. We performed this analysis in two replicates and we found that for 12 targeted genes, the number of egg masses and / or galls was significantly reduced in a reproducible manner. For the 4 other genes, we also observed reductions but they were either not reproducible or not statistically significant.

These 16 genes (and more particularly the 12 genes that led to significant and reproducible diminution of infestation) constitute novel promising candidates for the development of new, more specific and efficient control methods. These genes are absent from non-target species, they are present in at least two plant-damaging species and we have demonstrated their effect on plant-parasitism efficiency. For this ensemble of reasons, beside the publication of our results in PLoS Pathogens (Danchin et al., 2013), a patent was deposited to protect the use of these genes as targets for the development of more specific control methods that would be more specific to RKN and safer for the environment.

## 2.2 Genome evolution in an asexually-reproducing animal

Although some singularities observed in the genome of *M. incognita* could be tentatively linked to adaptations to a plant-parasitic lifestyle, others were more probably the result of ongoing evolution in the absence of meiosis and sexual reproduction. Indeed, *M. incognita* has a mitotic parthenogenetic mode of reproduction (asexual reproduction without meiosis). The two aspects are not completely dissociated and it can be imagined that genomic consequences of asexual mitotic reproduction might also benefit to the parasitic ability of the nematode. This counter-intuitive idea is reinforced by the observed larger host spectrum and more global geographical distribution of obligate parthenogenetic root-knot nematode species. In this section, we will discuss two singular aspects of the *M. incognita* genome that may be related to their strict asexual mode of reproduction: a peculiar re-arranged genome structure and a high proportion of repeats and mobile elements. Both these elements might provide mechanisms for genomic plasticity that could compensate those that are missing in the absence of sexual reproduction. However, these mechanisms, by themselves do not provide a fully satisfactory explanation for the paradoxical higher parasitic success of asexual root-knot nematodes. This higher success might be related to the origin of the peculiar genome structure discussed in section 3.2. Indeed, the asexually reproducing root-knot nematodes might have undergone hybridization and it is postulated that hybrids can feature phenotypic characteristics that exceeds those of the progenitors (transgressive segregation) (Rieseberg et al., 1999).

### 2.2.1 A genome mainly composed of pairs of regions re-arranged along scaffolds.

Assembly of the genome of *M. incognita* reached 86 Mb (Abad et al., 2008). The assembled genome was much bigger than the expected size based on measures of the nuclear DNA content. Flow cytometry and biochemical analyses have led to an estimated haploid genome size of ~50 Mb (Leroy et al., 2003; Pableo and Triantaphyllou, 1989). The haploid nuclear DNA contents of *M. javanica*, *M. arenaria* and *M. hapla* have also been estimated to be ~50 Mb (Pableo and Triantaphyllou, 1989). Sequencing and assembly of the facultative meiotic parthenogenetic *M. hapla* yielded a 54 Mb haploid genome, in good agreement with independent experimental estimates. The reason why the *M. incognita* genome assembly was almost twice the estimated size probably lies in its peculiar reproductive mode itself. Because this nematode reproduces asexually and without meiosis, reduction of the genome to a haploid state never happens. In the absence of selective pressure for chromosome homology and pairing (no meiosis), the equivalent of paternal and maternal homologous chromosomes may have independently accumulated mutations and re-arrangements to a point that they are now substantially divergent. Consequently, these divergent parts, that used to be homologous, will not fuse during genome assembly and will be assembled separately as two diverged versions or copies. If the divergence spans the whole genome, then we expect a genome assembly to be ca. twice the estimated haploid size. Although substantial sequence divergence can also exist between allelic copies in sexual species, this problem is resolved by producing extensively inbred lines prior to genome sequencing. These inbred lines are basically homozygous at all loci and homologous chromosomes fuse into a unique same sequence during genome assembly.

We thus tried to assess whether some homology and variability exist between genome regions in the *M. incognita* assembly that could explain the bigger assembly size as compared to predicted haploid genome size. In collaboration with Genoscope, all predicted protein-coding genes in the *M. incognita* genome were compared against themselves at the protein level to identify genes present in two or

more copies. Information of positions allowed retrieving collinear series of gene copies present on different scaffolds (Figure 8). Genomic regions containing series of consecutive collinear gene copies (at least 3) were aligned. Alignment of these regions allowed estimating an average nucleotide divergence of 7%. More than half of the assembly (~55 Mb) is covered by genomic regions in two copies, and this includes all larger scaffolds. Scaffolds that were not present as two copies were in general gene-poor and rich in repetitive elements (Abad et al., 2008). Conserved syntenies using genes as markers will fail to identify the copies for these scaffolds. It is thus well possible that the whole genome of *M. incognita* is in fact present as two copies with substantial nucleotide divergence. Presence of regions in two copies is probably the explanation for the bigger than expected genome assembly.

Examination of the *M. incognita* genome organization, allowed realizing that no large scaffold can be aligned on its whole length with another large scaffold (Figure 8). Instead, we observe series of synteny breakpoints. These breakpoints indicate that there is no chromosome-scale homology (i.e. no chromosome can be aligned on its whole length with another chromosome). This genome organization is consistent with the absence of meiosis. Indeed, with such a genome structure, chromosome cannot pair during the meiosis and it is not possible to separate the genome into two (haploid) sets of homologous chromosomes. This observed re-arranged genome structure in copies with an average nucleotide divergence of 7% is certainly the result of the peculiar mode of reproduction of *M. incognita* (without sex and without meiosis). Supporting the idea that this genome structure is linked to strict asexual reproduction, the genome of *M. hapla*, a related root-knot nematode able to reproduce sexually neither shows a bigger than expected genome nor presence of divergent copies at the whole genome scale (Bird et al., 2009). Availability of the genomes of two other mitotic parthenogenetic root-knot nematodes (*M. arenaria* and *M. javanica*) will allow verifying whether a similar structure in diverged copies is found. The origin and the functional consequences of the peculiar genome structure observed in *M. incognita* will be discussed in the perspectives (3.2 and 0). Determining if, more generally, in animals that lack meiosis and sexual reproduction, a common signature can be found at the genomic level is something that will need comparison with other models. There are only a few examples of animal lineages with strict asexual reproduction and no reported meiosis (Danchin et al., 2011). These animal lineages recently attracted my attention because they would provide very interesting comparisons with the obligate parthenogenetic root-knot nematodes and may allow identifying genomic signatures of long term asexual reproduction. These elements will be further discussed in the perspectives but preliminary results have already been obtained from the bdelloid rotifers.

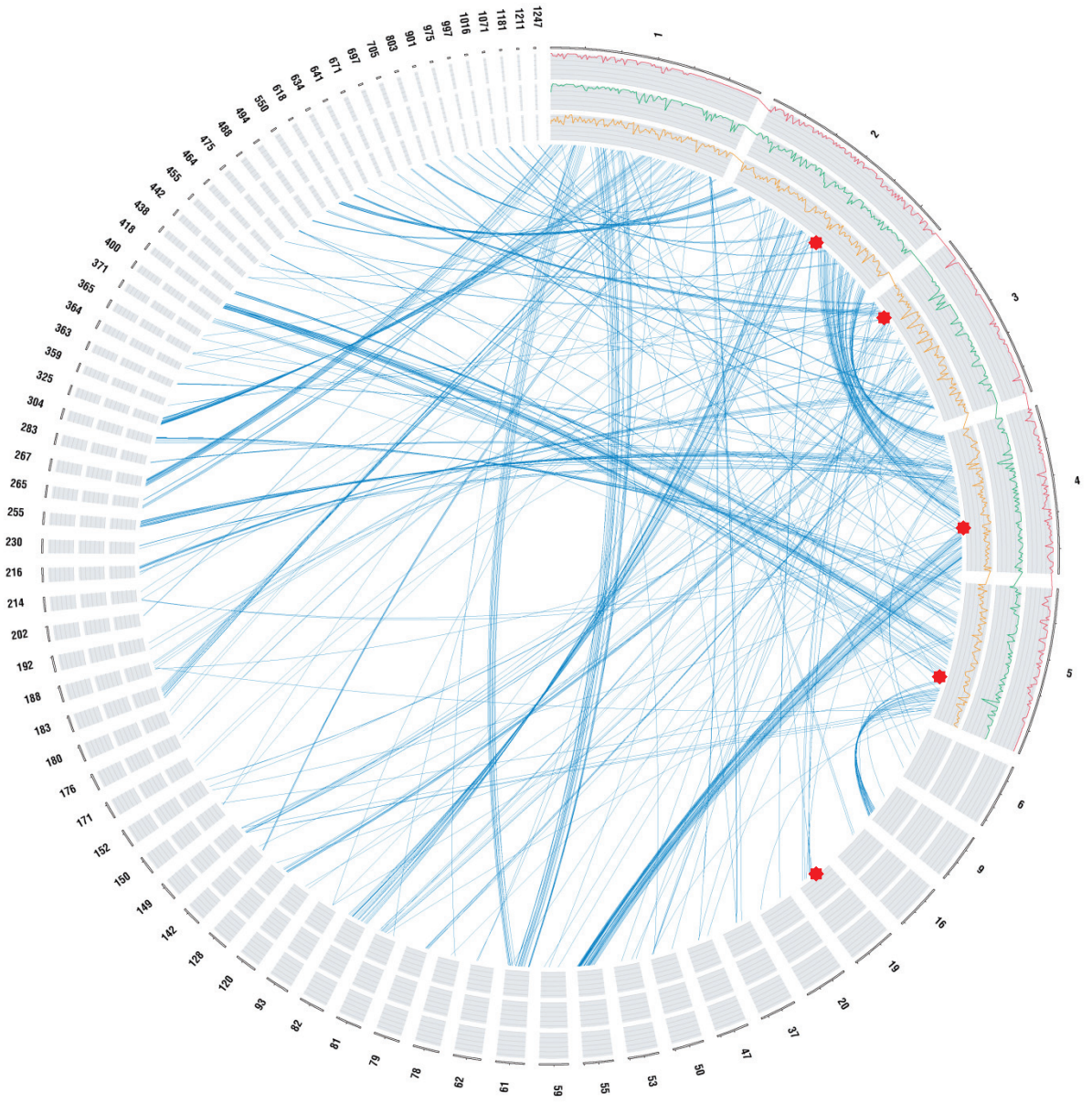


**Figure 8. *M. incognita* peculiar genome structure**

The five largest supercontigs are shown with plots of gene density (orange curve), similarity to *C. elegans* at protein level (green curve) and EST density (pink curve). These are linked to 70 other supercontigs through bi-directional best matching protein pairs (blue lines). EST-rich scaffolds (for example 2 and 4) show a high number of links. 8-branches red stars indicate synteny breakpoints. There are many potential other breakpoints but, unlike big scaffolds, scattering of the synteny over small supercontigs does not allow confirming a synteny breakpoint.

Bdelloid rotifers are small aquatic creatures that present two interesting singularities (i) they are considered as evolutionary scandals (Judson and Normark, 1996; Neiman et al., 2009) because they have been surviving without sexual reproduction and without meiosis for millions of years (ii) they survive repeated cycles of desiccation and rehydration. This is of course the first point that is more directly connected to my research interest although the two might be linked. Having participated to the 14th Evolutionary Biology Meeting in Marseilles, I have followed the talk of Karine Van Doninck

from the University of Namur, Belgium, who presented the genome sequencing project of the bdelloid rotifer *Adineta vaga*. Discussing our mutual interests on the genomics of asexually-reproducing animals, we decided to collaborate and wrote together a book chapter about ancient asexual animals (Danchin et al., 2011). Following this initial collaborative reflection, I have been involved in the genome annotation consortium of *A. vaga*. In particular, I have been involved in the functional annotation of genes acquired via LGT and in the study / interpretation of the genome structure. Interestingly, several parallels could be found with the genome of *M. incognita*. First, a high proportion of genes acquired via LGT was found (~8% of protein-coding genes), a proportion unparalleled so far in animal and higher than the proportion found in root-knot nematodes (~3% of protein-coding genes (Paganini et al., 2012)). Interestingly, an independent analysis at the transcriptome level yielded a very similar estimate of ~8-9% genes acquired via lateral transfer in another bdelloid rotifer, *A. ricciae*. This transcriptome analysis also provided strong evidence for transcription of the acquired genes, suggesting they are functional. Although in nematodes LGT were apparently associated with evolution of plant parasitism, it is possible that in bdelloid rotifers, acquisition of foreign genes via LGT provides genome plasticity in the absence of sexual reproduction (Flot et al., 2013). This argument does not hold for root-knot nematodes as most LGT events observed in obligate parthenogenetic species are also found in facultative and obligate sexual species. The second parallel with the genome of *M. incognita* concerns the genome structure itself and is more relevant to the present chapter. Similarly to *M. incognita*, the genome of *A. vaga* is composed of re-arranged genomic copies with substantial nucleotide divergence. The average nucleotide divergence between allelic-like regions is 4.4% as compared to 7% in *M. incognita* but the genome of *A. vaga* is made of two times two-copies (tetraploid) while the genome of *M. incognita* is rather present in two copies (diverged diploid). Another similarity to the genome of *M. incognita* lies in the absence of chromosome scale conserved synteny. In *A. vaga* also, no long scaffold could be aligned on its whole length to another long scaffold, indicating an absence of chromosome-scale homology. Instead, we observed series of synteny breakpoints within the different scaffolds (Figure 9). This genome structure, as the one observed in *M. incognita*, is incompatible with pairing of homologous chromosomes as required in classical meiosis. This structure confirms at a genomic level the absence of meiosis both in *M. incognita* and in *A. vaga*. Furthermore, as this structure is observed in two lineages of animals that have independently evolved a mitotic parthenogenetic mode of reproduction, this could represent a common genomic signature of the absence of meiosis and sexual reproduction.

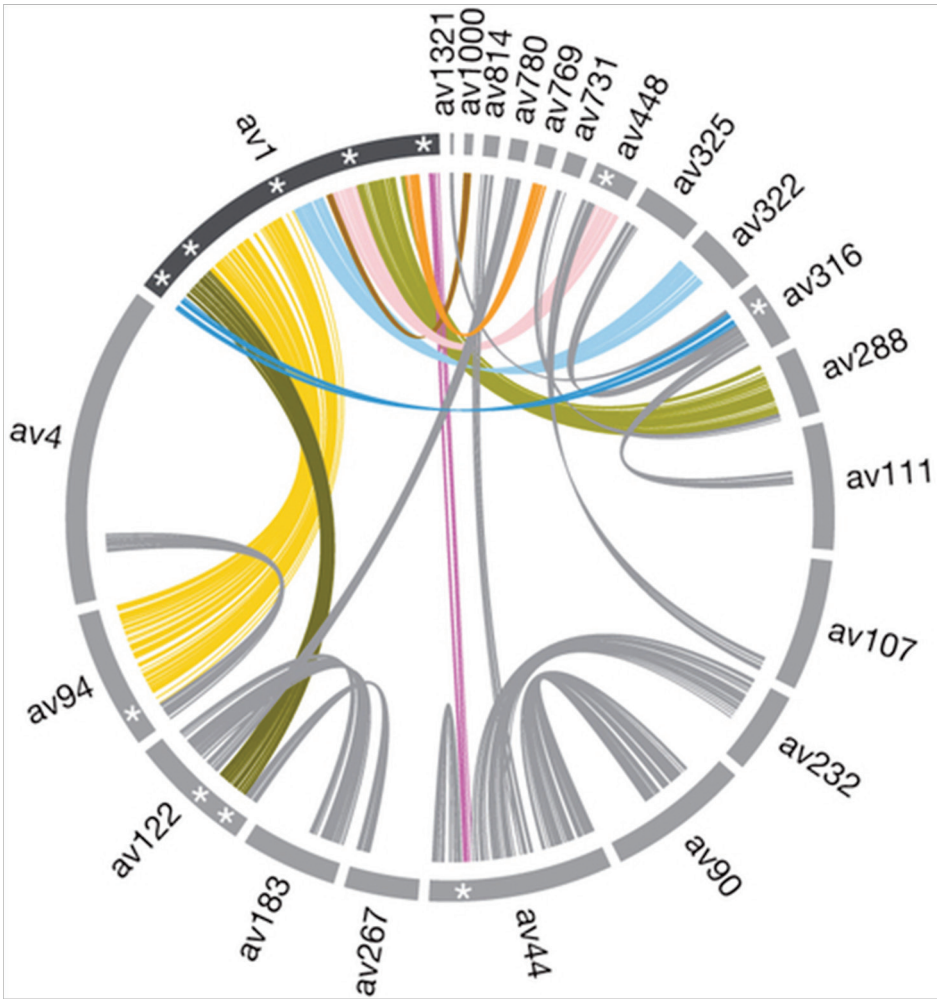



**Figure 9. Re-arranged genome of *Adineta vaga* showing absence of chromosome-scale synteny**

The longest scaffold (av1) is represented together with scaffolds that present genomic regions aligning with those on scaffold1. Colored lined represent pairs of genes that are reciprocal best blast matches. Stars represent synteny breakpoints in scaffolds. Long scaffolds like av1 cannot be lined-up with another long scaffold on its whole length, indicating the absence of chromosome-scale homology.

### 2.2.2 A high proportion of mobile elements and repeats.

During the initial annotation of the *M. incognita* genome, we found that as much as 36.7% of the genome was covered by repetitive and transposable elements (TE) (Abad et al., 2008). This proportion is higher than the proportion observed in other nematode genomes (Figure 10). For example, TE and repeats cover 16.5% and 22.4% of the genomes of *C. elegans* and *C. briggsae*, respectively (Stein et al., 2003), 18% in the animal parasite *Trichinella spiralis* (Mitreva et al., 2011), 14-15 % in the human parasite *Brugia malayi* (Ghedin et al., 2004, 2007; Scott and Ghedin, 2009) and 22% in the pine wilt disease vector *Bursaphelenchus xylophilus* (Kikuchi et al., 2011). This high proportion of repetitive and mobile elements may be a characteristic of root-knot nematode genomes. However, contradicting this idea, TE and repeats occupy only 18.3% of the *Meloidogyne hapla* genome (Opperman et al., 2008).

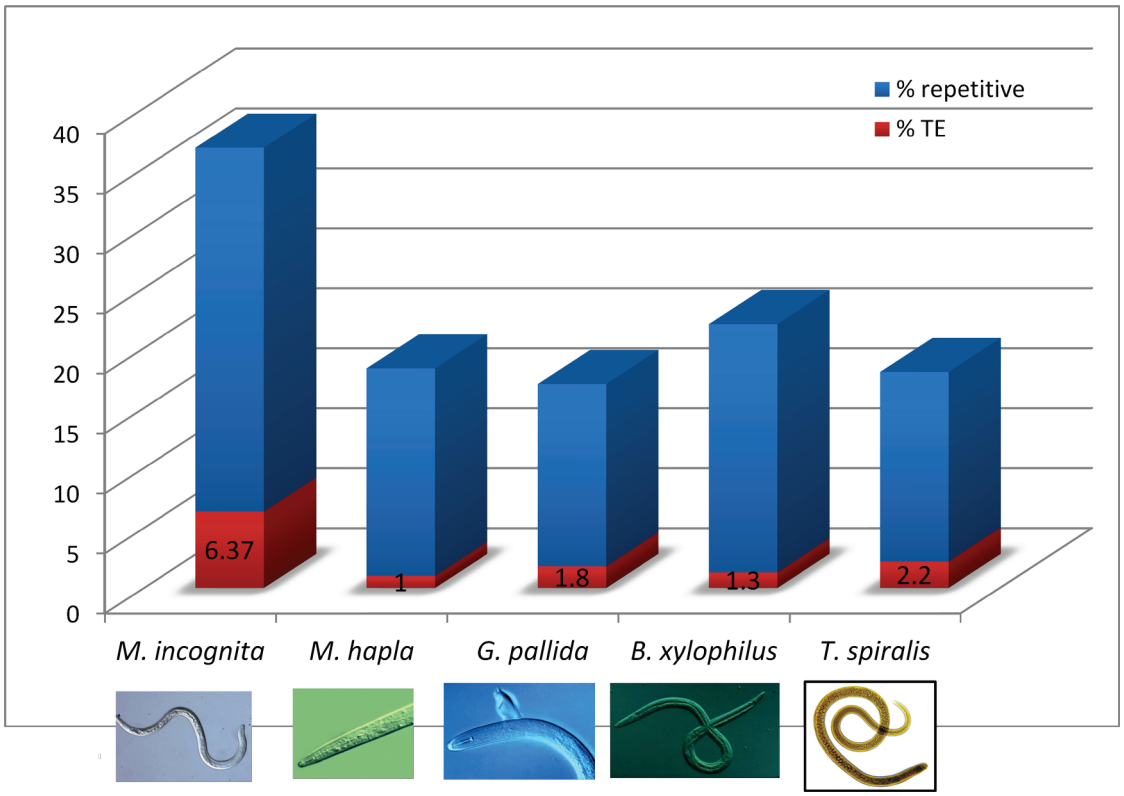



**Figure 10 Proportion of repetitive elements in nematode genomes**

Histograms representing the proportion of the genome (in %) covered by repeats (blue), including transposable elements (red) in various nematodes. In this histogram, the proportion of repetitive DNA ranges from 17 % (*Globodera pallida*) to 22 % (*Bursaphelenchus xylophilus*) but reaches 36.7% in the root-knot nematode *M. incognita*.

Hence, rather than representing a specificity of root-knot nematode genomes, the high proportion of TE and repeats in *M. incognita* might, in fact, be a consequence of strictly asexual and ameiotic reproduction. Supporting this hypothesis, it has been proposed that, in the absence of a mechanism to control their proliferation, TE and repeats may invade the genomes of asexually-reproducing animals (Arkhipova and Meselson, 2005). In species with sexual reproduction, individuals carrying a high copy number of TE as well as their offspring can be counter-selected, maintaining TE to a controlled low number in populations. Mathematical simulations have suggested that in the absence

of excision mechanisms, TE can proliferate in populations following a ratchet mechanism, generation after generation (Dolgin and Charlesworth, 2006). Interestingly, these theoretical models can be tested in some animal lineages. For instance, in parasitoid wasps, it has been shown that TE are more abundant in Wolbachia-induced asexual lineages than in sexual lineages of the same species (Kraaijeveld et al., 2012). However, partly contradicting theoretical models, the authors have shown that this proliferation was not true for all categories of TE. For instance, in that particular case, class II transposable elements (DNA transposons), that move via a "cut and paste" mechanism were more prone to proliferation than class I elements (retrotransposons) that transpose via a "copy and paste" mechanism. The same authors suggested that this category-specific proliferation might be a side-effect of Wolbachia-induced parthenogenesis rather than due to asexual reproduction itself. It is interesting to note here that in the case of TE horizontal transfers similar preference for DNA transposons over non LTR retrotranposons (such as LINE) has been observed (Schaack et al., 2010).

Whether TE and repeats also cover a high proportion of the genomes of *M. javanica* and *M. arenaria*, two mitotic parthenogenetic species closely related to *M. incognita* will be interesting to determine. Analysis of the genome of the mitotic parthenogenetic bdelloid rotifer *A. vaga* showed a surprisingly low proportion of TE. Transposable elements occupy only 3% of the *A. vaga* genome and while a high diversity of TE was found, they are generally present at very low copy numbers (Flot et al., 2013). For comparison, TE cover more than 6% of the *M. incognita* genome. The low copy number suggests that proliferation of TE is under control in the genome of *A. vaga* and that a genome rich in TE and repeats is not necessarily a common signature of a strictly asexual mode of reproduction. This observation does not necessarily contradicts the idea that the abundance of TE and repeats in *M. incognita* is linked to its mode of reproduction and analysis of other root-knot nematode genomes will allow determining whether this is a singularity of *M. incognita* or if this is systematically associated to mitotic parthenogenetic root-knot nematodes.

Regardless whether the abundance in TE represents a specificity linked to the peculiar mode of reproduction, we can hypothesize that, if still active, these elements might provide some plasticity despite absence of sex. Supporting this hypothesis, in the mitotic parthenogenetic species *M. javanica*, a transposon, named Tm1, has been described recently and an homolog with predicted full-length transposase has been found in the *M. incognita* genome while no homolog with an intact transposase could be found in the sexually-reproducing relative *M. hapla* (Gross and Williamson, 2011). What is even more interesting is that the Cg-1 gene, which does not code for a protein but whose deletion is associated to strains of *M. javanica* able to break tomato resistance, has been identified within one of these Tm1 transposons.

Hence, it is well possible that these mobile elements have played or still play a role in the genome of root-knot nematodes and can have functional consequences, including at the host spectrum level. Whether their activity more generally provide a mechanism of plasticity in the absence of sexual reproduction remains to be determined at a larger scale.

# 3 Perspectives, projects and future directions

Examination of the genome of the root-knot nematode *M. incognita* has initially revealed several major points and opened many perspectives that can all be further explored and enriched by the foreseen availability of additional genomic and transcriptomic data from plant-parasitic nematodes. Here I discuss these perspectives and the future directions that will drive my main future projects. These projects range from applied aspects such as the identification of new gene targets for the development of more efficient anti-nematode methods to more basic research aspects such as the evolution and adaptation in organisms that reproduce without sex and meiosis. Applied and basic research are interconnected in these different projects since the peculiar genome organization associated to asexual reproduction probably influence the biological traits such as plant parasitism itself.

## 3.1 New target genes for the development of specific control methods against root-knot nematodes

During our NEMATARGETS project (see section 2.1.2), we found 5,530 proteins that were absent from non-target species, possibly conserved in other plant-damaging species, and supported by transcriptional data (Danchin et al., 2013). We focused on effector-like proteins because; as genes suspected to be involved in parasitism they were evident targets. We identified 190 novel candidate effectors. A total of 15 could be experimentally tested and we showed that the silencing of 11 of these genes led to significant and reproducible reduction of the infection success on tomato plants. However, because most of these proteins possess no known domain and lack homology in species other than root-knot nematodes, we have no idea about their exact function. In-situ hybridization experiments included localization in secretory glands, suggesting a possible secretion by the nematode in plant tissue. Yet, considering they are actually secreted in the plant, their possible mode of action and plant targets are unknown. The same applies to the putative transcription factor we have identified and whose corresponding gene silencing also leads to reduction of infestation. Thus, in collaboration with the group of Bruno Favery in our research team, we will try to identify plant targets of these putative effectors and secreted transcription factors. *In planta* localization and experiments such as yeast 2 hybrids will provide valuable information to further characterize these novel candidates at the functional level. Identification of plant genes and functions targeted by the nematodes can help guiding plant breeding and selection strategies and lead to new resistant cultivars.

Besides the putative transcription factor and the effector-like proteins, there is absolutely no indication about the putative processes for the rest of the identified proteins (>5,000) that are absent from non-target species and supported by transcriptional evidence. Because they are specific to plant-damaging species, they also represent a pool for the discovery of novel target genes for the development of safer and more specific control methods. In addition to effectors, there is a variety of other processes that can affect success of parasitism when disrupted. For example, genes involved in the ontogeny of key organs for parasitism such as the stylet or the esophageal glands can represent interesting targets. Similarly, genes encoding proteins involved in sensory processes constitute promising targets. Indeed, it is commonly admitted that plant-parasitic nematodes are able to detect plant compounds in the soil that are able to guide them to the roots. Interfering with this process may also potentially lead to significant reduction of nematode infection. Hence, one of the future

directions I will explore will consist in pruning this set of >5,000 candidate genes toward various different and more reduced subsets of interest. To help achieving this, we can take into account additional information such as the genes expression patterns obtained from the RNA-seq data we have generated and from upcoming new data. For instance, genes that are upregulated during parasitic stages constitute interesting candidates. Similarly, genes that are differentially expressed between males and females during the developmental life cycle can be of high interest. Indeed, only female are able to lead to an offspring and repeat the infection cycle to the next generation. Males are dead ends at the parasitism point of view. Thus, identifying gene networks and signalization pathways involved in male / female differentiation could lead to methods for manipulating this process toward more males and reduced infestation. Comparisons with the genomes and transcriptomes of other plant-parasitic nematodes can also provide valuable information. Indeed, if candidate parasitism genes are conserved in other plant-parasitic nematode species yet absent from non target species, they might represent interesting targets for broader and more generic control of these pests.

Overall, the comparative genomics approach we have undertaken in NEMATARGETS, has led to the identification of novel genes that could be used as targets for the development of new control methods against root-knot nematodes. Their experimental validation as genes involved in the parasitic success and their specificity to plant-damaging species has led to the registration of a patent. Besides the few genes that were experimentally tested and that putatively function like known effectors, our comparative genomics screen also led to a promising pool of new genes of as yet unknown function, specific to plant-damaging species. Further pruning of this set of genes, coupled with experimental assays, will not only reveal additional interesting candidates for the development of new and more specific control methods, but will also provide new insights into the molecular basis of the interactions between plants and nematodes.

## 3.2 Origin of the genome structure observed in *M. incognita*

So far, two main hypotheses have been formulated to explain the observed *M. incognita* genome structure in two copies. The first category of hypotheses suggests that the observed copies result from former homologous paternal and maternal chromosomes that independently accumulated mutations since the loss of meiosis and requirement for chromosomal pairing (the so-called White / Meselson effect, resulting in basically two genomes in one (Mark Welch and Meselson, 2000; White, 1973). This hypothesis implies that a facultative sexual ancestor, that was also able to undergo meiotic parthenogenesis, lost the capacity to perform meiosis and became unable to reproduce sexually. This species was still able to survive and produce offspring thanks to its capacity to perform parthenogenesis. The scaffolds now observed as re-arranged and divergent copies in the genome would have derived from ancestral paternal and maternal homologous chromosomes that progressively accumulated modifications, independently one from the other in the absence of selective pressure for pairing of homologous chromosomes. Transition from a meiotic parthenogenesis to mitotic parthenogenesis, though requires considerable modifications and remains to be explained. The second category of hypotheses supposes that mitotic parthenogenetic root-knot nematodes, such as *M. incognita*, result from inter-species hybridization. Interestingly, asexuality in invertebrates, vertebrates and plants is often associated with hybridization (Bullini, 1994; Delmotte et al., 2003; Hörandl, 2009; Kearney, 2005, 2003; Simon et al., 2003), suggesting that

the two events are linked. One possible explanation of this apparent link is that hybridization make meiosis very complicated or impossible due to combination of possibly different karyotypes into one single nucleus. In this case, only species that have been able to circumvent meiosis and perform mitotic parthenogenesis survived. Under the hybrid theory, the presently observed divergence between the *M. incognita* genomic copies would result from pre-existing divergence between the two donor homeolog genomes further amplified by independent evolution since the loss of meiosis. According to this hypothesis, the two donor nematode species are necessarily at least sexual and it is more parsimonious to consider that they were also facultative meiotic parthenogenetic. The way the hybrid offspring evolved from facultative sexual parents towards obligate mitotic parthenogenesis also remains to be established. The two hypotheses are consistent with observation of a re-arranged genome, lacking chromosome-scale homology and mainly present in two copies with substantial sequence divergence. Though, several aspects distinguish the two categories of hypotheses, in terms of expectation. The first hypotheses (inheritance from a single facultative sexual ancestor that lost meiosis) implies that loss of meiosis and subsequent independent evolution of the two former homologous genomes (paternal and maternal) must be old enough to have led to ~7% average nucleotide divergence and to considerable observed re-arrangements. Indeed, this within-genome divergence level is substantially higher than the divergence levels observed between genomes of individuals within wild populations. For instance, this is higher than the 3.7 % nucleotide divergence level observed in the amphioxus (Putnam et al., 2008) and the 4.0 % level in the sea urchin (Sodergren et al., 2006), both considered as highly heterozygous animals. Hence, even if we postulate that the facultative sexual ancestor of mitotic parthenogenetic RKN had a comparable high heterozygosity level; sufficient time must have passed to reach the 7 % divergence observed today between the two copies of the genome. If this was the case, loss of meiosis must be old and we expect that meiosis genes are degenerated in *M. incognita*. A comparison of meiosis genes between *M. incognita* and *M. hapla* revealed no sign of decay in the mitotic parthenogenetic species (personal observations). Furthermore, while they do not genetically participate to the genome of the offspring (Triantaphyllou, 1981), males are still observed in *M. incognita* and the other mitotic parthenogenetic species. These two elements somewhat contradict the hypothesis of an old loss of sex, meiosis and long-lasting separate evolution of former paternal and maternal chromosomes. The alternative category of hypotheses (hybridization) does not necessarily require an old loss of meiosis and separate evolution of homeologs. Indeed, if the level of divergence between the two progenitor species that gave rise to the mitotic parthenogenetic species was already high, the time to reach the observed 7% divergence can be much shorter. Furthermore, other observations are in favor of a hybridization hypothesis. Although high average divergence is observed between gene copies (regardless whether they represent homeologs or ancient alleles), phylogenetic analyses indicate that they do not cluster according to recognized morphological species. Instead, nearly identical copies can be found between different species (e.g. between *M. incognita*, *M. javanica* and *M. arenaria*) whereas second copies are specific to each mitotic species (Lunt, 2008). This observation strongly suggests that one genome copy might be of very recent common origin in the different mitotic parthenogenetic species while the other is species-specific. Such an observation would be compatible with multiple hybridizations with one common species being involved in each event having given rise to the different mitotic parthenogenetic nematodes. A last independent point, that appears to support the hybridization hypothesis, comes from observations at the mitochondrial DNA level. Early analyses in 1997 have shown that the level of nucleotide diversity between individuals in asexual root-knot nematodes (*M. incognita*, *M. javanica* and *M. arenaria*) is ten times lower than in

the facultative parthenogenetic species *M. hapla* (Hugall et al., 1997). This suggests that the asexual species have emerged much more recently. Discordances in the levels of nucleotide divergence between mitochondrial and nuclear genes have been observed in asexual root-knot nematodes (Fargette et al., 2010). These discordances can also be explained by hybridization events. Furthermore, high similarity between fragments of mtDNA in *M. incognita*, *M. javanica* and *M. arenaria* has led to the hypothesis that several independent hybridization events occurred with a same female parental donor lineage and different males to give rise to these different species (Lunt, 2008).

Following up genome sequencing efforts in root-knot nematodes, we are currently trying to assemble the whole mtDNA genomes of *M. incognita*, *M. javanica* and *M. arenaria*. Typical nematode mtDNA genomes are ~15kb long, and contain 36 genes, including 12 protein-coding genes, 22 tRNA genes and 2 rRNA genes. At the moment, we have been able to reconstruct ~8-9 kb of mtDNA for each of these 3 asexual species. Our initial comparisons at the mtDNA level show an average nucleotide divergence level of ~2% between *M. incognita*, *M. javanica* and *M. arenaria*. This is much lower than the 7% average divergence observed between copies of the genome in a single species (*M. incognita*). Mitochondrial DNA (mtDNA) accumulates mutation much more rapidly than nuclear DNA, making it a marker to distinguish closely related species. There is an important discrepancy between the high (7%) within-species divergence at the nuclear nucleotide level and the low (2%) inter-species divergence at the mtDNA level. Combined with absence of decay of meiosis or sex-related genes in mitotic species, the hypothesis of hybridization appears as the most likely. Further efforts in assembling and annotating the mtDNA of the asexual root-knot nematodes and that of the facultative parthenogenetic *M. hapla* will allow performing phylogenetic analyzes based on concatenated protein-coding genes. This will provide a more precise estimate of the divergence level between the asexual species and possibly dating these events.

Interestingly, the amount of genetic material in a nucleus of *M. incognita*, *M. javanica* and *M. arenaria* has been estimated to be almost the same (~100 Mb at diploid state) and equivalent to the amount found in the facultative sexual *M. hapla* (Lapp and Triantaphyllou, 1972; Leroy et al., 2003; Pableo and Triantaphyllou, 1989). Hence, there is no apparent genome doubling in the asexual RKN and if hybridization did actually happen, this was most probably an homoploid hybridization, a phenomenon already observed in animals (Mavárez and Linares, 2008).

Determining the events at the origin of the peculiar genome structure, observed in *M. incognita*, might appear as a purely basic research interest without evident agronomic outcome. However, if it turns out that the asexual root-knot nematodes, considered so far as the most dangerous at an agricultural level, arose from hybridization events, this information has to be seriously taken into consideration. This would suggest that new hybrid species with a broader host spectrum and more dangerous to the worldwide agriculture might potentially emerge. Hybridizations between different facultative sexual root-knot nematode species have already been obtained in lab conditions (van der Beek and Karssen, 1997). Although in this particular case, the obtained hybrids were unable to produce offspring; other combinations might give rise to successful hybrids in the wild.

## 3.3 Biological consequences of a genome structure in copies

One clear consequence of the peculiar genome structure, evoked previously, is the lack of chromosome-scale homology consistent with the absence of meiosis and also observed in the bdelloid rotifer *A. vaga* (Flot et al., 2013). Besides these structural aspects, another consequence is that a number of protein-coding genes are present as two copies that substantially diverge in their sequences. Whether this sequence divergence has functional implications would be very interesting to determine. Indeed, it is possible that sequence divergence includes non-synonymous mutations (that change the amino-acid in the encoded protein). Altered protein sequence can potentially lead to change in the biochemical function exactly as observed for paralogous genes. Whether this phenomenon takes place at the whole genome level and allows genomic plasticity in the absence of sexual recombination is an interesting hypothesis. Measuring divergence at the biochemical function level between gene copies at the whole genome scale appears not currently feasible. However, another functional aspect that can be investigated at the whole genome scale is expression divergence between copies. Interestingly, whole genome scale expression divergence between gene copies resulting from polyploidozation (including as a result of hybridization) has recently been shown in plants, using RNA-seq techniques (Pont et al., 2011; Roulin et al., 2013). Similar approaches could be undertaken to assess whether divergence in expression can be observed between putative homeologous gene copies in *M. incognita* at a large scale. Interestingly, we have generated RNA-seq transcriptomic data for different *M. incognita* developmental life stages. Preliminary analysis allowed identifying a set of gene copies resulting from the genome structure with divergent expression (personal communication). These preliminary results remain to be further refined and, for instance, the prevalence of potential functional divergence between gene copies at the whole genome scale as not yet been determined. In the bdelloid rotifer *A. vaga*, ratios of non-synonymous (Ka) vs. synonymous (Ks) mutations have been calculated between gene copies resulting from the tetraploid genome structure. Interestingly, cases of Ka / Ks > 1, indicative of positive selection were identified in a substantial proportion. These ratios > 1 indicate potential neo or sub-functionalization between gene copies although this remains to be confirmed experimentally. Altogether, these observations suggest that the peculiar genome structure in copies observed in mitotic parthenogens may allow plasticity at the gene expression level and potentially at the protein function level, despite absence of sexual recombination.

I am currently the coordinator of an ANR (Agence Nationale de la Recherche) funding we have recently obtained (programme Jeunes Chercheurs 2013), to investigate the functional consequences of the peculiar genome structure observed in the root-knot nematode *M. incognita*. In this project, we will first establish a list of *M. incognita* gene copies that result from the peculiar "duplicated" genome structure. This will provide an estimate of the total impact of the genome structure on the protein-coding space. We will align all these gene pairs and estimate the rates of synonymous and non-synonymous mutations (Ka / Ks). From this data, we will first check whether some copies show signs of positive selection (Ka / Ks > 1). Traces of positive selection represent an indication for potential functional divergence at the molecular level between the products of the gene copies. To further explore potential functional divergence, we will analyze the expression patterns of these copies and check whether divergence in expression can be observed between copies that diverge at the sequence level. To determine expression patterns of gene copies, we will generate RNA-seq data for different developmental life stages of *M. incognita*. Previously obtained RNA-seq data were generated mainly to obtain transcriptional evidence on genes predicted from the genomes. The

newly generated data will include multiple biological replicates per extracted life stage to provide the necessary statistical support to test for expression divergence. A positive correlation between sequence divergence and expression divergence has already been shown in the model nematode *C. elegans* (Castillo-Davis et al., 2004; Conant and Wagner, 2004). A similar correlation has also been observed in other species including some that have underwent whole genome duplication such as the budding yeast (Gu et al., 2002). However, positive correlation between sequence and expression divergence does not appear to be a systematic rule and other studies have found no correlation between these two parameters (Li et al., 2005). Combining expression data to sequence similarity data, our project will allow determining whether there is a correlation between these two features in *M. incognita* gene copies. We obviously are mindful that expression is not controlled by the coding sequence itself but by surrounding regulatory elements. Therefore, we will also analyze conservation / divergence of putative regulatory elements around gene copies presenting divergent expression. If we identify expression divergence between gene copies during the development of *M. incognita*, this will represent a first indication of a functional consequence for the observed peculiar genome structure. Such a phenomenon has not yet been shown in an animal at the whole genome level and the root-knot nematodes represent an interesting opportunity and model in that sense.

Another aspect, probably even more intriguing, is that strict parthenogenetic root-knot nematodes have a wider host range and distribution than the facultative sexuals. This observation is counter-intuitive, considering the postulated advantages of sexual reproduction in terms of adaptation to a changing environment or to different hosts (Meirmans, 2009; Rice and Friberg, 2009). One of our hypotheses is that the peculiar genome structure, through the presence of genes in divergent copies, represents a pool for emergence of functional divergence and allows plasticity in the absence of sexual reproduction and inter-individual genetic exchanges. In order to test this hypothesis, our project aims at generating RNA-seq data from *M. incognita* infestation on various different host plants. In particular, we will rear *M. incognita* on two kinds of host plants: (I) plants that are compatible hosts to both *M. incognita* and *M. hapla* (e.g. tomato, pepper) and (II) plants that are compatible hosts to *M. incognita* only. We will extract the transcriptomes of *M. incognita* infective larvae after several generations on type I and type II hosts. The transcriptomes will be sequenced with RNA-seq techniques, including multiple biological replicates and we will compare expression patterns of gene copies upon infestation of different plant hosts. Identification of *M. incognita* gene copies differential expression between hosts that are compatible vs. incompatible with *M. hapla* will constitute a first series of evidence for a potential link between the "duplicated" genome structure and the larger host spectrum. Differential expression of one gene copy relative to the other, upon infestation of different plant hosts, will be a strong indication for potential functional divergence linked to the genome structure with impact on an economically important aspect of the biology of asexual root-knot nematodes. In a broader and more basic research perspective, this will also bring important results to a major evolutionary biology question, unresolved so far: how an animal can evolve and adapt in the absence of sexual reproduction? Interestingly, as previously introduced, a genome structure including regions in divergent copies has also been revealed in the genome of the ancient asexual rotifer *A. vaga* (Flot et al., 2013). Such a genome structure, with resulting divergent gene copies might represent a common signature of survival and adaptation without sexual reproduction.

## 3.4 Genome evolution and adaptation in asexually-reproducing animal lineages

Asexual reproduction is traditionally viewed as an evolutionary dead in animals and supposed to lead to rapid extinction. Indeed, it is postulated that asexuals generally occupy shallow positions in the animal tree of life and have more limited ecological niches and geographical distributions than sexual relatives. However, a comprehensive review of age estimates of asexual animal lineages compared to their sexual relatives showed that the majority of asexuals were probably more than 500,000 years old which contradicts the common assumption for a rapid evolutionary dead end (Neiman et al., 2009). Among these example lineages, some have been estimated to be very old (several million years old) or to occupy broader ecological niches than their sexual relatives. For instance, asexual root-knot nematodes have a broader geographical distribution and infest more plant hosts than their facultative sexual relatives; and bdelloid rotifers have been surviving without sex and meiosis for more 30-40 million years. Genome analysis of the root-knot nematode *M. incognita* and of the bdelloid rotifer *A. vaga*, two species that reproduce without meiosis and without sex, have both revealed a peculiar genome structure that comprises copies with substantial nucleotide divergence. This genome singularity might represent a common adaptation to survival without sexual reproduction in animals. The underlying hypothesis being that the resulting gene copies are prone to neo-functionalization and constitute a reservoir for the emergence of new functions (see previous section). This mechanism would allow adaptability in the absence of sexual recombination and genetic exchange between individuals. Although this mechanism might represent a factor of genomic plasticity it is unknown whether it is more general in asexual animals.

We thus proposed a project to "France Génomique" sequencing program in 2013 in order to investigate more broadly the evolutionary processes in the genomes of asexually-reproducing animals. As coordinator of this project, I have set up a consortium of European laboratories working on the evolution of parthenogenetic animals. This consortium is composed of the seven following laboratories:

- The laboratory of Isa Schön and Koen Martens in Belgium has set up a research program around species that evolve without sexual reproduction and proposed many concepts and theories based on actual biological data, mainly using non-marine ostracods as model group. They are specialists of one of the model species of ancient asexual reproduction (*Darwinula stevensoni*) and of one ostracod species with mixed reproduction (*Eucypris virens*).

- Our research team in "Institut Sophia Agrobiotech" has coordinated genome sequencing and analysis of the first strict asexually reproducing animal ever sequenced (*M. incognita*) and has ongoing projects to explore genome evolution in this species despite absence of sex. In this specific project, our main objective was to estimate the divergence between individuals within a population of *M. incognita*. These individuals can virtually be considered as clones, genetically identical to their progenitor female, but it is unknown at which rate mutations accumulate between individuals. Depending on the number of generations substantial genetic variation can be expected. Whole genome amplification and single individual sequencing, followed by reads alignment on a reference genome can detect SNPs and provide information on the level of inter-individual polymorphism.

- The laboratory of Karine Van Doninck in Belgium has coordinated genome sequencing and analysis of a bdelloid rotifer species that can be considered as one of the model systems for ancient asexual scandals (*A. vaga*). Karine Van Doninck also worked during years on the asexual darwinulid ostracods with Isa Schön and Koen Martens.

- The laboratory of Tanja Schwander in Switzerland has discovered and investigated a genus of insects (Timema stick insects) that contains both sexual and asexuals (recent and older) and proposed many concepts and theories about their evolution that could now be verified at a large scale with the genomes.

- The IGEPP laboratory (INRA Rennes) has led the genome sequencing and analysis consortium of the parthenogenetic pea aphid and has proposed models of genome evolution and population genomics that can all be tested and refined on further whole genome data.

- The laboratory of Stefan Scheu in Germany is a leading group studying one of the few phyla considered as very ancient asexuals, the oribatid mites.

- The IRISA laboratory (INRIA Rennes) develops algorithms and software for genome assembly specially dedicated to genomes considered as challenging due to high proportions of duplications and repeats.

Our project main aim is to characterize genome evolution in the absence of recombination and identify genomic features of asexual animals that may allow them to persist over evolutionary times in the absence of the benefits of sex. We will generate genome data from divergent species that have lost sexual reproduction independently and at different timescales to identify common trends of genomic adaptation to asexuality. Genome sequencing of related sexual lineages will further allow pinpointing differences in genome contents and genomic selection landscapes under sex vs. asex.

One of our main questions is to assess whether there are common genomic signatures of the absence of meiosis and recombination. We propose to investigate this question at different levels:

- At the structural level, we will explore whether genomic singularities like duplicated regions, palindromes, re-arranged former homologous chromosomes, inversions, fusions etc. are common in the genomes of asexuals as compared to sexual animals.
- At the genome content level, we will assess whether features like horizontal gene transfers as well as mobile and other repetitive elements show higher proportions in the genomes of asexuals.
- At the population level, we will estimate the rate of accumulated deleterious, neutral and non-synonymous mutations within and between populations of asexuals as compared to sexuals.

In this project we have included six animal lineages in which losses of sexual reproduction happened at different times (Figure 11). Loss of sexual reproduction in the bdelloid rotifer *A. vaga* is estimated to have happened 40 Myr ago (Danchin et al., 2011), in the darwinulid ostracod *D. stevensoni* it is estimated to be 20 Myr ago (Straub, 1952) and in *Timema* stick insects up to 1.5 Myr ago (Schwander et al., 2011). The time of loss of sexual reproduction in root-knot nematodes was estimated to have occurred between 17-80Myr ago (Hugall et al., 1997; Judson and Normark, 1996) but asexuality may

be the result of a much more recent inter-species hybridization event, as previously discussed in section 3.2. In aphids, and the ostracod species complex *Eucypris virens*, loss of asexuality is apparently more recent and may vary between and within species from thousands of years to a few generations (Schon et al., 2000). This full range of different ages for the loss of sexuality, from very recent to very old, will provide a unique opportunity to assess whether there are gradual genomic consequences to sex loss.

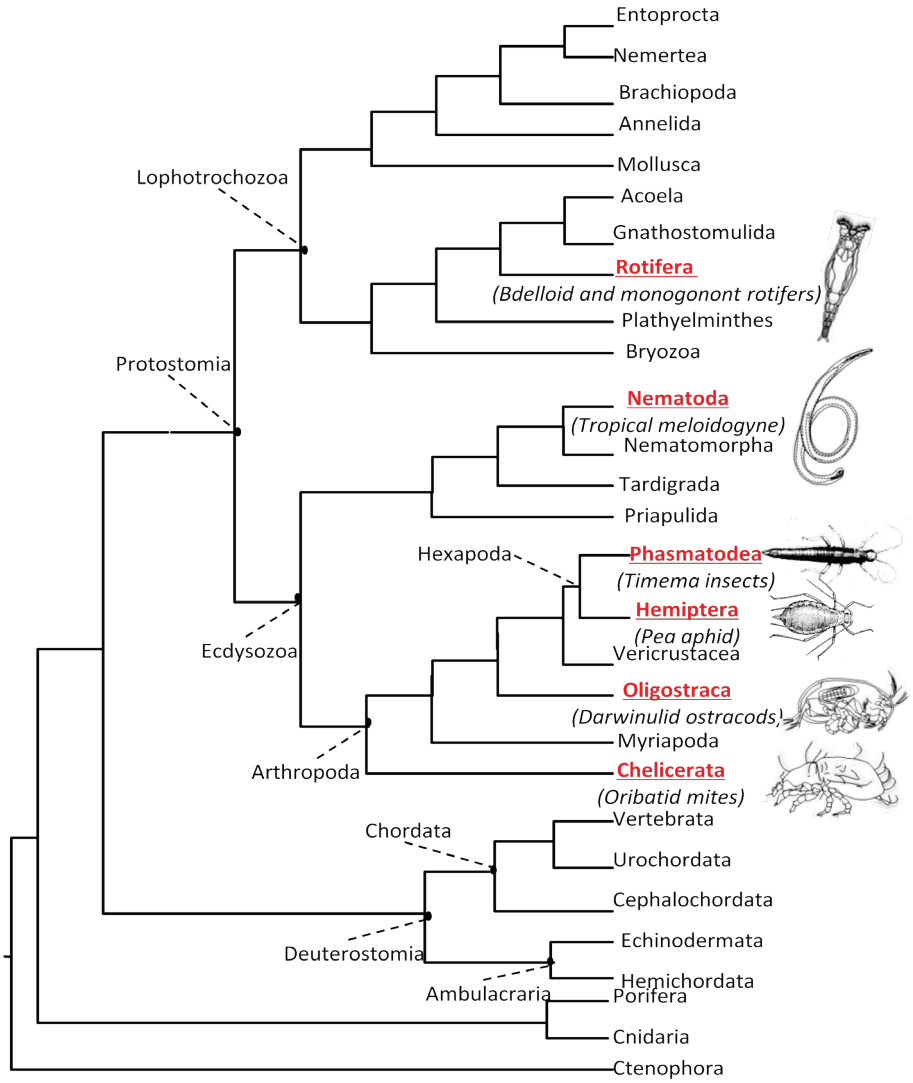



**Figure 11. Relative phylogenetic position of the six animal lineages containing asexual species**

In this simplified animal phylogeny, animal lineages that contain "more or less ancient" sexual species are highlighted in red. Four of the 6 lineages are arthropods.

Another challenging question is how the genomes of asexually reproducing animals evolve and adapt in the absence of recombination and genetic exchanges between individuals. This point will be investigated for the root-knot nematodes through the ANR-funded project discussed in section 0. For the other asexual animals, though, this point needs to be explored. This part of the project will focus on investigating the accumulation of deleterious and neutral mutations in the genome and changes in gene expression patterns using both low-coverage whole genome sequencing and RNA-seq. Mapping of reads to reference genomes will allow elucidating how asexually reproducing animals

evolve and adapt as a response to new environmental challenges. In the case of aphid species, the new environmental challenge will consist in switching the pest from a usually compatible host to a challenging new host. In the case of the bdelloid rotifers, the challenge will consist of cycles of desiccation followed by rehydration.

How asexual animal lineages can persist over evolutionary time and how they evolve and adapt in response to environmental changes and challenges is a major unresolved question in evolutionary biology. We can expect that future models of genome evolution in asexuals, emerging from this multi-species, multi-timescale comparative project, will be considered as a reference landmarks for the whole evolutionary genomics community.

While a mechanism of plasticity is necessary for adaptation and evolution, one of the other postulated advantages of sexual reproduction is that it prevents the progressive, step by step accumulation of deleterious mutations. In asexual animals, it is hypothesized that accumulation of mutations leads to genome decay through processes described by Muller's ratchet (Muller, 1964) and Kondrashov's hatchet (Kondrashov, 1988). Long term survival in asexuals probably requires a process that either prevents from the accumulation of mutations or purge deleterious mutations. This aspect, not included in the 2013 "France Génomique" project will also be very interesting to explore. In the bdelloid rotifer *A. vaga*, it has been proposed that extensive gene conversion can provide a mechanism to avoid progressive accumulation of deleterious mutations. Gene conversion either restores the mutated copy to the original version or make the mutated version present at an homozygous state, which can then be subject to elimination from the population through selection (Flot et al., 2013). Interestingly, a role of gene conversion for countering the accumulation of deleterious mutations in ancient asexuals has also been proposed as soon as 1998 for the ancient asexual darwinulid ostracods and evidences for gene conversion in at least one genomic position have been shown in *Darwinula stevensoni* (Schön and Martens, 1998, 2003). Also supporting this possibility, high levels of gene conversions have been observed in *Daphnia pulex*, a microcrustacean species that was re-considered as a recent asexual (Tucker et al., 2013). Future projects will have to investigate whether a similar hypothetical mechanism of deleterious mutations purge is at play in other asexual animal lineages too, including in root-knot nematodes.

## 3.5 Upcoming plant-parasitic nematode genomes and transcriptomes

Our group is currently leading the genome sequencing and annotation projects for two root-knot nematode species, *M. javanica* and *M. arenaria*. Like *M. incognita*, these two "tropical" plant parasitic species reproduce asexually and without meiosis. Accessing these two new genomes will allow verifying whether the singularities observed in the genome of *M. incognita* are further shared with other root-knot nematodes. For example, it will be very interesting to determine whether a peculiar genome structure, made of divergent copies is also found in these species. This will provide further evidence for an association of a re-arranged genome in diverged copies and an asexual mitotic mode of reproduction. Besides confirming that such a genome structure might represent a common signature of asexuality, several other aspects will also become investigable. An evident feature that we will investigate is the level of conserved synteny between the genomes of the three tropical root-knot nematodes. An initial raw comparison of the genomes of *M. hapla* and *M. incognita* revealed no extensive conservation of synteny but the three tropical species are much

more closely related one another than they are to *M. hapla*. Substantial conservation of synteny can thus be expected. Furthermore, comparison of the genome structure and level of divergence between the three obligate parthenogenetic species will also allow provide crucial elements to decipher the origin of the asexual root-knot nematodes and their evolutionary relationships. Distinguishing between the hypotheses of a hybrid origin vs. an old loss of sexuality accompanied by independent divergent of ancient allelic regions can be aided by the comparative analysis of the 3 genomes. If, for instance, a nearly identical genome copy is found between the 3 species while each second copy seems species-specific, this would strongly suggest a hybridization hypothesis. Indeed, in this case, it can be clearly hypothesized that one genome results from one common parental donor and the other ones from different donors through multiple hybridization events. Other important aspects that can be validated through access of these additional root-knot nematode genomes, including the abundance of transposable elements (TE). These elements were found to span a higher proportion of the *M. incognita* genome (36%) than that of *M. hapla* (12%). However, estimates of the TE content has been obtained using different methods in the two root-knot nematode genomes. A first important step would be to re-annotate the two genomes for TEs using exactly the same methods and parameters. Validation of the TE-richness in *M. incognita* is important because we have hypothesized that this characteristic might represent a signature of asexual reproduction. If a similar abundance was found in the genomes of *M. javanica* and *M. arenaria* but not in *M. hapla*, this would further support the idea that TE-richness is linked to strict asexual reproduction in root-knot nematodes. Hence, one of our upcoming projects will be to proceed with systematic annotation of TE in the different available root-knot nematode genomes to check for their abundance, distribution and diversity. As previously discussed, these TE might play a role in genome plasticity with possible incidence on important life traits such as the parasitism success.

Besides genomic data, our group has also generated transcriptomic data not only for root-knot nematodes but also for other plant-parasitic nematodes. RNA-seq data generated as part of the NEMATARGETS program for different developmental life stages of *M. incognita* provided transcriptional support to new candidate parasitism genes. These data also allowed determining that more than 96% of genes acquired by LGT in *M. incognita* are transcribed in at least one stage, reinforcing their successful domestication by the nematode genome (personal communication). Besides *M. incognita*, we have obtained RNA-seq data for *Pratylenchus coffeae* (clade 12), *Ditylenchus dipsaci* (clade 12), *Bursaphelenchus xylophilus* (clade 10) and *Xiphinema index* (clade 2). Interestingly, while *M. incognita* is a sedentary endoparasite, able to induce the development of a feeding site (giant cells), the two other clade 12 nematodes are migratory endoparasites and do not induce feeding sites. Comparison of the transcriptomes of migratory vs. sedentary parasites might reveal genes specifically induced in the development of the feeding site. This is an important perspective since sedentary endoparasites are the nematodes causing most of the damages to agriculture. The availability of the genome of the cyst nematode *Globodera pallida*, sequenced at the Wellcome Trust Sanger Center, will provide very interesting data for that question. Indeed, cyst nematodes are endoparasites that independently evolved the ability to induce a feeding site. Unlike the giant cells in root-knot nematodes that result from multiple cell replications without division, cyst nematode create a syncitium made of fusion of neighboring cells. Comparisons of genomes and transcriptomes of cyst and root-knot nematodes to those of migratory relatives will allow determining whether similar gene networks and pathways have been recruited independently for the evolution of this ability. *B. xylophilus* and *X. index* will allow investigating evolutionary question at a

deeper phylogenetic resolution. They belong to different clades (other than clade 12) and have evolved the ability to parasitize plants independently from root-knot and cyst nematodes. Hence, their comparisons to the genomes and transcriptomes of clade 12 nematodes will allow determining whether similar genomic modifications occurred as a result of adaptation to plant parasitism. The genome sequence of *B. xylophilus* is already available (Kikuchi et al., 2011) and the RNA-seq data we have generated can provide additional transcriptional support for gene prediction. A comparative genomics and transcriptomics analysis including the root-knot nematodes, the cyst nematode *G. pallida*, the upcoming genomes of *P. coffeae* and *Radopholus similis* (both being analyzed at NC State University in the US) as well as the genomes of non plant-parasitic nematodes will constitute an important landmark towards deciphering the genomic signatures of adaptation to plant parasitism. To this end, I have constituted an international consortium gathering project leaders of the different genome sequencing projects for plant-parasitic nematodes. We have recently written a review paper that proposed to identify whether common genomic signatures could be identified in the currently available genome data as a response to adaptation to plant parasitism (Bird et al., 2014). In addition to the acquisition of parasitism genes via horizontal gene transfers, we have remarked that nematode peptides that act as plant mimics were commonly found in both root-knot and cyst nematodes. However, it appears that this plant mimics are not homologs and have emerged twice independently in root-knot nematodes and cyst nematodes. Further supporting this idea, we have remarked that very few effector proteins were homologs between the different plant-parasitic nematode lineages, suggesting evolutionary convergence toward adaptation to a plant parasitism in nematodes. Another feature common to several plant-parasitic nematode genomes was the observed reduction of defense / detoxification arsenals. Plant parasitic nematodes lack many antifungal and antibacterial genes found in other nematodes. It is hypothesized that some of these genes have been lost because surviving within plant tissue protects from the usual predators and pathogens encountered in the other ecological niches. Due to lack of data from other clades, it is important to note that these conclusions were based exclusively on data for clade 12 and clade 10 nematodes (Figure 1). Interestingly, one of the species we are studying, *X. index* is a clade 2 plant-parasitic nematode and is thus much more distantly related than all the other nematodes mentioned above. Despite an initial EST analysis that allowed isolating ca. 1,400 genes (Furlanetto et al., 2005), there is no ongoing genome project for *X. index* to date. The RNA-seq analysis we have performed generated ~40 million illumina reads that we assembled in ~75,000 contigs. This represents a substantial dataset for exploring the gene content in this species too. Efforts are also being deployed in other laboratories to obtain "omics" data from clade 1 parasitic nematodes. Nothing yet is known for plant parasites of this clade at the sequence level. Once all the current genome projects are published, we plan to perform a large scale de novo comparative analysis including all the recent genomic and transcriptomic efforts, expanding the consortium I had initially built up with other interested colleagues. We expect that this comprehensive analysis, covering all the clades that include plant-parasitic species, will provide a general view of the multiple independent adaptations to plant parasitism in nematodes at the genome level.

# 4 Concluding remarks

Our comparative genomics efforts in plant-parasitic nematodes have allowed identification of different singularities that highlight evolutionary events and adaptations related to the emergence of plant parasitism. One major part of our research pointed to the importance of lateral gene transfers as key evolutionary events that not only promoted adaptation towards parasitism but significantly contributed to the composition of the nematode genome itself. At a more translational point of view, mining the genomes of root knot nematodes led to the identification of a series of novel genes specific to plant-damaging species. Inactivation of these genes yielded significant reduction of nematode infestation. This opened new genome-based perspectives for the development of safer and more specific control strategies. These strategies might be further developed in the future, driven by the advent of RNAi technology on the market which promises specific targeting of pest genes (Kupferschmidt, 2013). An interesting perspective concerning gene repertoire-based approaches studying adaptation to parasitism would be to compare the results obtained on plant-parasitic nematodes to those on animal parasites as well as on fungal, bacterial and oomycetes plant pathogens. We have already started to establish some parallels between the genomes of plant parasitic nematodes and those of other plant pathogens or animal parasites (Bird et al., 2014). These parallels include the secretion of parasitism genes in the host that manipulate important functions, cases of horizontal gene transfers and occasional signs of genome reduction associated with loss of some immunity genes. Further developing these comparisons will allow deciphering whether there are some common genomic features associated to adaptation to parasitism.

Besides our gene-based research we have also analyzed the genomes of plant-parasitic nematodes at a more structural point of view. The identification of a peculiar genome architecture in *M. incognita*, made of diverged pairs of genomic regions, opened a whole panel of hypotheses and research topics. Putting this genome structure in perspective with its ameiotic and asexual mode of reproduction as well as its host spectrum will constitute an important part of my future research. This topic also allowed me to start new collaborations on the broader aspect of animal genome evolution in the absence of sexual reproduction. Development of these collaborations and topics will probably result in emergence of new theories about genome evolution and adaptation in animals in the absence of sexual reproduction. These theories will hopefully be tested with the foreseen availability of new genomes from animals with this peculiar mode of reproduction.

# 5 Curriculum Vitae

## Personal details

**Etienne G.J. Danchin**

e-mail : etienne.danchin@sophia.inra.fr

French nationality. Born on July 12, 1976 in Lille (France).

http://edanchin.free.fr

## Qualifications

2004 PhD in Bioinformatics, Structural Biology and Genomics. Research field: **Comparative and evolutionary genomics**. Title: Reconstruction of ancestral genomic regions by comparative analysis of evolutionary conserved syntenies. Towards reconstructing the genome of the ancestor of all Bilaterian species (Urbilateria).University of Aix-Marseille II, Marseilles, France. Funded by the French research Ministry.

2001 Master degree of Bioinformatics, Structural Biology and Genomics. (Marseilles, France) Universities of Aix-Marseille I & II.

2000 D.U. P.R.E.M.I.C.E (One year University Diploma for Masters in Science's preparation for insertion and creativity in companies) Specialization: Immunology. Optional modules: **Business Management** (Superior Business school of Marseilles), **Communication**, **Databases**. University of Aix-Marseille II, Marseilles, France.

**TOEIC** (Test Of English For International Communication); level: 2+ (**Advanced Working Proficiency**).

1999 Master degree of Cell Biology and Physiology (specialized in Molecular and Cell Genetics). University of Aix-Marseille II, Marseilles, France.

1998 License degree of Cell Biology and Physiology. University of Aix-Marseille II, Marseilles, France.

## Career overview

Since April 2012: **Principal Investigator** (Chargé de Recherche 1ére classe) at INRA in « Institut Sophia Agrobiotech», Sophia-Antipolis. **Comparative and evolutionary genomics of plant-parasitic nematodes.**

Octobre 2007 – Avril 2012: **Full-time researcher** (Chargé de Recherche 2éme classe) at INRA in «Biotic Interactions and Plant Health» à Sophia-Antipolis. **Comparative and evolutionary genomics of plant-parasitic nematodes.**

April 2005 – September 2007 : **Postdoct** at "Architecture et Fonction des Macromolécules Biologiques (AFMB) laboratory, UMR 6098 - CNRS, Universités d'Aix-Marseille I et II). Research activity: **Functional Annotation** of Genomes, **Comparative and Evolutionary Genomics, Evolutionary and Adaptive history of Enzyme repertoires, Phylogeny**.

October 2004 – March 2005: **Assistant Professor** at Aix-Marseille I University, France. Phylogenomics laboratory, EA EGEE 3781. Research and courses in **Comparative** and **Evolutionary genomics**, in **Phylogenomics**, and **Bioinformatics**.

September 2001 – September 2004: **PhD student** in Bioinformatics in the **Phylogenomics Laboratory** (EA Evolution, Genome, Environment) (Supervisor Dr. Pierre Pontarotti) at Aix-Marseille I University. **Search for evolutionary conserved gene clusters and reconstruction of ancestral genomic regions.**

September 2000 – July 2001: **Internship at INSERM (unit 119)** for Master diploma in Bioinformatics, at Phylogenomics Laboratory (Head: Dr. Pierre Pontarotti). Main focus: G**enomic data analysis, Bioinformatics**; large scale genomes comparisons.

October 1999 - June 2000: **Internship at INSERM (unit 119)** in Phylogenomics Laboratory (Head: Dr. Pierre Pontarotti). Main focus: **genomic data analysis, Bioinformatics,** degenerate PCR, genes **cloning** in the Amphioxus.

## Publications

**Publications in international peer-reviewed journals:**

**47 publications** including three in **Nature**, three in **Nature Biotechnology** and one in **PNAS**; main topic: comparative genomics, evolutionary biology and bioinformatics.
**10** as first author, **7** as last author.
More than **2,200 citations** 2004-2013; h-index = **17**

**2014**

- 47) Castagnone-Sereno P., Danchin E.G.J. **Parasitic success without sex – the nematode experience.** Journal of Evol. Biol. (In Press).
- 46) Bird D.M., Jones J.T., Opperman C.H., Kikuchi T., Danchin E.G.J. **Signatures of adaptation to plant parasitism in nematode genomes.** Parasitology (In Press).

**2013**

- 45) Danchin EGJ, Arguel M-J, Campan-Fournier A, Perfus-Barbeoch L, Magliano M, Rosso M-N, Da Rocha M, Da Silva C, Nottet N, Labadie K, Guy J, Artiguenave F, Abad P. **Identification of novel target genes for safer and more specific control of root-knot nematodes from a pan-genome mining.** PLoS Pathogens.
- 44) Flot JF, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EGJ, Hejnol A, Henrissat B,..., Pontarotti P, Weissenbach J, Wincker P, Jaillon O, Van Doninck K. **Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga***. Nature.
- 43) Jones J. T., Haegeman A., Danchin E.G.J., Gaur H.S., Helder J., Jones M., Kikuchi T., Manzanilla-Lopez R., Palomares-Rius J., Wesemael W., Perry R.N. **Top 10 plant parasitic nematodes in molecular plant pathology.** Mol Plant Path.
- 42) Castagnone-Sereno P., Danchin E.G.J., Perfus-Barbeoch L. and Abad P. **Diversity and evolution of root-knot nematodes, genus Meloidogyne: new insights from the genomic era.** Annual Review of Phytopathology.

**2012**

- 40) Paganini J, Campan-Fournier A, Da Rocha M, Gouret P, Pontarotti P, Wajnberg E, Abad P, Danchin EGJ. **Contribution of Lateral Gene Transfers to the Genome Composition and Parasitic Ability of Root-Knot Nematodes**. PLoS One.
- 39) Arias MC, Danchin EGJ, Coutinho PM, Henrissat B, Ball S. **Eukaryote to gut bacteria transfer of a glycoside hydrolase gene essential for starch breakdown in plants**. Mobile Genetic Elements.
- 38) Jaouannet M, Perfus-Barbeoch L, Deleury E, Magliano M, Engler G, Vieira P, Danchin EGJ, Da Rocha M, Coquillard P, Abad P, Rosso MN. **A root-knot nematode secreted protein is injected into giant-cells and targeted to the nuclei**. New Phytol.
- 37) Danchin E.G.J, Rosso MN. **Lateral gene transfers have polished animal genomes: lessons from Nematodes**. Front. Cell. Inf. Microbiol.
- 36) Bignon C, Coutard B, Danchin E.G.J, Oubelaid R, Canard B. **Single pH buffer refolding screen for protein from inclusion bodies.** Protein Expression and Purification.

**2011**

- 35) Danchin E.G.J. **What Nematode genomes tell us about the importance of horizontal gene transfers in the evolutionary history of animals ?** Mobile Genetic Elements.
- 34) Amselem J., Cuomo C.A., van Kan J.A.L., Viaud M., Benito E.P., Couloux A., Coutinho P.M, de Vries R.P.,...,Danchin E.G., Da Silva C.,...,Rollins J.A., Lebrun MH, Dickman M. **Genomic analysis of the necrotrophic fungal pathogens Sclerotinia sclerotiorum and Botrytis cinerea.** PLoS Genetics.

- 33) Hok S., Danchin E.G.J., Allasia V., Panabières F., Attard A., Keller H. **An Arabidopsis (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease**. Plant, Cell & Environment.
- 32) Haegeman A., Jones J., Danchin E.G.J. **Horizontal gene transfer in nematodes: a catalyst for plant parasitism?** MPMI.
- 31) Vens C., Rosso MN., Danchin E.G.J. **Identifying discriminative classification-based motifs in biological sequences.** Bioinformatics.
- 30) Marino D., Andrio E., Danchin E.G.J., Oger E., Gucciardo S., Lambert A., Puppo A., Pauly N. **A Medicago truncatula NADPH oxidase is involved in symbiotic nodule functioning**. New Phytol.
- 29) Vieira P., Danchin E.G.J., Neveu C., Crozat C., Jaubert S., Hussey R.S., Engler G., Abad P., de Almeida-Engler J., Castagnone-Sereno P., Rosso M.N. **The plant apoplasm is an important recipient compartment for nematode secreted proteins.** J Exp Bot.
- 28) Castagnone-Sereno P., Deleury E., Perfus-Barbeoch L., Danchin E.G.J., Abad P. **Data-mining of the Meloidogyne incognita degradome and comparative analysis of proteases in nematode.** Genomics.

**2010**

- 27) Castagnone-Sereno P, Danchin E.G.J., Deleury E, Guillemaud T, Malausa T, Abad P. **Genome-wide survey and analysis of microsatellites in nematodes, with a focus on the plant-parasitic species Meloidogyne incognita**. BMC Genomics.
- 26) Danchin E.G.J., Rosso M-N., Vieira P., de Almeida-Engler J., Coutinho P, Henrissat B., Abad P. **Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes.** PNAS.
- 25) Ma L., van der Does C., Borkovich K.A., Coleman J.J., ... , Danchin E.G.J., ... , Galagan J., Cuomo C.A., Kistler H.C., Rep M. **Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium.** Nature.

**2009**

- 24) Lopez-Rascol V, Levasseur A, Chabrol O, Gruseas S, Gouret P, Danchin EGJ, Pontarotti P. **CASSIOPE: an expert system for conserved regions searches.** BMC Bioinformatics.
- 23) Coleman J.J, Rounsley S.D, Rodriguez-Carres M, Kuo A, ... , Danchin EGJ, Henrissat B, Coutinho P.M, ... , Covert S.F, Temporini E, VanEtten H.D. **The genome of Nectria haematococca: contribution of supernumerary chromosomes to gene expansion.** Plos Genetics.
- 22) Bird D.M, Williamson V.M, Abad P, McCarter J, Danchin EGJ, Castagnone-Sereno P, Opperman C.H.. **The genomes of root-knot nematodes.** Annu. Rev. Phytopatol.

**2008**

- 21) Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EGJ,..., Rosso MN, Schiex T, Smant G, Weissenbach J, Wincker P. **Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita.** Nat Biotechnol.
- 20) Darbo E, Danchin EGJ, Mc Dermott MF, Pontarotti P. **Evolution of major histocompatibility complex by "en bloc" duplication before mammalian radiation.** Immunogenetics.
- 19) Espagne E, Lespinet O, Malagnac F, ... , de Vries RP, Battaglia E, Coutinho PM, Danchin EGJ, Henrissat B, ... , Debuchy R, Wincker P, Weissenbach J, Silar P. **The genome sequence of the model ascomycete fungus Podospora anserina.** Genome Biol.
- 18) Martinez D, Berka RM, Henrissat B, Saloheimo M, ... , Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, ... , Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS. **Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).** Nat Biotechnol.
- 17) Martin F, Aerts A, Ahrén D, Brun A, Danchin E.G.J,...,Coutinho P.M,..., Henrissat B,...,Tuskan G & Grigoriev I. V. **The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis.** Nature.

**2007**

- 16) Benoit I, Danchin E.G.J, Bleichrodt R.J., de Vries R.P. **Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity.** Biotechnology Letters.
- 15) Da Lage J-L, Danchin E.G.J, Casane D. **Where do animal alpha-amylases come from? An interkingdom trip.** FEBS Letters.
- 14) Levasseur A, Orlando L, Bailly X, Milinkovitch M.C, Danchin E.G.J, Pontarotti P. **Conceptual bases for quantifying the role of environmental changes on gene evolution: the participation of positive selection and neutral evolution**. Biological Reviews.
- 13) Pel H.J., de Winde J.H.,..., Coutinho P.M, Danchin E.G.J.,..., Henrissat B,..., Visser J., Stam H. **Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88.** Nat Biotechnol.

**2006**

- 12) Stam M.R., Danchin E.G.J., Rancurel C., Coutinho P.M., Henrissat B. **Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins.** PEDS.
- 11) Asther M, Record E, Gimbert I, Sigoillot J-C, Levasseur A, Danchin E.G.J, Coutinho P.M., Monot F. **Des champignons qui carburent: de l'étude du génome à la production industrielle.** Biofutur, 28 (269).
- 10) Danchin E.G.J, Levasseur A, Lopez-Rascol V, Gouret P, Pontarotti P. **The use of evolutionary biology concepts for genome annotation**. J. Exp. Zoology Part B: Mol. and Dev. Evol.
- 9) Danchin E.G.J, Gouret P, Pontarotti P. **Eleven ancestral gene families lost in mammals and vertebrates while otherwise universally conserved in animals.** BMC Evolutionary Biology.

**2005**

- 8) Balandraud N, Gouret P, Danchin E.G.J, Blanc M, Zinn D, Roudier J, Pontarotti P. **A rigorous method for multigenic families' functional annotation: the peptidyl arginine deiminase (PAD) proteins family example.** BMC Genomics.
- 7) Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin E.G.J. **FIGENIX: Intelligent automation of genomic annotation: expertise integration in a new software platform.** BMC Bioinformatics.

**2004**

- 6) Danchin EGJ, Pontarotti P. **Towards reconstructing the bilaterian ancestral pre-MHC region.** Trends Genet.
- 5) Danchin EG, Pontarotti P. **Statistical evidence for a more than 800 Million years old evolutionary conserved genomic region in our genome.** Journal of Molecular Evolution.
- 4) Meglécz E, Petenian F, Danchin E, Coeur D'Acier A, Rasplus JY, Faure E. **High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera : Parnassius apollo and Euphydryas aurinia**. Molecular Ecology
- 3) Danchin E, Vitiello V, Vienne A, Richard O, Gouret P, McDermott MF, Pontarotti P. **The Major Histocompatibiliy Complex Origin.** Immunological reviews.

**2003**

- 2) Vienne A, Shiina T, Abi Rached L, Danchin E, Vitiello V, Cartault F, Inoko H, P Pontarotti. **Evolution of the proto-MHC ancestral region: more evidence for the plesiomorphic organisation of human chromosome 9q34 region.** Immunogenetics.
- 1) Danchin EG, Abi-Rached L, Gilles A, Pontarotti P. **Conservation of the MHC-like region throughout evolution.** Immunogenetics.

**Book chapters:**

**4,** including 3 as first author**.** Topics: conserved syntenies, genome annotation and comparative genomics, evolutionary biology.

- 4) Danchin EGJ, Flot JF, Perfus-Barbeoch L and Van Doninck K. **Genomic Perspectives on the Long-Term Absence of Sexual Reproduction in Animals.** In " Evolutionary Biology: Concepts, Biodiversity, Macroevolution and Genome Evolution" (Editor: Pierre Pontarotti for Springer-Verlag). 2011
- 3) Coutinho PM, Rancurel C, Stam M, Bernard T, Couto FM, Danchin EGJ and Henrissat B. **Carbohydrate-Active Enzymes Database: Principles and Classification of Glycosyltransferases**. In "Bioinformatics for Glycobiology and Glycomics: An Introduction" (Editors: Claus-Wilhelm von der Lieth, Thomas Luetteke and Martin Frank for Wiley Press). 2009
- 2) Danchin EGJ, Laetitia Perfus-Barbeoch. **The Genome Sequence of Meloidogyne incognita Unveils Mechanisms of Adaptation to Plant-Parasitism in Metazoa.** In " Evolutionary Biology: Concept, Modeling, and Application " (Editor: Pierre Pontarotti for Springer-Verlag). 2009
- 1) Danchin EGJ, Gaucher EA, Pontarotti P. **Computational reconstruction of ancestral genomic regions from evolutionarily conserved gene clusters.** In “Ancestral Sequence Reconstruction” (Editor: David Liberles, for Oxford University Press). 2007

## Posters and seminars

**43 oral presentations** and **posters**.

**12 invited seminars**

## Editorial functions

Associate Editor for "BMC Evolutionary Biology" (Biomed Central, ISSN: 1471-2148)

Review Editor for "Frontiers in Evolutionary and Genomic Microbiology" (Frontiers, ISSN: 1664-302X)

Section Editor "Evolutionary Biology" for "Biology Direct" (Biomed Central, ISSN: 1745-6150)

## Reviewer for international journals

Nature Communications
Trends in Parasitology
PLoS ONE
BMC Evolutionary Biology
Journal of Molecular Evolution
Molecular Plant Pathology
Mobile Genetic Elements
Comptes rendus Biologies
Journal of Theoretical Biology
Immunogenetics
PLoS Computational Biology
Nematology
BMC Genomics
Proceedings of the Royal Society, Biological Sciences
GENE
Insect Biochemistry and Molecular Biology
Cambridge University Press Books
Phytochemistry
Plant Molecular Biology

## Grants and projects coordination

- 2014 - French national program **France Génomique** "Large-scale DNA sequencing projects", project "GALA" (Genomics of Asexual Lineages of Animals). Project **writer**, **coordinator** and principal investigator. International project including a collaboration with 7 European laboratories. Topic: Genome evolution in asexually-reproducing animals. Project rated "A" but not granted for this call.

- 2014 - 2017 - French national program **ANR JCJC (young scientist) 2013**, project ASEXEVOL. Project **writer**, **coordinator** and principal investigator. Topic "Functional consequences of the *M. incognita* peculiar genome structure." Project granted, starting in 2014.

- 2014 - 2015 - Grant from "**Université de Nice - Sophia Antipolis**" to fund a 12 month postdoctoral position. Complement to the project ASEXEVOL. Project **writer**, **coordinator** and principal investigator. Project granted, starting in 2014.

- 2011 - 2014 - **INRA** program "Plant Health and Environment" AAP-SPE 2011. **Project coordinator**. co-writer with Laetitia Zurletto MCF, Université de Nice. Topic : Comparative analysis of mitochondrial genomes of root-knot nematodes". Project granted and ongoing.

- 2009 - **CNRS** Interdisciplinary program "**Maladies Infectieuses Emergentes**" MIE. Project **writer**, **coordinator** and principal investigator. Project including a collaboration with two laboratories from "Aix-Marseille Université" and CNRS. Topic: Systematic search for horizontal gene transfers in the genomes of root-knot nematodes. Project granted, finished in 2012 : 3 publications.

- 2008 - 2014. French national sequencing program **GENOSCOPE**. Project co-writer, **principal investigator** on bioinformatics aspects. Topic: genome sequencing and assembly of the root-knot nematodes *M. javanica* et *M. arenaria*, deeper sequencing of *M. incognita*. Project granted, genomes recently assembled and annotation ongoing.

- 2008 - 2013. French national program **ANR GENOPLANTE**, project "Nematargets". Project co-writer, **principal investigator** and supervisor of the bioinformatics genome mining aspects. Project granted, finished in 2013 : 5 publications and one patent.

## Chairman for meetings and conferences

Chairman "Nematode Genomics and Transcriptomics" session at the 31st international symposium of the European Society of Nematologists (ESN) in Adana, Turkey. 23-27 September 2012.

Chairman "Fungal Genomics" session at 9èmes Rencontres de Phytopathologie - Mycologie de la Société Française de Phytopathologie in Aussois (73), France. 16-20 January 2012.

## Scientific committee member for conferences

13th edition of "Journées Ouvertes Biologie Informatique et Mathématiques" (JOBIM), Rennes, France, 3-6 July 2012.

Annual "Evolutionary Biology Meeting in Marseilles" (EBM), Marseille, France.

## Teaching / supervisor

**Member of Thesis jury of** :

- Dr. Katarzyna Rybarcyk-Mydłowska. «Phylogenetic relationships within major nematode clades based on multiple molecular markers», defended October 7, 2013 at University of Wageningen, Netherlands.
- Dr. Eduardo De Paula Costa. «Algorithms for analyzing biological sequences», defended June 4, 2013 at KU Leuven, Belgium.
- Dr. Phuong Thi LE, «Lateral Transfer of Sequences» defended March 14, 2013 at Université de la Méditerranée, Faculté de Médecine de la Timone, Marseilles, France.
- Dr. Kalliopi Georgiades, « Phylogenomics of pathogenic Bacteria» defended September 08, 2011 at Université de la Méditerranée, Faculté de Médecine de la Timone, Marseilles, France.
- Dr. Vincent Lombard, «Structuration et exploration d'information génomique et fonctionnelle des enzymes actives sur les glucides» defended 12 May 2011 at Université de la Méditerranée, faculté des Sciences de Luminy, Marseilles, France.
- Dr. Virginie Lopez-Rascol, «L'évolution moléculaire : de la comparaison vers la reconstruction » defended 14 November 2008 at l'Université de Provence, centre Saint-Charles à Marseilles, France.

**Supervision**:

- Morgane Demeocq, Master I Biology, Bio-Info-Maths speciality, Université de Nice - Sophia Antipolis, February - June 2014.
- Dr. Romain Blanc-Mathieu, postdoc, ANR project ASEXEVOL, January 2014 - June 2016.
- Nicolas Nottet, Master II Biology, Bio-Info-Maths speciality, Université de Nice - Sophia Antipolis, January-June 2013.
- Lauriane Massardier, Master I Biology, Bio-Info-Maths speciality, Université de Nice - Sophia Antipolis, January - July 2012.
- Vivien Deshaies, Master I Bioinformatics, Université de Rouen, March – June 2011.
- Martine Da Rocha, Master II Bioinformatics, Université de Rouen March 2010 – June 2011.
- Amandine Campan-Fournier, research engineer ANR January 2010 –September 2011.
- Dr. Celine Vens, postdoc, KU Leuven October 2009 – September 2010.

**Teaching:**

Master of Biology, Facultés Universitaires Notre Dame de la Paix (FUNDP), Namur, Belgium, 14 december 2013. Two hours course on genomic adaptations to plant-parasitism in nematodes.

European Master of Science in Nematology (EUMAINE, Erasmus Mundus Master), Ghent, Belgium, le 3 July 2009. One hour course on comparative genomics of plant-parasitic nematodes.

Practical course on bioinformatics tools for genome annotation of plant-parasitic nematodes given to the members of the European COST Action 872 « Exploiting Genomics to Understand Plant-Nematode Interactions ». 19 - 21 November 2008, at INRA, Sophia-Antipolis, France.

Course at INRA Research School: *La Phylogénomique : une aide à l'étude des grandes fonctions du vivant*. Carry le rouet, France, 12-14 Décembre 2006.

2000 - 2005 82 hours of course given to Licence and Master students from "Aix-Marseille Université" on comparative genomics, genome evolution and bioinformatics, mainly as an assistant professor (ATER, 2004).

# 6 Bibliography


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., et al. (2008). Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nat. Biotechnol. *26*, 909–915.

Agrios, G.N. (2005). Plant Pathology, 5th Edition (Burlington, USA: Elsevier Academic Press).

Aguinaldo, A.M., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., and Lake, J.A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature *387*, 489–493.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389–3402.

Amselem, J., Cuomo, C.A., van Kan, J.A.L., Viaud, M., Benito, E.P., Couloux, A., Coutinho, P., de Vries, R.P., Dyer, P.S., Fillinger, S., et al. (2011). Genomic analysis of the necrotrophic fungal pathogens Sclerotinia sclerotiorum and Botrytis cinerea. PLoS Genet.

Andersson, J.O. (2005). Lateral gene transfer in eukaryotes. Cell Mol Life Sci *62*, 1182–1197.

Arkhipova, I., and Meselson, M. (2005). Deleterious transposable elements and the extinction of asexuals. Bioessays *27*, 76–85.

Baldwin, J.G., Nadler, S.A., and Adams, B.J. (2004). Evolution of plant parasitism among nematodes. Annu Rev Phytopathol *42*, 83–105.

Van der Beek, J.G., and Karssen, G. (1997). Interspecific Hybridization of Meiotic Parthenogenetic Meloidogyne chitwoodi and M. fallax. Phytopathology *87*, 1061–1066.

Bert, W., Karssen, G., and Helder, J. (2011). Phylogeny and Evolution of Nematodes. In Genomics and Molecular Genetics of Plant-Nematode Interactions, J. Jones, G. Gheysen, and C. Fenoll, eds. (Dordrecht: Springer Netherlands), pp. 45–59.

Bird, D.M., Williamson, V.M., Abad, P., McCarter, J., Danchin, E.G., Castagnone-Sereno, P., and Opperman, C.H. (2009). The genomes of root-knot nematodes. Annu Rev Phytopathol *47*, 333–351.

Bird, D.M., Jones, J.T., Opperman, C.H., Kikuchi, T., and Danchin, E.G.J. (2014). Signatures of adaptation to plant parasitism in nematode genomes. Parasitology *FirstView*, 1–14.

Blaxter, M. (2007). Symbiont genes in host genomes: fragments with a future? Cell Host Microbe *2*, 211–213.

Blaxter, M. (2011). Nematodes: The Worm and Its Relatives. PLoS Biol *9*, e1001050.

Blaxter, M.L., Ley, P.D., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., et al. (1998). A molecular evolutionary framework for the phylum Nematoda. Nature *392*, 71–75.

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E., Nesbo, C.L., Case, R.J., and Doolittle, W.F. (2003). Lateral gene transfer and the origins of prokaryotic groups. Annu Rev Genet *37*, 283–328.

Boukharov, A.A. [US/US], Du, Z. [US/US], Guo, L. [US/US], Hresko, M.C. [US/US], Kovalic, D.K. [US/US], Li, Z. [US/US], Lu, M. [US/US], McCarter, J.P. [US/US], Miller, N.M. [US/US], Vaudin, M. [GB/GB], et al. (2007). Identification and use of target genes for control of plant parasitic nematodes (US).

Bullini, L. (1994). Origin and evolution of animal hybrid species. Trends Ecol. Evol. *9*, 422–426.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res *37*, D233–8.

Castagnone-Sereno, P., Danchin, E.G., Perfus-Barbeoch, L., and Abad, P. (2013). Diversity and Evolution of Root-Knot Nematodes, Genus Meloidogyne: New Insights from the Genomic Era. Annu Rev Phytopathol *51*, 130517121443006.

Castillo-Davis, C.I., Hartl, D.L., and Achaz, G. (2004). cis-Regulatory and protein evolution in orthologous and duplicate genes. Genome Res *14*, 1530–1536.

Coleman, J.J., Rounsley, S.D., Rodriguez-Carres, M., Kuo, A., Wasmann, C.C., Grimwood, J., Schmutz, J., Taga, M., White, G.J., Zhou, S., et al. (2009). The genome of Nectria haematococca: contribution of supernumerary chromosomes to gene expansion. PLoS Genet *5*, e1000618.

Conant, G.C., and Wagner, A. (2004). Duplicate genes and robustness to transient gene knock-downs in Caenorhabditis elegans. Proc Biol Sci *271*, 89–96.

Daimon, T., Taguchi, T., Meng, Y., Katsuma, S., Mita, K., and Shimada, T. (2008). Beta-fructofuranosidase genes of the silkworm, Bombyx mori: insights into enzymatic adaptation of B. mori to toxic alkaloids in mulberry latex. J Biol Chem *283*, 15271–15279.

Danchin, É.G.J. (2011). What Nematode genomes tell us about the importance of horizontal gene transfers in the evolutionary history of animals. Mob. Genet. Elem. *1*, 269–273.

Danchin, E.G.J., and Rosso, M.-N. (2012). Lateral gene transfers have polished animal genomes: lessons from Nematodes. Front. Cell. Infect. Microbiol. *2*.

Danchin, E., Flot, J.-F., Perfus-Barbeoch, L., and Van Doninck, K. (2011). Genomic perspectives on the long-term absence of sexual reproduction in animals. In Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution, P. Pontarotti, ed. (Berlin Heidelberg: Springer),.

Danchin, E.G., Rosso, M.N., Vieira, P., de Almeida-Engler, J., Coutinho, P.M., Henrissat, B., and Abad, P. (2010). Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc Natl Acad Sci U A *107*, 17651–17656.

Danchin, E.G.J., Arguel, M.-J., Campan-Fournier, A., Perfus-Barbeoch, L., Magliano, M., Rosso, M.-N., Da Rocha, M., Da Silva, C., Nottet, N., Labadie, K., et al. (2013). Identification of Novel Target Genes for Safer and More Specific Control of Root-Knot Nematodes from a Pan-Genome Mining. PLoS Pathog *9*, e1003745.

Delmotte, F., SABATER-MUñOZ, B., PRUNIER-LETERME, N., Latorre, A., Sunnucks, P., Rispe, C., and Simon, J.-C. (2003). Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. Evolution *57*, 1291–1303.

Dieterich, C., Clifton, S.W., Schuster, L.N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R., Godfrey, J., Minx, P., et al. (2008). The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism. Nat Genet *40*, 1193–1198.

Dolgin, E.S., and Charlesworth, B. (2006). The Fate of Transposable Elements in Asexual Populations. Genetics *174*, 817–827.

Doyle, E.A., and Lambert, K.N. (2002). Cloning and characterization of an esophageal-gland-specific pectate lyase from the root-knot nematode Meloidogyne javanica. Mol Plant Microbe Interact *15*, 549–556.

Dubon, G. (2011). Des nématodes dérobent des gènes a des bactéries. Réussir Fruits Légumes 42.

Dunning Hotopp, J.C. (2011). Horizontal gene transfer between bacteria and animals. Trends Genet *27*, 157–163.

Elsworth, B., Wasmuth, J., and Blaxter, M. (2011). NEMBASE4: The nematode transcriptome resource. Int. J. Parasitol. *41*, 881–894.

Espagne, E., Lespinet, O., Malagnac, F., Da Silva, C., Jaillon, O., Porcel, B.M., Couloux, A., Aury, J.M., Segurens, B., Poulain, J., et al. (2008). The genome sequence of the model ascomycete fungus Podospora anserina. Genome Biol *9*, R77.

Fargette, M., Berthier, K., Richaud, M., Lollier, V., Franck, P., Hernandez, A., and Frutos, R. (2010). Crosses prior to parthenogenesis explain the current genetic diversity of tropical plant-parasitic Meloidogyne species (Nematoda: Tylenchida). Infect. Genet. Evol. *10*, 806–813.

Flot, J.-F., Hespeels, B., Li, X., Noel, B., Arkhipova, I., Danchin, E.G.J., Hejnol, A., Henrissat, B., Koszul, R., Aury, J.-M., et al. (2013). Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. Nature *500*, 453–457.

Fu, Z., Agudelo, P., and Wells, C.E. (2012). Differential Expression of a β-1,4-Endoglucanase Induced by Diet Change in the Foliar Nematode *Aphelenchoides fragariae*. Phytopathology *102*, 804–811.

Furlanetto, C., Cardle, L., Brown, D.J., and Jones, J.T. (2005). Analysis of expressed sequence tags from the ectoparasitic nematode Xiphinema index. Nematology *7*, 95–104.

Ghedin, E., Wang, S., Foster, J.M., and Slatko, B.E. (2004). First sequenced genome of a parasitic nematode. Trends Parasitol. *20*, 151–153.

Ghedin, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., Allen, J.E., Delcher, A.L., Guiliano, D.B., Miranda-Saavedra, D., et al. (2007). Draft genome of the filarial nematode parasite Brugia malayi. Science *317*, 1756–1760.

Gilbert, C., and Cordaux, R. (2013). Horizontal Transfer and Evolution of Prokaryote Transposable Elements in Eukaryotes. Genome Biol. Evol. *5*, 822–832.

Gilbert, C., Chateigner, A., Ernenwein, L., Barbe, V., Bézier, A., Herniou, E.A., and Cordaux, R. (2014). Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. Nat. Commun. *5*.

Gogarten, J.P., and Townsend, J.P. (2005). Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol *3*, 679–687.

Gouret, P., Vitiello, V., Balandraud, N., Gilles, A., Pontarotti, P., and Danchin, E.G.J. (2005). FIGENIX: Intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics *6*, 198.

Gouret, P., Thompson, J.D., and Pontarotti, P. (2009). PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics *10*, 298.

Gross, S.M., and Williamson, V.M. (2011). Tm1: A Mutator/Foldback Transposable Element Family in Root-Knot Nematodes. PLoS ONE *6*, e24534.

Gu, Z., Nicolae, D., Lu, H.H., and Li, W.H. (2002). Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet *18*, 609–613.

Gubanov, N.M. (1951). Giant nematoda from the placenta of Cetacea; Placentonema gigantissima nov. gen., nov. sp. Dokl. Akad. Nauk SSSR *77*.

Haegeman, A., Joseph, S., and Gheysen, G. (2011a). Analysis of the transcriptome of the root lesion nematode Pratylenchus coffeae generated by 454 sequencing technology. Mol Biochem Parasitol *178*, 7–14.

Haegeman, A., Jones, J.T., and Danchin, E.G. (2011b). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? Mol Plant Microbe Interact *24*, 879–887.

Haegeman, A., Mantelin, S., Jones, J.T., and Gheysen, G. (2012). Functional roles of effectors of plant-parasitic nematodes. Gene *492*, 19–31.

Haegeman, A., Bauters, L., Kyndt, T., Rahman, M.M., and Gheysen, G. (2013). Identification of candidate effector genes in the transcriptome of the rice root knot nematode Meloidogyne graminicola. Mol. Plant Pathol. *14*, 379–390.

Hamilton, J.P., Neeno-Eckwall, E.C., Adhikari, B.N., Perna, N.T., Tisserat, N., Leach, J.E., Levesque, C.A., and Buell, C.R. (2011). The Comprehensive Phytopathogen Genomics Resource: a web-based resource for data-mining plant pathogen genomes. Database Oxf. *2011*, bar053.

Hörandl, E. (2009). Geographical Parthenogenesis: Opportunities for Asexuality. In Lost Sex, I. Schön, K. Martens, and P. Dijk, eds. (Springer Netherlands), pp. 161–186.

Hugall, A., Stanton, J., and Moritz, C. (1997). Evolution of the AT-rich mitochondrial DNA of the root knot nematode, Meloidogyne hapla. Mol Biol Evol *14*, 40–48.

Jaouannet, M., Perfus-Barbeoch, L., Deleury, E., Magliano, M., Engler, G., Vieira, P., Danchin, E.G., Da Rocha, M., Coquillard, P., Abad, P., et al. (2012). A root-knot nematode-secreted protein is injected into giant cells and targeted to the nuclei. New Phytol *194*, 924–931.

Jaubert, S., Laffaire, J.B., Abad, P., and Rosso, M.N. (2002). A polygalacturonase of animal origin isolated from the root-knot nematode Meloidogyne incognita. FEBS Lett *522*, 109–112.

Jones, J.T., Kumar, A., Pylypenko, L.A., Thirugnanasambandam, A., Castelli, L., Chapman, S., Cock, P.J., Grenier, E., Lilley, C.J., Phillips, M.S., et al. (2009). Identification and functional characterization of effectors in expressed sequence tags from various life cycle stages of the potato cyst nematode Globodera pallida. Mol Plant Pathol *10*, 815–828.

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., et al. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. Mol. Plant Pathol. *14*, 946–961.

Judson, O.P., and Normark, B.B. (1996). Ancient asexual scandals. Trends Ecol Evol *11*, 41–46.

Kearney, M. (2005). Hybridization, glaciation and geographical parthenogenesis. Trends Ecol. Evol. *20*, 495–502.

Kearney, M.R. (2003). Why is sex so unpopular in the Australian desert? Trends Ecol. Evol. *18*, 605–607.

Kikuchi, T., Cotton, J.A., Dalzell, J.J., Hasegawa, K., Kanzaki, N., McVeigh, P., Takanashi, T., Tsai, I.J., Assefa, S.A., Cock, P.J., et al. (2011). Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen Bursaphelenchus xylophilus. PLoS Pathog *7*, e1002219.

Koltai, H., Spiegel, Y., and Blaxter, M.L. (1997). Regulated use of an alternative spliced leader exon in the plant parasitic nematode Meloidogyne javanica. Mol. Biochem. Parasitol. *86*, 107–110.

Kondrashov, A.S. (1988). Deleterious mutations and the evolution of sexual reproduction. Nature *336*, 435–440.

Kraaijeveld, K., Zwanenburg, B., Hubert, B., Vieira, C., De Pater, S., Van Alphen, J.J.M., Den Dunnen, J.T., and De Knijff, P. (2012). Transposon proliferation in an asexual parasitoid. Mol. Ecol. *21*, 3898–3906.

Kumar, S., Schiffer, P.H., and Blaxter, M. (2012). 959 Nematode Genomes: a semantic wiki for coordinating sequencing projects. Nucleic Acids Res. *40*, D1295–1300.

Kupferschmidt, K. (2013). A Lethal Dose of RNA. Science *341*, 732–733.

Lambshead, P.J.D. (1993). Recent developments in marine benthic biodiversity research.

Lambshead, P.J.D., and Boucher, G. (2003). Marine nematode deep-sea biodiversity – hyperdiverse or hype? J. Biogeogr. *30*, 475–485.

Lapp, N.A., and Triantaphyllou, A.C. (1972). Relative DNA Content and Chromosomal Relationships of some Meloidogyne, Heterodera, and Meloidodera spp. (Nematoda: Heteroderidae). J. Nematol. *4*, 287–291.

Ledger, T.N., Jaubert, S., Bosselut, N., Abad, P., and Rosso, M.N. (2006). Characterization of a new beta-1,4-endoglucanase gene from the root-knot nematode Meloidogyne incognita and evolutionary scheme for phytonematode family 5 glycosyl hydrolases. Gene *382*, 121–128.

Leroy, S., Duperray, C., and Morand, S. (2003). Flow cytometry for parasite nematode genome size measurement. Mol Biochem Parasitol *128*, 91–93.

De Ley, P. (2006). A quick tour of nematode diversity and the backbone of nematode phylogeny. WormBook Online Rev. C Elegans Biol. 1–8.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res *13*, 2178–2189.

Li, W.H., Yang, J., and Gu, X. (2005). Expression divergence between duplicate genes. Trends Genet *21*, 602–607.

Lindblom, T.H., and Dodd, A.K. (2006). Xenobiotic detoxification in the nematode Caenorhabditis elegans. J Exp Zool. Comp Exp Biol *305*, 720–730.

Lunt, D.H. (2008). Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. BMC Evol Biol *8*, 194.

Ma, L.J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., et al. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. Nature *464*, 367–373.

Mark Welch, D., and Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science *288*, 1211–1215.

Martin, F., Aerts, A., Ahren, D., Brun, A., Danchin, E.G., Duchaussoy, F., Gibon, J., Kohler, A., Lindquist, E., Pereda, V., et al. (2008). The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. Nature *452*, 88–92.

Martin, J., Abubucker, S., Heizer, E., Taylor, C.M., and Mitreva, M. (2011). Nematode.net update 2011: addition of data sets and tools featuring next-generation sequencing data. Nucleic Acids Res.

Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J., Chertkov, O., Coutinho, P.M., Cullen, D., et al. (2008). Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). Nat Biotechnol *26*, 553–560.

Maupas, E. (1901). Modes et formes de reproduction des nematodes. Arch. Zool. Exp. Gen. *8*, 463–624.

Mavárez, J., and Linares, M. (2008). Homoploid hybrid speciation in animals. Mol. Ecol. *17*, 4181–4185.

Mayer, W.E., Schuster, L.N., Bartelmes, G., Dieterich, C., and Sommer, R.J. (2011). Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. BMC Evol Biol *11*, 13.

McCarter, J.P. (2009). Molecular Approaches Toward Resistance to Plant-Parasitic Nematodes. In Cell Biology of Plant Nematode Parasitism, R.H. Berg, and C.G. Taylor, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 239–267.

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

Meirmans, S. (2009). The Evolution of the Problem of Sex. In Lost Sex, I. Schön, K. Martens, and P. Dijk, eds. (Springer Netherlands), pp. 21–46.

Mitchum, M.G., Hussey, R.S., Baum, T.J., Wang, X., Elling, A.A., Wubben, M., and Davis, E.L. (2013). Nematode effector proteins: an emerging paradigm of parasitism. New Phytol.

Mitreva, M., Jasmer, D.P., Zarlenga, D.S., Wang, Z., Abubucker, S., Martin, J., Taylor, C.M., Yin, Y., Fulton, L., Minx, P., et al. (2011). The draft genome of the parasitic nematode Trichinella spiralis. Nat. Genet. *43*, 228–235.

Mitreva-Dautova, M., Roze, E., Overmars, H., de Graaff, L., Schots, A., Helder, J., Goverse, A., Bakker, J., and Smant, G. (2006). A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode Meloidogyne incognita. Mol Plant Microbe Interact *19*, 521–529.

Muller, H.J. (1964). The Relation of Recombination to Mutational Advance. Mutat Res *106*, 2–9.

Neiman, M., Meirmans, S., and Meirmans, P.G. (2009). What can asexual lineage age tell us about the maintenance of sex? Ann. N. Y. Acad. Sci. 185–200.

Nicol, P., Gill, R., Fosu-Nyarko, J., and Jones, M.G.K. (2012). de novo analysis and functional classification of the transcriptome of the root lesion nematode, Pratylenchus thornei, after 454 GS FLX sequencing. Int. J. Parasitol. *42*, 225–237.

Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., et al. (2008). Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. Proc Natl Acad Sci U A *105*, 14802–14807.

Pableo, E.C., and Triantaphyllou, A.C. (1989). DNA Complexity of the Root-knot Nematode (Meloidogyne spp.) Genome. J Nematol *21*, 260–263.

Paganini, J., Campan-Fournier, A., Da Rocha, M., Gouret, P., Pontarotti, P., Wajnberg, E., Abad, P., and Danchin, E.G.J. (2012). Contribution of Lateral Gene Transfers to the Genome Composition and Parasitic Ability of Root-Knot Nematodes. PLoS ONE *7*, e50875.

Parkinson, J., Whitton, C., Schmid, R., Thomson, M., and Blaxter, M. (2004). NEMBASE: a resource for parasitic nematode ESTs. Nucleic Acids Res *32*, D427–30.

Pel, H.J., de Winde, J.H., Archer, D.B., Dyer, P.S., Hofmann, G., Schaap, P.J., Turner, G., de Vries, R.P., Albang, R., Albermann, K., et al. (2007). Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88. Nat Biotechnol *25*, 221–231.

Plasterk, R.H.A. (2011). Surely You're Joking, Mr. Brenner. Science *331*, 32–32.

Poinar, G.O. (2011). The Evolutionary History of Nematodes: As Revealed in Stone, Amber and Mummies (BRILL).

Poinar, G., Kerp, H., and Hass, H. (2008). Palaeonema phyticum gen. n., sp. n.(Nematoda: Palaeonematidae fam. n.), a Devonian nematode associated with early land plants. Nematology *10*, 9–14.

Pont, C., Murat, F., Confolent, C., Balzergue, S., and Salse, J. (2011). RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (Triticum aestivum L.). Genome Biol *12*, R119.

Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature *453*, 1064–1071.

Qin, L., Kudla, U., Roze, E.H., Goverse, A., Popeijus, H., Nieuwland, J., Overmars, H., Jones, J.T., Schots, A., Smant, G., et al. (2004). Plant degradation: a nematode expansin acting on plants. Nature *427*, 30.

Raoult, D. (2010). The post-Darwinist rhizome of life. Lancet *375*, 104–105.

Rice, W.R., and Friberg, U. (2009). A Graphical Approach to Lineage Selection Between Clonals and Sexuals. In Lost Sex, I. Schön, K. Martens, and P. Dijk, eds. (Springer Netherlands), pp. 75–97.

Rieseberg, L.H., Archer, M.A., and Wayne, R.K. (1999). Transgressive segregation, adaptation and speciation. Heredity *83*, 363–372.

Rodelsperger, C., and Sommer, R.J. (2011). Computational archaeology of the Pristionchus pacificus genome reveals evidence of horizontal gene transfers from insects. BMC Evol Biol *11*, 239.

Rosso, M.N., Favery, B., Piotte, C., Arthaud, L., De Boer, J.M., Hussey, R.S., Bakker, J., Baum, T.J., and 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. Mol Plant Microbe Interact *12*, 585–591.

Rosso, M.N., Hussey, R.S., Davis, E.L., Smant, G., Baum, T., Abad, P., and Mitchum, M. (2012). Nematode Effector Proteins: Targets and Functions in Plant Parasitism. In Effectors in Plant-Microbe Interactions, F. Martin, and S. Kammoun, eds. (Wiley-Blackwell),.

Roulin, A., Auer, P.L., Libault, M., Schlueter, J., Farmer, A., May, G., Stacey, G., Doerge, R.W., and Jackson, S.A. (2013). The fate of duplicated genes in a polyploid plant genome. Plant J. *73*, 143–153.

Rybarczyk-Mydlowska, K., Maboreke, H.R., van Megen, H., van den Elsen, S., Mooyman, P., Smant, G., Bakker, J., and Helder, J. (2012). Rather than by direct acquisition via lateral gene transfer, GHF5 cellulases were passed on from early Pratylenchidae to root-knot and cyst nematodes. BMC Evol Biol *12*, 221.

Rybarczyk-Mydłowska, K., Mooyman, P., van Megen, H., van den Elsen, S., Vervoort, M., Veenhuizen, P., van Doorn, J., Dees, R., Karssen, G., Bakker, J., et al. (2012). Small subunit ribosomal DNA-based phylogenetic analysis of foliar nematodes (Aphelenchoides spp.) and their quantitative detection in complex DNA backgrounds. Phytopathology *102*, 1153–1160.

Sasser, J.N., and Freckman, D.W. (1987). A world perspective on nematology; the role of society. In Vistas on Nematology, J.A. Veech, and D.W. Dickerson, eds. (Hyattsville, USA: Society of Nematologists), pp. 7–14.

Schaack, S., Gilbert, C., and Feschotte, C. (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol *25*, 537–546.

Schon, I., Gandolfi, A., Di Masso, E., Rossi, V., Griffiths, H.I., Martens, K., and Butlin, R.K. (2000). Persistence of asexuality through mixed reproduction in Eucypris virens (Crustacea, Ostracoda). Hered. Edinb *84 ( Pt 2)*, 161–169.

Schön, I., and Martens, K. (1998). DNA repair in ancient asexuals — a new solution to an old problem? J. Nat. Hist. *32*, 943–948.

Schön, I., and Martens, K. (2003). No slave to sex. Proc Biol Sci *270*, 827–833.

Schwander, T., Henry, L., and Crespi, B.J. (2011). Molecular evidence for ancient asexuality in timema stick insects. Curr Biol *21*, 1129–1134.

Scott, A.L., and Ghedin, E. (2009). The genome of Brugia malayi — All worms are not created equal. Parasitol. Int. *58*, 6–11.

Simon, J.-C., Delmotte, F., Rispe, C., and Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. Biol. J. Linn. Soc. *79*, 151–163.

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., et al. (1998). Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. Proc Natl Acad Sci U A *95*, 4906–4911.

Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess, D.R., Burke, R.D., et al. (2006). The genome of the sea urchin Strongylocentrotus purpuratus. Science *314*, 941–952.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The Genome Sequence of Caenorhabditis briggsae: A Platform for Comparative Genomics. PLoS Biol *1*, e45.

Straub, E. (1952). Mikropaläontologische Untersuchungen im Tertiär zwischen Ehingen und Ulm a. d. 433–523. *66*, 453–523.

The C. elegans Genome Sequencing Consortium (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012–2018.

Triantaphyllou, A.C. (1981). Oogenesis and the chromosomes of the parthenogenetic Root Knot Nematode Meloidogyne incognita. J. Nematol. 95–104.

Trudgill, D.L. (1995). Origins of root-knot nematodes (Meloidogyne spp., nematoda) in relation to their cultural control. Phytoparasitica *23*, 191–194.

Trudgill, D.L., and Blok, V.C. (2001). Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. Annu Rev Phytopathol *39*, 53–77.

Trudgill, D.L., Blok, V.C., Fargette, M., Phillips, M.S., and Bradshaw, J. (1996). The possible origins of genetic variability within the plant parasitic nematodes Meloidogyne and Globodera spp. Agric. Zool. Rev. *7*.

Tucker, A.E., Ackerman, M.S., Eads, B.D., Xu, S., and Lynch, M. (2013). Population-genomic insights into the evolutionary origin and fate of obligately asexual Daphnia pulex. Proc. Natl. Acad. Sci. *110*, 15740–15745.

Vens, C., Rosso, M.N., and Danchin, E.G. (2011). Identifying discriminative classification-based motifs in biological sequences. Bioinformatics *27*, 1231–1238.

Watanabe, H., and Tokuda, G. (2010). Cellulolytic Systems in Insects. Annu. Rev. Entomol. *55*, 609–632.

White, M.J.D. (1973). Animal cytology and evolution (Cambridge, UK: Cambridge University Press).

Winnenburg, R., Urban, M., Beacham, A., Baldwin, T.K., Holland, S., Lindeberg, M., Hansen, H., Rawlings, C., Hammond-Kosack, K.E., and Kohler, J. (2008). PHI-base update: additions to the pathogen host interaction database. Nucleic Acids Res *36*, D572–6.

# 7 Selected Publications


- Danchin EGJ, Arguel M-J, Campan-Fournier A, Perfus-Barbeoch L, Magliano M, Rosso M-N, Da Rocha M, Da Silva C, Nottet N, Labadie K, Guy J, Artiguenave F, Abad P. **Identification of novel target genes for safer and more specific control of root-knot nematodes from a pan-genome mining.** PLoS Pathogens. **2013**

- Paganini J, Campan-Fournier A, Da Rocha M, Gouret P, Pontarotti P, Wajnberg E, Abad P, Danchin EGJ. **Contribution of Lateral Gene Transfers to the Genome Composition and Parasitic Ability of Root-Knot Nematodes**. PLoS One. **2012**

- Danchin E.G.J., Rosso M-N., Vieira P., de Almeida-Engler J., Coutinho P, Henrissat B., Abad P. **Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes.** PNAS. **2010**

- Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EGJ,..., Rosso MN, Schiex T, Smant G, Weissenbach J, Wincker P. **Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita.** Nat Biotechnol. **2008**

OPEN ACCESS Freely available online

PLOS | PATHOGENS

# Identification of Novel Target Genes for Safer and More Specific Control of Root-Knot Nematodes from a Pan-Genome Mining

**Etienne G. J. Danchin[1,2,3]*, Marie-Jeanne Arguel[1,2,3☯], Amandine Campan-Fournier[1,2,3☯], Laetitia Perfus-Barbeoch[1,2,3], Marc Magliano[1,2,3], Marie-Noëlle Rosso[1,2,3], Martine Da Rocha[1,2,3], Corinne Da Silva[4], Nicolas Nottet[1,2,3], Karine Labadie[4], Julie Guy[4], François Artiguenave[4], Pierre Abad[1,2,3]**

**1** INRA, UMR 1355 ISA, Institut Sophia Agrobiotech, Sophia-Antipolis, France, **2** CNRS, UMR 7254 ISA, Institut Sophia Agrobiotech, Sophia-Antipolis, France, **3** Université de Nice Sophia-Antipolis, UMR ISA, Institut Sophia Agrobiotech, Sophia-Antipolis, France, **4** CEA-Institut de Génomique, GENOSCOPE, Centre National de Séquençage, Evry, France

## Abstract

Root-knot nematodes are globally the most aggressive and damaging plant-parasitic nematodes. Chemical nematicides have so far constituted the most efficient control measures against these agricultural pests. Because of their toxicity for the environment and danger for human health, these nematicides have now been banned from use. Consequently, new and more specific control means, safe for the environment and human health, are urgently needed to avoid worldwide proliferation of these devastating plant-parasites. Mining the genomes of root-knot nematodes through an evolutionary and comparative genomics approach, we identified and analyzed 15,952 nematode genes conserved in genomes of plant-damaging species but absent from non target genomes of chordates, plants, annelids, insect pollinators and mollusks. Functional annotation of the corresponding proteins revealed a relative abundance of putative transcription factors in this parasite-specific set compared to whole proteomes of root-knot nematodes. This may point to important and specific regulators of genes involved in parasitism. Because these nematodes are known to secrete effector proteins *in planta*, essential for parasitism, we searched and identified 993 such effector-like proteins absent from non-target species. Aiming at identifying novel targets for the development of future control methods, we biologically tested the effect of inactivation of the corresponding genes through RNA interference. A total of 15 novel effector-like proteins and one putative transcription factor compatible with the design of siRNAs were present as non-redundant genes and had transcriptional support in the model root-knot nematode *Meloidogyne incognita*. Infestation assays with siRNA-treated *M. incognita* on tomato plants showed significant and reproducible reduction of the infestation for 12 of the 16 tested genes compared to control nematodes. These 12 novel genes, showing efficient reduction of parasitism when silenced, constitute promising targets for the development of more specific and safer control means.

**Citation:** Danchin EGJ, Arguel M-J, Campan-Fournier A, Perfus-Barbeoch L, Magliano M, et al. (2013) Identification of Novel Target Genes for Safer and More Specific Control of Root-Knot Nematodes from a Pan-Genome Mining. PLoS Pathog 9(10): e1003745. doi:10.1371/journal.ppat.1003745

**Editor:** Elodie Ghedin, University of Pittsburgh, United States of America

**Received** March 18, 2013; **Accepted** September 19, 2013; **Published** October 31, 2013



**Funding:** This work was financially supported by a grant of the French National Research Agency (ANR) under the Genoplante program (project NEMATARGETS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: etienne.danchin@sophia.inra.fr

☯ These authors contributed equally to this work.

## Introduction

Plant-parasitic nematodes (PPN) cause significant damage to agriculture throughout the world. A global survey in 1987 evaluated crop losses at $78–125 billion per year [1]. More recent direct global estimates are not available, but when the increase in agricultural productivity is taken into account, the extrapolated 2001 loss for crops totaled $118 billion (11% of production) [2]. The current figure is thus probably much higher. Measures such as growing resistant crop varieties and the use of nematicides are extensively employed to control PPN infections. Billions of Euros have been spent annually on soil fumigants and other nematicides. Current and previous chemical controls against nematodes are not only costly but they are highly toxic and hazardous, and involve application of environmentally unacceptable compounds. Such toxicological problems and environmental damage caused by nematicides have led to banning of the most efficient chemicals that were commonly used so far (EC directive 2007/619/EC). In the absence of alternative control methods or development of specific and environmentally safe molecules, severe crop losses within major sectors of the agricultural industry are a distinct possibility. Indeed, nematode problems recently re-emerged in some areas where the use of traditional nematicides had been abandoned for a short while [3,4]. Therefore, novel control measures are urgently needed. The identification of PPN-specific genes expressed during the interaction with the plant host is one of the most promising approaches for identification of new anti-parasitic strategies.

Infective PPN larvae in the soil are nearly microscopic worms, virtually invisible to the naked eye. Although a few nematode

## Author Summary

Plant-parasitic nematodes are annually responsible for more than $100 billion crop yield loss worldwide and those considered as causing most of the damages are root-knot nematodes. These nematodes used to be controlled by chemicals that are now banned from use because of their poor specificity and high toxicity for the environment and human health. In the absence of sustainable alternative solutions, new control means, more specifically targeted against these nematodes and safe for the environment are needed. We searched in root-knot nematode genomes, genes conserved in various plant-damaging species while otherwise absent from the genomes of non target species such as those of chordates, plants, annelids, insect pollinators and mollusks. These genes are probably important for plant parasitism and their absence from non-target species make them interesting candidates for the development of more specific and safer control means. Further bioinformatics pruning of this set of genes yielded 16 novel candidates that could be biologically tested. Using RNA interference, we knocked down each of these 16 genes in a root-knot nematode and tested the effect on plant parasitism efficiency. Out of the 16 tested genes, 12 showed a significant and reproducible diminution of infestation when silenced and are thus particularly promising.


species feed on above ground plant parts, such as leaves, stems, flowers, and seeds, the majority of these parasites feed on underground parts of plants, including roots, bulbs, and tubers. Most PPN feed on root tissue and damage their host mainly by stunting the root system, resulting in reduced water uptake and by promoting microbial infections through wound sites or by serving as vectors for pathogenic viruses. Some nematode species exhibit a hit-and-run strategy, remaining migratory during their plant root-associated life cycle. An increase in complexity of host-parasite interactions is observed in sedentary parasite species with their enhanced capacity to manipulate host plant genes in their favor [5]. These endoparasitic nematodes settle down after an initial migratory phase and assume a sedentary life style while transforming plant cells into complex feeding structures. Nematodes of this category represent the most damaging species for crops. Some of these nematodes have a relatively specialized host range (e.g. cyst nematodes *Heterodera* and *Globodera* genus) while others are able to reproduce on thousands of unrelated host plant species (e.g. root-knot nematodes, *Meloidogyne* spp.).

Because root-knot nematodes represent the most economically-important PPN, they constitute the most explored group of species and can now be considered as one of the most advanced models for understanding mechanisms of plant parasitism in nematodes. As with other PPN, they have a syringe-like stylet that is used to pierce and penetrate plant cell walls, to release esophageal secretions into the host tissue and to take up nutrients. During their infective life-cycle root-knot nematode larvae penetrate plant root tissue and migrate along the vascular cylinder. By injecting secretions into plant cells, they induce the formation of a feeding site indispensable for their development. As a consequence of the formation of these feeding structures, root-knots or galls are observed as symptoms of the infestation. Plant nutrient and water uptake are substantially reduced by the resulting damage to the root system, and infested plants are therefore weak and give low yields. Once the feeding structure is established, female nematodes continue their development and eventually become pear-shaped and produce hundreds to thousands of eggs. These eggs are then extruded as an egg-mass, protected within a gelatinous matrix, at the outer surface of the root.

Mining the genomes of root-knot nematodes [6,7,8] through an evolutionary and comparative genomics approach, we searched genes conserved in various plant-damaging species while otherwise absent from the genomes of non target species such as those of chordates, plants, annelids, insect pollinators and mollusks. We identified a set of root-knot nematode genes absent from non-target species but present in several plant-damaging organisms. Further bioinformatics pruning of this set of genes yielded new candidates that were silenced using RNA interference (RNAi). Upon silencing experiments, 75% of the candidates induced a significant and reproducible diminution of infestation and are thus particularly promising for the development of new and more specific control strategies.

## Results

### Elimination of root-knot nematode genes shared by non-parasitic species

Our main objective was to identify root-knot nematode (RKN) genes that could be used as targets for the development of new control means against these pests. As we absolutely wanted to minimize the risk of collateral effects and preserve non-targeted species, we systematically discarded RKN genes that had putative homologs in non-target species (Figure 1 and methods).

To select RKN proteins without predicted homologs in non-target species, we first performed an OrthoMCL [9] analysis comparing all predicted proteins in *M. incognita* and *M. hapla* (34,780 proteins) with the whole proteomes of 23 other species (Figure 2). This step was aimed at eliminating RKN proteins having evident orthologs in fully-sequenced non-target genomes and to substantially reduce the number of proteins that will be subsequently compared against the NCBI's nr library. We selected, in priority, species whose whole genomes have been annotated to a quality level allowing a reliable prediction of the ensemble of protein-coding genes. Our selection of species comprised 4 other nematodes, 5 insects, 9 vertebrates (including mammals, ray-finned fishes, amphibian and sauropsida), 4 fungi and 1 plant. Among selected species, we included two plant-pathogenic fungi (*Magnaporthe grisea* and *Fusarium graminearum*), one nematode parasite of animals (*Brugia malayi*) and two insects that feed on living plant tissue (*Acyrthosiphon pisum* and *Bombyx mori*). The 18 other species were blacklisted and whenever a RKN protein had a predicted ortholog in these blacklisted species, the protein was discarded from the rest of the analysis. According to OrthoMCL, a total of 15,181 RKN proteins had a predicted ortholog in at least one blacklisted species and were thus eliminated. The rest of RKN proteins (19,599) had no predicted ortholog in any of the blacklisted species and passed this first filter. Among these proteins, a total of 2,446 were redundant between *M. incognita* and *M. hapla*. To avoid redundancy, and because subsequent biological assays will be performed in *M. incognita*, we kept as representative the *M. incognita* versions. At the end of this first filtering step, a total of 17,153 *Meloidogyne* proteins were kept for further analysis.

Although, with a total of 25 species representing >500,000 proteins, the OrthoMCL analysis we performed is far from negligible, this only represents a limited sample of the whole sequence biodiversity available in public databases. Thus, using a BLASTp [10] analysis, we compared the 17,153 RKN proteins that passed the OrthoMCL filter against the NCBI's nr library. Applying a similar filter as was applied to the OrthoMCL results, we systematically eliminated RKN proteins having putative

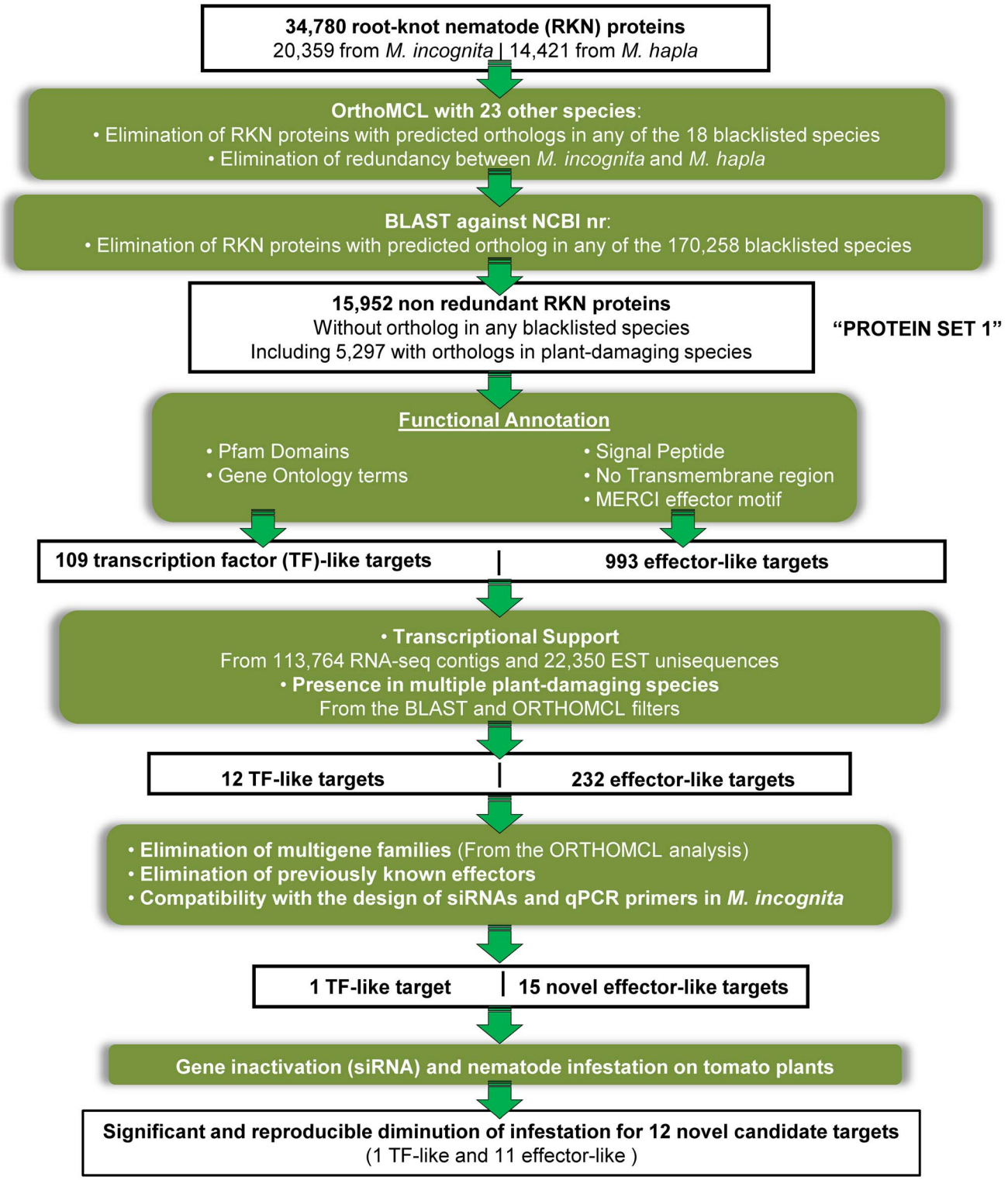



**Figure 1. Whole RKN proteomes filtering pipeline for identification of novel targets.** Pipeline illustrating the main filtering steps, from the two RKN whole protein sets, that allowed identification of novel and non-redundant targets for the development of specific and safer control methods.
doi:10.1371/journal.ppat.1003745.g001

orthologs in non-target, blacklisted species. Those that had no putative ortholog in any of the blacklisted species or returned no significant similarity at all in any other species, were kept for subsequent analysis. Because there is no comprehensive database indicating the lifestyles of the plethora of species with a sequence in the nr library, we generated a list of blacklisted taxa (methods). In

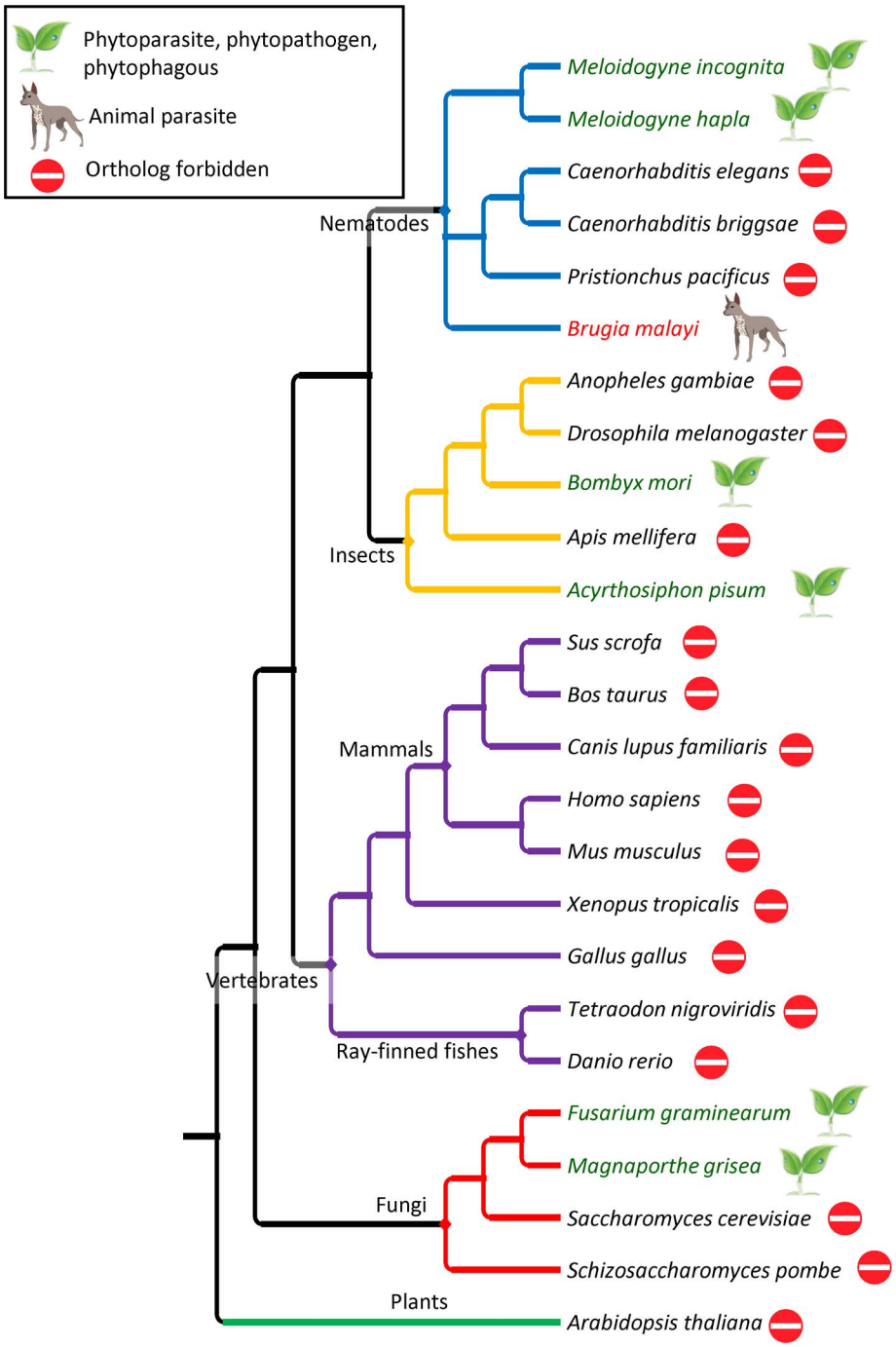



**Figure 2. Phylogenetic tree of selected species for OrthoMCL comparison.** The relative phylogenetic position and simplified taxonomy of the 25 species included in the OrthoMCL comparison of whole proteomes. The topology is according to the NCBI's taxonomy, except within the nematode and insecta lineages which are according to ref [45] and to ref [46], respectively. Species that are known plant-parasites plant-pathogens or phytophagous are highlighted in green and with a plant symbol. The animal-parasitic nematode *B. malayi* is highlighted in red and with a dog symbol. All the other species, in black, with a "wrong way" sign, are blacklisted.
doi:10.1371/journal.ppat.1003745.g002

total, our blacklist included 170,258 species covering 4 whole clades (annelida, chordata, mollusca and viridiplantae) in addition to the 18 species already blacklisted in the OrthoMCL analysis. Overall, a total of 10,105 RKN proteins did not return any significant BLASTp hit in nr using the thresholds we had set (methods). More than half of these proteins (5,536) also had no predicted ortholog in the OrthoMCL analysis and were thus considered as potentially orphan or restricted to RKN at this stage. In contrast, 1,201 RKN proteins returned significant BLASTp hits in at least one blacklisted species and were discarded. In total, 15,952 RKN sequences were kept and constituted our protein set 1. This set 1 represents RKN proteins predicted to be absent from blacklisted species and possibly present in other plant-damaging species.

## Conservation of root-knot nematode proteins in plant-damaging species

We assessed whether part of the RKN proteins absent from non-target species were present in other plant-damaging species. The rationale of this analysis is that the more a gene is shared between plant pests while absent from other species, the more it is likely to be involved in core interaction processes with the plant. To assess conservation in plant-damaging species, we filtered the results of both the OrthoMCL and BLASTp analyses. In the OrthoMCL analysis, two plant-pathogenic fungi were included as well as two insects that feed on plant. A total of 4,398 RKN proteins had predicted orthologs in, and only in, these plant-damaging species. Similarly to the list of blacklisted species for the BLASTp filtering, we built up a list of 28,054 potentially plant-damaging species in the NCBI's taxonomy (methods). We identified 1,252 RKN proteins that returned significant BLASTp hits with at least one plant-damaging species. After removing redundancy between the OrthoMCL and BLASTp analyses, we obtained a non-redundant list of 5,297 RKN proteins absent from non-target species but present in at least two plant-damaging species.

## Functional annotation of root-knot nematode proteins and further filtering

To gain functional insight on the proteins that appeared restricted to RKN and other plant-damaging species, we searched and retrieved a series of functional annotations. This included a search for signal peptides for secretion, a search for transmembrane regions, a search for known protein domains and associated functional annotations. We also assessed whether corresponding genes had transcriptional support.

**Predicted functions of RKN proteins.** Out of the 15,952 RKN proteins in set 1 (i.e. that passed both the OrthoMCL and BLASTp filters), only 3,835 or 24% have been assigned a Pfam domain. This is in contrast with the two whole RKN proteomes. Indeed, a total of 10,379 *M. incognita* proteins out of 20,359 [6] and 7,151 *M. hapla* proteins out of 14,421 [7] have been assigned at least one Pfam domain, representing 51% and 49.6% of their respective proteomes (methods).

Echoing the scarcity of Pfam domains assigned to proteins in set 1, only 2,255 proteins (or 13.8%) out of the 15,952 present in set 1 have been assigned a Gene Ontology (GO) term. By comparison, GO terms were assigned to 6,881 (33.8%) and 4,673 (32.4%) of *M. incognita* and *M. hapla* whole proteomes, respectively (methods).

**Transcription-related proteins were more abundant in RKN-restricted proteins.** Using a without *a priori* approach, we searched functional categories present in protein set 1. We compared the relative abundance of the different GO terms between protein set 1 and the two RKN whole proteomes. We remarked that despite substantially different numbers of predicted proteins in *M. incognita* and *M. hapla*, the relative abundance of GO terms was very similar in the two whole proteomes (Table S1, Figure 3). The higher number of protein models in *M. incognita* is due to its bigger genome with a peculiar structure, mainly constituted of regions in two copies with substantial divergence [7].

In contrast, we noticed that, compared to the two RKN whole proteomes, several GO terms were substantially over-represented or under-represented in set 1, for the three different GO ontologies, "biological process", "molecular function" and "cellular component" (Figure 3). For instance, in the "biological process" ontology, we remarked an over-representation of the term "transcription" in set 1 (~15.6%) compared to RKN proteomes (8–9%, p-values $6.12E^{-17}$ - $1.67E^{-13}$). The term "regulation of biological processes" was also more frequent in set 1 (~11.7%) compared to the RKN proteomes (8–9%, p-values $5.84E^{-6}$ - $8.59E^{-4}$). Conversely, some terms were less frequent in set 1 as compared to the whole RKN protein sets. For example, the term "translation" represented only ~1.3% of GO terms in set 1, while it represented ~3.7–3.8% in whole RKN proteomes (p-values $3.21E^{-8}$ - $3.65E^{-8}$).

In the "molecular function" ontology, mirroring observations on the "biological process" ontology; we remarked an over-representation of the term "Transcription factor activity" in set 1 (~6.5%) compared to the whole RKN proteomes (3.3–3.7%, p-values $8.81E^{-11}$ - $2.53E^{-9}$). Besides transcription-related terms, we also noticed an over-representation of the terms "receptor activity" (~4.3% vs. 2.0–2.1%, p-values $6.17E^{-11}$ - $2.89E^{-7}$) and "peptidase activity" (~5.7% vs. 3.8–4.2%, p-values $6.67E^{-5}$ - $6.69E^{-4}$).

In the "cellular component" ontology, we noted that the "nucleus" term was over-represented (23.3%) in set 1 compared to whole RKN proteomes (13.8–14.6%, p-values $3.65E^{-7}$ - $1.38E^{-6}$). Interestingly, more than half (109) of the 190 proteins annotated as localized in the nucleus in set 1 are also annotated as transcription factors in the "molecular function" ontology. Hence the over-representation of the nucleus localization in set 1 is essentially due to the relative abundance of putative transcription factors.

Because putative transcription factors specific from RKN and other plant-damaging organisms constitute interesting potential targets, we searched, in set 1, proteins that were annotated with the term "transcription" in the "biological process" ontology, with the term "transcription factor activity" in the "molecular function" ontology and with the term "nucleus" in the "cellular component" ontology. We found a total of 109 RKN proteins that cumulated these three annotations (Figure 4).

## Effector-like proteins specific to phytoparasites

Root-knot nematodes and other plant parasites secrete, into plant tissue, proteins that support successful parasitism. In nematodes, these proteins, called effectors are generally produced in esophageal gland cells and secreted *via* a syringe-like stylet in plant tissue. Several RKN effectors have been characterized so far and shown to support parasitism by playing roles in different key processes such as degradation of the plant cell wall, suppression of plant defense, manipulation of plant cells to produce feeding structures or interaction with plant signaling pathways [11,12,13]. Because these genes are directly involved in successful parasitism, they naturally constitute interesting targets to develop new control measures. Provided that these proteins are specific to parasitic species they can lead to the development of more targeted and specific control measures. In an *a priori*-based approach, we searched within protein set 1, those presenting the same characteristics than known effectors. Typically, effector proteins bear a signal peptide for secretion and no transmembrane region. We also had identified previously, using the MERCI software [14], a set of protein motifs that are frequent in known effector proteins but absent from housekeeping proteins in RKN. Among protein set 1, we found a total of 3,311 proteins that possessed a signal peptide, 2,453 that possessed an effector MERCI motif and 13,856 that had no predicted transmembrane region. Overall, out of the 15,952 proteins present in set 1, we found 993 proteins that cumulated all these 3 criteria and thus have the same characteristics than canonical effectors.

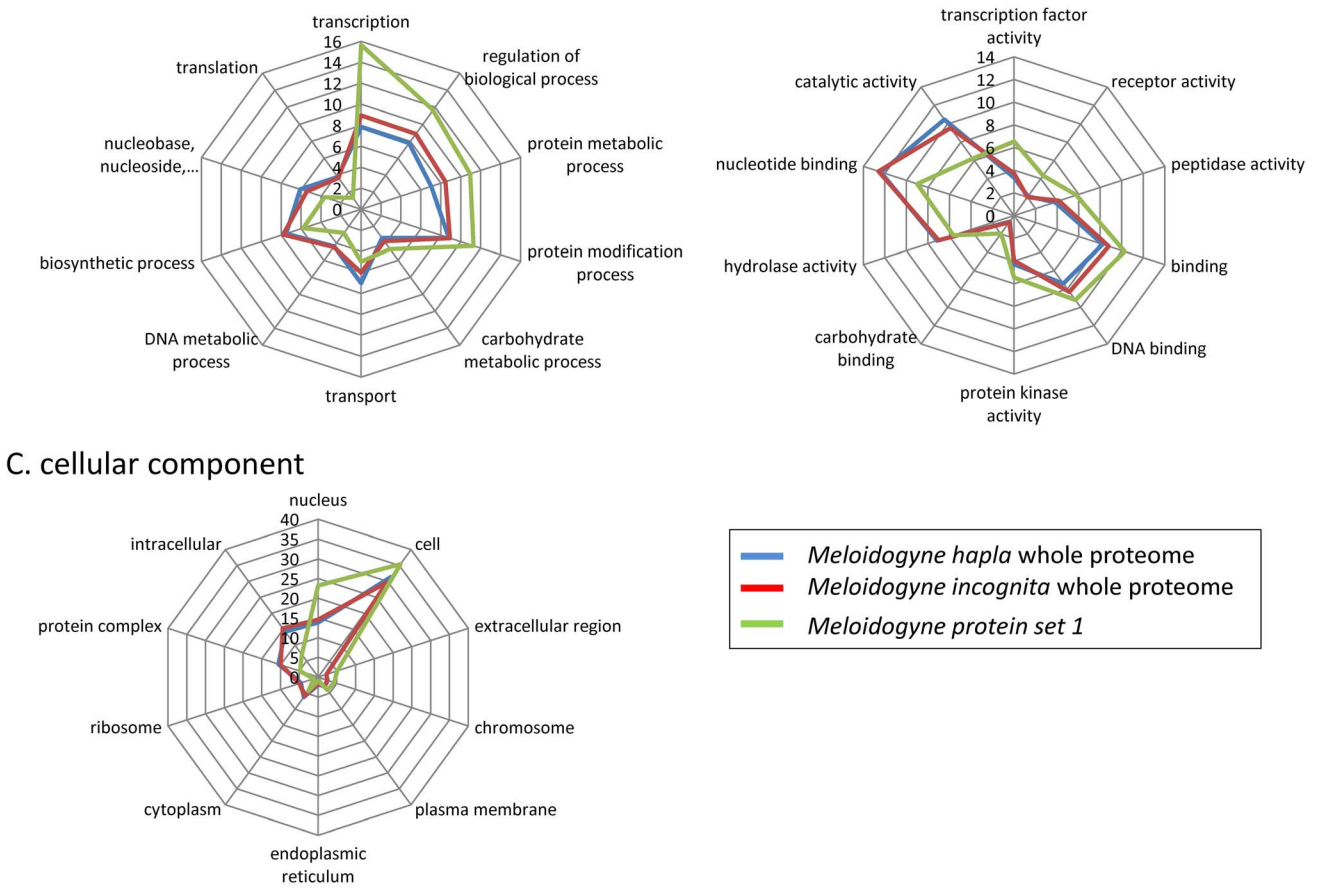



**Figure 3. Gene Ontology terms relative abundance in candidate targets compared to whole proteomes.** Kiviat diagram representing, the relative abundance of Gene Ontology (GO) terms, in percent for the whole *M. incognita* and *M. hapla* proteomes (in blue and red, respectively) as well as for the proteins that passed the OrthoMCL and BLAST filters (protein set 1 in green). (A) Relative abundance of GO terms assigned to whole RKN proteomes and protein set 1in the Biological Process category. (B) Relative abundance of GO terms assigned to whole RKN proteomes and protein set 1 in the Molecular Function category. (C) Relative abundance of GO terms assigned to whole RKN proteomes and protein set 1 in the Cellular Component category. In the three categories, the ten GO terms that presented the most different relative abundance (in percent) in protein set 1 compared to the whole RKN proteomes are presented.
doi:10.1371/journal.ppat.1003745.g003

## Transcriptional support of putative novel targets and further pruning

Because the *M. incognita* and *M. hapla* proteins have been deduced from the gene models predicted as part of automated genome annotations [6,8], set 1 may contain a proportion of proteins deduced from wrongly or over-predicted genes. To minimize the risk of functionally analyzing proteins representing false predictions, we required two additional criteria. (i) The protein must be present in at least two different plant-damaging organisms (including the two RKN species) and, (ii) the corresponding gene must be supported by transcriptomic data from RKN. We had previously assembled the ensemble of available *M. incognita* EST data, as described in [15]. This represented a total of 63,816 ESTs assembled in 22,350 distinct unisequences. Although substantial, this dataset can still be viewed as relatively limited. To complete this relatively scarce transcriptomic dataset, we generated RNA-seq transcriptome sequencing for six different developmental life stages of *M. incognita* (Table 1 and methods). RNA-seq generated more than 190 million reads in total that were assembled in 137,733 contigs (methods). Combined with available ESTs, this dataset is likely to encompass a significant proportion of the diversity of transcripts in a RKN. Out of the 15,952 proteins in set 1, a total of 5,530 had a corresponding CDS sequence that received significant transcriptional support from RKN ESTs or RNA-seq data (methods).

From the set of 109 putative transcription factors identified during the functional annotation, a total of 12 were supported by expression data and were present in at least two plant-damaging species (Figure 1).

From the set of 993 effector-like proteins, 232 were present in at least two plant-damaging species and were transcriptionally supported by alignments with Meloidogyne ESTs or RNA-seq data (Figure 1, Table S2). Among these 232 effector-like proteins, we found 42 previously reported RKN effectors, including SXP/RAL-2 like proteins [16], Venom Allergen-like Proteins (VAP) [17], Chorismate mutases [18], Cathepsin L-like protease 1 (MiCpl1) [19] as well as 32 plant cell wall-degrading enzymes, encompassing cellulases, xylanases, pectate lyases and expansin-

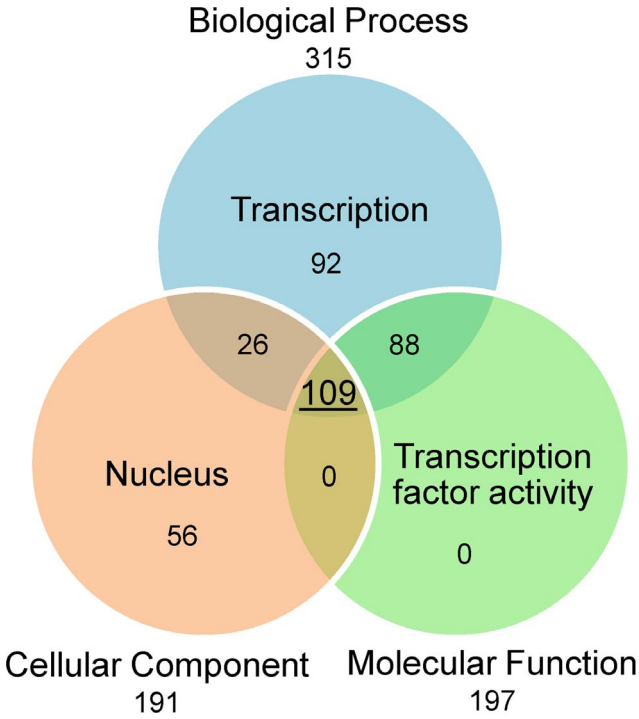



**Figure 4. Gene Ontology annotation of putative transcription factors.** This Venn diagram represents the number of proteins that cumulate transcription-related terms in their Gene Ontology annotation in protein set 1. A total of 315 proteins have been assigned the term 'Transcription' in their Biological Process G.O. annotation. A total of 191 proteins have been assigned the term 'Nucleus' in their Cellular Component G.O. annotation. A total of 197 proteins have been assigned the term 'Transcription factor activity'' in their Molecular Function G.O. annotation. Overall, a total of 109 proteins in set 1 cumulate these three transcription factor-related G.O. terms.
doi:10.1371/journal.ppat.1003745.g004

like proteins [20]. Finding previously known and characterized Meloidogyne effectors among our list of predicted effectors constituted an important validation of our approach. Because the main aim of our genome mining approach was to find novel potential targets we were exclusively interested in the 190 remaining effector-like proteins. Out of these 190 novel effector-like proteins, only 25 different Pfam domains were found in 46 proteins. Because they all received transcriptional support from *M. incognita* and have a homolog in at least one additional plant-damaging species, we can rule out the hypothesis that they are the product of over-prediction due to gene calling software.

## Experimental validation of targets by *M. incognita* infestation on tomato plants after gene silencing

Having identified novel putative transcription factors and effector-like proteins, present in plant-damaging species but absent from blacklisted ones, we wanted to experimentally validate their potential as amenable targets for the development of new control methods. Basically, we targeted selected genes one by one using small interfering RNAs (siRNA) on *M. incognita* infective J2 larvae, and infected host tomato plants with treated larvae. Six weeks after inoculation, we compared the numbers of galls and egg masses in siRNA-treated and control nematodes, as described in the methods.

Starting from the 12 putative transcription factors and 190 novel effector-like RKN proteins, we further pruned the list according to the following criteria. Because we perform biological assays on *M. incognita*, we first discarded proteins from *M. hapla* that had no ortholog in *M. incognita*. To avoid potential compensation of the silencing effect by gene copies performing similar function, we also removed all proteins that were encoded by multigene families. We ended up with a list comprising one putative transcription factor and 39 non-redundant effector-like proteins found in *M. incognita*, present in at least one other plant-damaging species, transcriptionally supported and without a homolog in a blacklisted species (Figure 1). We examined the corresponding coding sequences for compatibility with the design of specifically-matching siRNAs and the design of quantitative PCR primers (methods). We were able to design specific siRNA as well as specific PCR primers for the putative transcription factor (Minc07817) as well as for 15 out of the 39 genes encoding effector-like proteins. These 16 protein-coding genes were all present both in the *M. incognita* and *M. hapla* genomes. A total of 13 of the corresponding proteins do not have any predicted Pfam-A domain and, hence, no indication of the potential molecular function they may be involved in is available. One of the proteins (Minc03866) had a predicted C-type lectin domain and another (Minc03313) had an Astacin (peptidase family M12A) domain.

**Table 1.** Samples used for RNA-seq and resulting contigs.

| Life stage | Reads | Reads used for assembly* | Contigs |
|---|---|---|---|
| Eggs | 25,958,384 | 12,467,657 | 26,570 |
| Early sedentary | 29,278,684 | 13,295,793 | 18,485 |
| Parasitic sedentary | 33,750,791 | 15,558,421 | 18,045 |
| Stage 3 and 4 larvae | 25,126,758 | 12,859,902 | 8,272 |
| Adult female | 26,177,405 | 14,953,875 | 13,071 |
| Adult male | 25,464,740 | 16,995,332 | 29,321 |
| Mixed stages | 24,901,111 | 12,082,564 | 23,969 |

*after trimming, collapsing and cleaning as described in methods.
doi:10.1371/journal.ppat.1003745.t001

**Effect of siRNA on the 16 novel target genes.** To test whether the designed siRNA interfere with the expression of each of the 16 novel target genes, we performed Real-Time quantitative PCR (qPCR) experiments on soaked J2s (methods).

Six siRNAs induced significant reduction in the corresponding targeted transcripts (Minc00801, Minc01632, Minc02483, Minc08335, Minc09526, Minc17987) 24 h after soaking treatment, compared to their expression level in control samples (Table 2, Figure 5). On the other hand, Minc08013, Minc08014 and Minc12224 siRNAs induced a diminution that was not reproducible between independent qPCR replicate experiments (data not shown). Surprisingly, 7 siRNAs (Minc03313, Minc03866, Minc05001, Minc10706, Minc14652, Minc17713, Minc07817) induced significant and reproducible increase in transcript abundance when qPCR was performed 24 h after soaking. A similar 'bounce' effect, that could be due to a response of the nematode RNAi machinery to increased siRNA quantities within the cells, had already been reported in PPN after gene silencing [21,22,23]. To test this 'bounce' effect, we performed qPCR analysis 16 h after soaking on Minc07817, Minc03866 and Minc05001. Confirming siRNA knockdown and bounce effect, Minc03866 was significantly silenced 16 h after soaking. However, Minc05001 expression was reduced but not significantly and Minc07817 expression was still higher than control (Figure S1). Overall, our qPCR results indicate that the designed siRNAs, were

**Table 2.** Results of infestation, qPCR, and *in-situ* hybridization.

| Gene | Trans. level variation | Reduct. of infestation* | Gall # reduction* | Egg mass # reduction* | ISH localization |
|---|---|---|---|---|---|
| Minc00801 | Down @24 h | Yes | Not repro | Yes | Intestinal |
| Minc01632 | Down @24 h | Yes | Yes | Yes | Ubiquitous |
| Minc02483 | Down @24 h | Yes | Yes | Yes | Nerve ring |
| Minc03313 | Up @24 h | Yes | Not repro | Yes | No signal |
| Minc03866 | Up @24 h, Down @16 h | Yes | Yes | Yes | Subv. gland cells |
| Minc05001 | Up @24 h, Not repro @16 h | Yes | Yes | Yes | No signal |
| Minc07817 | Up @24 h, Up @16 h | Yes | Not repro | Yes | No signal |
| Minc08013 | Not repro | Yes | Yes | Not repro | No signal |
| Minc08014 | Not repro | Not repro | Not repro | Not repro. | Unchecked |
| Minc08335 | Down @24 h | Yes | Not repro | Yes | Ubiquitous |
| Minc09526 | Down @24 h | Yes | Yes | Yes | No signal |
| Minc10706 | Up @24 h | Not repro. | Not repro | Not repro. | Unchecked |
| Minc12224 | Not repro | Yes | Not sign. | Yes | Ubiquitous |
| Minc14652 | Up @24 h | Not repro | Not repro | Not repro. | Unchecked |
| Minc17713 | Up @24 h | Yes | Not repro | Yes | Intestin |
| Minc17987 | Down @24 h | Not repro | Not repro | Not sign. | Unchecked |

*statistically significant and reproducible reduction.
doi:10.1371/journal.ppat.1003745.t002

efficient to knockdown 7 out of the 16 target genes at the tested time points.

**Effect of siRNA soaking on nematode infestation.** For each of the 16 newly identified targets, we tested whether nematodes soaked with matching siRNAs showed a significant reduction in the numbers of galls and/or egg masses on infected plants as compared to control nematodes (methods).

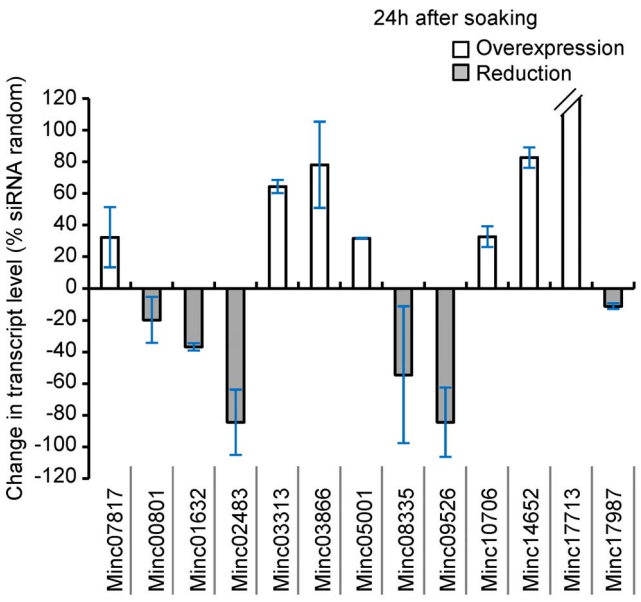



**Figure 5. Transcript abundance percentage change in siRNA soaked J2s relative to control.** siRNAs induced significant change in the targeted transcripts expression level. Transcript level for each of the targeted gene was measured by qPCR 24 h after soaking treatment and compared to their transcript level in control sample treated with siRNA targeting no sequence in the *M. incognita* genome (siRNA random).
doi:10.1371/journal.ppat.1003745.g005

Overall, 12 out of the 16 siRNA-treated samples (Minc00801, Minc01632, Minc02483, Minc03313, Minc03866, Minc05001, Minc08013, Minc08335, Minc09526, Minc12224, Minc17713 and Minc07817) showed a significant and reproducible reduction in the number of galls or egg masses after nematode infestation compared to control nematodes (Table 2, Figure 6, Table S3).

Genes that presented the highest reproducible reduction in number of galls were Minc01632 (43.67% to 57.48%), Minc03866 (38.36% to 41.20%), Minc05001 (33.03% to 51.54%) and Minc08013 (33.03% to 51.54%). Genes presenting the most important reproducible effects on the reduction of egg mass numbers were Minc01632 (54.17% to 70.09%), Minc02483 (39.54% to 60.85%, and Minc09526 (37.08% to 39.74%). A total of 5 targeted genes (Minc01632, Minc02483, Minc03866, Minc05001, Minc09526) showed significant and reproducible reductions of both the number of galls and egg masses (Table 2).

To check whether the observed effect on plant infestation was due to a toxicity of the siRNA treatment itself, we measured the viability and motility of the nematodes 1 hour and 16 hours after siRNA treatment. Comparison with control nematodes revealed that there was no significant and replicable toxicity effect (Figure S2).

***In situ* hybridizations of the 16 novel candidate targets.** Among the 16 siRNA experiments (15 on effector-like genes and one on the putative transcription factor), 12 yielded significant and reproducible reduction of the number of galls or egg masses. Because all these genes are novel candidates and nothing is known about their possible functions, we investigated whether information could be gained from their expression localization using *in-situ* hybridization (methods). Results of *in-situ* hybridizations could be grouped in 5 different categories (Figure S3, Table 2). (i) Genes with ubiquitous expression detected (Minc01632, Minc08335, Minc12224). (ii) Genes with expression detected specifically in secretory subventral gland cells

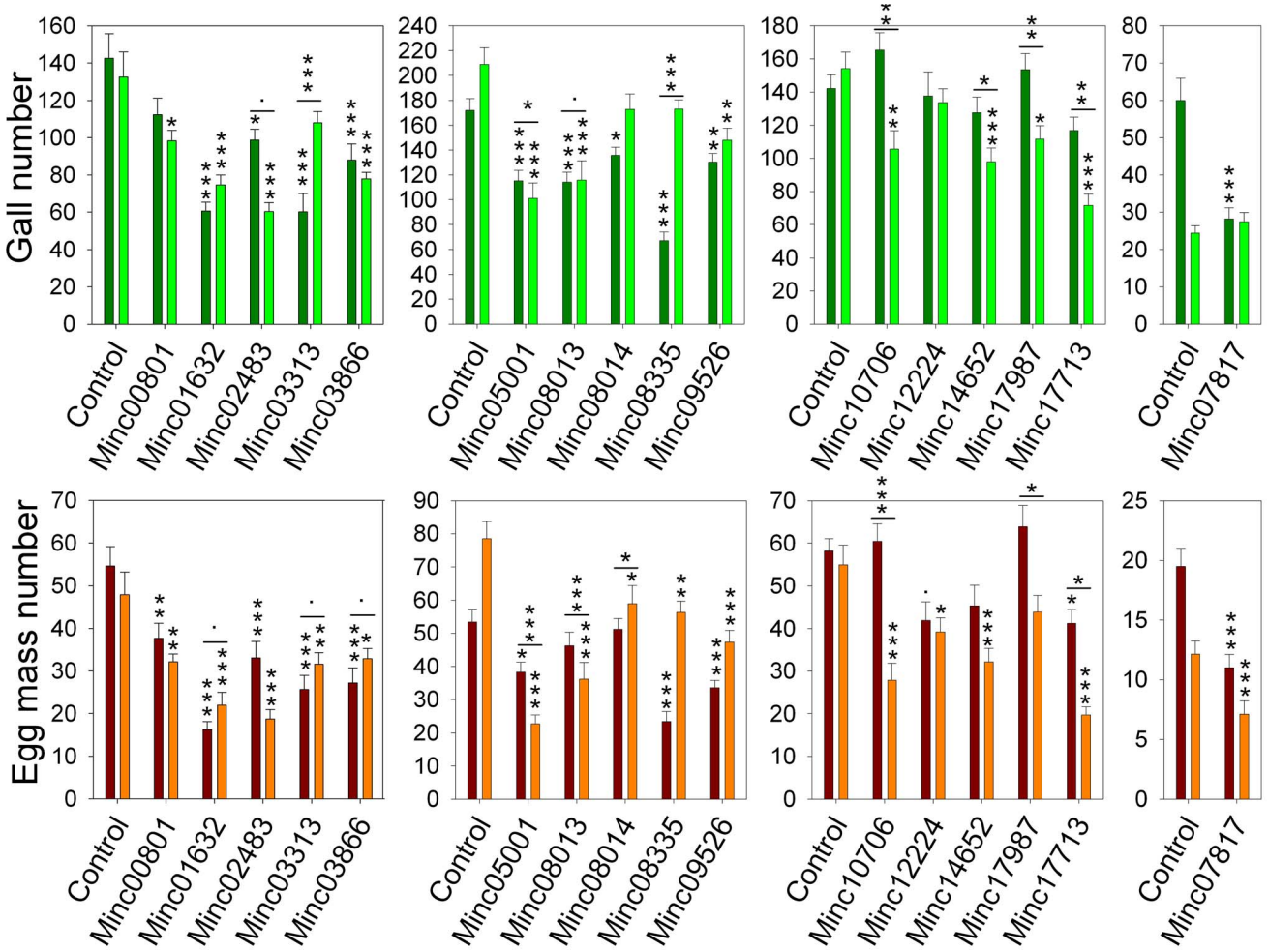



**Figure 6. Effect of siRNA on nematode infestation.** Controls are J2 larvae soaked with siRNA targeting no sequence in the *M. incognita* genome, accession numbers indicate M. incognita genes targeted by siRNAs. Infection tests were performed in duplicate represented as dark color bars for the first replicate and light color bars for the second replicate. Variations of number of galls and of numbers of egg masses, measured six weeks after inoculation, are represented in the top panels in green and in the bottom panel in red/orange, respectively. Error bars were calculated by standard error of the mean (SEM). P-Value above each bar indicates measures statistically different from controls. A significant change between the two duplicates is represented by p-values above a line spanning duplicates. P-value signification codes are as follows: 0.0001 '***', 0.001 '**', 0.01 '*', 0.05 '.'.

doi:10.1371/journal.ppat.1003745.g006

(Minc03866). (iii) Genes with expression detected specifically in the intestinal tract (Minc00801, Minc17713). (iv) Genes with expression detected specifically in the circumpharyngeal nerve ring (CNR), a structure surrounding the metacorpus, a pump-like organ (Minc02483). (v) Genes that returned no detectable signal (Minc03313, Minc05001, Minc08013, Minc09526, Minc07817).

**Putative expression pattern of the 16 novel candidate targets.** Because we have generated RNA-seq data for 6 developmental life stages in *M. incognita*, information on the genes expression pattern can be obtained in addition to raw transcriptional support. Aligning these RNA-seq reads separately for each life stage on the *M. incognita* genome (methods) allowed estimating the relative expression level of the 16 novel candidate targets in the 6 developmental life stages (Table S4). Interestingly, although there was no evident common expression pattern, all 16 genes were expressed at a substantially high level in at least one parasitic stage (early sedentary, parasitic sedentary, J3 and J4 larvae and adult female). This is consistent with the measured effect on infestation following soaking of siRNA at the infective J2 larvae stage.

## Discussion

### Orphan genes and genes shared with other plant-damaging species

Out of the 34,780 predicted proteins from the *M. incognita* and *M. hapla* whole proteomes, we have eliminated a total of 15,181 proteins because they had predicted orthologs in at least one of the 18 blacklisted species, based on OrthoMCL. In comparison, our taxonomic BLASTp analysis against the NCBI's nr library allowed elimination of only 1,201 further RKN sequences. This result suggests that our OrthoMCL filtering was able to eliminate most of the RKN proteins having potential orthologs in non-target species. Despite our selection of 23 species compared to the RKN is far from representing a significant portion of the whole biodiversity available, it constituted a stringent filter, probably because representatives from various different lineages, ranging from fungi to vertebrates, were included. This OrthoMCL filter also allowed us to dramatically reduce the number of proteins to be compared with the nr library in subsequent BLASTp

comparison. The 1,201 sequences eliminated at the taxonomic BLASTp step probably consisted of gene families not represented among the 23 compared species. Besides allowing elimination of proteins having orthologs in non-target species, the OrthoMCL and BLASTp filters also allowed identification of RKN genes shared by several plant-damaging species. A total of 5,297 non redundant RKN proteins were present in at least two plant-damaging species but absent in non-target species, according to OrthoMCL and BLASTp filters. These proteins, apparently restricted to plant-damaging species, may be involved in core mechanisms common to several of these agricultural pests. Another point of interest revealed by the OrthoMCL and BLAST analyses is the set of potential orphan genes in RKN. A total of 5,536 non-redundant RKN proteins neither returned predicted orthologs in the OrthoMCL analysis nor had any significant BLASTp hits, in other species. These apparently RKN-restricted proteins can represent true orphans but may also be the result of possible artifacts due to over-predictions made by gene calling software in RKN genomes. However, 949 of the corresponding genes received transcriptional support from EST or RNA-seq data and are thus unlikely to be the results of over-predictions. Similarly, 2,416 of these orphan genes are present both in the *M. incognita* and *M. hapla* genomes and it appears improbable that these genes have been over-predicted twice independently in two distinct genomes using distinct gene calling strategies. These genes, apparently restricted to RKN and otherwise orphan, may be involved in processes specific to RKN such as the fine interactions between the nematode and the plant host (e.g. induction of a feeding site in the plant) or in the ontogeny of specialized organs (e.g. gland cells or protrusible stylet).

Not only are those genes candidate targets for new treatments against RKN, but also fundamental genes to better understand adaptation to a plant-parasitic life.

Whether these genes are true orphans can be questioned when considering the relative scarcity of omics data available for plant-parasitic nematodes in general. Our OrthoMCL analysis included only two proteomes of plant-parasitic nematode species (*M. incognita* and *M. hapla*) and to date, no whole proteome for a phytoparasitic nematode species is present in the NCBI's nr database. Hence, these genes may have orthologs in other plant-parasitic nematode species. Availability of further whole genomes, transcriptomes and deduced proteomes from additional phytoparasitic nematodes in the future will allow us to decipher whether some of these genes are shared with other plant-parasitic species and may, consequently, be involved in core processes linked to this lifestyle.

## Nature of filters and novel candidate target genes

The series of filters we have set up in our bioinformatics pipeline resulted in a very stringent screening of the two whole RKN proteomes. We have first eliminated all proteins that had potential orthologs in a series of blacklisted species that must be preserved if new nematode control means, targeting these genes, were developed. We next ran two strategies in parallel to identify novel candidates in RKN proteomes that would be more clearly amenable for development of new control methods.

The first strategy was an *ab initio* data-driven one. Because we noticed an over-abundance of putative transcription factors in the set of RKN proteins absent from blacklisted species, we focused on this category. We identified 12 putative transcription factors absent from blacklisted species and supported by transcriptional evidence. If these proteins actually function as transcription factors, they may be involved in regulation of genes involved in RKN-specific functions such as parasitism genes or modulate the expression of host plant genes. One of those putative transcription factors was present as a single copy gene in *M. incognita* and was compatible with the design of specific siRNA and qPCR primers and thus amenable for biological assays.

The second strategy we used was an *a priori* based one. Because effector proteins secreted by nematodes are known to be important in their plant-parasitic ability, we searched proteins that featured the same characteristics and identified a list of 232 putative effectors. Validating our *a priori* strategy, we retrieved 42 proteins that were previously described as known effectors in the literature. Obviously, not all effectors previously described so far were found. This is mainly for the following reasons: (i) several known effectors do not possess an N-terminal signal peptide and/or a MERCI effector-motif (ii) some nematode effectors have homologs in blacklisted species. Because we were mainly interested in the discovery of novel potential targets, we focused our analysis on the 190 remaining novel effector-like proteins not present in blacklisted species. A total of 39 corresponding genes were not redundant in *M. incognita* and present in at least one other plant-damaging species. Out of these 39 genes, 15 were compatible with the design of siRNAs and qPCR primers and thus amenable for further biological assays.

During infestation tests on tomato plants, out of the 16 novel candidates identified (15 effector like and 1 putative transcription factor), 12 turned out to show significant and reproducible reduction in the number of egg masses or galls when treated with anti-candidate siRNAs.

Overall, our strategy was not to build a comprehensive list of candidate genes that might produce the most severe phenotypes on nematodes. In contrast, the originality of our approach was to focus from the beginning on genes that were present in plant-damaging species but absent from non-target "blacklisted" species. We thus produced a stringent and restrictive list of candidates that cumulated a series of characteristics that made them the most promising candidates for the development of safer and more-specific control methods.

## Infestation assays

Overall, after siRNA soaking, we measured significant and reproducible effect on infection on 12 targeted genes. This effect was measured by a diminution in the number of galls or egg masses. Reduction in the number of galls implies that fewer nematodes have managed to induce a feeding structure. A reduction in the number of egg masses signifies that fewer female nematodes have managed to complete their development until the production of egg masses, a necessary step to propagate the infection at the next generation. In 6 cases, we measured a reproducible and significant reduction in the number of galls. Interestingly, 5 out of these 6 cases also led to significant and reproducible diminution of the number of egg masses. This observation makes sense since reduction in the capacity of nematodes to form galls will have direct downstream impact on the number of egg masses produced. Interestingly, targeting gene Minc01632 was responsible for both the most important reproducible diminution of the number of galls and of the number of egg masses. The corresponding protein is 155 amino-acids long and has neither significant similarity in the NCBI's nr database nor predicted protein domain, as most of the 16 identified novel targets.

In contrast, observing significant reduction of the number of egg masses does not necessarily require upstream reduction of the number of galls. Indeed, if the siRNA-targeted gene has functional consequences in processes that take place between the formation of galls and the production or extrusion of eggs we should observe

a significant reduction in the number of egg masses but not in the number of galls. This is indeed what we observed for 6 targeted genes (Minc00801, Minc03313, Minc08335, Minc12224, Minc17713 and Minc07817). While reduction of the number of egg masses was significant and reproducible; reduction in number of galls was either not reproducible or did not reach the significance threshold.

Overall, we observed no correlation between reduction of infestation and a measurable effect on nematode motility or viability. This indicated that the effect on infestation was globally not due to a toxicity of the siRNA treatment. For instance, genes that showed among the most important reduction in the numbers of egg masses or galls (e.g. Minc01632>40% reduction or Minc09526~40% reduction) did not show substantial diminution of viability or mobility 1 h or 16 h after soaking. We can thus deduce that the reduced infestation observed is generally not a consequence of reduced motility but more likely results from modification in other processes important for parasitism. Because the genes we have targeted are mostly specific to RKN and not shared by many species, we expected no systematic effect on viability or motility as opposed to evolutionarily conserved housekeeping genes [24].

## Effect of RNA interference on transcripts levels

Treatments with siRNAs had reproducible significant effects on target transcript levels in 13 out of the 16 samples tested (Table 2). Twenty-four hours after soaking, six genes showed a diminution of the transcript abundance while 7 yielded an increase of transcript level. Because we suspected a possible bounce effect, we randomly picked 3 of these 7 genes and measured transcript abundance at an earlier time point (16 h). One of the tested genes (Minc03866) showed a significant and reproducible diminution of transcripts level at this time point. It is possible that some of the six other genes that showed an increase of transcripts level at 24 h may also present an initial decrease at an earlier time point. Such bounce phenomenon has already been described in plant-parasitic nematodes [21,22,23]. Interestingly, the 13 siRNAs yielding effects on transcript level encompass 10 out of the 12 cases of reproducible significant reduction of infestation. Furthermore, for 7 genes (Minc00801, Minc01632, Minc02483, Minc03866, Minc08335, Minc09526, and Minc07817), following the siRNA treatment, there is both a significant and reproducible diminution of the abundance of transcripts and of the infestation of nematodes. Intriguingly, for two genes, there is significant and reproducible diminution of infestation but no significant effect on transcripts level. Investigating earlier or later time points may reveal significant effects. Alternatively, the corresponding mRNA may be sequestered away from the translation machinery without being itself degraded. Such a mechanism of translation repression without mRNA degradation has already been documented in plants and animals [25].

## *In situ* hybridizations

We performed *in situ* hybridization assays on the 12 genes that yielded significant reduction of infestation to try to gain information on their putative functions. Because 11 of the 12 tested genes share characteristics with known RKN effectors, it could be expected that they show transcription localization patterns similar to the known effectors. Canonical effectors are transcribed in secretory gland cells for injection by the nematode in plant tissue. We found one gene expressed specifically in the subventral gland cell (Minc03866). This gene could well encode an effector protein eventually secreted in plant tissue during infestation. Interestingly, when targeted *via* siRNA, this gene returned one of the strongest effect on reduction of infestation. Ubiquitous expression, which includes the secretory gland cells, was observed for 3 genes and these genes could be multi-functional, including possibly effectors depending on whether they are eventually secreted *in planta* or not. A total of 5 genes returned no detectable signal and although they may function as effector, there is no further supporting data from *in situ* hybridization assays. For the three other genes, expression localization does not support a possible secretion *in planta*, at least at the observed J2 stage. One gene (Minc02483) shows an expression localization specifically on nerve tissue surrounding a region called the metacorpus. The metacorpus acts as a pump to inject secretion or to take up nutrients from the nematode syringe-like stylet. It is possible that the gene expressed in the surrounding nerve cells may be involved in correct functioning of this pump. siRNA targeted against this gene led to the second strongest reduction effect on the number of galls and egg masses. The two other genes have an expression restricted to the intestinal tract and their targeting by siRNAs leads to significant and reproducible reduction of the number of egg masses. Lacking any known protein domain, it would be too speculative to predict any function for the corresponding gene products.

## Potential for the development of novel control methods

Using soaking experiments with siRNAs targeting each of the 16 identified novel genes, we noticed a significant and reproducible diminution of infestation in 12 cases. These results were obtained by inoculating infective J2 larvae after one hour soaking in a solution containing a siRNA concentration of 0.05 mg/ml. Although siRNA delivery *via* soaking can be relatively efficient because of systemic propagation of the RNA interference, levels of inactivation can vary and duration of the effect is poorly known [26]. Thus, it is possible that some of the genes we have identified would show significant reduction of infestation only when targeted at later stages of the nematode life cycle. Unfortunately, J2 infective larvae is the only free-living stage that can be targeted with soaking approaches, the rest of RKN life cycle takes place within plant tissue. However, delivery of siRNA at later stages can be imagined by development of transgenic plants expressing these interfering RNAs. Because the nematode feeds on root cells, siRNA can be actively delivered and it can be hypothesized that this mode of delivery is more efficient than passive soaking. Supporting this idea, genetically modified plants that express interfering RNAs have already proved to be efficient at reducing nematode infestation, at least in laboratory conditions (for review [26,27]). Before such transgenic plants can possibly reach the market, their bio-safety must be assessed. Because the genes we have identified are absent from species concerned by bio-safety risk (including human and the host plant), development of such transgenic plants is a promising potential application towards novel control methods of RKN. Alternatively, and because genetically-modified plants are not well accepted, especially in Europe, agrochemical approaches to develop new compounds specifically targeting one or several of the genes we have identified can be considered. However, this probably represents a less evident and straightforward strategy.

## Conclusion

Overall, our bioinformatics RKN genome screen has led to the identification of a series of genes present in multiple plant-damaging species that probably play important roles in successful parasitic interactions. In the absence of genetic tools to test the effect of gene knock-out in RKN, we have opted for an RNAi gene knock-down strategy. Despite potential limitations in detecting

physiological effects, siRNA treatment yielded significant and reproducible reduction of infestation in 12 out of the 16 testable cases. Overall, 5 siRNAs yielded both a diminution of egg masses and galls and this diminution was correlated with a diminution of the abundance of transcripts in the corresponding gene. These 5 genes probably represent the most promising targets for the development of novel efficient control means more specific and safe for the environment.

## Materials and Methods

### Bioinformatics

**ORTHOMCL comparative analysis.** We compared the whole protein sets of *M. incognita* and *M. hapla* against those of 23 other species using OrthoMCL [9] with default parameters to detect putative orthologs based on a reciprocal best blast hit approach. Criteria for species selection are explained in the results section. We used in house Perl scripts to extract RKN genes that had either no predicted orthologs in the 23 other selected species at all or orthologs only in parasites, phytopathogens or phytophagous species. To avoid redundancy between *M. incognita* and *M. hapla* proteins, we only kept *M. incognita* sequences as representative whenever proteins from both the two RKN were present in a same OrthoMCL cluster. All against all BLASTp analysis of the 25 protein sets have been performed on a computational grid (ProActive PACA grid: http://proactive.inria.fr/pacagrid/).

**Taxonomic BLASTp analysis.** BLASTp analyses were performed with an e-value threshold of 0.01 and without low-complexity filter against the NCBI's nr database at the protein level. BLASTp hits were considered as significant based on adapted percent identity and percent of alignment length thresholds. These thresholds were determined by examining the average lowest percent identity and query length coverage from one-to-one orthologs between RKN and the other species obtained during the OrthoMCL analysis (Table S5). We distinguished two different cases depending on whether or not the subject species is a Metazoan. For Metazoan subjects, BLASTp hits were considered as significant if they aligned with at least 40% identity on at least 70% of the RKN query protein length. For non-metazoan subjects, BLASTp hits were considered significant if they aligned with at least 30% identity on at least 50% of the query length.

Similarly to the OrthoMCL analysis, we eliminated every RKN protein that presented a significant BLASTp hit with at least one blacklisted species. For the BLASTp analysis, we blacklisted 4 different whole taxa, representing a total of 170,258 species:

- Chordata: 46,011 species
- Annelida: 4,551 species
- Mollusca: 11,932 species
- Viridiplantae: 107,764 species

All RKN proteins that did not return a significant hit in one of these species or that did not return any significant hit at all were kept for the rest of the analysis.

In parallel, we tagged RKN proteins that were not eliminated and returned a significant hit in at least one "plant-damaging species" as shared with another plant parasite.

**Constitution of a database of "plant-damaging species".** We established a list of species that are known plant-pathogens, plant-parasites or known to feed specifically on plant material, using available databases dedicated to plant-interacting organisms as well as prior knowledge on phylogenetic clades containing plant-pathogens or plant-parasites.

We first retrieved the list of species present in the "Comprehensive Phytopathogen Genomics Resource" (CPGR) database [28]. This list contains 806 referenced species and strains, including 63 bacteria, 56 fungi, 16 nematodes, 12 oomycetes, 36 viroids and 623 viruses. For some species, several strains are listed and refer to the same NCBI's taxonomy identifiers (TaxIDs). In total, 794 distinct corresponding TaxIDs could be listed. Besides CPGR, we retrieved the list of species having plant as hosts in the "Pathogen Host Interaction database" (PHI-base) [29]. This allowed retrieval of 54 species, including 46 fungi, 3 bacteria and 5 oomycetes. Elimination of redundancy between the lists of species extracted from the CPGR and from PHI-base led to a total of 788 different species (distinct TaxIDs), including 63 bacteria, 50 fungi, 14 nematodes, 8 oomycetes, 619 viruses and 34 viroids.

Because we aimed at producing the most comprehensive possible list of plant-damaging species, we completed the non-redundant list extracted from CPGR and PHI-base with whole nodes from the NCBI's tree of life:

- Plant-parasitic nematodes, 4 nodes: Tylenchida (879 species, e.g. root-knot nematodes, cyst nematodes etc.), Nordiidae (26 species, e.g. *Pungentus thornei*, *Longidorella parva* etc.), Longidoridae (130 species, e.g. *Xiphinema index*, *Longidorus sylphus* etc.) Trichodoridea (30 species, e.g. *Trichodorus primitivus*, *Paratrichodorus minor*).
- Plant-parasitic insects, 1 node: Aphididae (448 species, e.g. *Acyrthosiphon pisum*) and phytophagous insects (mostly), 1 node Lepidoptera (24,551 species, e.g. *Spodoptera frugiperda*).
- Plant-pathogenic Oomycetes, 4 nodes: Pythium (484 species, e.g. *Pythium ultimum*), Phytophtora (363 species, e.g. *Phytophthora ramorum*), Peronosporaceae (312 species, e.g. *Hyaloperonospora arabidopsidis*), Albugo (17 species, e.g. *Albugo candida*)+2 species: *Aphanomyces cochlioides*, *Aphanomyces euteiches*.
- Plant-parasitic Trypanosomatidae, 1 node: Phytomonas (82 species, e.g. *Phytomonas serpens*).

Overall, the list we have constituted contains 834 distinct TaxIDs (species and nodes), including 18 from nematodes, 16 from oomycetes, 79 from fungi, 65 from bacteria, 1 from trypanosomatida, 2 from insects, 619 from viruses and 34 from viroids. This represents a total of 28,054 plant-damaging species.

***In silico* functional annotation.** Known protein domains in the 34,780 RKN proteins were searched using HMMER 3.0 [30] against the Pfam-A database of manually curated HMM profiles [31]. Corresponding Gene Ontology (GO) terms assigned to Pfam domains have been retrieved using a Perl script developed for this occasion. For comparative purpose between the two RKN proteomes and between the set of RKN proteins that passed both the OrthoMCL and taxonomic Blast filters (protein set 1), we mapped all the G.O. terms to the generic GO-slim ontology containing only parent terms. Consequently, the different datasets are all annotated at the same granularity level which allows direct comparisons. To map the general GO terms to the GO-slim ontology, we used the Perl module GO-Perl and the Perl script map2slim. We used a Fisher's exact test to assess whether observed differences in relative abundances of GO terms were statistically significant.

**Searching effector-like proteins.** Signal peptides for secretion were searched using SignalP 3.0 [32] with both neural network and hmm based methods. Whether predicted signal peptides were supported by one or both methods was reported in a dedicated database. We also searched transmembrane regions using TMHMM [33] with default parameters and stored the results in our database. Domains specifically frequent in known

RKN effectors, previously identified using the MERCI software [14] were searched in the two whole RKN protein sets.

**Transcriptional support.** To assess whether a gene was supported by transcriptional data, we used several different sources of evidence. Basically, we aligned the coding sequences (CDS) corresponding to the RKN proteins to different transcriptomic libraries using BLASTn. We aligned the CDS to 22,350 distinct *M. incognita* EST contigs generated as described in [15] and to 137,733 contigs generated from the RNA-seq transcriptome sequencing of 6 different *M. incognita* developmental life-stages as described below. Both for EST or RNA-seq contigs, we set the e-value threshold to 1e-20. To have comparable bit scores and e-values between the different sequence libraries, we manually set the size of the database to $z = 60{,}000{,}000$ sequences. We considered a CDS as transcriptionally supported, provided that it returned alignments with at least 98% identity on at least 80% of the EST or RNA-seq contig length.

**Processing, mapping and assembly of RNA-seq data.** All the sequence libraries were assembled *de novo* using Velvet/Oases software after elimination of reads of low quality. Adaptators were removed and reads redundancy at 100% identity level was eliminated (collapsing) before assembly. Reads longer than 25 bp were assembled with velvet_1.0.15/oases_0.1.18. Number of predicted contigs (i.e. transcripts) ranged from ~8,500 for *M. incognita* parasitic stage to ~30,000 for adult males. Assembled reads for the different developmental life stages are available for download at the following URL: http://www7.inra.fr/meloidogyne_incognita/genomic_resources/downloads. Individually for the 6 developmental life stages, cleaned reads were aligned to the *M. incognita* genome using Bowtie2 [34] and Tophat2 [35]. Gene expression patterns were deduced from the aligned reads, using Cufflinks [36], according to the protocol published in [37] and presented as RPKM values in Table S4.

## Biological experiments

**Sample preparation and RNA-seq Illumina sequencing.** A total of 6 different life stage samples (Table 1) were collected from tomato roots (*Solanum esculentum* cv. St Pierre) by incubation in 10% (v/v) Pectinex (Novozymes, Bagsvaerd, Denmark) and 5% (v/v) Celluclast BG (Realco, Louvain-la-neuve, Belgium) for 3 hours, respectively 10, 40 and 60 days after inoculation. Males were collected as previously described [38,39]. All the other samples were purified from root debris by sucrose gradient centrifugations. RNA isolation using TRIzol Reagents (Invitrogen, Carlsbad, CA, USA) was done according to protocol available from Invitrogen and re-suspended in 10 µl RNase-free water. Purity and concentration of the RNA was determined on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and cDNA was only produced from high RNA quality (RIN>7). Reverse transcription was carried out using the Ovation pico WTA System (NuGEN Technologies, Inc, San Carlos, CA, USA).

The cDNAs were sonicated separately to a 150- to 600-bp size range using the S2 covaris instrument (Covaris, Inc., USA). Single end libraries were prepared following Illumina protocol (Illumina DNA sample kit). Briefly, fragments were end-repaired, then 3′-adenylated, and Illumina adapters were added. Ligation products of 350–400 bp were gel-purified and size-selected DNA fragments were PCR-amplified using Illumina adapter-specific primers. Libraries were purified and then quantified using a Qubit Fluorometer (Life technologies) and libraries profiles were evaluated using an Agilent 2100 bioanalyzer (Agilent Technologies, USA). Each library was sequenced using 76 base-length read chemistry in a single flow cell on the Illumina GA IIx (Illumina, USA).

**siRNA design and siRNA treatment for viability tests and infection assays.** siRNAs (Figure S4) were prepared using the Silencer siRNA Construction Kit (AM1620, Ambion, Austin, TX) and siRNA yields were determined using a NanoDrop 2000 spectrophotometer (NanoDrop Products, Wilmington, DE). For control, we used a siRNA designed to have no sequence similarity in the *M. incognita* genome [40] as confirmed by BLASTn searches. siRNAs were conserved at −80°C in 2 µg aliquots until use. Eggs of *M. incognita* were collected from tomato plants (*Solanum esculentum* cv. St Pierre) cultured in greenhouse. Eggs were collected as described by Rosso *et al.* [38] and J2s were hatched in water. About 10,000 J2s were soaked in 40 µl final volume of spring water in the presence of 0.05 mg/ml siRNA for 1 hour. Worms were washed twice with water by centrifugation at 10,000 g for 1 min and suspended in 100 µl of water. For infection assays, roots of 24 tomato plants aged of four weeks were each inoculated with 250 *M. incognita* infective J2 larvae, previously washed and oxygenated over night in spring water as recently described [41]. Two replicates of the infection assays were performed at three weeks intervals. Galls and egg masses were counted six weeks after inoculation. Statistical analyses were performed by an ANOVA test using the R software. For viability/motility assay, J2s were also soaked with siRNAs for one hour. The number of dead J2s was counted under microscope observation on 100 individuals, one hour and 16 hours after soaking and washing. We also analyzed movement quality and rapidity for 25 individuals. For movement quality we observed extremity and whole length movement and the quality of movement was noticed by rapidity of undulation or absence of undulation.

**Quantitative-PCR.** RNA was isolated from approximately 500 J2s soaked for 1 hour in 0.05 mg/ml siRNA and incubated for 16 h or 24 h in water, using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA). Purity and concentration of the RNA was determined on a NanoDrop_2000 spectrophotometer (NanoDrop Products, Wilmington, DE). Reverse transcription was carried out using the iScript cDNA Synthesis Kit (Bio-Rad laboratories, Marnes la Coquette, France). The primers for qPCR were determined using primer3 software [42] and synthesised by Eurogentec (Seraing, Belgium) (Figure S4). The cDNA was diluted 10 times, and 5 µL was used per PCR reaction. Seven and a half microliters of 2× SYBR Green Master Mix (Eurogentec, Liege, Belgium), 0.2 µL each of 1001µM of forward and reverse primer, and 4.61µL of water were added to the cDNA. Thermocycling was carried out with one cycle at 95°C for 15 min, followed by 40 cycles of 95°C for 15 sec and 56°C for 1 min and 72°C for 30 sec. The dissociation curve of the final products was checked to ascertain the presence of a single amplification product. qPCR was performed on triplicate samples of each cDNA. Among the three tested candidate reference genes, i.e. *M. incognita* polygalacturonase (*Mi-pg-1*, Minc18543), Glyceraldehyde 3-phosphate dehydrogenase (GAPDH, Minc10963) and Actin α (Minc06773), GAPDH was determined as the most stable reference using Genorm algorithm [43] (Data not shown) and was selected as reference gene. For normalization, the threshold cycle values of GAPDH amplifications ($CT_{GAPDH}$) were subtracted from the threshold cycle values of the analyzed genes (CTexp). Transcript levels in arbitrary units (AU) were calculated with the formula: $AU = 100 * 2^{(CTexp - CTGAPDH)}$. Figure 5 shows results from two independent replicates.

**Transcript analyses by *in situ* hybridization.** Sense and antisense probes were synthesized from each target gene with specific primers designed with Primer3 software [42] and synthesised by Eurogentec (Seraing, Belgium) (Figure S4). As controls, we used the polygalacturonase *Mi-pg-1* (Minc18543) gene

for esophageal gland-specific labeling and the *GAPDH* Minc10963 gene for ubiquitous labeling. *In situ* hybridizations were conducted as described previously [44]. 10,000 J2s were hybridized with DIG-labeled specific probes at 40°C over night for each target transcript.

## Supporting Information

**Figure S1 Transcript level measured 16 h after soaking to test bounce effect.** Expression level was measured for Minc07817, Minc03866 and Minc05001 and compared to nematodes treated with control siRNA 16 h after soaking to test siRNA knock-down effect at earlier time point.
(PPTX)

**Figure S2 Effects of siRNAs on motility and viability of nematodes.** Controls are J2 larvae soaked with siRNA targeting no sequence in the *M. incognita* genome, accession numbers indicate *M. incognita* genes targeted by siRNAs. Viability was assessed 1 h (A) and 16 h (B) after soaking by counting the number of dead nematodes (black bars), the number of individuals with movements restricted to extremities (red bars), movements on the whole length of the body (dark pink bars) and fast undulations (light pink bars) under microscope observation. Three independent replicates were analyzed. Error bars represent standard error of the mean. Variability from one replicate to another was generally too high and frequently higher than from one siRNA to another which precluded any statistical test from finding significant differences.
(TIF)

**Figure S3 *In situ* hybridizations by categories.** Localization of transcripts for the 12 genes that yielded significant and reproducible reduction of infestation. (I) Ubiquitous expression: Minc01632 Minc08335 and Minc1224. As a positive control, an antisense probe targeting NADPH transcripts was used. (II) Expression localized to subventral secretory gland cells: Minc03866. As a positive control, antisense probe targeted against Mi-PG1 transcript was used. (III) Expression localized to the intestinal tract: Minc00801, Minc17713. (IV) Expression is localized in the circumpharyngeal nerve ring: Minc02483. As a negative control, we used a sense probe designed on polygalacturonase gene Mi-PG1, known to be expressed in subventral gland cells. (V) No detectable signal.
(PPTX)

**Figure S4 Positions of qPCR primers and siRNAs on the 16 tested genes.** The positions of qPCR primers (forward in red, reverse in blue) as well as siRNAs (in green) are reported along the exon/intron structures of the 16 genes used for infestation assays and transcript level analyzes. The scale bars above each gene model represents a length of 100 nucleotides.
(PPTX)

**Table S1 Gene Ontology terms in *M. incognita*, *M. hapla* and in protein set 1.** For the 3 ontologies, (A) 'biological process', (B) 'molecular function', (C) 'cellular components of the gene ontology, we report the abundance of GO-slim general ontology terms in *M. hapla*, *M. incognita* and protein set 1. Raw abundance and proportion of total terms in a given ontology for a given set are indicated. The last raw represents the difference in proportion between protein set 1 and the average proportion in the whole RKN proteomes. A color gradient from enriched terms in red toward depleted terms in green accompanies the values.
(XLSX)

**Table S2 232 effector-like proteins supported by transcriptional data.** We list the 232 effector-like proteins that received transcriptional support and associated information. We include, the RKN species (*M. hapla* or *M. incognita*), the protein accession number, the length in amino-acids of the corresponding protein, the presence in plant-damaging species according to OrthoMCL and BLASTp analyses, the source of transcriptional support, presence of a known Pfam protein domain, and, whether the protein is a known effector.
(XLS)

**Table S3 Percent reduction in number of egg masses and galls after siRNA treatment.** For the 16 genes tested with siRNA treatment, we report the percent reduction of the number of egg masses or of the number of galls. We list gene accession numbers and the category (effector-like or putative transcription factor). Values of percent reduction in the numbers of galls or egg masses compared to control are accompanied with standard error of the mean (SEM) values and a significance code for the associated p-value.
(XLS)

**Table S4 Expression pattern of the 16 novel candidate targets according to RNA-seq data.** The expression level for the 16 novel candidate target genes in each of the 6 developmental life stages (from left to right: eggs to adults) according to RNA-seq data. Values correspond to RPKM (reads per kilobase per million mapped reads) obtained by aligning RNA-seq cleaned reads to the *M. incognita* genome. A color gradient from red to green indicate low to high RPKM values for each gene (row), individually.
(XLSX)

**Table S5 Percent identity/length OrthoMCL.** Average percent identity and query protein length of 1-to-1 orthologs between species used in the OrthoMCL analysis are indicated. We separated the dataset into two pools (i) closely related (metazoan) species, (ii) distantly related (non-metazoan) species. Lowest obtained values are represented in red.
(XLS)

## Acknowledgments

The authors want to thank Nathalie Marteu (INRA, Sophia Antipolis, France) for growing plant and nematode materials. We are also grateful to René Feyereisen and Franck Panabières (both INRA, Sophia Antipolis, France) who helped setting up the list of plant-damaging species in arthropods and oomycetes, respectively. This work benefited from computational power and infrastructure of ProActive PACA grid.

## Author Contributions

Conceived and designed the experiments: EGJD PA FA ACF LPB MNR. Performed the experiments: MJA ACF CDS KL JG MDR EGJD MM LPB NN. Analyzed the data: MJA ACF EGJD MM LPB MNR CDS NN. Contributed reagents/materials/analysis tools: CDS KL JG MNR. Wrote the paper: EGJD LPB PA.

## References

1. Sasser JN, Freckman DW (1987) A world perspective on nematology; the role of society. In: Veech JA, Dickerson DW, editors. Vistas on nematology. Hyattsville, USA: Society of Nematologists. pp. 7–14.

2. McCarter JP (2009) Molecular Approaches Toward Resistance to Plant-Parasitic Nematodes. In: Berg RH, Taylor C, editors. Cell Biology of Plant Nematode Parasitism. Springer Berlin Heidelberg. pp. 239–267.

3. Djian-Caporalino C (2012) Root-knot nematodes (Meloidogyne spp.), a growing problem in French vegetable crops. EPPO Bulletin 42: 127–137.
4. Wesemael WML, Viaene N, Moens M (2011) Root-knot nematodes (Meloidogyne spp.) in Europe. Nematology 13: 3–16.
5. Castagnone-Sereno P, Danchin EG, 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: 203–20.
6. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EG, et al. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nat Biotechnol 26: 909–915.
7. Bird DM, Williamson VM, Abad P, McCarter J, Danchin EG, et al. (2009) The genomes of root-knot nematodes. Annu Rev Phytopathol 47: 333–351.
8. Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, et al. (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.
9. Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178–2189.
10. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
11. Haegeman A, Mantelin S, Jones JT, Gheysen G (2012) Functional roles of effectors of plant-parasitic nematodes. Gene 492: 19–31.
12. Rosso MN, Hussey RS, Davis EL, Smant G, Baum T, et al. (2012) Nematode Effector Proteins: Targets and Functions in Plant Parasitism. In: Martin F, Kammoun S, editors. Effectors in Plant-Microbe Interactions: Wiley-Blackwell.
13. Mitchum MG, Hussey RS, Baum TJ, Wang X, Elling AA, et al. (2013) Nematode effector proteins: an emerging paradigm of parasitism. New Phytol 199 :879–94.
14. Vens C, Rosso MN, Danchin EG (2011) Identifying discriminative classification-based motifs in biological sequences. Bioinformatics 27: 1231–1238.
15. Jaouannet M, Perfus-Barbeoch L, Deleury E, Magliano M, Engler G, et al. (2012) A root-knot nematode-secreted protein is injected into giant cells and targeted to the nuclei. New Phytol 194: 924–931.
16. Tytgat T, Vercauteren I, Vanholme B, De Meutter J, Vanhoutte I, et al. (2005) An SXP/RAL-2 protein produced by the subventral pharyngeal glands in the plant parasitic root-knot nematode Meloidogyne incognita. Parasitol Res 95: 50–54.
17. Wang X, Li H, Hu Y, Fu P, Xu J (2007) Molecular cloning and analysis of a new venom allergen-like protein gene from the root-knot nematode Meloidogyne incognita. Exp Parasitol 117: 133–140.
18. Huang GZ, Dong RH, Allen R, Davis EL, Baum TJ, et al. (2005) Two chorismate mutase genes from the root-knot nematode *Meloidogyne incognita*. Molecular Plant Pathology 6: 23–30.
19. Neveu C, Abad P, Castagnone-Sereno P (2003) Molecular cloning and characterization of an intestinal cathepsin L protease from the plant-parasitic nematode Meloidogyne incognita. Physiological and Molecular Plant Pathology 63: 159–165.
20. Danchin EG, Rosso MN, Vieira P, de Almeida-Engler J, Coutinho PM, et al. (2010) Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc Natl Acad Sci U S A 107: 17651–17656.
21. Dubreuil G, Magliano M, Deleury E, Abad P, Rosso MN (2007) Transcriptome analysis of root-knot nematode functions induced in the early stages of parasitism. New Phytol 176: 426–436.
22. Bakhetia M, Urwin PE, Atkinson HJ (2008) Characterisation by RNAi of pioneer genes expressed in the dorsal pharyngeal gland cell of Heterodera glycines and the effects of combinatorial RNAi. Int J Parasitol 38: 1589–1597.
23. Dalzell JJ, Warnock ND, Stevenson MA, Mousley A, Fleming CC, et al. (2010) Short interfering RNA-mediated knockdown of drosha and pasha in undifferentiated Meloidogyne incognita eggs leads to irregular growth and embryonic lethality. Int J Parasitol 40: 1303–1310.
24. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, et al. (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature 421: 231–237.
25. Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. Nat Rev Genet 10: 94–108.
26. Rosso MN, Jones JT, Abad P (2009) RNAi and functional genomics in plant parasitic nematodes. Annu Rev Phytopathol 47: 207–232.
27. Atkinson HJ, Lilley CJ, Urwin PE (2012) Strategies for transgenic nematode control in developed and developing world crops. Curr Opin Biotechnol 23: 251–256.
28. Hamilton JP, Neeno-Eckwall EC, Adhikari BN, Perna NT, Tisserat N, et al. (2011) The Comprehensive Phytopathogen Genomics Resource: a web-based resource for data-mining plant pathogen genomes. Database (Oxford) 2011: bar053.
29. Winnenburg R, Urban M, Beacham A, Baldwin TK, Holland S, et al. (2008) PHI-base update: additions to the pathogen host interaction database. Nucleic Acids Res 36: D572–576.
30. Eddy SR (2011) Accelerated Profile HMM Searches. PLoS Comput Biol 7: e1002195.
31. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, et al. (2008) The Pfam protein families database. Nucleic Acids Res 36: D281–288.
32. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340: 783–795.
33. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2: 953–971.
34. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357–359.
35. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14: R36.
36. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511–515.
37. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562–578.
38. Rosso MN, Favery B, Piotte C, Arthaud L, De Boer JM, et al. (1999) Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. Mol Plant Microbe Interact 12: 585–591.
39. Snyder DW, Opperman CH, Bird DM (2006) A Method for Generating Meloidogyne incognita Males. J Nematol 38: 192–194.
40. Dalzell JJ, McMaster S, Fleming CC, Maule AG (2010) Short interfering RNA-mediated gene silencing in Globodera pallida and Meloidogyne incognita infective stage juveniles. International Journal for Parasitology 40: 91–100.
41. 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.
42. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365–386.
43. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: RESEARCH0034.
44. Jaubert Sp, Laffaire J-B, Abad P, Rosso M-Nl (2002) A polygalacturonase of animal origin isolated from the root-knot nematode Meloidogyne incognita. FEBS Letters 522: 109–112.
45. van Megen H, van den Elsen S, Holterman M, Karssen G, Mooyman P, et al. (2009) A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology 11: 927–950.
46. Trautwein MD, Wiegmann BM, Beutel R, Kjer KM, Yeates DK (2012) Advances in insect phylogeny at the dawn of the postgenomic era. Annu Rev Entomol 57: 449–468.

OPEN ACCESS Freely available online

PLOS ONE

# Contribution of Lateral Gene Transfers to the Genome Composition and Parasitic Ability of Root-Knot Nematodes

**Julien Paganini[1☯], Amandine Campan-Fournier[2,3,4☯], Martine Da Rocha[2,3,4], Philippe Gouret[1], Pierre Pontarotti[1], Eric Wajnberg[2,3,4], Pierre Abad[2,3,4], Etienne G. J. Danchin[2,3,4]***

**1** Aix-Marseille Université, Centre National de la Recherche Scientifique, LATP, UMR 7353, Evolution Biologique et Modélisation, Marseille, France, **2** INRA, Institut National de la Recherche Agronomique, UMR 1355 ISA, Institut Sophia Agrobiotech, Sophia-Antipolis, France, **3** CNRS, Centre National de la Recherche Scientifique, UMR 7254 ISA, Institut Sophia Agrobiotech, Sophia-Antipolis, France, **4** Université de Nice Sophia Antipolis, Institut Sophia Agrobiotech, Sophia-Antipolis, France

## Abstract

Lateral gene transfers (LGT), species to species transmission of genes by means other than direct inheritance from a common ancestor, have played significant role in shaping prokaryotic genomes and are involved in gain or transfer of important biological processes. Whether LGT significantly contributed to the composition of an animal genome is currently unclear. In nematodes, multiple LGT are suspected to have favored emergence of plant-parasitism. With the availability of whole genome sequences it is now possible to assess whether LGT have significantly contributed to the composition of an animal genome and to establish a comprehensive list of these events. We generated clusters of homologous genes and automated phylogenetic inference, to detect LGT in the genomes of root-knot nematodes and found that up to 3.34% of the genes originate from LGT of non-metazoan origin. After their acquisition, the majority of genes underwent series of duplications. Compared to the rest of the genes in these species, several predicted functional categories showed a skewed distribution in the set of genes acquired via LGT. Interestingly, functions related to metabolism, degradation or modification of carbohydrates or proteins were substantially more frequent. This suggests that genes involved in these processes, related to a parasitic lifestyle, have been more frequently fixed in these parasites after their acquisition. Genes from soil bacteria, including plant-pathogens were the most frequent closest relatives, suggesting donors were preferentially bacteria from the rhizosphere. Several of these bacterial genes are plasmid-borne, pointing to a possible role of these mobile genetic elements in the transfer mechanism. Our analysis provides the first comprehensive description of the ensemble of genes of non-metazoan origin in an animal genome. Besides being involved in important processes regarding plant-parasitism, genes acquired via LGT now constitute a substantial proportion of protein-coding genes in these nematode genomes.

**Citation:** Paganini J, Campan-Fournier A, Da Rocha M, Gouret P, Pontarotti P, et al. (2012) Contribution of Lateral Gene Transfers to the Genome Composition and Parasitic Ability of Root-Knot Nematodes. PLoS ONE 7(11): e50875. doi:10.1371/journal.pone.0050875

**Editor:** Carlos Eduardo Winter, Universidade de São Paulo, Brazil

**Received** July 11, 2012; **Accepted** October 25, 2012; **Published** November 30, 2012



**Funding:** This project received funding from the CNRS multidisciplinary program 'Maladies Infectieuses Emergentes'. JP's salary was funded by the CNRS multidisciplinary program 'Maladies Infectieuses Emergentes'. AC-F's salary was funded by the ANR program plant-genomics NEMATARGETS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: etienne.danchin@sophia.inra.fr

☯ These authors contributed equally to this work.

## Introduction

Recognized cases of lateral gene transfers (LGT) in animals are relatively scarce compared to the plethora of examples in prokaryotes. Furthermore, most reported cases of gene transfers from a non-metazoan donor species to an animal host genome are not clearly linked to an identified biological process or life trait in the receiver species. Thus, it is difficult to assess whether LGT events have played an important evolutionary role in animal genomes [1]. Nevertheless, a few studies have shown an evident role of transferred gene products in the receiver animal organisms [2,3,4]. These examples involve gene transfers from non-metazoan eukaryotes to animals but not from prokaryotes to animals. This is surprising because much more cases of transfers from bacteria to animals than between eukaryotes have been reported so far [5]. Yet, several cases of LGT of bacterial origin and with significant functional consequences in animal have been reported from plant-parasitic nematodes. These nematodes represent an important economic threat as they are annually responsible for over 100 billion Euros loss in crop plants yields [6]. A recent review showed that genes acquired via LGT in these nematodes are involved in key parasitism processes such as modulation of plant defense, establishment of a feeding structure or degradation of the plant cell wall [7]. For instance, a whole repertoire of genes for the degradation of the plant cell wall has been acquired by several independent LGT events from different bacterial sources, followed by gene duplications [8]. Cases of LGT in plant-parasitic nematodes have been so far essentially identified indirectly by searching candidate parasitism genes and do not result from systematic and comprehensive genome scans. Consequently, there is currently no estimation of the total contribution of LGT to the genome composition and biology of plant-parasitic nematodes. Root-knot nematodes (*Meloidogyne* genus) are the most widespread

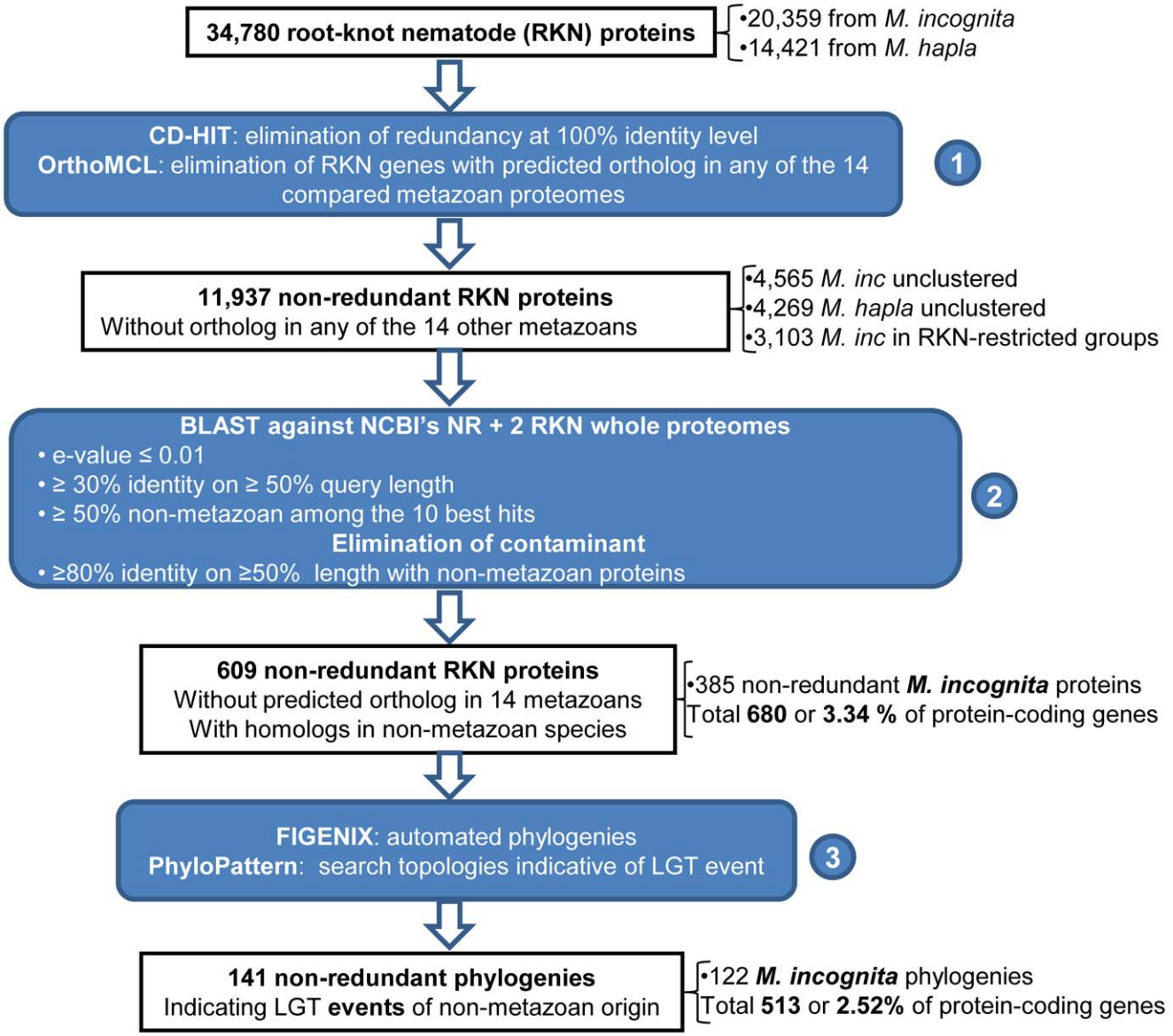



**Figure 1. Schematic pipeline for detection of lateral gene transfers.** This simplified representation highlights the three main steps used to detect potential lateral gene transfers of non-metazoan origin in root-knot nematodes. The three bioinformatics steps are represented within blue rectangles while initial, intermediate and final results are represented within white rectangles. Starting from 34,780 root-knot nematode proteins, step 1 consisted in eliminating redundancy at 100% identity and detecting orthologs in proteomes of 14 other metazoan species. Step 2 consisted in "blasting" all proteins that passed step 1 against the NCBI's NR database completed with the whole proteomes of the two root-knot nematodes. Only proteins that returned at least 50% of non-metazoan hits among their 10 best blast hits were kept at this stage. Proteins that showed more than 80% identity with non-metazoan hits on at least half of their length were considered as contaminants and eliminated. All proteins passing step 2 were sent for automated phylogenetic analysis using FIGENIX pipeline. Topologies compatible with a lateral gene transfer were automatically searched among all generated trees using PhyloPattern. At the end of step 2 and of step 3, the total number of *M. incognita* protein-coding genes of non-metazoan origin, (including gene duplicates) and the proportion of the whole gene set are indicated in bold.
doi:10.1371/journal.pone.0050875.g001

and damaging of these plant parasites. A previous analysis of ESTs from three root-knot nematodes, *Meloidogyne javanica*, *M. incognita* and *M. hapla*, and comparison to the genomes of *C. elegans* and *D. melanogaster*, has provided a first "without *a priori*" overview of putative LGT events in these nematodes [9]. However, EST data only offer a fractional representation of the whole set of protein-coding genes in a given species and many potential LGT events may be missed in the absence of an available whole genome sequence.

Here, taking advantage of the availability of two root-knot nematodes whole genome sequences [10,11], we have systematically searched potential LGT events of non-metazoan origin using a comparative genomics analysis with 16 species coupled with an automated phylogenetic reconstruction and tree topology scan. Using a phylogenetic approach not only allowed confirming similarity-based prediction of LGT but also examining the fate of genes after their transfer, including their duplication pattern. Our approach allowed retrieving all cases of LGT of non-metazoan origin in root-knot nematodes reported so far in the literature as

well as new candidate cases of LGT not identified before. An analysis of the domain composition and putative functions of these genes indicates that they are preferentially involved in functions related to degradation, modification and metabolism of carbohydrates and proteins, reflecting the parasitic lifestyle of root-knot nematodes. Examination of the topologies of phylogenetic trees showed that the majority of genes acquired via LGT underwent series of duplications after their transfer. Overall, we show that up to 3.34% of protein-coding genes originate from LGT of non-metazoan origin in a root-knot nematode genome. Genes acquired via LGT do not appear to form clusters in the genome but the density of transposable elements is higher around genes acquired via LGT. Bacterial genes, including from notorious plant-pathogens sharing the same hosts as root-knot nematodes were frequently found as the most closely relatives. Finally, we discuss the hypothetical mechanisms involved in these LGT events and their evolutionary importance, both in the making of an animal genome and the emergence of plant-parasitism in nematodes.

## Results

### Identification of Lateral Gene Transfers in Root-knot Nematodes

From pooled root-knot nematode whole proteomes (14,421 proteins from *M. hapla* and 20,359 from *M. incognita*), we identified a total of 11,937 non-redundant proteins that had no predicted ortholog in any of the 14 other compared metazoan genomes, based on an OrthoMCL [12] analysis (Figure 1, Table S1). Although these genes might actually have no evident homolog in metazoan species, our selection of 14 metazoan proteomes from various lineages (Figure S1) cannot be considered as fully representative of the spectrum of diversity present in animals. Furthermore, the presence and degree of conservation of these proteins in non-metazoan species needs to be assessed. We thus compared the 11,937 proteins apparently specific from root-knot nematodes against the NCBI's non-redundant (NR) database using BLASTp [13]. Proteins that returned no significant hit in NR with the parameters we have set (methods) were discarded from the analysis because it is not possible to state whether (i) they actually originate from LGT event of as yet unidentified source or (ii) represent over-predicted gene models or (iii) represent true orphan genes restricted to root-knot nematodes. Because we are interested in transfers of genes from non-metazoan species to root-knot nematodes, we specifically selected root-knot nematode proteins that returned at least 50% of non-metazoan hits among their 10 best blast hits (Figure 1). A total of 609 non-redundant proteins satisfied this criterion and were considered as potentially originating from LGT of non-metazoan origin (Table S2).

These proteins were sent for automatic phylogenetic analysis using FIGENIX [14,15] and topologies supporting potential lateral gene transfer events were searched using PhyloPattern within the DAGOBAH framework [16] (Figure 1, Figure 2). Phylogenetic trees were successfully constructed for 490 out of the 609 protein-coding genes. A total of 141 proteins yielded phylogenetic trees with topologies supporting LGT events (methods, Figure 2). The 141 trees can be consulted interactively in the I.O.D.A. database [17] (http://ioda.univ-provence.fr). The rest of the proteins (468) either did not return a tree topology compatible with the searched phylogenetic pattern(s) (349 cases) or due to an insufficient number of BLAST hit did not allow construction of a phylogenetic tree (119 cases). The corresponding genes were considered as possibly acquired via lateral gene transfer but without phylogenetic support (Table S2).

An analysis of the literature allowed us to establish a list of 15 distinct cases of genes or gene families acquisition via LGT in root-knot nematodes (Table 1). Interestingly, all these previously reported cases of LGT were retrieved in our systematic genome scan and all received a phylogenetic support. Hence, as a validation, our approach allowed retrieval of all previously published cases of candidate LGT events in *M. hapla* and *M. incognita*, indicating a good sensitivity. Concerning the specificity, we found all the previously reported cases within a set of only 141 non-redundant proteins that passed the BLAST filter and returned phylogenetic trees with topologies supporting LGT. Given that the number of root-knot nematodes proteins initially used in entry of the pipeline is 20,359 and 14,421 for *M. incognita* and *M. hapla*, respectively, our method can be considered highly specific.

### Functions of Genes Acquired via LGT

In each of the two whole root-knot nematode proteomes, half of the proteins have been assigned at least one Pfam domain (50.9% and 49.59% for *M. incognita* and *M. hapla*, respectively). Based on these domains assignments, a total of 6,881 and 4,673 proteins were associated at least one Gene Ontology (GO) term, in *M. incognita* and *M. hapla*, respectively (Methods). All the GO terms were mapped to at least one parent term in the generic GO-slim ontology (Table S3). The distribution of GO terms for the three ontologies (biological process, cellular component and molecular function) were very similar between *M. incognita* and *M. hapla* whole proteomes, indicating a similar global qualitative distributions of putative functions in these two species despite different gene numbers (Table S3). The higher number of protein-coding genes in *M. incognita* compared to *M. hapla* is due to the peculiar structure of the *M. incognita* genome, mainly composed of pairs and triplets of similar yet divergent regions [10,18]. The similarity of GO-terms distribution in *M. incognita* and *M. hapla* suggests that the frequency of gene copy retention in *M. incognita* has been homogeneous across the different functional categories.

We more specifically compared the distributions of GO-terms assigned to the 609 non-redundant candidate LGT proteins that passed both the OrthoMCL and BLAST filters to those of the whole root-knot nematode proteomes. Out of these 609 candidate LGT proteins, 335 (~55%) were assigned at least one Pfam domain. Corresponding gene ontology terms could be assigned to 234 of these proteins (~38% of the candidate LGT). For the three ontologies, the distributions of GO terms of the candidate LGT proteins were substantially different from those of the whole root-knot nematode proteomes (Figure 3, Table S3).

In the 'biological process' ontology, differences of highest amplitudes included an over-representation of the 'carbohydrate metabolic process', 'protein metabolic process' and 'protein modification process' terms in the set of LGT-acquired genes compared to the whole proteomes (Figure 3A ~10% vs. ~3–4% of annotated proteins for 'carbohydrate metabolic process', ~16% vs. 7–8% for 'protein metabolic process' and ~18% vs. ~9% for 'protein modification process'). In the 'carbohydrate metabolic process' category, several genes previously reported as acquired via LGT and known to encode plant cell wall-degrading or modifying enzymes were retrieved, including 12 GH5 cellulases, 3 GH28 polygalacturonases and 2 GH43 candidate arabinanases (Table S2). Besides these known LGT cases, other enzymes not previously described and potentially targeting plant polysaccharides were identified, including a putative starch-binding CBM20-bearing protein, a mannose 6p isomerase that can be involved in modification of the plant cell wall or a GH25 enzyme annotated in Pfam as possibly active on cell wall macromolecules. The 'protein metabolic process' category contained a majority of

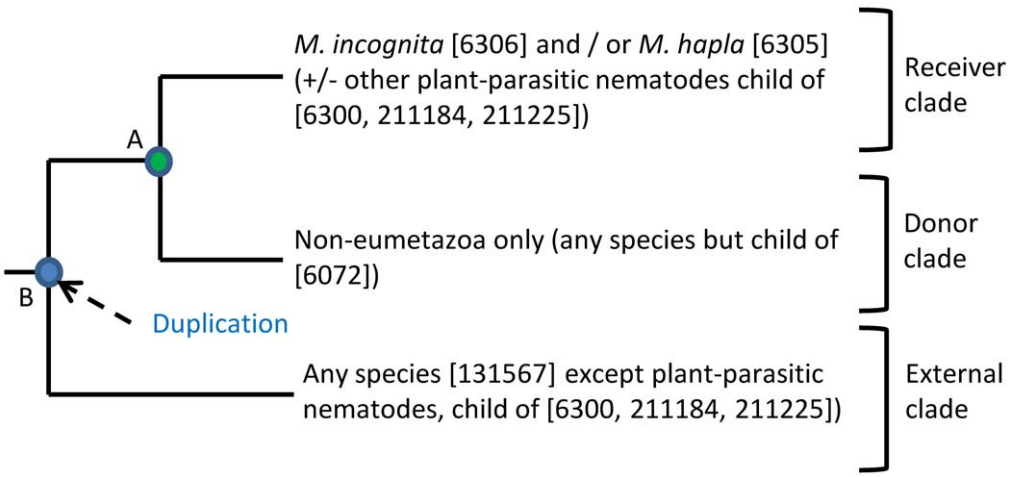



**Figure 2. Phylogenetic pattern searched to identify lateral gene transfers.** Schematic representation of the phylogenetic patterns searched with PhyloPattern [16] to identify trees harboring a topology indicating a lateral gene transfer of non-metazoan origin in root-knot nematode genomes. Basically, the topology searched is composed of three main clades. In every clade, species or taxonomic division authorized or forbidden as well as their NCBI's taxonomy identifiers are indicated. The "receiver clade" must contain at least one sequence from *M. incognita* or from *M. hapla* and possibly from other species provided that these species are plant-parasitic nematodes. The "donor clade" can contain any species but eumetazoan (e.g. bacteria, fungi, plant, …). The external clade can contain any species but plant-parasitic nematodes. Presence of a node "A" connecting the receiver clade and the donor clade to the exclusion of the external clade is required and constitutes a minimal phylogenetic support for LGT. Strong support for LGT was assigned when, additionally, a node "B", defined as follows was found. This node "B" must connect node "A" to the external clade and this node must be detected as a duplication node due to presence of at least one non metazoan species in the external clade.
doi:10.1371/journal.pone.0050875.g002

peptidases of different families, the most abundant being lon proteases that belong to MEROPS peptidase family S16. Phylogenetic trees indicate putative fungal and bacterial origins for these peptidases. Although these enzymes may be involved in degradation of plant proteins including for detoxification none has yet been experimentally characterized so far. Finally, the 'protein modification process' category mainly consisted of protein kinases. None had previously been reported and all those supported by a phylogeny indicate a candidate protist origin. Although 6 protein kinases have a predicted signal peptide and could be secreted by the nematodes, their precise role remains to be determined. In contrast, the term 'regulation of biological processes' was under-represented in the set of candidate LGT genes (~3% vs. ~8% of annotated proteins). Thus, overall, it appears that proteins with putative functions involved in carbohydrate metabolism as well as in protein metabolism and modification are over-represented in the set of proteins putatively acquired via LGT.

**Table 1.** 15 distinct previously reported case of LGT in root-knot nematodes.

| Gene/gene family | From ref. | Best phylogenetic support | RKN protein accession number |
|---|---|---|---|
| GH28 polygalacturonase | [7,18,39] | A+B | Minc18543b |
| PL3 pectate lyase | [7,40,41] | A+B | Minc01522c |
| GH43 candidate arabinase | [7,10,11] | A+B | Minc10639 |
| GH5 cellulase | [7,10,18,42,43,44] | A+B | Minc18711 |
| GH30 xylanase | [7,10,18,45] | A+B | Minc18650 |
| Expansin-like protein | [7,41] | A | Minc10987 |
| GH32 candidate invertase | [7,10,18] | A | Mh_Contig1358:39..3354 |
| Chorismate mutase | [7,10,11,46] | A | Minc10536 |
| Cyanate lyase | [7,11] | A+B | Minc06015 |
| VB5 pantothenate | [7] | A+B | Minc14603 |
| VB7 biotin | [7] | A+B | Minc09512+ Minc09513* |
| NodL | [7,9] | A+B | Mh_Contig222:75571..76264 |
| MI00426 Glutamine synthetase (GSI) | [9] | A | Minc08077 |
| MI01644 L-threonine aldolase | [9] | A+B | Mh_Contig2499:3012..6351 |
| MI00109 candidate phosphoribosyltransferase | [9] | A | Minc16723 |

*Gene models Minc09512 and Minc09513 have to be fused to form a full length protein.
doi:10.1371/journal.pone.0050875.t001

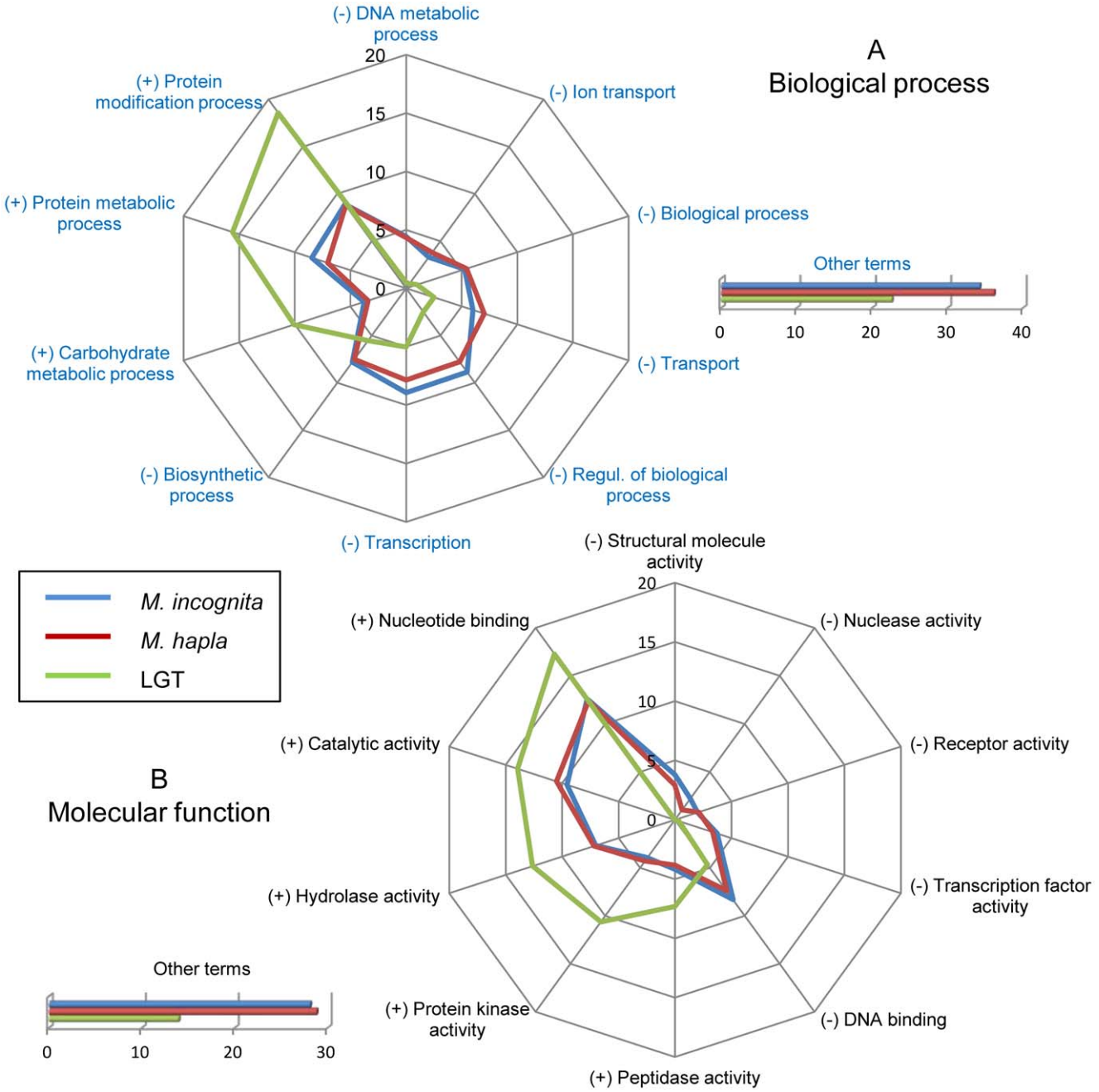



**Figure 3. Functional categories with deviating abundance in the set of proteins acquired via LGT events.** Kiviat diagram representing, the relative abundance of Gene Ontology (G.O.) terms, in percent for the whole *M. incognita* and *M. hapla* proteomes as well as for the 609 non-redundant root-knot nematode proteins originating from lateral gene transfer events. Distribution of G.O. terms in *M. incognita* and *M. hapla* are represented in blue and red, respectively. Distribution of G.O. terms in proteins acquired via LGT is represented in green. (A) Relative abundance of the G.O. terms assigned to root-knot nematode and LGT proteins in the Biological Process category. (B) Relative abundance of the G.O. terms assigned to root-knot nematode and LGT proteins in the Molecular Function category. In the two categories, the ten G.O. terms that presented the most different relative abundance (in percent) in LGT-acquired proteins in comparison to the whole root-knot nematode proteomes are presented.
doi:10.1371/journal.pone.0050875.g003

In the 'molecular function' ontology, we noted an over-representation of the protein kinase (~11% vs. ~4%), hydrolase (~13% vs. ~7%), catalytic (14% vs. 10%) and peptidase (7% vs. 4%) activities in candidate LGT-acquired proteins compared to the whole proteomes (Figure 3B). This reinforces and mirrors the over-representation of proteins involved in carbohydrate/protein degradation and metabolic processes in the "biological process" ontology. Curiously, we also found a slight over-representation of the 'nucleotide binding' term (17% vs 12%) but this was essentially due to the abundance of ATP-dependent protein kinases and peptidases in the LGT set. In contrast, we noted an under-representation of proteins annotated as 'transcription factor' (<0.5% VS ~3%) or 'regulation of biological processes' (3% vs. 8%) indicating that these are not frequent functions of genes acquired via LGT or retained after transfer in these nematodes. In the 'cellular component' ontology, we remark a clear over-

representation of proteins annotated as present in the 'extracellular component'. Almost 30% of annotated proteins in the set of candidate LGT are annotated as such while only ~2% of proteins are predicted to be in the extracellular component in the whole root-knot nematode proteomes (Table S3).

## Genes Acquired via LGT are Prone to Duplications

A previous phylogenetic analysis of genes encoding cell wall-degrading enzymes in plant-parasitic nematodes has shown that several genes underwent duplications after their acquisition via LGT and now form multigene families [8]. The same analysis showed that most of the duplications started before the separation of the different nematode lineages and, at least in root-knot nematodes, gene duplications have continued independently in the genomes of *M. incognita* and *M. hapla* after their separation from a common ancestor. In order to assess whether such a pattern of duplications is frequent after acquisition of a gene via LGT, we analyzed the 141 phylogenies indicative of an LGT. Out of these 141 phylogenies, 92 contain genes both from the *M. incognita* and *M. hapla* genomes, indicating they have been acquired at least in a common ancestor of the two nematodes (Table S4).

Using PhyloPattern (methods), we searched genes that underwent duplications since their acquisition in a common ancestor of the two root-knot nematodes. We found that after their acquisition, 76 of these genes (83%) underwent duplications either in one, both or a common ancestor of the two nematodes. Interestingly, in 79% of cases (60 out of 76), duplications have started before the separation of *M. incognita* and *M. hapla* and thus occurred in a common ancestor of the two species (Table S4). Duplications continued independently after the separation of the two lineages in 43 cases out of 60 (72%). In contrast, in no more than 16 cases, duplications occurred only after the separation of the two lineages in one or both Meloidogyne species. This observation indicates that the vast majority of genes acquired via LGT underwent duplications and most of these duplications (79%) started early in a common ancestor of the species.

We also assessed whether, at a large scale, and regardless phylogenetic support, genes putatively acquired via LGT have a higher tendency for species-specific duplications that the rest of protein-coding genes in the genomes of root-knot nematodes. We analyzed results of the OrthoMCL clustering to determine the number of species-specific duplications or in-paralogs (methods). Out of the 609 non-redundant genes putatively acquired via LGT, 403 were not clustered in any OrthoMCL group. These genes are thus present as single copies specific from *M. incognita* or from *M. hapla*. Out of the 206 remaining LGT genes in OrthoMCL groups, a total of 149 groups contain at least two genes from *M. incognita* or at least two from *M. hapla*. We discarded 26 LGT candidates present as two copies in *M. incognita* while in one single copy in *M. hapla*. Indeed, these copies might result from the genome structure of *M. incognita* and were not considered as having undergone "true" species-specific duplications. Overall, a total of 123 candidate LGT genes were present in at least three copies in *M. incognita* or at least two copies in *M. hapla* (Table S5). Hence, 59.71% of the 206 LGT genes present in OrthoMCL groups have undergone species-specific duplications since the separation of the *M. incognita* and *M. hapla* lineages from their common ancestor. Duplications ranged from multigene families of size 2 to 25 in a single species (Table S5). While some genes underwent duplications both in *M. incognita* and *M. hapla* after their separation from a common ancestor, most gene duplications observed were asymmetric. For instance, LGT gene Minc09058 is present in 25 copies in *M. incognita* and no ortholog was found in *M. hapla* (Table S5). In contrast, LGT gene Minc18743 is present in one single copy in *M. incognita* while it is present in 6 copies in *M. hapla*. Hence, there are no systematic tendencies for a given gene to be equally duplicated and fixed in both root-knot nematode species. To assess whether the proportion of lineage-specific duplication is different for LGT genes than for the remainder of the genes in root-knot nematodes, we calculated the number of in-paralogs in the whole genome of *M. incognita*. In this nematode, 20,359 gene models have been predicted [10] and 15,365 genes are present in 7,647 OrthoMCL groups. As for LGT genes, we discarded groups containing *M. incognita* single-copy genes and those containing two copies in *M. incognita* but a single copy in *M. hapla*. Overall, a total of 2,137 OrthoMCL groups out of 7,647 (27.94%) contain at least three *M. incognita* in-paralogs and represent species-specific duplications. In comparison, in the set of genes acquired via LGT, the proportion of OrthoMCL groups with in-paralogs is more than twice as high. This observation suggests that genes acquired via lateral transfer are prone to duplications that continue independently in different species after their acquisition in a common ancestor.

## Contribution of LGT of Non-metazoan Origin to the Genome of a Root-knot Nematode

In root-knot nematodes, we identified a total of 609 non-redundant genes with no predicted ortholog in 14 other metazoan species and that returned more than 50% non-metazoan hits in blast searches. Because a majority of genes underwent duplications after their acquisition, the estimation of their total abundance in extant genomes has to take duplications into account. Out of the 609 non-redundant genes acquired via LGT, a total of 385 are from *M. incognita* and 202 of these do not cluster in any OrthoMCL group. These genes are thus present in single copy in *M. incognita* and are absent in the 14 other metazoan species compared. In contrast, 183 *M. incognita* genes are in OrthoMCL groups with at least another gene either from *M. incognita* or from *M. hapla*. Considering that *M. incognita* genes present in multiple copies in a same OrthoMCL group underwent duplications after their acquisition via LGT, the total number of genes of non-metazoan origin in *M. incognita* is 680 (202 singletons and 478 gene copies present in 183 groups, Table 2). This represents ~3.34% of the 20,359 protein-coding genes in *M. incognita*.

We also estimated the proportion of genes of non-metazoan origin that, besides OrthoMCL and Blast support, also received phylogenetic support for LGT. Out of the 141 non-redundant phylogenies compatible with LGT in root-knot nematode, 122 contain at least one *M. incognita* gene in the LGT subtree. To account for gene duplications after acquisition via LGT, we counted the total number of *M. incognita* genes per acceptor subtree. Overall, we enumerate a total of 513 or 2.52% of *M. incognita* protein-coding genes with phylogenetic support for acquisition via LGT (Table 2, Table S4).

Hence, we estimate that genes of non-metazoan origin represent between ~2.52% and ~3.34% of protein-coding genes in a root-knot nematode, depending on whether or not phylogenetic support is required.

## Distribution of LGT Candidates Along a Root-knot Nematode Genome

Considering duplications after transfer, in the genome of *M. incognita*, as much as 680 genes are of non-metazoan origin. We analyzed the distribution of these genes on the 2,817 *M. incognita* scaffolds and showed that 38 clusters contain three or more putative LGT genes separated by no more than 50 kb, representing a total of 161 LGT candidates (Table S6). The five largest

**Table 2.** Total number of LGT genes in the genome of *M. incognita*.

| | OrthoMCL+Blast | | Phylogenetic inference | |
|---|---|---|---|---|
| Copies | # MCL Groups | # Genes | # trees | # Genes |
| 1 | 65* | 267** | 34 | 34 |
| 2 | 59 | 118 | 27 | 54 |
| 3 | 23 | 69 | 10 | 30 |
| 4 | 16 | 64 | 9 | 36 |
| 5 | 6 | 30 | 7 | 35 |
| 6 | 2 | 12 | 2 | 12 |
| 7 | 5 | 35 | 3 | 21 |
| 8 | 2 | 16 | 2 | 16 |
| >8 | 5 | 69 | 28 | 275 |
| **Total** | **183** | **680** | **122** | **513** |

*65 groups containing a single *M. incognita* protein and at least one *M. hapla* gene.
**65 genes in OrthoMCL groups + 202 single copies that were not clustered in any OrthoMCL group.
doi:10.1371/journal.pone.0050875.t002

clusters contain 16, 8, 7, 7 and 6 genes putatively acquired via LGT on scaffolds 90, 53, 85, 91, and 69, respectively (Table S6). One of the largest clusters comprises 7 genes encoding pectate lyases of family PL3 [19] on scaffold 53. Interestingly, a similar cluster, consisting of 4 genes encoding PL3s was found in the genome of *M. hapla* [11], suggesting that they derive from a common ancestral cluster that predates the separation of *M. incognita* and *M. hapla* lineages. The five largest clusters all consist of multiple copies of a same or a few different genes. This is attested by membership to the same OrthoMCL groups and/or a same Pfam domain annotation. The largest cluster in *M. incognita*, consists of 16 genes present on scaffold 90. This cluster is no exception to the rule and these 16 genes belong to only 6 distinct OrthoMCL groups and 14 have the same predicted Pfam Kinase domain.

The longest cluster containing only different genes putatively acquired by LGT consists of 5 genes on scaffold 154. The five genes have different predicted Pfam domains and none belong to a same OrthoMCL group. However, only two out of the five genes have a phylogenetic support for possible LGT. Hence, overall, the vast majority of clusters of candidate LGT genes consist of copies of a same or a few distinct genes not of aggregation of multiple independently-acquired genes from distinct families.

## Putative Donors for LGT are Mainly Soil Bacteria

From the 141 phylogenetic trees indicative of an LGT event, we reported the ensemble of species present in the putative donor sub-trees. Donor sub-trees consist in monophyletic groups composed exclusively of non-metazoan species and holding the closest outgroup position relative to the plant-parasitic nematode receiver group.

Bacteria were present in donor sub-trees in 72 out of the 141 phylogenetic trees and represented the most frequent taxonomic division. In 57 cases the donor sub-tree contained only bacteria while in 15 cases, species from other kingdoms were present besides bacteria (Figure 4, Table S7). Interestingly, in many occasions, bacterial species found in the donor clades are soil bacteria, including notorious plant pathogens (e.g. *Ralstonia*

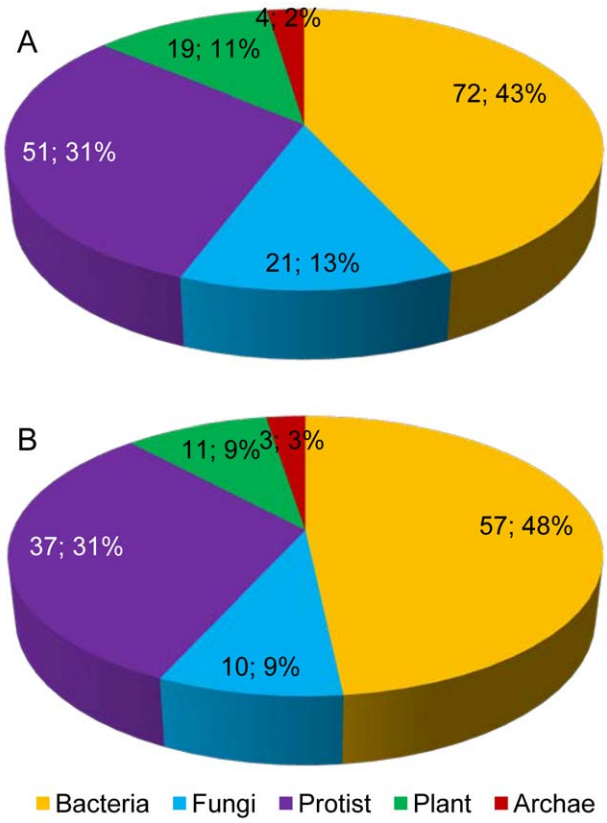



**Figure 4. Taxonomic distribution of candidate donors for LGT events.** (A) Number and proportion of phylogenetic trees that contain at least one species of a given taxonomic division in the closest donor clade. (B) Number and proportion of phylogenetic trees that contain only species of a given taxonomic division in the closest donor clade. The different possible taxonomic divisions are Bacteria, Fungi, Protists, Plants and Archaea. Note that Protist is not a monophyletic division and encompasses several distinct eukaryotic lineages.
doi:10.1371/journal.pone.0050875.g004

*solanacearum*, *Xanthomonas oryzae*, *Xanthomonas campestris*, *Pseudomonas syringae*) plant symbionts (e.g. *Sinorhizobium meliloti*, *Methylobacterium nodulans*, *Mesorhizobium loti*), or more generally species known to dwell in the rhizosphere, the region of soil surrounding plant roots (e.g. *Burkholderia ambifaria*, *Agrobacterium radiobacter*, *Flavobacterium johnsoniae*).

The next most frequent category in the donor sub-trees is protist, a non-monophyletic group of mostly unicellular eukaryotic microorganisms. In 51 out of the 141 phylogenetic trees, protists were present in the donor sub-tree. In 37 cases protists were present alone whereas in 14 cases they were associated with species from other kingdoms. Ciliophora, amoebozoa, apicomplexa and oomycetes were the most frequently present protists. Interestingly, while Oomycetes were found in donor sub-trees in five distinct cases, this is systematically due to one single species, *Phytophtora infestans*, a known plant-pathogen. Similarly, some of the Amoebozoa found in donor sub-trees are known soil-dwelling slime molds (e.g. *Polysphondilium palladium*, *Dictyostelium discoideum)*. For Ciliophora, in contrast, most retrieved species dwell in freshwater although some of them have a more ubiquitous habitat, including soil. Finally, in the apicomplexa category, mainly animal parasites are found (e.g. *Babesia bovis*, *Toxoplasma gondii*). However, any conclusion on the habitat of these species must be taken with

caution because the biodiversity in protists has been so far much less extensively sampled than those of bacteria and fungi.

Fungi were identified in the donor clades in a total of 21 out of 141 trees and represent in fact the second most frequent monophyletic category. In 10 cases, only fungi are found in the donor clade while in 11 cases, fungi are found associated with species from different kingdoms. Among the fungal species identified in the donor clades, many notorious plant pathogens were found (e.g. *Magnaporthe grisea*, *Sclerotinia sclerotiorum*, *Botrytis cinerea*) as well as soil-dwelling fungi (e.g. *Chaetomium globosum*, *Aspergillus fumigatus*).

Plants were also identified as potential donors in 19 phylogenetic trees. In 11 cases the donor sub-tree contained only plants while association with species from other kingdoms was observed in 8 cases. Plant species identified ranged from the unicellular green alga *Chlamydomonas reinhardtii* that possesses a ubiquitous habitat, including soil, to plant that are compatible hosts of root-knot nematodes (e.g. *Oryza sativa*, *Nicotiana attenuata*).

Finally, we also identified Archaea in donor clades of 4 phylogenetic trees and 3 of these clades contained only Archae. All archaeal species identified were hyperthermophilic and thus do not share an evident common habitat with root-knot nematodes. Though, once again the sampled biodiversity in Archea is not deep enough to allow concluding on the habitat of the potential donors.

## Several LGT Genes Crucial for Parasitism have Homologs on Bacterial Plasmids

Mobile genetic elements such as plasmids or bacteriophages are commonly involved in LGT events between bacteria. Because bacteria have been frequently found in the donor sub-trees in our analysis, it is interesting to search whether some of the genes putatively acquired via LGT in nematodes are present on microbial mobile genetic elements. Out of the 609 non-redundant candidate LGT genes in root-knot nematodes, a total of 146 returned blast e-values $<0.001$ with proteins present on microbial mobile genetic elements (methods). Overall, 117 of these protein-coding genes (80%) were found on known bacterial plasmids while 19 were found on prophages and 10 on bacterial viruses (Table S8A). We further focused our analysis on 32 root-knot nematode proteins that aligned with at least 30% identity on at least half of their length with proteins present on microbial mobile elements (Table S8B).

Out of these 32 proteins, 28 are plasmid-borne in candidate donor bacteria. Interestingly, among the plasmid-borne proteins, 9 are present in the list of 15 previously reported clear cases of LGT from the literature (Table S8B, Table 1). These 9 proteins encompass a Chorismate mutase, thought to be involved in modulation of plant defense, aVB5 pantothenate, aVB7 biotin, and a GH32 candidate invertase all probably involved in nutrient processing, a NodL suspected to play a role in the establishment of the root-knot nematodes feeding structure, two GH30 xylanases involved in degradation of plant polysaccharides [7] and a candidate phosphoribosyltransferase as well as a candidate L-threonine aldolase both of as yet unknown function in nematodes [9]. Hence, several genes that, according to the literature, play important functions in plant-parasitism have homologs borne by bacterial mobile elements.

For the 32 genes putatively transferred via plasmid or prophage vectors, a total of 21 different potential donor bacteria are found. Interestingly and in line with the previous section on potential donors, 16 of these bacterial species are known to dwell in the rhizosphere (e.g. *Ralstonia or Rhizobium*), the same habitat than root-knot nematodes (Table S8B).

## The Density of Transposable Elements is Higher in the Vicinity of Genes Acquired via LGT

Transposable elements (TEs) can jump from one position to another in a genome as well as between the genomes of different species across the kingdom boundaries and are even known to mediate transfer of genes within a species genome through a hitchhiking-like process [20]. Previous report of LGT to metazoan receiver species, in the bdelloid rotifer *Adineta vaga* [21] and in the necromenic nematode *Pristionchus pacificus* [22] have both shown a genomic environment rich in transposable elements (TEs) around genes acquired via LGT. To evaluate the density of TEs around LGT genes in root-knot nematodes, we counted the number of TEs present in genomic windows of size 200, 500, 1,000 and 2,000 bp around genes acquired via LGT and around the rest of protein-coding genes in the *M. incognita* genome (methods). We found that the density of TEs was significantly higher around genes acquired via LGT for all genomic windows of size $\geq$500 (Table 3, Figure 5). Hence, similarly to *A. vaga* and *P. pacificus*, the genomic environment around LGT genes is rich in TEs in root-knot nematodes.

In the bdelloid rotifer *Adineta vaga*, genes acquired via LGT were more frequently found in telomeric regions rich in TEs [21]. We checked whether in the genome of *M. incognita*, transferred genes had a tendency to accumulate at the tips of scaffolds. Although with 2,817 scaffolds we are far from the estimated 30–40 chromosomes in *M. incognita*, we observed no sign for grouping of candidate LGT genes at scaffold ends.

## GC Content and Codon Usage

Because genes acquired via LGT originate from species that can feature codon usage and GC content markedly different from those of the receiver species, these genes might have kept characteristics of their genome of origin. For instance, in the necromenic nematode *P. pacificus*, genes acquired via LGT from insects have kept codon usage more closely related to those of insect donors than that of the "endogenous" *P. pacificus* genes [22].

We thus compared the codon usage and GC content of genes acquired via LGT in *M. incognita* to the rest of protein-coding genes (methods). With an average GC content of 31.4%, the *M. incognita* whole genome is globally GC poor. We measured an average GC content for a protein-coding gene in *M. incognita* of 36.21% (excluding genes acquired via LGT). By comparison, the average GC content for the 680 genes acquired via LGT is 36.47% (Figure 6A). We also generated codon usage tables for genes acquired via LGT and for the rest of the *M. incognita* gene set (Figure 6B, Table S9). The two codon usage tables were very similar with an average difference in frequency of codon usage for a given encoded amino-acid of 0.02 (2%). Only two sets of codons differed by more than 5% in frequency, Cystein codons and STOP codons.

Hence, genes acquired via LGT in *M. incognita* cannot be differentiated from the rest of protein-coding genes based on their GC content or codon usage.

# Discussion

Our analysis represents the first comprehensive pan-genomic search for LGT events in nematodes with phylogenetic validation. Previous reports have shown that different genes acquired by LGT in plant-parasitic nematodes play important roles in the parasitism process and hence, have been functionally significant [7]. Our systematic search for LGT events of non-metazoan origin overall confirms their link to functions important for parasitism but also provides the first estimate of the total contribution of LGT to the making of a metazoan animal genome.

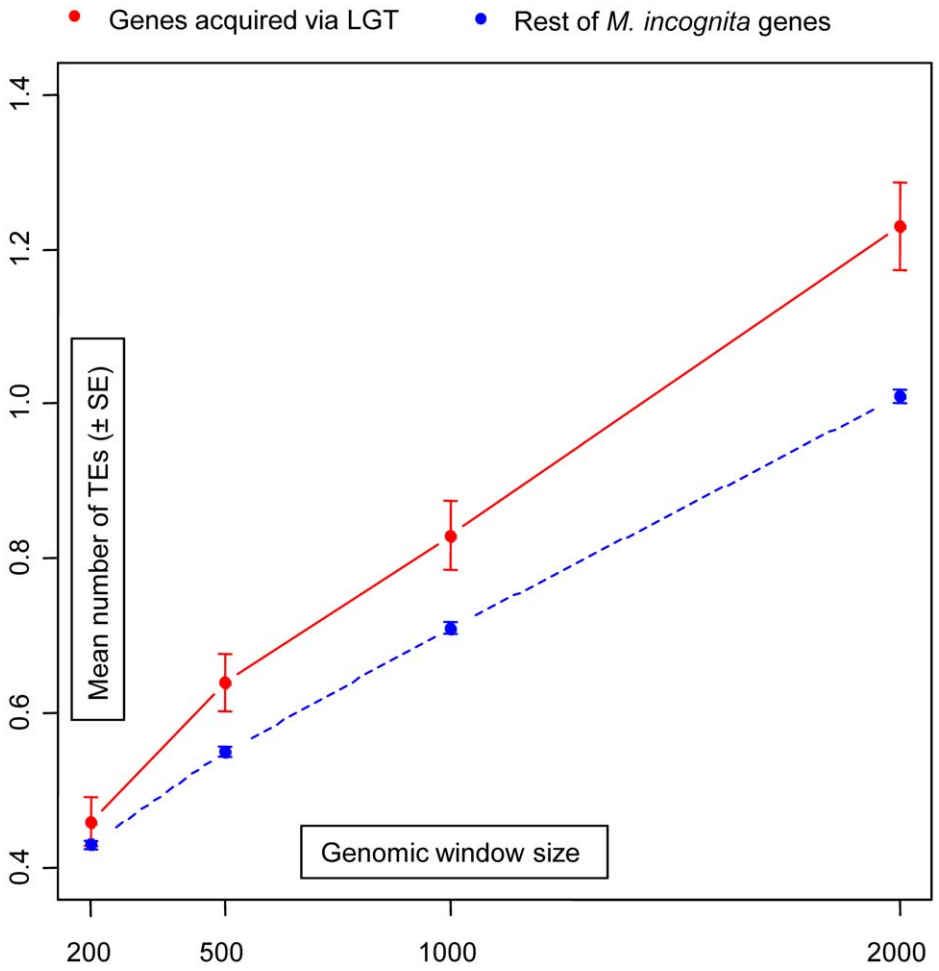



**Figure 5. Density of transposable elements around genes acquired via LGT in *M. incognita*.** Plot of the mean number of transposable elements (± standard error) in genomic windows of size 200, 500, 1,000 and 2,000 bp around genes acquired via LGT (red) and around the rest of protein-coding genes (blue) in the genome of *M. incognita*.
doi:10.1371/journal.pone.0050875.g005

## Contribution of LGT to the Genomes of Root-knot Nematodes

In a single root-knot nematode species, *M. incognita*, we have estimated that between 513 and 680 genes are of non-metazoan origin, depending on whether or not phylogenetic support is required. This represents between ~2.52% and ~3.34% of the protein-coding genes in this nematode, a substantial proportion. For comparison, this is more than the whole proportion of genes encoding carbohydrate-active enzymes (420 genes, 2.06%) or peptidases (334 genes, 1.64%) in the *M. incognita* genome [10].

**Table 3.** Density of transposable elements around genes acquired via LGT and around the rest of protein-coding genes in *M. incognita*.

| Window size | Gene type | Mean # of TE in window | Std error | $X^2$ (1 df)* | p-value |
|---|---|---|---|---|---|
| 200 | LGT | 0.46 | 0.032 | 1.05 | 0.306 |
| | Other | 0.43 | 0.006 | | |
| 500 | LGT | 0.64 | 0.037 | 5.96 | 0.015 |
| | Other | 0.55 | 0.007 | | |
| 1,000 | LGT | 0.83 | 0.044 | 8.07 | 0.005 |
| | Other | 0.71 | 0.008 | | |
| 2,000 | LGT | 1.23 | 0.057 | 17.27 | <0.001 |
| | Other | 1.01 | 0.009 | | |

*result of the chi-square test with one degree of freedom.
doi:10.1371/journal.pone.0050875.t003

A – GC content

| Type | Mean GC% | SE |
|---|---|---|
| LGT | 36.47 | +/- 0.06 |
| Rest | 36.21 | +/- 0.06 |

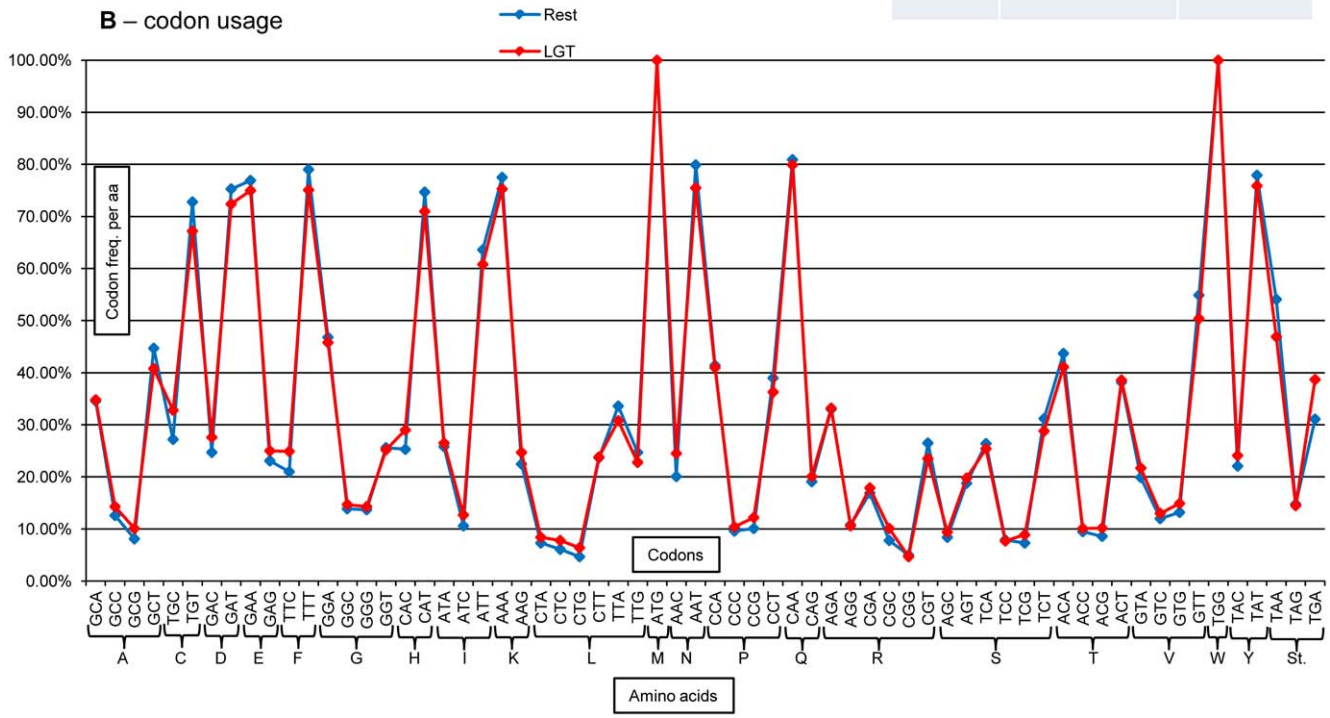



**Figure 6. GC percent and codon usage of LGT genes in *M. incognita*.** (A) Average and standard error (SE) percent of GC nucleotides in genes acquired via lateral gene transfer (LGT) and in the rest of protein-coding genes in *M. incognita* (Rest). (B) Comparison of the relative percent of codon usage per coded amino acids between genes acquired via lateral gene transfer (LGT) and the rest of protein-coding genes (Rest). X-axis: the 64 different codon and their associated 20 amino-acids and Stop. Y-axis: the percentage of usage of a given codon for a given encoded amino-acid. doi:10.1371/journal.pone.0050875.g006

To date, a whole genome inventory of genes acquired via LGT in a nematode has been established in only one other species, the plant-pathogenic agent of pine wilt disease, *Bursaphelenchus xylophilus*. Although no extensive functional analysis nor information of their genomic distribution and environment have been performed, this analysis revealed that between 24 and 223 genes were possibly acquired via LGT in this nematode, depending on whether phylogenetic or BLAST support were taken into account [23]. Representing between ~0.13% and ~1.25% of the 18,074 predicted protein-coding genes, the proportion in *B. xylophilus* is lower than in *M. incognita* but remains significant. We can expect that exploration of forthcoming genomes of other nematodes, including plant and animal parasites will bring new evidences for the importance of LGT in the making of a metazoan genome.

Although already representing a substantial proportion of protein-coding genes, the set of predicted LGT-acquired genes we present in our analysis is probably a minimal estimate of the actual number of such acquired genes. Indeed, we used stringent criteria in order to eliminate as much as possible false positives but necessarily may have missed some true positives. For instance, a gene that would have been transferred once in plant-parasitic nematodes and once in another metazoan would be either eliminated by our OrthoMCL filters or yield a phylogenetic tree not compatible with the searched LGT PhyloPattern. Similarly, we focused our analysis on transfers from non-metazoan donors to nematode receivers. Thus, we do not consider here transfers of metazoan origin in these nematodes. As unlikely as such event might appear, it was recently proposed that several genes in the necromenic nematode *P. pacificus* have been transferred from insect donors via LGT [22]. The total number of genes that do not originate from a common ancestor via vertical inheritance is thus probably even higher than currently estimated in plant-parasitic nematodes.

In this analysis, we have considered that genes have been transferred from non-metazoan donor to metazoan root-knot nematodes. Although it can be hypothesized that transfers may have occurred in the other direction (metazoan to non-metazoan), this appears substantially less likely for the following reasons. First, except root-knot nematodes and their close plant-parasitic relatives, no other nematode possess these genes while in the donor clades the diversity of represented phyla is generally high. Hence, this would require that the genes have been "invented" in root-knot nematodes then transferred to multiple non-metazoan species, independently, an unlikely hypothesis. Second, LGT genes all possess one or more spliceosomal introns and in many cases, closest relatives are bacterial. It intuitively appears more difficult to transfer an eukaryotic multi-exon gene in a bacterial species than an intronless bacterial gene in an eukaryotic genome followed by intron gains. Third, at least in the 52 phylogenetic trees that present a "duplication" node B, the direction of the transfer is explicitly from non-metazoan to metazoa because in these cases root-knot nematode genes are nested inside clades of non-

metazoan species. In contrast, transfers in the other direction would require non-metazoan genes to be nested within plant-parasitic nematode clades.

## Contribution of LGT to the Biology of Root-knot Nematodes

Root-knot nematodes possess a spear-like structure named "stylet", connected to esophageal and dorsal gland cells. Proteins secreted through this stylet in plant tissue are called effectors and play essential roles in the parasitic interaction [24]. Most gene products of LGT cases reported so far in root-knot nematodes are effectors. Our whole genome scan for LGT events allowed retrieving all the 15 distinct cases of LGT reported so far in the literature (Table 1). The effectors originating from these LGT events are involved in processes such as modulation of plant defense, establishment of a feeding structure, nutrient processing or degradation of the plant cell wall [7]. Besides these 15 known cases, all the other genes we detected as acquired via LGT had not been described previously in the literature and the possible function they play in root-knot nematodes has been so far unexplored. Search for Pfam protein domains and deduction of Gene Ontology terms allowed retrieving information on the putative functions for approximately 40% of the 609 non-redundant genes of non-metazoan origin. Compared to the rest of protein-coding genes in root-knot nematodes, we observed an over-representation of proteins putatively involved in carbohydrate metabolism, in protein metabolism and modification as well as in hydrolase, catalytic or peptidase activities. Functions related to carbohydrate and protein metabolism, catalytic and hydrolase activities point to degradation of the plant cell wall and nutrient processing. A possible role in detoxification can be proposed for sequences annotated with the "protein modification" term. We also remarked that proteins predicted to be secreted or in the extracellular compartment were more frequent in the LGT set.

This ensemble of observations suggests that genes acquired via LGT mainly encode proteins secreted by the nematode and involved in degradation of carbohydrates and proteins, in nutrient processing, metabolism and detoxification processes. Overall, these predicted functions make sense considering the plant-parasitic lifestyle of the root-knot nematodes and echoes the functional roles assigned to the previously reported cases of LGT in these species.

In contrast, we remarked that functions related to regulation of biological processes or transcription were under-represented in the set of genes acquired via LGT. Although there is no *a prior*' reason for a functional category to be more prone to LGT than another, there are reasons for a category to be more easily fixed at the population level and then at a species level. Indeed, genes that provide a selective advantage through transfer of a new capability may have benefited from positive selection and consequently accelerated fixation. For a gene to be positively selected, it must first be functional in the receiver species. Genes involved in core basic biochemical functions, such as degradation of a carbohydrate or a protein, intuitively appear more likely to be able to perform the same exact function in a distant receiver organism than genes involved in processes such as fine regulation of gene expression or of biological processes in multicellular organisms.

## Duplications after Acquisition via LGT

Our analysis revealed a high propensity for duplications after acquisition of genes via lateral transfer. The majority of duplications we observed started before the separation of the *M. incognita* and *M. hapla* lineages and continued independently in the two root-knot nematode genomes. We had previously observed a comparable tendency for at least three gene families encoding plant cell wall-degrading enzymes in these nematodes [8]. Our whole genome analysis suggests the phenomenon is more general after an LGT event. Interestingly, importance of duplications, including some that started before the separation of the different nematode lineages analyzed and that continued independently after their separation has also been reported in necromenic nematodes of the *Pristionchus* genus [25]. The presence of LGT genes as multigene families suggests that positive selective pressure may have favored individuals with multiple copies of laterally-acquired genes. Observation that duplications constitute an adaptive mechanism to overcome a change or a stress in the environment at the time scale of a few generations has lead to propose an evolutionary model of adaptive radiation for the origin of new gene functions [26]. Under this model, duplication of sub-optimal genes may allow emergence of new gene variants with more optimal or divergent function through neo-functionalization [26]. Duplications can also allow specialization and partition of function through sub-functionalization. Interestingly, it has recently been shown that spontaneous gene duplications occur at a much higher rate than point mutations in *C. elegans*, suggesting that early adaptive genomic changes could be supported more by advantageous duplications than by beneficial mutations, in particular when dosage is under selective pressure [27]. In the case of an LGT event of non-metazoan origin, it is likely that the transferred gene was initially poorly adapted to the nematode genome's GC content, codon usage and regulatory elements. Consequently, individuals with multiple copies of the transferred genes probably presented a higher probability for the emergence of a fully functional gene or for expression of the gene product at a sufficient level. Similarly, presence of the acquired gene in multiple copies may have favored the emergence of new functions. For this ensemble of reasons, individuals harboring multiple copies of genes acquired via LGT may have been positively selected generation after generation leading to the eventual presence of multigene families in root-knot nematode genomes after transfer of a single gene.

Overall, the most common fate of gene copies after duplication is their loss, often via pseudogenization [28]. In teleost fishes, it has been estimated that only 15% of gene copies were maintained in a functional form after whole genome duplication [29]. It is therefore potentially interesting to search pseudogenes in multi-gene families acquired via LGT. A recent analysis has allowed identification of candidate pseudogenes encoding altered cellulases acquired via LGT in the peanut pod nematode *Ditylenchus africanus* but no significant traces of pseudo-cellulases were found in the genomes of *M. incognita* and *M. hapla* [30]. This apparent absence of pseudogenes suggests that most genes present in multiple copies in extant Meloidogyne genomes have initially arisen from old duplications and have been fixed early; the possible old pseudogenes probably accumulated too much mutations since their initial decay to be differentiated from the intergenic DNA in present genomes.

## The Nature of Putative Donors and Possible Transfer Vectors

Although identification of the exact donor in an LGT event appears extremely challenging, analysis of the phylogenetic trees provides information on the nature of potential donors at the kingdom and phylum levels. Overall, we remarked that bacteria constituted the most frequent group of potential donors and many bacteria from the rhizosphere, including plant-pathogens and plant-symbionts were present in these donor clades. Similarly, a number of plant-pathogenic fungi were also identified in the

donor clades. Species present in the rhizosphere share the same environment than root-knot-nematode and thus appear as interesting candidate donors, particularly because several of the genes known to have been acquired via LGT in these nematodes are involved in parasitic interactions with the plant.

Even if these species share a common environment and similar lifestyles, in association with plants, the mechanism of transfer itself remains elusive. In the case of plant cell wall-degrading enzymes, acquisition through feeding on plant-associated bacteria was favored over hypothesis of acquisition from bacterial endosymbionts. Indeed, no homologs of the genes acquired via LGT had been identified in known nematode endosymbiont while many are found on bacteria that are plant-parasites or plant-symbionts [31].

Regardless the origin, different mechanisms of transfer, including bacterial secretion systems or possible intermediates such as viruses, transposable elements or plasmids have been evoked [32]. Our whole genome analysis showed that the genomic environment around LGT genes is rich in transposable elements and that several bacterial homologs of the transferred genes are plasmid-borne.

It is tempting to propose that plasmids may have served as vectors for the transmission of genes of bacterial origin in the genome of the root-knot nematodes. The genes putatively transferred via plasmids include some previously characterized as important for parasitism processes, such as GH30 xylanases involved in the degradation of the plant cell wall or chorismate mutases involved in plant defense modulation. Plasmids are already known to support LGT and acquisition of new capabilities such as antibiotic resistance between bacteria. Our findings suggest that plasmids may have also played a significant role in the transfer of bacterial genes of functional importance to the genomes of root-knot nematode. The mechanism of transfer from bacterial plasmids to the genome of nematodes may be similar to the transfer of genes via Agrobacterium Ti plasmid to nuclear genome of plant cells. As for any case of putative lateral gene transfer, it is important to ascertain that genes do not result from a contamination. This hypothesis can be ruled out in our case for several reasons. (i) the 32 genes that present significant similarities with plasmid-borne bacterial genes have between one and 13 spliceosomal introns. (ii) the sequence similarities between these nematode genes and their bacterial plasmid counterparts range at best between 30–40% identity, far from the level expected for a contamination. (iii) these genes are assembled in the root-knot nematode genomes in the vicinity of true nematode genes presenting significant similarities with genes in *C. elegans*.

Similarly to reported cases of LGT in two other animals, the bdelloid rotifer *Adineta vaga* and the necromenic nematode *Pristionchus pacificus*, we observed a preferential distribution of LGT genes in regions rich in transposable elements (TEs) in root-knot nematode genomes. Thus, a possible role of TEs as hitchhiking vector in the mechanism of transfer can be hypothesized. Supporting this possibility, it has been shown that a DNA transposon has undergone repeated lateral transfers in different tetrapod species, including human [33]. An alternative hypothesis is that TEs do not play any role in the mechanism of transfer but that some regions in root-knot nematode genomes are more tolerant to both the accumulation of TEs and integration of genes of foreign origin.

### LGT Events have been Multiple and Probably Ancient

Analyses of our phylogenetic tree topologies have indicated that multiple different species and kingdoms were positioned in donor clades, suggesting that there is not a single or low number of donors but a multitude of possible species. Consistent with these observations we did not identify clear genomic clusters of different genes acquired via LGT. All the biggest LGT genomic clusters consist of repeats of a same or a few genes that underwent duplications after their acquisition via LGT. Overall, these features suggest that the clusters of LGT genes observed in the *M. incognita* genome result from multiple cis-duplications and not from "en-bloc" co-transfers from a same donor or from multiple independent transfers in a hotspot of integration of foreign genes. Previous observations on genes acquired via LGT and involved in degradation of the plant cell wall had also shown no evidence for clustering of different gene families in a same genomic region in *M. incognita* [8].

Overall, we remarked that genes acquired via LGT showed GC content and codon usage very similar to those of the other protein coding genes in *M. incognita* despite putative origins in a multitude of evolutionary distant donors. Based on these characteristics, they cannot be distinguished from typical endogenous root-knot nematode genes. This suggests that, in general, transfer events have been sufficiently ancient to have allowed adaptation to the codon usage and GC content of a typical *M. incognita* protein-coding gene. An alternative hypothesis is that the only gene transfers that have been successful are those that involve donor genes featuring GC content and codon usage similar to those of the receiver species. However, given the multitude and diversity of putative donors, this hypothesis appears unlikely. Furthermore, in the necromenic nematode *P. pacificus*, genes acquired via LGT of insect origin showed a codon usage more similar to those of insect donors than to those of the other nematode genes [22]. This indicates that there is no prerequisite in terms of similarity in codon usage or GC content for an LGT to occur.

## Conclusions

Overall, our root-knot nematode pan-genomic analysis shows that, even if LGT events are not as prevalent as in prokaryotes, they also have significantly contributed both to the genome composition and biology in these metazoan animals. Representing up to 3.34% of protein-coding genes, predicted and known functions of genes acquired via LGT indicate a clear link with different processes crucial for plant-parasitism. Hence, LGT events have probably played an important role in the emergence of this capability in nematodes. Further comprehensive whole genome search for LGT events in other metazoan species will probably allow assessing whether evolutionary and biological importance of LGT is a specificity of nematodes or whether the phenomenon is more general in metazoan species.

# Methods

## Determination of Groups of Orthologous Metazoan Proteins

The whole sets of predicted proteins from the root-knot nematodes *Meloidogyne incognita* and *Meloidogyne hapla* were compared to those of 14 other metazoan species using the OrthoMCL [12] software with default parameters. The 14 other metazoan species compared are *Branchiostoma floridae*, *Brugia malayi*, *Bombyx mori*, *Caenorhabditis briggsae*, *Caenorhabditis elegans*, *Ciona intestinalis*, *Drosophila melanogaster*, *Homo sapiens*, *Mus musculus*, *Nematostella vectensis*, *Pristionchus pacificus*, *Strongylocentrotus purpuratus*, *Trichoplax adhaerens* and *Tribolium castaneum* (Figure S1, Table S1). Prior to OrthoMCL comparisons, redundancy was eliminated in each metazoan proteome using the program CD-HIT [34] set to keep only one representative protein (the longest) in clusters of 100% identical proteins. All *Meloidogyne* proteins that clustered with at least another metazoan species in OrthoMCL groups were

discarded from the analysis. In root-knot nematode-restricted OrthoMCL groups, only one representative protein sequence per group was kept. The longest *M. incognita* protein was used as reference in all root-knot nematode-restricted OrthoMCL groups except those that contained no *M. incognita* protein but at least one *M. hapla* protein. In these cases, the longest protein from *M. hapla* was selected as a reference.

## BLAST Filtering for Identification of Candidate Lateral Gene Transfers

All predicted proteins that passed the OrthoMCL filter were used as queries for a BLASTp search against a custom database that consisted in the NCBI's non-redundant (nr) library completed by the whole proteomes of *M. incognita* and *M. hapla*. An e-value cut-off of 0.01 and an alignment covering at least 30% of the query length were required. All proteins that returned at least 50% non-metazoan hits among their ten best blast hits were considered as putatively acquired via lateral gene transfer. The NCBI's tree of life and Taxonomy IDs associated to protein sequences were used as reference for the taxonomy. To avoid eliminating LGT events that occurred before the separation of different plant-parasitic nematode species or that gave rise to multigene families in these nematodes, BLAST hits that returned Taxonomy IDs corresponding to the three lineage containing plant-parasitic nematodes (Tylenchida: Taxonomy ID: 6300, Triplonchida: Taxonomy ID: 211184 and Dorylaimina: Taxonomy ID: 211225), were not considered in the count of metazoan hits.

## Check for Possible Contaminations

As a gene resulting from bacterial contamination would yield a BLAST result pattern exactly identical to one of a true case of LGT, we further searched and eliminated among the proteins that passed the BLAST filter, those that presented more than 80% sequence identity on more than half of their length (query) with non-metazoan genes.

## Phylogenetic Analyses and Detection of Topologies Compatible with Lateral Gene Transfers

Each *Meloidogyne* protein that passed both the OrthoMCL and the BLAST filters were sent to automated phylogenetic analysis using the FIGENIX [14,15] platform. Phylogenetic analyses performed with FIGENIX use three different reconstruction methods (neighbor joining, maximum parsimony and maximum likelihood) with bootstrap replications to provide a fusion-tree with support values. Tree topologies corresponding to a potential LGT event were automatically searched among the ensemble of produced phylogenetic trees, using the program PhyloPattern [16]. The pattern searched consisted in the presence of at least a node 'A' partitioning the tree in two sub-clades, one monophyletic clade containing only *M. incognita* and/or *M. hapla* and possibly other plant-parasitic nematodes and another distinct clade containing only non-eumetazoan species (any NCBI's taxid but none descending from 6072, Eumetazoa). Strong phylogenetic support was assigned when an additional node 'B' considered as a duplication node and branching to external species was found (Figure 2).

## Functional Annotation of Candidate LGT-acquired Genes

The whole proteomes of *M. incognita* and *M. hapla* were scanned against the Pfam [35] database of HMM protein domains using the PfamScan perl script and the HMMER package. Every root-knot nematode protein sequence was compared to the Pfam-A library (ver. 24.0) of manually curated HMMs using default parameters. Using the Pfam2GO association file, gene ontology terms were assigned to proteins on the basis of their Pfam domain composition. Using the map2slim perl script from the go-perl module, we mapped the initially assigned GO terms to their parent terms in the generic GO-slim ontology. This allowed direct comparison of GO terms at a same granularity level between the different proteins from the two root-knot nematode proteomes, including those originating from LGT.

## Duplications after Transfer Estimated by Phylogenetic Patterns

The number of genes that underwent duplications after transfer was estimated by searching all trees that contained at least two genes from *M. hapla* or at least three genes from *M. incognita* (to correct effects due to the *M. incognita* genome structure in two copies) in the node A subtree corresponding to an LGT event. To detect gene duplications that occurred before the separation of the two lineages, we used PhyloPattern and searched trees that contained both *M. incognita* and *M. hapla* sequences and duplication nodes prior to speciation nodes separating the two species. To detect gene duplications that occurred after the separation of *M. incognita* and *M. hapla* lineages, we searched node 'A' subtrees that contained duplication nodes at the base of *M. incognita*-restricted or *M. hapla*-restricted monophyletic groups, using PhyloPattern.

## Lineage-specific Duplications after Transfer Estimated from OrthoMCL Data

We detected in-paralogs or lineage-specific duplications in *M. incognita* and *M. hapla* whole genomes and LGT-acquired genes based on the OrthoMCL analysis conducted with 16 metazoan species. Ever since a gene was present in an OrthoMCL group that contained at least two genes from *M. incognita* or at least two genes from *M. hapla*, the gene was considered as having underwent lineage-specific duplication after the separation of the two nematode species from their last common ancestor. Because a substantial proportion of the genome of *M. incognita* is present in two copies compared to the genome of *M. hapla*, OrthoMCL groups containing two genes from *M. incognita* and a single gene from *M. hapla* were not considered as duplicated.

## Distributions of Transposable Elements and Genes Acquired via LGT

The positions of all gene models on the *M. incognita* scaffolds, including those acquired via LGT, as well as the positions of transposable elements, were retrieved from the GFF files generated at the occasion of the initial annotation of the genome [10]. Annotation of transposable elements (TEs) in the *M. incognita* genome was performed using the REPET pipeline [36]. Using the genome GFF files, we counted the number of TEs in windows of 200, 500, 1,000 and 2,000 bp, flanking LGT genes on the one hand, and flanking the rest of protein-coding genes, on the other hand. $Chi^2$ tests were used to compare the distribution of TEs density around LGT genes and the rest of protein-coding genes for the four genomic window sizes.

## Clusters of Genes Acquired via LGT on *M. incognita* Scaffolds

In a same OrthoMCL group, in case several genes from *M. incognita* and/or *M. hapla* were present, we only kept one root-knot nematode gene as representative of the group. Because we were interested in the distribution and enumeration of all potentially laterally acquired genes in the genome of *M. incognita*, we had to take into account the different in-paralogous copies. We assumed

that if a representative gene was predicted as potentially resulting from LGT, all the in-paralogous copies resulting from species-specific duplications were equally likely to have been acquired via LGT. We mapped a total of 661 genes, including in-paralogs on the 2,817 *M. incognita* scaffolds. Information about the position of the genes on the different scaffolds was extracted from GFF files generated during the initial annotation of the *M. incognita* genome [10]. We extracted all clusters consisting of at least 3 genes potentially acquired via LGT (including in-paralogs) and distant of less than 50 kb on a same scaffold.

### GC Content and Codon Usage

To measure the GC content and codon usage of protein-coding genes in *M. incognita*, we extracted the corresponding CDS sequences from the GFF files generated at the occasion of the initial annotation of the genome [10]. We used the program geecee from the EMBOSS software package [37] to calculate the GC content of every CDS from predicted LGT gene as well as for the rest of protein-coding genes. We generated codon usage tables for LGT genes and for the rest of protein-coding genes using the cusp program from the EMBOSS package. We then used the codcmp program from EMBOSS to compare codon usage in LGT genes and in the rest of protein-coding genes.

### Mobile Genetic Elements

The 609 non-redundant protein sequences corresponding to genes putatively acquired via lateral gene transfer were compared to the set of proteins present on mobile genetic elements in the ACLAME database [38]. The protein fasta sequences present on known bacterial plasmids, prophages and phages were downloaded from the ACLAME web site (version 0.4, http://aclame.ulb.ac.be). The fasta file containing 122,154 proteins was formatted as a Blast database. The 609 root-knot nematode LGT proteins were searched using BlastP against this library using an e-value cut-off of 0.001. We further filtered nematode proteins that aligned with at least 30% identity on at least 50% of their length and of 50% of the subject length with proteins from the ACLAME database. These proteins were considered as having significant similarities with proteins present on bacterial mobile genetic elements.

## Supporting Information

**Figure S1 Schematic phylogeny of the 16 metazoan species compared at the proteome level.** This tree represent the relative phylogenetic position of the 16 metazoan species compared, including the two root-knot nematodes, *Meloidogyne incognita* and *Meloidogyne hapla*. Besides these two root-knot nematodes, species compared comprise other nematodes (in green), insects (in blue), chordates (in red), urochordates (in orange), cnidaria (in violet) and placozoa (in dark red). Names of the main phylogenetic divisions are given at the corresponding nodes.
(TIF)

**Table S1 List, source and number of proteins in the 16 metazoan species compared.** Names of the 16 metazoan species compared are indicated in the first column, followed by the taxonomic group, the source and version of the proteome retrieved, the number of predicted proteins as well as the number of unique proteins after elimination of redundancy with CD-HIT [34].
(DOC)

**Table S2 Detailed list of the 609 non-redundant root-knot nematode proteins considered as acquired via LGT.** For each protein predicted to have been acquired via LGT of non-metazoan origin, a series of information is listed. Accession numbers, presence of a signal peptide (Y), the number of spliceosomal exons are indicated in columns 1, 2 and 3, respectively. The presence of a Pfam protein domain is indicated in columns 4. Gene Ontology (G.O.) terms associated to the protein based on the Pfam domains composition for the three categories 'Biological Process', 'Molecular Function' and 'Cellular Component', are given in columns 5, 7 and 9, respectively. The G.O. slim terms that were the most over-represented in the set of genes acquired via LGT for the 'Biological Process' and 'Molecular Function' ontologies are indicated in columns 6 and 8, respectively. The column 10 'family' indicates whether the gene belongs to a known family. If this family corresponds to one previously reported LGT case in the literature (listed in Table 1), the whole line is put in bold. The putative substrate/activity of the protein/enzyme is indicated in column 11. Phylogenetic support (presence of a node 'A' or 'A' and 'B') is given in column 12. If none of the searched phylogenetic patterns were retrieved 'No' is indicated in this column whereas in cases no phylogenetic tree at all could be constructed 'No tree' is indicated. The last three columns indicate, respectively, the list of species in the donor phylogenetic subtree, the simplified taxonomy of the species found, and the short taxonomy (i.e. Bacteria, Fungi, Protist, Plant or Archae). N/A: not applicable.
(XLS)

**Table S3 Gene Ontology terms assigned to the whole proteomes of root-knot nematodes and to those originating from LGT.** For each Gene Ontology (G.O.) category (A) Biological Process, (B) Molecular Function, (C) Cellular Component, the number of occurrence and abundance (in percent) of G.O. terms are given for the whole *M. incognita* and *M. hapla* proteomes as well as for the proteins considered as acquired via LGT in root-knot nematodes. For comparison purpose, the relative abundance of a G.O. term in proteins acquired via LGT compared to the average abundance observed in the two root-knot nematode is indicated in percent in the last column. Relative abundance follows a heat map color code ranging from red gradient for terms more abundant (+) in LGT proteins to green gradients for terms less abundant (–) in LGT proteins. Yellow gradient indicates a similar abundance of the term in LGT proteins compared to the whole root-knot nematode proteomes.
(XLSX)

**Table S4 Phylogenetically-inferred duplications of LGT genes before and after the separation of the two root-knot nematodes.** This table lists for each of the 141 phylogenetic trees indicating an LGT event, the following information. The accession number of the reference Meloidogyne species used as query for phylogenetic reconstruction. The list of species in the receiver clade. The number of gene copies in *M. hapla* and in *M. incognita*. Whether the gene is observed in one or both the two root-knot nematodes. Whether the gene as underwent duplications either before or after the separation of *M. incognita* and *M. hapla*. Whether the gene as underwent duplications before the separation. Whether the gene as underwent duplications after the separation. The last column indicates the phylogenetic support for the LGT event ('A' or 'A' + 'B').
(XLS)

**Table S5 OrthoMCL-based inference of LGT gene duplications after the separation of the two root-knot**

**nematodes.** This table lists the 123 non-redundant proteins clustered in OrthoMCL groups that contain at least three *M. incognita* proteins or at least two *M. hapla* proteins. Accession numbers of the selected reference proteins are given in the first column. The total numbers of genes from M. incognita and M. hapla in the OrthoMCL group are given in columns 2 and 3, respectively. The total number of root-knot nematode proteins in a given OrthoMCL group and whether a phylogenetic support was assigned is indicated in columns 4 and 5, respectively. Accession numbers of all *M. incognita* and *M. hapla* proteins present in a given OrthoMCL group are listed in the last two columns.
(XLSX)

**Table S6 Genomic clusters of genes acquired via LGT in *Meloidogyne incognita*.** This table lists the 161 genes present in 38 genomic clusters composed of at least 3 genes distant of less than 50 kb on a same scaffold. For each gene, the accession number is given, followed that the scaffold and genomic coordinates on the scaffold (columns 1–4). Presence of a Pfam protein domain and grouping in an OrthoMCL group are indicated in columns 5 and 6. In a given genomic cluster, a same OrthoMCL group is represented with a same color. The size (in number of genes) of every genomic cluster is given in the last column.
(XLS)

**Table S7 Putative donors of LGT in root-knot nematodes.** Putative donors for the 141 distinct LGT cases inferred by phylogenetic analysis are listed in this table. Accession numbers of the reference proteins are given in the first column. The nature of the phylogenetic support for LGT is indicated in column 2. Species present in the donor clade are listed in column 3. A simplified taxonomy presenting the clades donor species belong to is given in column 4. Taxonomic division (Bacteria, Fungi, Protist, Plant or Archae) is indicated in the last column.
(XLSX)

**Table S8 LGT proteins with significant hits in bacterial mobile genetic elements.** (A) The list of 146 non-redundant LGT proteins that have a Blast hit $<0.01$ with mobile genetic elements in ACLAME. First column: accession number, second column: number of exons, third column: best blast hit in ACLAME, last column: whether the protein has a highly significant hit (at least 30% identity on at least 50% of the query length). (B) Details on the 32 LGT proteins that have a highly significant hit in ACLAME. Columns 1–3 are identical to (A). Column 4: description of the ACLAME best blast hit, column 5: the species holding the mobile genetic element, column 6: whether a phylogenetic tree support the LGT event, last column: whether the protein belongs to a family previously described as acquired via LGT in the literature.
(XLSX)

**Table S9 Codon usage of LGT genes and of the rest of protein-coding genes in *M. incognita*.** LGT columns show values for genes acquired via LGT whereas "Rest" columns display values for the rest of protein-coding genes. Fraction represents the percent usage of a given codon for a given encoded amino-acid, the sum is thus always = 1 for each amino-acid. Frequency represents the overall average percent usage of a given codon for the whole set of genes (LGT or Rest) the total sum for the 64 codon is = 1. Occurrence: the number of times a codon has been observed in the set of genes acquired via LGT or in the rest of protein-coding genes.
(XLSX)

## Acknowledgments

We thank Florin-Alexandru Bratu, Denis Caromel and Cédric Morandin from INRIA Sophia-Antipolis for help in deploying and running parts of the bioinformatics analyses on the computational ProActive PACA Grid.

## Author Contributions

Conceived and designed the experiments: EGJD JP AC-F MDR PG PP. Performed the experiments: EGJD JP AC-F MDR PG EW. Analyzed the data: EGJD PP JP EW PA. Contributed reagents/materials/analysis tools: PG EW. Wrote the paper: EGJD PP PA.

## References

1. Blaxter M (2007) Symbiont genes in host genomes: fragments with a future? Cell Host Microbe 2: 211–213.
2. Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, et al. (2008) Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug Elysia chlorotica. Proc Natl Acad Sci U S A 105: 17867–17871.
3. Pierce SK, Fang X, Schwartz JA, Jiang X, Zhao W, et al. (2011) Transcriptomic evidence for the expression of horizontally transferred algal nuclear genes in the photosynthetic sea slug, Elysia chlorotica. Mol Biol Evol.
4. Moran NA, Jarvik T (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science 328: 624–627.
5. Dunning Hotopp JC (2011) Horizontal gene transfer between bacteria and animals. Trends Genet.
6. Agrios GN (1997) Plant pathology. Plant pathology. San Diego USA: Academic Press. pp. xvi +635 pp.
7. Haegeman A, Jones JT, Danchin EG (2011) Horizontal gene transfer in nematodes: a catalyst for plant parasitism? Mol Plant Microbe Interact 24: 879–887.
8. Danchin EG, Rosso MN, Vieira P, de Almeida-Engler J, Coutinho PM, et al. (2010) Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. Proc Natl Acad Sci U S A 107: 17651–17656.
9. Scholl EH, Thorne JL, McCarter JP, Bird DM (2003) Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. Genome Biol 4: R39.
10. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EG, et al. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nat Biotechnol 26: 909–915.
11. Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, et al. (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.
12. Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178–2189.
13. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
14. Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, et al. (2005) FIGENIX: Intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 6: 198.
15. Paganini J, Gouret P (2012) Reliable Phylogenetic Trees Building: A New Web Interface for FIGENIX. Evol Bioinform Online 8: 417–421.
16. Gouret P, Thompson JD, Pontarotti P (2009) PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC Bioinformatics 10: 298.
17. Levasseur A, Paganini J, Dainat J, Thompson JD, Poch O, et al. (2012) The chordate proteome history database. Evol Bioinform Online 8: 437–447.
18. Bird DM, Williamson VM, Abad P, McCarter J, Danchin EG, et al. (2009) The genomes of root-knot nematodes. Annu Rev Phytopathol 47: 333–351.
19. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 37: D233–238.
20. Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25: 537–546.
21. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. Science 320: 1210–1213.
22. Rodelsperger C, Sommer RJ (2011) Computational archaeology of the Pristionchus pacificus genome reveals evidence of horizontal gene transfers from insects. BMC Evol Biol 11: 239.
23. Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, et al. (2011) Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen Bursaphelenchus xylophilus. PLoS Pathog 7: e1002219.
24. Rosso MN, Hussey RS, Davis EL, Smant G, Baum T, et al. (2012) Nematode Effector Proteins: Targets and Functions in Plant Parasitism. In: Martin F, Kammoun S, editors. Effectors in Plant-Microbe Interactions: Wiley-Blackwell.

25. Mayer WE, Schuster LN, Bartelmes G, Dieterich C, Sommer RJ (2011) Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. BMC Evol Biol 11: 13.
26. Francino MP (2005) An adaptive radiation model for the origin of new gene functions. Nat Genet 37: 573–577.
27. Lipinski KJ, Farslow JC, Fitzpatrick KA, Lynch M, Katju V, et al. (2011) High Spontaneous Rate of Gene Duplication in *Caenorhabditis elegans*. Current Biology 21: 306–310.
28. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155.
29. Brunet FG, Roest Crollius H, Paris M, Aury JM, Gibert P, et al. (2006) Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol Biol Evol 23: 1808–1816.
30. Haegeman A, Kyndt T, Gheysen G (2010) The role of pseudo-endoglucanases in the evolution of nematode cell wall-modifying proteins. J Mol Evol 70: 441–452.
31. Danchin EGJ, Rosso M-Nl (2012) Lateral gene transfers have polished animal genomes: lessons from Nematodes. Frontiers in Cellular and Infection Microbiology 2.
32. Danchin ÉGJ (2011) What Nematode genomes tell us about the importance of horizontal gene transfers in the evolutionary history of animals. Mobile Genetic Elements 1: 1–5.
33. Pace JK, 2nd, Gilbert C, Clark MS, Feschotte C (2008) Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci U S A 105: 17023–17028.
34. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658–1659.
35. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, et al. (2008) The Pfam protein families database. Nucleic Acids Res 36: D281–288.
36. Flutre T, Duprat E, Feuillet C, Quesneville H (2011) Considering transposable element diversification in de novo annotation approaches. PLoS One 6: e16526.
37. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.
38. Leplae R, Lima-Mendez G, Toussaint A (2010) ACLAME: a CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Res 38: D57–61.
39. Jaubert S, Laffaire JB, Abad P, Rosso MN (2002) A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. FEBS Lett 522: 109–112.
40. Doyle EA, Lambert KN (2002) Cloning and characterization of an esophageal-gland-specific pectate lyase from the root-knot nematode *Meloidogyne javanica*. Mol Plant Microbe Interact 15: 549–556.
41. Huang G, Dong R, Allen R, Davis EL, Baum TJ, et al. (2005) Developmental expression and molecular analysis of two *Meloidogyne incognita* pectate lyase genes. Int J Parasitol 35: 685–692.
42. Rosso MN, Favery B, Piotte C, Arthaud L, De Boer JM, et al. (1999) Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. Mol Plant Microbe Interact 12: 585–591.
43. Bera-Maillet C, Arthaud L, Abad P, Rosso MN (2000) Biochemical characterization of MI-ENG1, a family 5 endoglucanase secreted by the root-knot nematode *Meloidogyne incognita*. Eur J Biochem 267: 3255–3263.
44. Ledger TN, Jaubert S, Bosselut N, Abad P, Rosso MN (2006) Characterization of a new beta-1,4-endoglucanase gene from the root-knot nematode *Meloidogyne incognita* and evolutionary scheme for phytonematode family 5 glycosyl hydrolases. Gene 382: 121–128.
45. Mitreva-Dautova M, Roze E, Overmars H, de Graaff L, Schots A, et al. (2006) A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode *Meloidogyne incognita*. Mol Plant Microbe Interact 19: 521–529.
46. Lambert KN, Allen KD, Sussex IM (1999) Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*. Mol Plant Microbe Interact 12: 328–336.

# Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes

Etienne G. J. Danchin[a,1], Marie-Noëlle Rosso[a], Paulo Vieira[a], Janice de Almeida-Engler[a], Pedro M. Coutinho[b], Bernard Henrissat[b], and Pierre Abad[a]

[a]Institut National de la Recherche Agronomique, Université de Nice-Sophia Antipolis, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 1301, Intéractions Biotiques et Santé Végétale, F-06903 Sophia-Antipolis, France; and [b]Centre National de la Recherche Scientifique, Unité Mixte de Recherche 6098, Architecture et Fonction des Macromolécules Biologiques, Universités Aix-Marseille I and II, F-13009 Marseille, France

Edited* by Jeffrey I. Gordon, Washington University School of Medicine, St. Louis, MO, and approved August 31, 2010 (received for review June 18, 2010)

**Lateral gene transfer from prokaryotes to animals is poorly understood, and the scarce documented examples generally concern genes of uncharacterized role in the receiver organism. In contrast, in plant-parasitic nematodes, several genes, usually not found in animals and similar to bacterial homologs, play essential roles for successful parasitism. Many of these encode plant cell wall-degrading enzymes that constitute an unprecedented arsenal in animals in terms of both abundance and diversity. Here we report that independent lateral gene transfers from different bacteria, followed by gene duplications and early gain of introns, have shaped this repertoire. We also show protein immunolocalization data that suggest additional roles for some of these cell wall-degrading enzymes in the late stages of these parasites' life cycle. Multiple functional acquisitions of exogenous genes that provide selective advantage were probably crucial for the emergence and proficiency of plant parasitism in nematodes.**

evolution | gene transfer | duplication | plant parasites

Lateral gene transfer (LGT) is the transmission of genes between organisms by mechanisms other than vertical inheritance from an ancestor to an offspring. Although largely documented as an important evolutionary mechanism in prokaryotes (1), LGT in animals that have a separate germline and whose genome is segregated in a nucleus is poorly explored. Although some examples have been described (2–4), most concern transfers from endosymbiotic bacteria, and none provide a clear link between the activity of the transferred gene products and the biology of the host species. Thus, arguments are lacking to support a selective advantage that would have driven fixation of transferred genes at the level of a population or species. By contrast, in plant-parasitic nematodes, a series of genes encoding plant cell wall-degrading or -modifying enzymes, which are usually absent from animals, exhibit similarity to bacteria and may thus originate from LGT. These genes are transcriptionally active, their products have been biochemically characterized, they are secreted in plant tissues, and their inactivation impairs parasitism efficiency (5). The most damaging nematodes to agriculture worldwide belong to the suborder Tylenchina in clade IV that comprises root-knot nematodes and cyst nematodes, the two most-studied lineages (*SI Appendix*, Fig. S1). These nematodes are able to penetrate and migrate into plant tissue and establish sophisticated parasitic interactions with their hosts. Invasion of the root tissues by nematodes requires degradation of the plant cell wall protective barrier, constituted mainly of cellulose and hemicelluloses as well as pectin and its branched decorations. The first plant cell wall-degrading enzymes from an animal were characterized in cyst nematodes in 1998 (6). Ten years later, analysis of the genome of *Meloidogyne incognita*, the first genome analysis for a plant-parasitic nematode, revealed that the repertoire of cell wall-degrading enzymes in a single species is diverse and abundant with more than 60 genes covering six different protein families for the degradation of cell wall oligo- and polysaccharides (7). This unprecedented repertoire in an animal includes cellulases and xylanases for the degradation of cellulose and hemicelluloses as well as polygalacturonases, pectate lyases, and candidate arabinanases for the degradation of pectins. A set of expansin-like proteins that soften the plant cell wall completes this repertoire (Table 1). Here, we have systematically investigated the evolutionary history and traced back the origin of each family of cell wall-degrading or modifying proteins in plant-parasitic nematodes. We show that these proteins most likely originate from multiple independent LGT events of different bacterial sources. Cellulases, pectate lyases, and expansin-like proteins are encoded by multigenic families, and we show that massive gene duplications after acquisition via LGT account for their abundance.

## Results and Discussion

In plant-parasitic nematodes, polygalacturonases and pectate lyases participate in pectin degradation. Polygalacturonase activity is known in the glycoside hydrolase GH28 family, frequently found in bacteria, fungi, oomycetes, and plants (www.cazy.org; ref. 8). Although generally absent from animals, this enzyme has been characterized in two phytophagous insects: *Sitophilus oryzae* (9), for which an acquisition via LGT from fungi has been proposed (10), and *Phaedon cochleariae*, in which the enzyme may be encoded by a gut digestive symbiont (11). In nematodes, polygalacturonase activity has been suspected in *Ditylenchus dipsaci* since the 1970s (12), and GH28 enzymes have been isolated and biochemically characterized in *M. incognita* (13). We identified polygalacturonase genes only in root-knot nematodes. Our phylogenetic analysis (Fig. 1*A* and *SI Appendix*, Fig. S2) shows that these nematode GH28 enzymes form a highly supported group with a series of bacterial orthologs. A cluster of GH28 enzymes from the bacterium *Ralstonia solanacearum* is positioned in the middle of root-knot nematode GH28 enzymes. Interestingly, *R. solanacearum* is a plant-pathogenic soil bacterium that shares plant hosts with root-knot nematodes. Our tree topology suggests that at least one LGT event occurred between bacteria that are probably closely related to *Ralstonia* and these nematodes. Possibly, a second LGT occurred with a different bacterial donor species. As an outgroup to the root-knot nematodes/*Ralstonia* cluster, we observed a series of other bacteria. The closest relatives in eukaryotes are from plants but are much more distant. Homologs from fungi and oomycetes and from the two reported insects were too distantly related to be included in a phylogenetic analysis, suggesting distinct origins.

Author contributions: E.G.J.D. and J.d.A.-E. designed research; E.G.J.D. and P.V. performed research; P.M.C. and B.H. contributed new reagents/analytic tools; E.G.J.D., M.-N.R., J.d.A.-E., and P.A. analyzed data; and E.G.J.D., M.-N.R., and B.H. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: etienne.danchin@sophia.inra.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1008486107/-/DCSupplemental.

**Table 1. Plant cell wall-modifying proteins in plant-parasitic nematodes**

| Family | Activity | Presence in nematodes | Closest relative |
|---|---|---|---|
| GH28 | Polygalacturonase (EC 3.2.1.15) | RKN* | Ralstonia: *Ralstonia solanacearum*† |
| PL3 | Pectate lyase (EC 4.2.2.2) | RKN,* CN,* Aphelenchoidea* | Actinomycetales: *Clavibacter michiganensis*,† *Frankia* sp.,‡ *Actinosynnema mirum, Cellulomonas flavigena, Jonesia denitrificans, Streptomyces avermitilis, S. coelicolor,* |
| GH43 | Putative arabinanase (EC 3.2.1.99) | RKN, CN | Actinomycetales: *Streptomyces coelicolor, Thermomonospora curvata, Kineococcus radiotolerans* |
| GH5 (cel) | Cellulase (EC 3.2.1.4) | RKN,* CN,* Pratylenchidae, Anguinidae, Radopholinae, Aphelenchoidea | Coleoptera: *Apriona germari*, *Psacothea hilaris*, Bacteroidetes: *Cytophaga hutchinsonii* |
| GH5 (xyl) | Endo-1,4-β-xylanase (EC 3.2.1.8) | RKN* and Radopholinae* | Firmicutes: *Clostridium acetobutylicum* |
| EXPN | Loosening of plant cell wall (EC N/A) | RKN, CN,* Anguinidae, Aphelenchoidea, Dorylaimida (*clade I*) | Actinomycetales: *Amycolatopsis mediterranei*, *Actinosynnema mirum*, *Streptomyces lavendulae* |

Spectrum of the presence in nematodes is indicated according to the taxonomy in *SI Appendix*, Fig. S1. Species possessing the genes most closely related to those of plant-parasitic nematodes are in the last column. CN, cyst nematode; RKN, root-knot nematode.
*For nematodes, species in which activity has been experimentally shown; details on functional characterization and corresponding bibliographic references are indicated in *SI Appendix*, Table S1.
†Plant-pathogenic bacteria.
‡Plant-symbiotic bacteria.

In contrast to polygalacturonases, pectate lyases cleave α-1,4-galacturonan, the major component of pectin backbone, via β-elimination instead of hydrolysis. All pectate lyases characterized in plant-parasitic nematodes belong to polysaccharide lyase (PL) family 3. In root-knot nematodes, PL3s are present as multigenic families in both *M. incognita* and *Meloidogyne hapla* (7, 14). Functional PL3's have also been isolated in cyst nematodes (15, 16) and in Aphelenchoidea. Nematodes that belong to this last

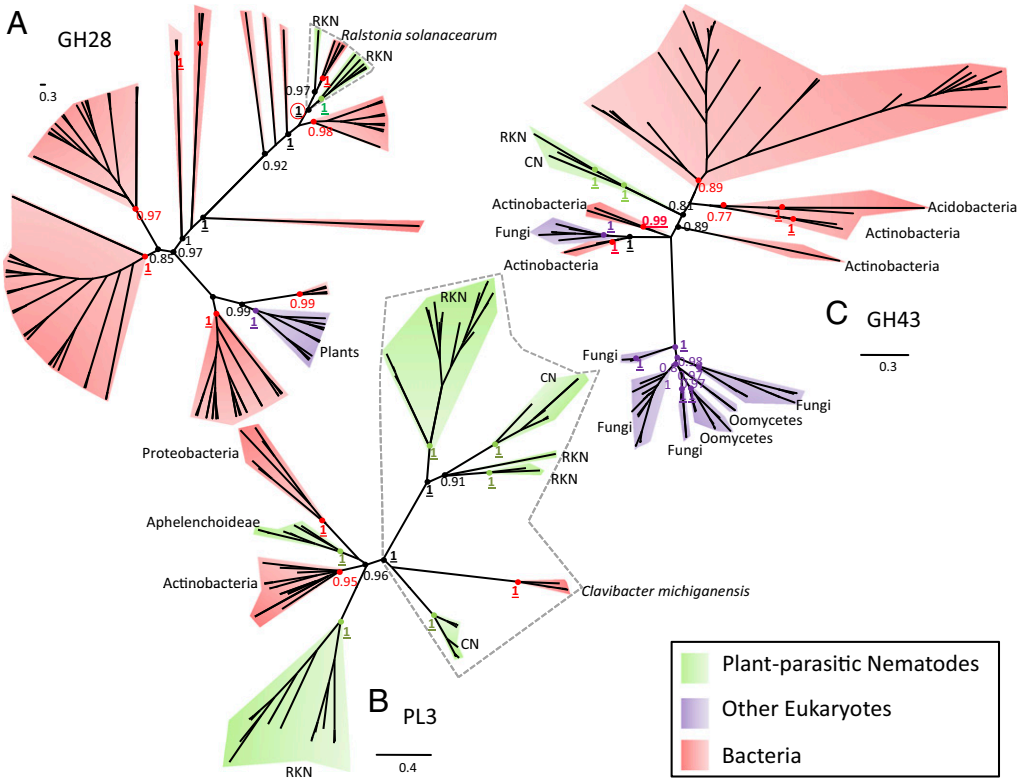



**Fig. 1.** Phylogenetic trees of pectin-modifying proteins. (*A*) GH28 polygalacturonases. (*B*) PL3 pectate lyases. (*C*) GH43 candidate arabinanases. Phylogenetic groups are color-coded according to their taxonomy. Posterior probability (PP) support values are indicated at corresponding nodes, and those supported by bootsrap values higher than 75 in maximum likelihood trees are underlined and in boldface type. RKN, root-knot nematode; CN, cyst nematode. Dashed lines delineate phylogenetic groupings of bacterial and plant-parasitic nematode genes; the corresponding PP value is circled.

group are plant-associated and hold an outgroup position relative to root-knot and cyst nematodes (17, 18) (*SI Appendix*, Fig. S1). Our phylogenetic reconstructions with homologs from nematodes, bacteria, oomycetes, and fungi provided strong support for separation of PL3's in two main clusters: one cluster grouping bacteria and nematodes and another cluster containing fungi and oomycetes *(SI Appendix*, Fig. S3). Inside the bacteria/nematode cluster, nematode PL3's are interspersed by two clusters of Actinobacteria. To gain a deeper insight into the bacteria/nematode cluster, we performed a phylogenetic analysis using only bacterial and nematode PL3's because they form a distinct monophyletic group. This tree (Fig. 1*B*) revealed evolutionary relations between the different plant-parasitic nematodes and bacterial clusters. Remarkably, *Clavibacter michiganensis*, the bacterium that possesses PL3's most closely related to the majority of those of root-knot and cyst nematodes, is a notorious plant parasite that shares host plants with these nematodes. This group of plant-parasitic nematodes may have acquired PL3's from an ancestor or a close relative of this bacterium. The non-monophyly of root-knot and cyst nematode PL3's suggests that a few independent LGT events gave rise to the different subfamilies. These distinct root-knot and cyst nematode clusters show that, in both lineages, a series of duplications followed the likely acquisition of ancestral PL3's via LGT and account for the abundance of this family.

Arabinans and arabinogalactans are the main components of pectin side chains. These chains can prevent access to the pectin backbone for cleavage by polygalacturonases or pectate lyases. We identified candidate arabinanases of the family GH43 in the root-knot nematode genomes and in the draft genomes of the cyst nematodes *Heterodera glycines* and *Globodera pallida*. No other significant similarity was found in animals, but a series of candidate homologs was identified in bacteria, oomycetes, and fungi. Root-knot and cyst nematode GH43 enzymes form a monophyletic group in our phylogenies (Fig. 1*C* and *SI Appendix*, Fig. S4). This suggests that a GH43 enzyme was present in the last common ancestor of these nematodes. Thus, the possibility of finding an enzyme from this family in other Tylenchina remains open. Actinomycetales GH43 enzymes appear to be the most closely related to nematode GH43 enzymes. This suggests that GH43 genes have been acquired in nematodes via LGT of bacterial origin (probably an ancestral or relative of Actinomycetales). Interestingly, a putative arabinogalactan endo-1,4-β-galactosidase (EC 3.2.1.89) from family GH53, unrelated to GH43, was found in ESTs from the cyst nematode *Heterodera schachtii*, and we identified only a homolog in the draft genome of *H. glycines* (19, 20), suggesting that it is restricted to cyst nematodes.

Cellulose is the most abundant biopolymer on earth, and cellulases secreted by plant-parasitic nematodes allow its breakdown during root invasion (5, 21). In clade IV nematodes, cellulases from a subfamily of family GH5 (8) and those from family GH45 have been characterized. GH5 cellulases have been reported in Tylenchina and in one Aphelenchoidea species (18), whereas GH45 cellulases were reported only in Aphelenchoidea (*SI Appendix*, Fig. S1). No nematode species has been found to harbor both families of cellulases, suggesting that they are mutually exclusive. We identified candidate homologs of GH5 cellulases in



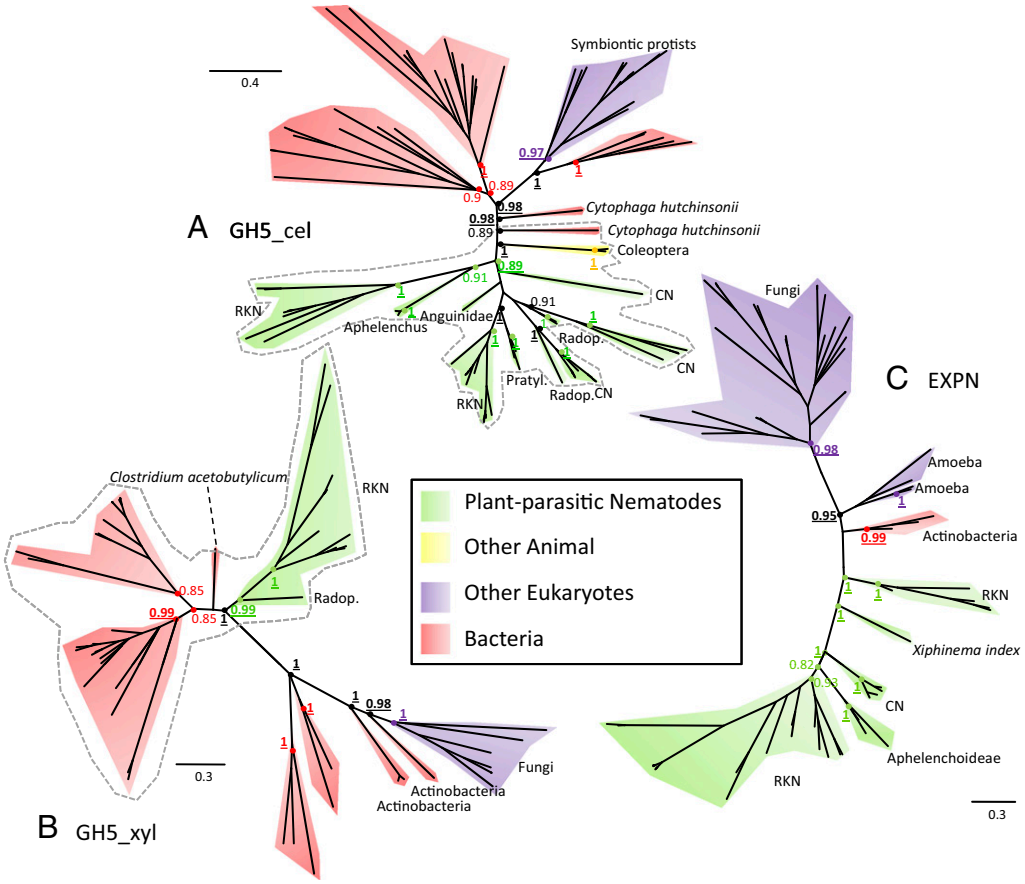



**Fig. 2.** Phylogenetic trees of cellulose/hemicellulose-modifying proteins and expansin-like proteins. (*A*) GH5 cellulases. Radop. stands for Radopholinae, Pratyl. for Pratylenchidae. (*B*) GH5 xylanases. (*C*) Expansin-like proteins. Phylogenetic groups are color-coded according to their taxonomy. Posterior probability (PP) support values are indicated at corresponding nodes, and those supported by bootstrap values higher than 75 in maximum likelihood trees are underlined and in boldface type. Groups supported by PP values higher than 0.8 were collapsed. RKN, root-knot nematode; CN for cyst nematode.

bacteria, in two insects that feed on plants, *Psacothea hilaris* (22) and *Apriona germari* (23), and in gut digestive symbionts of termites. Interestingly, as opposed to nematode cellulase genes, those found in insects are intronless. GH5 cellulases from clade IV nematodes form a highly supported monophyletic group (Fig. 2*A* and *SI Appendix*, Fig. S5). Their closest orthologs are found in the two insects, in *Cytophaga hutchinsonii*, and in other bacteria. Assuming that the two insect cellulase genes are endogenous, this topology suggests two hypotheses: (*i*) two LGT events occurred from a similar bacterial source in these insects and in clade IV nematodes or (*ii*) a cellulase gene was already present in the last common ancestor of these animals. The second hypothesis appears unlikely because their otherwise general absence in nematodes and insects would require many independent gene losses. An origin in the last common ancestor of eukaryotes is even less likely because GH5 enzymes found in plants and fungi belong to subfamilies distinct from that found in nematodes and are more distantly related than those of bacteria. In a set of nematode cellulases, a cellulose-binding CBM2 module is appended at the C-terminal end (*SI Appendix*, Fig. S5). This module is found neither in plants nor in fungi although association of CBM2 modules with GH5 modules is frequent in bacteria. This observation reinforces the hypothesis of the acquisition of nematode GH5 cellulases via LGT of bacterial origin. Notably, none of the two reported insect cellulases bear a CBM2 module.

Xylanases catalyze the degradation of xylose, the main constituent of hemicellulose. All xylanases reported in nematodes belong to a subfamily of family GH5 that is clearly distinct from that of the cellulases discussed above. We found candidate nematode xylanase only in root-knot nematodes and *Radopholus similis*, and our phylogenetic analysis showed that they form a monophyletic group (Fig. 2*B* and *SI Appendix*, Fig. S6). Nematode xylanases are nested among clusters of bacterial xylanases, and their closest ortholog is a protein from the soil bacteria *Clostridium acetobutylicum*. This suggests that an ancestral xylanase was acquired via LGT at least in the common ancestor of root-knot nematodes and *R. similis*. As the common ancestor of these species is also the ancestor of cyst nematodes (*SI Appendix*, Fig. S1), the absence of GH5 xylanase in these nematodes suggests that they were either secondarily lost in this lineage or not yet identified. Alternatively, enzymes from another family may perform the same function in cyst nematodes. This is the case for cellulases discussed previously that belong to different and mutually exclusive families (GH5, GH45) but perform the same enzymatic activity in nematodes.

Expansins (EXPNs) are not enzymes per se but loosen the noncovalent interactions between the constituents of the plant cell wall (24). They are typically plant proteins, but a functional EXPN has been characterized in cyst nematodes (25). Candidate EXPNs are present as multigenic families in root-knot nematodes (7, 26) and have also been found in other plant-associated nematodes (27, 28). Interestingly, we identified a homolog in *Xiphinema index*, a clade I plant-parasitic nematode (*SI Appendix*, Fig. S1). Our phylogenetic analysis shows that EXPNs from all plant-parasitic nematodes, including *X. index*, are grouped in a highly supported monophyletic group (Fig. 2*C* and *SI Appendix*, Fig. S7). Two hypotheses can be formulated concerning their origin. The first hypothesis says that a single acquisition occurred in the last common ancestor of clade IV and clade I nematodes with subsequent losses in the numerous nematode taxa not associated with plants. This appears unlikely and in contradiction with the current hypothesis of three independent emergences of plant parasitism in nematodes (29). The second hypothesis suggests that two LGT events from similar bacterial donor species occurred independently in clade I and clade IV nematode lineages. This is further supported by the noncongruence between the relative position of plant-parasitic nematode lineages in the EXPN phylogeny and the actual taxonomy (*SI Appendix*, Fig. S1).

A feature common to several GH5 cellulases and EXPNs is the presence of an appended CBM2 cellulose-binding module. In nematodes and bacteria, proteins only made of a single CBM2 module can be found. In root-knot and cyst nematodes, these proteins are called cellulose-binding proteins (CBPs). A CBP from the cyst nematode *H. schachtii* has been shown to interact with a plant pectin-methylesterase, promoting degradation of its cell wall (30). CBM2 modules are frequently present in bacteria and otherwise found only in a few mollusks associated with a GH9 cellulase module. Our similarity searches using nematode CBM2 modules as queries returned a series of bacterial CBM2's but none from mollusks, indicating that they are distantly related and probably have a distinct origin. All CBM2's found in nematodes form a monophyletic group, and most closely related CBM2 modules are those of actinobacteria (*SI Appendix*, Fig. S8). This suggests that, in plant-parasitic nematodes, CBM2's appended to other different modules have a common bacterial origin. We observed no clearly distinct groups separating CBM2's appended to EXPN modules from those appended to GH5 modules, suggesting that domain shuffling occurred. In root-knot nematodes, CBPs form a monophyletic group closely related to a group of CBM2's appended to GH5 cellulases. In contrast, CBPs of cyst nematodes are more closely related to CBM2's appended to EXPN modules. This suggests that root-knot and cyst nematode CBPs derive from CBM2-bearing cellulases and EXPNs, respectively. The secretion of CBM2-bearing cellulases in plants has been demonstrated along the migratory path of cyst nematodes that migrate intracellularly and destroy the walls of the cells that they travel through (31, 32). No *in planta* localization data have been published for these proteins in root-knot nematodes that migrate intercellularly without damaging plant cells. Our immunolocalization studies on tomato roots infected with *M. incognita* showed the secretion of CBM2-bearing proteins by root-knot nematodes during migratory stages (*SI Appendix*, Fig. S9). Interestingly, we also detected these proteins in later sedentary parasitism stages in eggs about to be extruded by the female and at its vulva region (Fig. 3 and *SI Appendix*, Fig. S10). This suggests a role for these proteins, probably in successful egg laying, a process necessary for spreading the offspring of nematodes to surrounding roots, which is crucial for their parasitic life cycle.

Although our phylogenetic analyses show that bacterial cell-wall–modifying proteins are the most closely related to those of nematodes, suggesting acquisition via LGT, the alternative hypothesis of descent from a common ancestor in eukaryotes, cannot be totally ruled out. To test the likelihood of this alternative hypothesis, we have compared constrained trees in which all eukaryotic homologs were put together in monophyletic groups to the unconstrained trees that we obtained. Statistical testing allowed us to reject with high confidence the topologies presenting monophyly of eukaryotes in all families except the EXPNs for which monophyly of fungal and nematode genes is only slightly less likely (*SI Appendix*, Table S2 *A–F*).

Our tree topologies show that the abundance of multigenic families (cellulases, pectate lyases, and EXPNs) in plant-parasitic nematodes is due to a series of duplications that started after acquisition by LGT events and before the separation of the different clade IV nematode lineages. Duplications that pursued independently after the separation of root-knot and cyst nematodes contibuted the most to this abundance. From the available plant-parasitic nematode genomes, only a few duplications appear to have continued at a species-specific level. All families of cell wall-modifying proteins feature gene structures with multiple introns that contradict the hypothesis of bacterial contamination. In families that are present in several clade IV nematode lineages, at least one intron position could be identified as shared between all lineages, suggesting that these introns were gained early after LGT and before the separation of these different lineages (*SI Appendix*, Fig. S11). In *M. incognita*, considering GC content and codon usage, LGT-acquired genes are indistinguishable from the other *M. incognita*

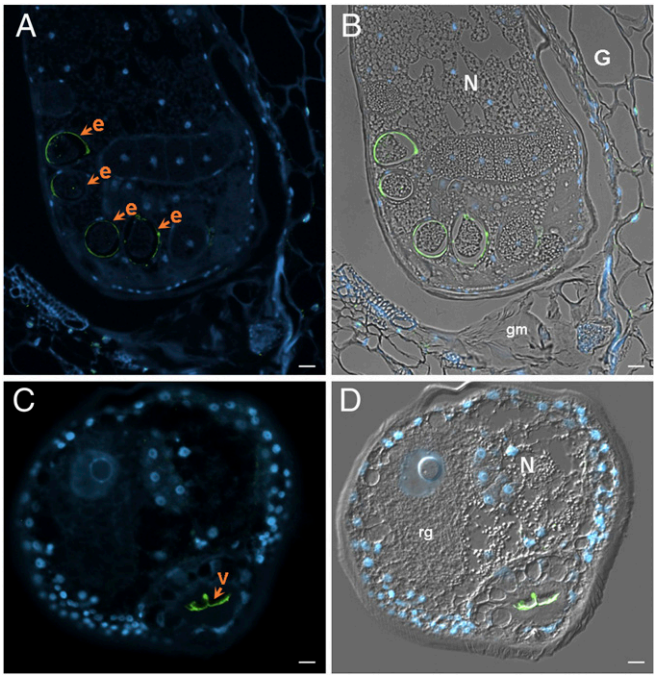



**Fig. 3.** Immunodetection of CBM2-bearing proteins within adult sedentary females of *M. incognita* during parasitism of tomato roots. (*A* and *B*) Gall containing an adult female, displaying CBM2-bearing proteins accumulated in the eggs (green and indicated by orange arrows) within the ovary. (*C* and *D*) Localization of CBM2-bearing proteins accumulated in the vagina of adult females (green and indicated by orange arrow in *C*). (*A* and *C*) Overlay images of CBM2-bearing proteins (green) and DAPI-stained nuclei (blue); (*B* and *D*) Overlay images of CBM2-bearing proteins (green), DAPI-stained nuclei (blue), and differential interference contrast (gray). N, nematode; e, egg; G, gall; gm, gelatinous matrix; v, vagina; rg, rectal gland. Scale bar: 10 μm. Control images are available in the *SI Appendix*, Fig. S10.

genes (*SI Appendix*, Fig. S12, and Tables S3 and S4). This ensemble of observations reinforces the hypothesis of ancient transfer.

Duplication events that led to multigenic families could have been under positive selective pressure as proposed for other species (33). Such duplications can promote subfunctionalization and neo-functionalization or increase the level of transcribed genes. The dual protein localization pattern of CBM2-bearing proteins, during both migratory and late sedentary stages of infection, may reflect neo- or subfunctionalization.

Our analysis shows that LGT events from several independent bacterial sources most likely gave rise to the six different gene families involved in plant cell-wall modification in clade IV plant-parasitic nematodes. Consistent with this hypothesis is the absence of genomic clusters grouping members of different families that could be a sign of cotransfer from an identical source (*SI Appendix*, Fig. S13). Four different groups of bacteria can be viewed as potential donors for these six cell-wall–modifying gene families. Three of these soil bacteria are notorious plant pathogens or are associated with symbiotic interactions within plant roots (Table 1 and *SI Appendix*, Fig. S14). Being sympatric with plant-parasitic nematodes, these soil bacteria satisfy the most elemental criterion to make LGT possible. However, how the genes are successfully transferred to the germline and then fixed in populations and species is not evident. Gene transfers from algae to the nuclear genome of their predator, a sea slug, have recently been shown (34). At least one example of a nematode that spends part of its life cycle in plant tissue and feeds on plant-symbiotic bacteria is known (35). The bacteria that they feed on possess plant cell-wall degradation genes, and this nematode belongs to the Cephalobidae family, a lineage closely related to clade IV plant-parasitic nematodes (*SI Appendix*, Fig. S1). Thus, acquisition through feeding can be hypothesized. An alternative hypothesis is gene transfers from endosymbiotic bacteria. Such events have been shown in several other animals, including nematodes, probably aided by physical proximity of endosymbionts and germline cells (2). Endosymbionts have been reported in several plant-parasitic nematodes (36–38), all at the vicinity of gametes or eggs. Although transfers appear more evident for endosymbionts, the presence of plant cell wall degradation genes is less likely in these bacteria.

At least one other example of gene transfer from bacteria to nematodes has been reported. Interestingly, this also involved a gene coding for an enzyme, although in this case the transfer probably took place in the last common ancestor of all nematodes (39) and appears not specifically linked to a given lifestyle. In animals other than nematodes, a similar case of massive gene transfers has been reported from bdelloid rotifers (4). Intriguingly, these transfers also concerned genes involved in the degradation of polysaccharides, and gains of introns were also reported. As for the nematode transfers, several transferred genes in these rotifers were indistinguishable from the endogenous genes, suggesting ancient acquisition. Gene transfers in bdelloid rotifers appear to have been fixed preferentially in telomeric regions, and transposable elements may play a role in successful transfer. These features were not observed in nematode transfers. Conversely, the massive duplications that we observed after transfer in nematodes have not been noted in bdelloid rotifers. Bacteria appear as common candidate donors in both reports, but fungi and plants are also candidate donors for bdelloid rotifers. In contrast to plant-parasitic nematodes, in bdelloid rotifers, transfers may have been facilitated by their peculiar life cycle during which they undergo desiccation involving DNA fragmentation and dispersal of membranes, which allows foreign DNA to be incorporated during recovery. Regardless of the possible mechanisms, both these cases highlight the potential importance of LGT in animals and suggest that this phenomenon may be more frequent than usually considered.

## Conclusions

We have shown that in plant-parasitic nematodes, a whole set of genes encoding proteins involved in the plant cell wall degradation was most likely acquired by LGT of bacterial origin. The function of the transferred gene products is directly linked to the capacity of these nematodes to parasitize plants. Selective advantage associated with transfer of these genes probably has driven their duplications and facilitated fixation in the different populations and species of plant-parasitic nematodes. Far from being negligible, these LGT events certainly have radically remolded evolutionary trends in recipient organisms, and similar roles in other animals can be expected to be discovered.

## Materials and Methods

**Phylogenetic Analyses.** Homologs of nematode plant cell-wall–degrading proteins were searched in public databases and checked for significance using the approach described in *SI Appendix*. Multiple alignments were done with MUSCLE (40). Phylogenetic analyses were performed using two approaches: a Bayesian method with mrBayes (41) using a mixture of models and a bootstrapped maximum-likelihood approach with RAxML (42) using the evolutionary models returning the highest posterior probabilities in Bayesian analyses. For both methods, we used an evaluation of the proportion of invariable rates as well as of the shape of the γ-distribution of evolutionary rates. The procedure is detailed in *SI Appendix*.

**Tree Selection Topology Tests.** We statistically tested the significance of the obtained tree topologies in comparison with alternative trees in which all eukaryotic homologs were constrained to form monophyletic groups using the program CONSEL (43). Details of the procedure are available in *SI Appendix*.

**GC Content, Codon Use, and Gene Localization.** We used the EMBOSS (44) software suite to calculate the codon usage and GC content of plant cell-wall

degradation genes and compared these values to those calculated for the rest of the protein-coding genes. Details of these analyses are available in the *SI Appendix*. Genome localization of plant cell wall degradation genes in *M. incognita* were obtained from GFF files of the genome sequence (7).

**Exon/Intron Structure Determination.** Information about the intron/exon structures was extracted from the literature and from the National Center for Biotechnology Information's GenBank. For sequences resulting from genome-sequencing projects, intron/exon structures were deduced from the alignment of protein models with the corresponding genome sequences using the procedure detailed in *SI Appendix*.

**Immunolocalizations.** Antibodies were raised against a peptide from the CBM2 module of *M. incognita* ENG1 cellulase. Dissected tomato roots infected with *M. incognita* were fixed and cut in sections. Slides harboring the nematode feeding sites were immunolabeled with a serum containing the anti-ENG1 antibodies. Slides were observed with a microscope equipped for epifluorescence and differential interference contrast optics. Full details of the procedure used are available in *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank H. Shimodaira for support on statistical tests; T. Guillemaud for providing access to a CPU cluster; and C. Vens, A. Campan, P. Castagnone-Sereno, and L. Perfus-Barbeoch for critical review and bioinformatics scripts. We are grateful for the support of European Cooperation in Science and Technology Action 872 and the Maladies Infectieuses Emergentes program from Centre National de la Recherche Scientifique for funding. P. Vieira is supported by a doctoral scholarship from Fundação para a Ciência e a Tecnologia, Portugal (BD\41339\2007).

1. Boucher Y, et al. (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet* 37:283–328.
2. Hotopp JC, et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317:1753–1756.
3. Nikoh N, Nakabachi A (2009) Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biol* 7:12.
4. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. *Science* 320:1210–1213.
5. Bakhetia M, Urwin PE, Atkinson HJ (2007) QPCR analysis and RNAi define pharyngeal gland cell-expressed genes of Heterodera glycines required for initial interactions with the host. *Mol Plant Microbe Interact* 20:306–312.
6. Smant G, et al. (1998) Endogenous cellulases in animals: Isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc Natl Acad Sci USA* 95:4906–4911.
7. Abad P, et al. (2008) Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. *Nat Biotechnol* 26:909–915.
8. Cantarel BL, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res* 37(Database issue):D233–D238.
9. Shen ZC, Reese JC, Reeck GR (1996) Purification and characterization of polygalacturonase from the rice weevil, Sitophilus oryzae (Coleoptera: Curculionidae). *Insect Biochem Mol Biol* 26:427–433.
10. Shen Z, et al. (2003) Polygalacturonase from Sitophilus oryzae: Possible horizontal transfer of a pectinase gene from fungi to weevils. *J Insect Sci* 3:24.
11. Girard C, Jouanin L (1999) Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, Phaedon cochleariae. *Insect Biochem Mol Biol* 29:1129–1142.
12. Muse BD, Moore LD, Muse RR, Williams AS (1970) Pectolytic and cellulolytic enzymes of two populations of Ditylenchus dipsaci on 'Wando' pea (Pisum sativum L.). *J Nematol* 2:118–124.
13. Jaubert S, Laffaire JB, Abad P, Rosso MN (2002) A polygalacturonase of animal origin isolated from the root-knot nematode Meloidogyne incognita. *FEBS Lett* 522: 109–112.
14. Opperman CH, et al. (2008) Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. *Proc Natl Acad Sci USA* 105:14802–14807.
15. Popeijus H, et al. (2000) Degradation of plant cell walls by a nematode. *Nature* 406: 36–37.
16. Vanholme B, et al. (2007) Molecular characterization and functional importance of pectate lyase secreted by the cyst nematode *Heterodera schachtii*. *Mol Plant Pathol* 8: 267–278.
17. Kikuchi T, Shibuya H, Aikawa T, Jones JT (2006) Cloning and characterization of pectate lyases expressed in the esophageal gland of the pine wood nematode Bursaphelenchus xylophilus. *Mol Plant Microbe Interact* 19:280–287.
18. Karim N, Jones JT, Okada H, Kikuchi T (2009) Analysis of expressed sequence tags and identification of genes encoding cell-wall-degrading enzymes from the fungivorous nematode Aphelenchus avenae. *BMC Genomics* 10:525.
19. Vanholme B, et al. (2006) Detection of putative secreted proteins in the plant-parasitic nematode Heterodera schachtii. *Parasitol Res* 98:414–424.
20. Vanholme B, Haegeman A, Jacob J, Cannoot B, Gheysen G (2009) Arabinogalactan endo-1,4-beta-galactosidase: A putative plant cell wall-degrading enzyme of plant-parasitic nematodes. *Nematology* 11:739–747.
21. Chen Q, Rehman S, Smant G, Jones JT (2005) Functional analysis of pathogenicity proteins of the potato cyst nematode Globodera rostochiensis using RNAi. *Mol Plant Microbe Interact* 18:621–625.
22. Sugimura M, Watanabe H, Lo N, Saito H (2003) Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle, Psacothea hilaris. *Eur J Biochem* 270:3455–3460.
23. Wei YD, et al. (2006) Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry longicorn beetle, Apriona germari. *Comp Biochem Physiol B Biochem Mol Biol* 145:220–229.
24. Carey RE, Cosgrove DJ (2007) Portrait of the expansin superfamily in Physcomitrella patens: Comparisons with angiosperm expansins. *Ann Bot* 99:1131–1141.
25. Qin L, et al. (2004) Plant degradation: A nematode expansin acting on plants. *Nature* 427:30.
26. Bird DM, et al. (2009) The genomes of root-knot nematodes. *Annu Rev Phytopathol* 47:333–351.
27. Haegeman A, et al. (2009) Expressed sequence tags of the peanut pod nematode Ditylenchus africanus: The first transcriptome analysis of an Anguinid nematode. *Mol Biochem Parasitol* 167:32–40.
28. Kikuchi T, et al. (2007) Expressed sequence tag (EST) analysis of the pine wood nematode Bursaphelenchus xylophilus and B. mucronatus. *Mol Biochem Parasitol* 155: 9–17.
29. Holterman M, et al. (2009) Small subunit rDNA-based phylogeny of the Tylenchida sheds light on relationships among some high-impact plant-parasitic nematodes and the evolution of plant feeding. *Phytopathology* 99:227–235.
30. Hewezi T, et al. (2008) Cellulose binding protein from the parasitic nematode Heterodera schachtii interacts with Arabidopsis pectin methylesterase: Cooperative cell wall modification during parasitism. *Plant Cell* 20:3080–3093.
31. Wang X, et al. (1999) In planta localization of a beta-1,4-endoglucanase secreted by Heterodera glycines. *Mol Plant Microbe Interact* 12:64–67.
32. Goellner M, Smant G, De Boer JM, Baum TJ, Davis EL (2000) Isolation of beta-1,4-endoglucanase genes from Globodera tabacum and their expression during parasitism. *J Nematol* 32:154–165.
33. Francino MP (2005) An adaptive radiation model for the origin of new gene functions. *Nat Genet* 37:573–577.
34. Rumpho ME, et al. (2008) Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug Elysia chlorotica. *Proc Natl Acad Sci USA* 105:17867–17871.
35. Westcott SW, Barker KR (1976) Interaction of Acrobeloides buetschlii and Rhizobium leguminosarum on Wando pea. *Phytopathology* 66:468–472.
36. Noel GR, Atibalentja N (2006) 'Candidatus Paenicardinium endonii', an endosymbiont of the plant-parasitic nematode Heterodera glycines (Nemata: Tylenchida), affiliated to the phylum Bacteroidetes. *Int J Syst Evol Microbiol* 56:1697–1702.
37. Haegeman A, et al. (2009) An endosymbiotic bacterium in a plant-parasitic nematode: Member of a new Wolbachia supergroup. *Int J Parasitol* 39:1045–1054.
38. Vandekerckhove TT, Willems A, Gillis M, Coomans A (2000) Occurrence of novel verrucomicrobial species, endosymbiotic and associated with parthenogenesis in Xiphinema americanum-group species (Nematoda, Longidoridae). *Int J Syst Evol Microbiol* 50:2197–2205.
39. Kondrashov FA, Koonin EV, Morgunov IG, Finogenova TV, Kondrashova MN (2006) Evolution of glyoxylate cycle enzymes in Metazoa: Evidence of multiple horizontal transfer events and pseudogene formation. *Biol Direct* 1:31.
40. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
41. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
42. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
43. Shimodaira H, Hasegawa M (2001) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
44. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 16:276–277.

nature biotechnology



# Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*

Pierre Abad[1–3], Jérôme Gouzy[4], Jean-Marc Aury[5–7], Philippe Castagnone-Sereno[1–3], Etienne G J Danchin[1–3], Emeline Deleury[1–3], Laetitia Perfus-Barbeoch[1–3], Véronique Anthouard[5–7], François Artiguenave[5–7], Vivian C Blok[8], Marie-Cécile Caillaud[1–3], Pedro M Coutinho[9], Corinne Dasilva[5–7], Francesca De Luca[10], Florence Deau[1–3], Magali Esquibet[11], Timothé Flutre[12], Jared V Goldstone[13], Noureddine Hamamouch[14], Tarek Hewezi[15], Olivier Jaillon[5–7], Claire Jubin[5–7], Paola Leonetti[10], Marc Magliano[1–3], Tom R Maier[15], Gabriel V Markov[16,17], Paul McVeigh[18], Graziano Pesole[19,20], Julie Poulain[5–7], Marc Robinson-Rechavi[21,22], Erika Sallet[23,24], Béatrice Ségurens[5–7], Delphine Steinbach[12], Tom Tytgat[25], Edgardo Ugarte[5–7], Cyril van Ghelder[1–3], Pasqua Veronico[10], Thomas J Baum[15], Mark Blaxter[26], Teresa Bleve-Zacheo[10], Eric L Davis[14], Jonathan J Ewbank[27], Bruno Favery[1–3], Eric Grenier[11], Bernard Henrissat[9], John T Jones[8], Vincent Laudet[16], Aaron G Maule[18], Hadi Quesneville[12], Marie-Noëlle Rosso[1–3], Thomas Schiex[24], Geert Smant[25], Jean Weissenbach[5–7] & Patrick Wincker[5–7]

**Plant-parasitic nematodes are major agricultural pests worldwide and novel approaches to control them are sorely needed. We report the draft genome sequence of the root-knot nematode *Meloidogyne incognita*, a biotrophic parasite of many crops, including tomato, cotton and coffee. Most of the assembled sequence of this asexually reproducing nematode, totaling 86 Mb, exists in pairs of homologous but divergent segments. This suggests that ancient allelic regions in *M. incognita* are evolving toward effective haploidy, permitting new mechanisms of adaptation. The number and diversity of plant cell wall–degrading enzymes in *M. incognita* is unprecedented in any animal for which a genome sequence is available, and may derive from multiple horizontal gene transfers from bacterial sources. Our results provide insights into the adaptations required by metazoans to successfully parasitize immunocompetent plants, and open the way for discovering new antiparasitic strategies.**

Plant-parasitic nematodes are responsible for global agricultural losses amounting to an estimated $157 billion annually. Although chemical nematicides are the most reliable means of controlling root-knot nematodes, they are increasingly being withdrawn owing to their toxicity to humans and the environment. Novel and specific targets are thus needed to develop new strategies against these pests.

The Southern root-knot nematode *Meloidogyne incognita* is able to infect the roots of almost all cultivated plants, making it perhaps the

[1]INRA, UMR 1301, 400 route des Chappes, F-06903 Sophia-Antipolis, France. [2]CNRS, UMR 6243, 400 route des Chappes, F-06903 Sophia-Antipolis, France. [3]UNSA, UMR 1301, 400 route des Chappes, F-06903 Sophia-Antipolis, France. [4]Laboratoire Interactions Plantes Micro-organismes, UMR441/2594, INRA/CNRS, Chemin de Borde Rouge, BP 52627, F-31320 Castanet Tolosan, France. [5]Genoscope (CEA), 2 rue Gaston Crémieux, CP5706, F-91057 Evry, France. [6]CNRS, UMR 8030, 2 rue Gaston Crémieux, CP5706, F-91057 Evry, France. [7]Université d'Evry, F-91057 Evry, France. [8]Plant Pathology Programme, SCRI, Invergowrie, Dundee DD2 5DA, UK. [9]CNRS, UMR 6098 CNRS and Universités d'Aix-Marseille I & II, Case 932, 163 Av. de Luminy, F-13288 Marseille, France. [10]Istituto per la Protezione delle Piante, Consiglio Nazionale delle Ricerche, Via G. Amendola 165/a, 70126 Bari, Italy. [11]INRA, Agrocampus Rennes, Univ. Rennes I, UMR1099 BiO3P, Domaine de la Motte, F-35653 Le Rheu Cedex, France. [12]INRA, UR1164 Unité de Recherche en Génomique et Informatique (URGI), 523 place des terrasses de l'Agora, F-91034 Evry, France. [13]Biology Department, Woods Hole Oceanographic Institution, Co-op Building, MS #16, Woods Hole, Massachusetts 02543, USA. [14]Department of Plant Pathology, North Carolina State University, 840 Method Road, Unit 4, Box 7903 Raleigh, North Carolina 27607,USA. [15]Department of Plant Pathology, Iowa State University, 351 Bessey Hall, Ames, Iowa 50011, USA. [16]Université de Lyon, Institut de Génomique Fonctionnelle de Lyon, Molecular Zoology team, Ecole Normale Supérieure de Lyon, Université Lyon 1, CNRS, INRA, Institut Fédératif 128 Biosciences Gerland, Lyon Sud, 46 allée d'Italie, F-69364 Lyon Cedex 07, France. [17]USM 501, Evolution des Régulations Endocriniennes, Muséum National d'Histoire Naturelle, 7 rue Cuvier, F-75005 Paris, France. [18]Biomolecular Processes: Parasitology, School of Biological Sciences, Medical Biology Centre, 97 Lisburn Road, Queen's University Belfast, Belfast BT9 7BL, UK. [19]Dipartimento di Biochimica e Biologia Molecolare "E. Quagliariello", University of Bari, Via Orabona 4, 70126 Bari, Italy. [20]Istituto Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Via G. Amendola, 122/D, 70126 Bari, Italy. [21]Department of Ecology and Evolution, University of Lausanne, UNIL-Sorge, Le Biophore, CH-1015 Lausanne, Switzerland. [22]Swiss Institute of Bioinformatics, quartier sorge, Bâtiment Genopode, CH-1015 Lausanne, Switzerland. [23]Plateforme Bioinformatique du Génopole Toulouse Midi-Pyrénées, GIS Toulouse Genopole, 24 Chemin de Borde Rouge, BP 52627, F-31320 Castanet Tolosan, France. [24]Unité de Biométrie et d'Intelligence Artificielle UR875, INRA, Chemin de Borde Rouge, BP 52627, F-31320 Castanet Tolosan, France. [25]Laboratory of Nematology, Wageningen University, Binnenhaven 5, 6709PD Wageningen, The Netherlands. [26]Institute of Evolutionary Biology, University of Edinburgh, Kings Buildings, Ashworth Laboratories, West Mains Road, Edinburgh EH9 3JT, UK. [27]INSERM/CNRS/Université de la Méditerranée, Centre d'Immunologie de Marseille-Luminy, 163 av. de Luminy, Case 906, F-13288, Marseille cedex 09, France. Correspondence should be addressed to P.A. (pierre.abad@sophia.inra.fr).

Received 30 April; accepted 25 June; published online 27 July 2008; doi:10.1038/nbt.1482

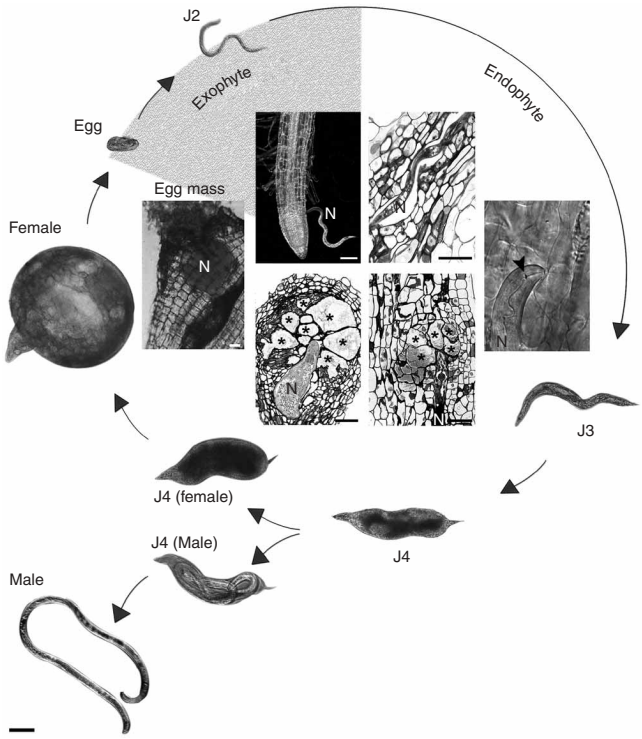



**Figure 1** The parasitic life cycle of *Meloidogyne incognita*. Infective second-stage juveniles (J2) penetrate the root and migrate between cells to reach the plant vascular cylinder. The stylet (arrowhead) connected to the esophagus is used to pierce plant cell walls, to release esophageal secretions and to take up nutrients. Each J2 induces the dedifferentiation of five to seven root cells into multinucleate and hypertrophied feeding cells (*). These giant cells supply nutrients to the nematode (N). The nematode becomes sedentary and goes through three molts (J3, J4, adult). Occasionally, males develop and migrate out of the roots. However, it is believed that they play no role in reproduction. The pear-shaped female produces eggs that are released on the root surface. Embryogenesis within the egg is followed by the first molt, generating second-stage juveniles (J2). Scale bars, 50 μm.

most damaging of all crop pathogens[1]. *M. incognita* is an obligatory sedentary parasite that reproduces by mitotic parthenogenesis[2]. Root-knot nematodes have an intimate interaction with their hosts. Within the host root, adult females induce the redifferentiation of root cells into specialized 'giant' cells, upon which they feed continuously (**Fig. 1**). *M. incognita* can infect *Arabidopsis thaliana*, making this nematode a key model system for the understanding of metazoan adaptations to plant parasitism[3,4] (**Supplementary Data**, section 1 online).

The phylum Nematoda comprises >25,000 described species, many of which are parasites of animals or plants[2]. As many as 10 million species may have yet to be described. Although the model free-living nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* have been the subjects of intensive study[5,6], little is known about the other members of this diverse phylum. These two free-living models will likely not illuminate the biology of nematode parasitism (**Supplementary Fig. 1** online), as shown by the substantial differences between their genome sequences and that of the human parasite *Brugia malayi*[7].

The genome sequence of *M. incognita* presented here provides insights into the adaptations required by metazoans to successfully parasitize and counter defenses of immunocompetent plants, and suggests new antiparasitic strategies.

## RESULTS

### General features of the *M. incognita* genome

The *M. incognita* genome was sequenced using whole-genome shotgun strategy. Assembly with Arachne[8] yielded 2,817 supercontigs, totaling 86 Mb (**Table 1**; **Supplementary Data**, section 2; **Supplementary Fig. 2**; **Supplementary Table 1** online)—almost twice the estimated genome size (47- to 51-Mb haploid genome)[9]. All-against-all comparison of supercontigs revealed that 648 of the longest (covering ~55 Mb) consist of homologous but diverged segment pairs (**Fig. 2**) that might represent former alleles (**Supplementary Data**, section 2; **Supplementary Figs. 3** and **4** online). About 3.35 Mb of the assembly constitutes a third partial copy aligning with these supercontig pairs. Average sequence divergence between the aligned regions is ~8% (**Fig. 3**). A combination of different processes may explain the observed pattern in *M. incognita*, including polyploidy, polysomy, aneuploidy and hybridization[10,11]; all are frequently associated with asexual reproduction. These observations are consistent with a strictly mitotic parthenogenetic reproductive mode, which can permit homologous chromosomes to diverge considerably, as hypothesized for bdelloid rotifers[12] (**Supplementary Data**, section 2.2). No DNA attributable to bacterial endosymbiont genome(s) was identified.

Noncoding DNA repeats and transposable elements represent 36% of the *M. incognita* genome (**Supplementary Data**, section 3; **Supplementary Figs. 5** and **6** and **Supplementary Tables 2** and **3** online). One repeat family with 283 members on 46 contigs encoded the nematode *trans*-spliced leader (SL) exon, SL1, of which 258 members were found associated with a satellite DNA[13] (**Supplementary Fig. 7** online). In nematodes, many mature mRNAs share this 5′ SL exon, and *trans*-splicing is also associated with resolution of polycistronic pre-mRNAs derived from operons. We identified 1,585 candidate

**Table 1 General features of the *Meloidogyne incognita* genome in comparison with the genomes of *B. malayi*[7] and *C. elegans*[5]**

| Features | *M. incognita* | *B. malayi* | *C. elegans* |
|---|---|---|---|
| Overall | | | |
| Estimated size of genome (Mb) | 47–51[a] | 90–95[a] | 100[a] |
| Total size of assembled sequence (Mb) | 86 | 88 | 100 |
| Number of scaffolds and/or chromosomes (chr.) | 2,817 | 8,180 | 6 chr. |
| G + C content (%) | 31.4 | 30.5 | 35.4 |
| Protein-coding regions | | | |
| Number of protein-coding gene models | 19,212 | 11,515 | 20,072 |
| Protein-coding sequence (% of genome) | 25.3 | 17.8 | 25.5 |
| Maximum/average protein length (amino acids) | 5,970/354 | 9,420/343 | 18,562/440 |
| Mean length of intergenic region (bp) | 1,402 | 3,783 | 2,218 |
| Gene density (genes per Mb) | 223 | 162 | 228 |
| Operon number | 1,585 | 926 | 1,118 |
| Percent of genes present in operon | 19 | 18 | 14 |

For *B. malayi* a gene count ranging from 14,500 to 17,800 was inferred after inclusion of genes in the unannotated portion of the genome[7]. For *C. elegans* the gene and protein count is according to Wormpep database (WS183 release).
[a]*M. incognita*: flow cytometry[9]; *B. malayi*: flow cytometry and clone-based[7]; *C. elegans* genome has been completely sequenced telomere to telomere (no gaps) and is exactly 100,291,840 bp[45].

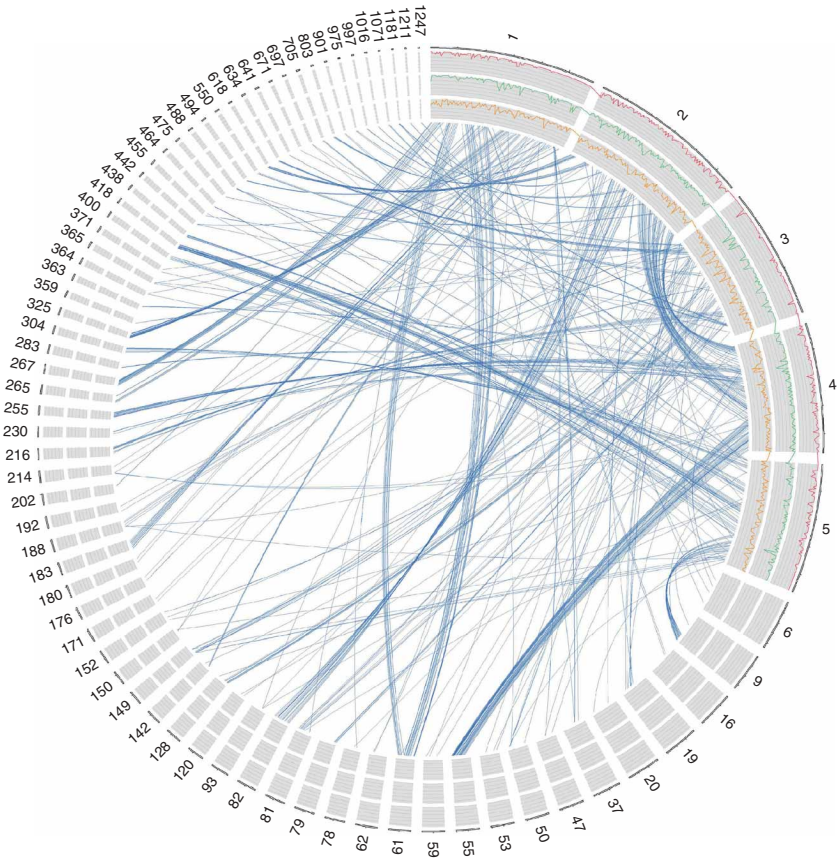


**Figure 2** Allelic-like relationships for the five largest supercontigs of the *M. incognita* assembly. The five largest supercontigs are shown with plots of gene density (orange curve), conservation with *C. elegans* at amino acid level (green curve) and EST density (pink curve). Blue lines represent most similar matches at the protein level between each predicted gene on these five supercontigs and 70 matching supercontigs.

*M. incognita* operons containing a total of 3,966 genes. The two longest operons contained ten genes each and are not allelic copies (**Supplementary Table 4** online). Operons are a dynamic component of nematode genome architecture, as different sets of genes were operonic in *M. incognita*, *C. elegans* and *B. malayi*, and only one operon was found to be strictly conserved between the three nematodes (**Supplementary Data**, section 4; **Supplementary Figs. 8** and **9**; **Supplementary Table 5** online).

## The gene content of a plant-parasitic nematode

The genome sequence was annotated using the integrative gene prediction platform EuGene[14], specifically trained for *M. incognita* (**Supplementary Data**, section 5; **Supplementary Table 6** online). We identified 19,212 protein-coding genes (**Table 1**). Due to the high variation between allelic-like copies (**Fig. 3**) potentially allowing functional divergence, all copies were considered to be different genes. Indeed, 69% of protein sequences were <95% identical to any other (**Supplementary Table 7** and **Supplementary Fig. 10** online). The protein-coding genes occupy 25.3% of the sequence at an average density of 223 genes $Mb^{-1}$, and 36% are supported by expressed sequence tags (ESTs). InterPro protein domains were identified in 55% of proteins and 22% were predicted to be secreted. Comparison of domain occurrence in *M. incognita* with that in *C. elegans* identified an increased abundance of 'pectate lyase', glycoside hydrolase family GH5 and peptidase C48 (SUMO) domains, and fewer chemoreceptor domains. We compared the domain content of the *M. incognita* protein set to those of *C. elegans*, *B. malayi*, *Drosophila melanogaster* and three fungi, of which two are plant pathogens. Thirty-two domains were detected only in *M. incognita*, and two additional domains were only shared between the two plant-pathogenic fungi and *M. incognita*. Functions assigned to the 34 domains specific to plant pathogens encompassed plant cell-wall degradation and chorismate mutase activity (see below). OrthoMCL[15] clustering of the same eight proteomes suggested that 52% of *M. incognita* predicted proteins had no ortholog in the other species. Among them, 1,819 proteins (of which 338 were supported by ESTs) are secreted and lack any known domain (**Supplementary Data**, section 6; **Supplementary Figs. 11** and **12**; **Supplementary Tables 8–10** online). The core complement of proteins in the phylum Nematoda is relatively small: ~23% of the ortholog groups were shared by *M. incognita*, *C. elegans* and *B. malayi* (**Supplementary Fig. 12b**).

## Identifying plant parasitism genes

Nematode proteins produced in and secreted from specialized gland cells into the host are likely to be important effectors of plant parasitism[4,16]. We identified gene products that might be involved in parasitic interaction, particularly those that might modify plant cell walls.

*M. incognita* has an unprecedented set of 61 plant cell wall–degrading, carbohydrate-active enzymes (CAZymes). Although a few such individual CAZymes had been identified previously in some plant-parasitic nematodes and in two insect species[4,16,17], they are absent from all other metazoans studied to date (**Table 2**; **Supplementary Data**, section 7.1; **Supplementary Tables 11–14** online). We identified 21 cellulases and six xylanases from family GH5, two polygalacturonases from family GH28 and 30 pectate lyases from family PL3. We also identified CAZymes not previously reported from metazoans, including two additional plant cell wall–degrading arabinases (family GH43) and two invertases (family GH32). Invertases catalyze the conversion of sucrose (an abundant disaccharide in plants) into glucose and fructose, which can be used by *M. incognita* as a carbon source. We also identified a total of 20 candidate expansins in *M. incognita*, which may disrupt noncovalent bonds in plant cell walls, making the components more accessible to plant cell wall–degrading enzymes[18]. This suite of plant cell wall–degrading CAZymes, expansins and associated invertases was probably acquired by horizontal gene transfer (HGT), as the most similar proteins (outside plant-parasitic nematodes) were bacterial homologs (**Supplementary Table 12**). *M. incognita* also has four secreted chorismate mutases[19], which most closely resemble bacterial enzymes. Chorismate mutase is a key enzyme in biosynthesis of aromatic amino acids and related products, and *M. incognita* may subvert host tyrosine-dependant lignification or defense responses.

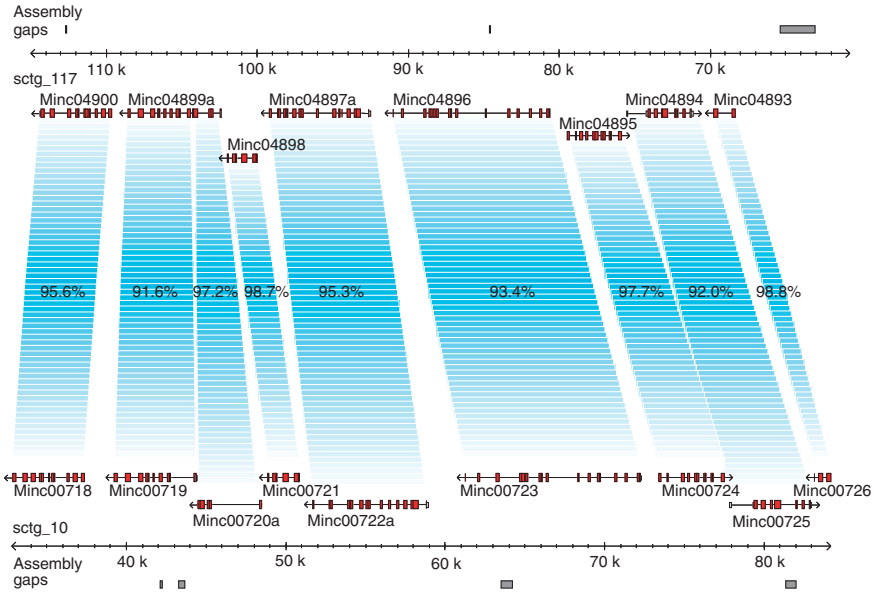



**Figure 3** Example of two allelic-like regions in the *Meloidogyne incognita* assembly. Exons are represented by red boxes and are linked together to form genes (arrows indicate the direction of transcription). Gray boxes show assembly gaps. Highly diverged allelic genes are linked together using blue boxes. Gene order is well conserved between the two allelic-like regions, with only minor differences in predicted gene structure. Percentages of sequence identity at the protein level between the two allelic-like regions are indicated.

Overall, these genes suggest a critical role of HGT events in the evolution of plant parasitism within root-knot nematodes.

Apart from genes restricted to *M. incognita*, we also identified gene families showing substantial expansion compared to *C. elegans*. Among the most notable idiosyncrasies in *M. incognita*, we identified more than 20 cysteine proteases of the C48 SUMO (small ubiquitin-like modifier) deconjugating enzyme family—four times the number in *C. elegans* (**Supplementary Data**, section 7.2; **Supplementary Table 15** online). As some phytopathogenic bacterial virulence factors are SUMO proteases[20], the proteolysis of sumoylated host substrates may be a general strategy used by pathogens to manipulate host plant signal transduction. The *M. incognita* genome also encodes nine serine proteases from the S16 sub-family (Lon proteases), whereas only three are identified in *C. elegans*. These proteases regulate type III protein secretion in phytopathogenic bacteria[21] and may have analogous roles in *M. incognita*.

We identified orthologs to other known candidate plant-parasitic nematode parasitism genes in the genome of *M. incognita*. As most of these gene families are also present in animal-parasitic nematodes and *C. elegans*, *M. incognita* members putatively involved in parasitism were probably recruited from ancestral nematode families (**Supplementary Data**, section 7.3; **Supplementary Table 16** online). Twenty-seven previously described *M. incognita*–restricted pioneer genes expressed in esophageal glands[22] were retrieved in the genome. Eleven additional copies were identified; all remain *Meloidogyne spp.* specific (**Supplementary Data**, section 7.4; **Supplementary Table 17** online). These secreted proteins of as-yet-unknown function are likely targets for novel intervention strategies, and warrant deeper investigation.

## Protection against environmental stresses

One aspect of plant defense responses is the production of cytotoxic oxygen radicals. However, *M. incognita* has fewer genes encoding superoxide dismutases and glutathione peroxidases than *C. elegans* (**Supplementary Data**, section 7.5; **Supplementary Table 18** online). More striking still was the reduction in glutathione *S*-transferases (GSTs) and cytochromes P450 (CYPs), enzymes involved in xenobiotic metabolism and protection against peroxidative damage. Whereas *C. elegans* has 44 GSTs, including representatives from the Omega, Sigma and Zeta classes[23], *M. incognita* possesses only 5 GSTs, all from the Sigma class. Sigma class GSTs are involved in protection against oxidants rather than xenobiotics. A comparable reduction in *gst* genes was observed in *B. malayi*[7]. Similarly, whereas *C. elegans* has 80 different *cyp* genes from 16 families[24], only 27 full or partial *cyp* genes, from 8 families, were identified in *M. incognita*. CYP35 and other families of xenobiotic-metabolizing P450s are absent from *M. incognita* (**Supplementary Data**, section 7.5; **Supplementary Table 18**).

We identified *M. incognita* orthologs of all genes of the innate immunity signaling pathways of *C. elegans*[25] except *trf-1*, which is part of the Toll pathway (**Supplementary Data**, section 7.5; **Supplementary Table 19** online). However, immune effectors such as lysozymes, C-type lectins and chitinases were much less abundant in *M. incognita* than in *C. elegans*. As previously observed in *B. malayi*[7], entire classes of immune effectors known from *C. elegans* were absent from *M. incognita*, including antibacterial genes such as *abf* and *spp*[26] and antifungal genes of several classes (*nlp*, *cnc*, *fip*, *fipr*)[25] (**Supplementary Data**, section 7.5; **Supplementary Table 19**). As plant parasites embedded in root tissues are protected from a variety of biotic and abiotic stresses, we speculate that the reduction and specialization of chemical and immune defense genes is a result of life in this privileged environment.

*C. elegans* has a broad range of unusual fucosylated N-glycan structures compared to other metazoans[27]. *M. incognita* has almost twice as many candidate fucosyltransferases as *C. elegans* (**Supplementary Data**, section 7.1; **Supplementary Table 14**). As suggested for animal-parasitic nematodes, multi-fucosylated structures on the surface of the nematode cuticle could help *M. incognita* to evade recognition[27].

**Table 2 *Meloidogyne incognita* enzymes with predicted plant cell wall–degrading activities, compared with those in *C. elegans* and *D. melanogaster***

| Substrate | Cellulose | Xylan | Arabinan | Pectin | | Other | |
|---|---|---|---|---|---|---|---|
| Family | GH5 (cel) | GH5 (xyl) | GH43 | GH28 | PL3 | EXPN | Total |
| *M. incognita* | 21 | 6 | 2 | 2 | 30 | 20 | 81 |
| *C. elegans* | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *D. melanogaster* | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Number of genes encoding enzymes with candidate activity on different substrate is listed in the three selected species. GH, glycoside hydrolases; PL, polysaccharide lyases; EXPN, expansin-like proteins, following the CAZy nomenclature (http://www.cazy.org/). A total of nine and two cellulose-binding modules of family CBM2 (bacterial type) were found appended to candidate expansins and cellulases, respectively.

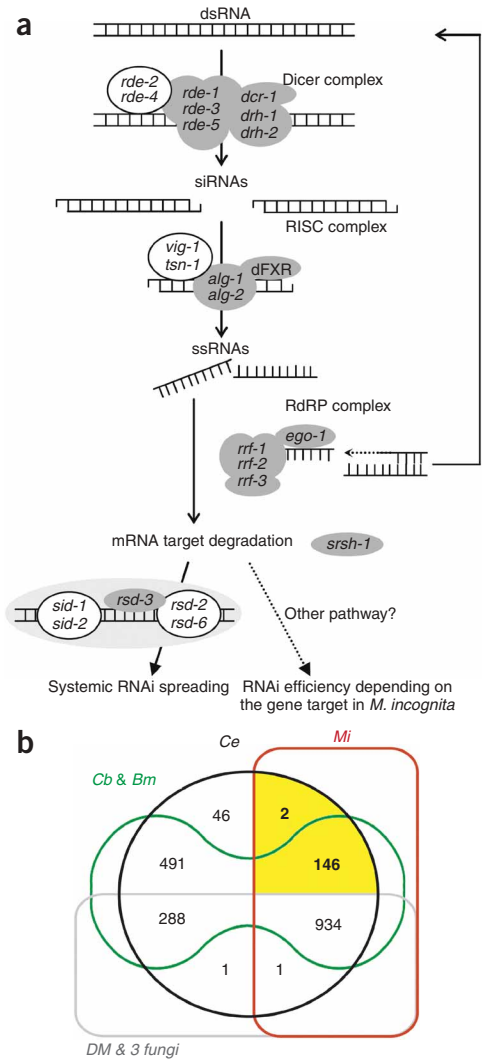



**Figure 4** RNAi pathway and lethal targets. (**a**) Comparison of the RNAi pathway genes of *C. elegans* and *M. incognita*. A gray background indicates that at least one homologous gene was found in *M. incognita*, and a white background indicates that no homologous gene was found in *M. incognita*. (**b**) Distribution of orthologs to *C. elegans* lethal RNAi genes (Ce, black) between *M. incognita* (Mi, red), *C. briggsae* and *B. malayi* (Cb & Bm, green), *D. melanogaster* and three fungi, *N. crassa*, *G. zea* and *M. grisea* (Dm & 3 fungi, gray) using OrthoMCL. A yellow background indicates 148 nematode-only gene clusters.



## Core biological processes

Nuclear receptors, kinases, G-protein coupled receptors (GPCRs) and neuropeptides encompass some of the gene products most extensively involved in core physiological, developmental and regulatory processes.

*C. elegans* has a surprisingly large number of nuclear receptors, but curiously lacks orthologs of many nuclear receptor types conserved in other animals[28]. Some of these conserved nuclear receptors are present in *B. malayi*[7]. Among the 92 predicted nuclear receptors in *M. incognita*, we identified orthologs of several known nematode nuclear receptors, although many of the nuclear receptors present in *B. malayi* and absent in *C. elegans* were also absent in *M. incognita* (**Supplementary Data**, section 7.6; **Supplementary Table 20** online). Many *C. elegans* nuclear receptors are classified as supplementary nuclear receptors (SupNRs), likely derived from a hepatocyte nuclear factor-4-like ancestor[29]. Orthologs of SupNRs were found in *M. incognita*, including a 41-member, *M. incognita*-specific expansion. Fourteen SupNRs are one-to-one orthologs between *B. malayi*, *M. incognita* and *C. elegans*, or conserved only between *M. incognita* and *C. elegans*, with secondary losses in *B. malayi* (**Supplementary Data**, section 7.6; **Supplementary Fig. 13** online). Thus the expansion of SupNRs started before the *Brugia-Meloidogyne-Caenorhabditis* split and has proceeded independently in *C. elegans* and *M. incognita*.

*M. incognita* has 499 predicted kinases compared to 411 in *C. elegans*[30] and 215 in *B. malayi*[7]. The kinases were grouped into 232 OrthoMCL clusters, 24 of which contained only nematode members, suggesting that they have nematode-specific functions. Four kinase families contained only *M. incognita* and *B. malayi* members, suggesting potential roles for these genes in parasitism. Finally, 66 kinase families, containing 122 genes, appear to be *M. incognita*-specific (**Supplementary Data**, section 7.7; **Supplementary Table 21** online). Seven percent (1,280) of all *C. elegans* genes are predicted to encode GPCRs that play crucial roles in chemosensation. These *C. elegans* genes have been divided into three serpentine receptor superfamilies and five solo families[31]. *M. incognita* has only 108 GPCR genes and these derive from two of the three serpentine receptor superfamilies and one of the solo families. These *M. incognita* chemosensory genes are commonly found as duplicates clustered on the genome, as observed in *C. elegans* (**Supplementary Data**, section 7.8; **Supplementary Fig. 14**; **Supplementary Table 22** online).

Neuropeptide diversity is remarkably high in nematodes, given the structural simplicity of their nervous systems. *C. elegans* has 28 Phe-Met-Arg-Phe-amide-like peptide (*flp*) and 35 neuropeptide-like protein (*nlp*) genes encoding ~200 distinct neuropeptides[32]. The identified neuropeptide complement of *M. incognita* is smaller: 19 *flp* genes and 21 *nlp* genes. However, two *flp* genes, *Mi-flp-30* and *Mi-flp-31*, encode neuropeptides that have not been identified in *C. elegans*, suggesting that they could fulfill functions specific to a phytoparasitic lifestyle (**Supplementary Data**, section 7.9; **Supplementary Table 23** online).

The XX-XO sex determination pathway in *C. elegans* is intimately linked to the dosage compensation pathway[33]. *M. incognita* reproduces exclusively by mitotic parthenogenesis, and males do not contribute genetically to production of offspring[11]. *M. incognita* also displays an environmental influence on sex determination: under less favorable environmental conditions far more males are produced. These males can arise due to sex reversal[34] and intersexual forms can be produced. *M. incognita* homologs of at least one member of each step of the *C. elegans* sex determination cascade were identified, including *sdc-1* from the dosage compensation pathway, *tra-1*, *tra-3* and *fem-2* from the sex determination pathway itself, and also downstream genes such as *mag-1* (which represses male-promoting genes) and *mab-23* (which controls male differentiation and behavior). In addition, a large family (~35 genes) of *M. incognita* secreted proteins, similar to the C2H2 zinc finger motif–containing *tra-1* from *C. elegans*, was identified (**Supplementary Data**, section 7.10; **Supplementary Table 24** online). It is therefore possible that *M. incognita* uses a similar genetic system for sex determination, but with the male pathway also modulated in response to environmental cues.

Taken together, these comparative analyses of genes, underpinning important traits, highlight the huge biodiversity in the phylum Nematoda. Idiosyncrasies identified in *M. incognita* may account for

its parasitic lifestyle and lead to the development of new control strategies directed against plant-parasitic nematodes.

### RNA interference and lethal phenotypes

RNA interference (RNAi) is a promising technology for the functional analysis of parasitic nematode genes. RNAi can be induced in *M. incognita* by feeding, with variable silencing efficiencies depending on the gene target[35,36]. *M. incognita* has many genes of the *C. elegans* RNAi pathway, including components of the amplification complex (*ego-1*, *rrf-1*, *rrf-2* and *rrf-3*). However, we found no homologs of *sid-1*, *sid-2*, *rsd-2* and *rsd-6*, which are genes involved in systemic RNAi and double-stranded RNA spreading to surrounding cells (**Fig. 4**, **Supplementary Data**, section 7.11; **Supplementary Table 25** online). These genes are also absent from *B. malayi*[7] and *Haemonchus contortus*[37], suggesting that systematic RNAi may spread through the action of novel or poorly conserved factors. We retrieved 2,958 *C. elegans* genes having a lethal RNAi phenotype and searched for orthologs in *M. incognita*. Among the 1,083 OrthoMCL families identified, 148 (containing 344 *M. incognita* genes) appear to be nematode specific (**Supplementary Data**, section 7.12). Because of their lethal RNAi phenotype and distinctive sequence properties, these genes provide an attractive set of new antiparasite drug targets.

## DISCUSSION

The genome of *M. incognita* has many traits that render it particularly attractive for studying the fundamentals of plant parasitism in the Nematoda. One remarkable feature is that most of the genome is composed of pairs of homologous segments that may denote former diverged alleles. This suggests that *M. incognita* is evolving without sex toward effective haploidy through the Meselson effect[38–40]. As the *M. incognita* genome is the first one sequenced and assembled for a strictly parthenogenetic species, we expect that its comparison with sexual nematode genomes will shed light on mechanisms leading to its peculiar structure. Functional divergence between ancient alleles of genes involved in the host-parasite interface could explain the extremely wide host range and geographic distribution of this polyphagous nematode. Analysis of the gene content of *M. incognita* revealed a suite of plant cell wall–degrading enzymes, which has no equivalent in any animal studied to date. The striking similarity of these enzymes to bacterial homologs suggests that these genes were acquired by multiple HGT events. Just as many instances of bacterial HGT involve sets of genes implicated in adaptations to new hosts or food sources, the candidate HGT events in *M. incognita* involve genes with potential roles in interactions with hosts. The alternative hypothesis—that these genes were acquired vertically from a common ancestor of bacteria and nematodes and lost in most eukaryote lineages—appears less parsimonious. Other singularities encompass *M. incognita*-restricted secreted proteins or lineage-specific expansions and/or reductions that may play roles in host-parasite interaction.

Transcriptional profiling, proteomic analysis and high throughput RNAi strategies are in progress and will lead to a deeper understanding of the processes by which a nematode causes plant disease. Combining such knowledge with functional genomic data from the model host plant *A. thaliana* should provide new insights into the intimate molecular dialog governing plant-nematode interactions and allow the further development of target-specific strategies to limit crop damage. Through the use of comparative genomics, the availability of free-living, animal- and plant-parasitic nematode genomes should provide new insights into parasitism and niche adaptation.

## METHODS

**Strain and DNA extraction.** We used the *M. incognita* strain 'Morelos' from the root-knot nematode collection held at INRA (Institut National de la Recherche Agronomique) Sophia Antipolis, France. Nematode eggs were collected in a sterile manner from tomato roots and checked for the presence of plant material contaminants. DNA was extracted as described in **Supplementary Methods**, section 8.1 online.

**Genome sequencing and assembly.** We obtained paired-end sequences from plasmid and BAC libraries with the Sanger dideoxynucleotide technology on ABI3730xl DNA analyzers. The 1,000,873 individual reads were assembled in 2,817 supercontigs using Arachne[8] (**Supplementary Methods**, section 8.2; **Supplementary Table 26** online).

**Genome structure, operons and noncoding elements.** The assembled genome was searched for repetitive and non-coding elements. Scaffolds were aligned to determine pairs and triplets of allelic-like regions. Gene positions along scaffolds were used to predict clusters of genes forming putative operons (**Supplementary Methods**, section 8.3–8.7).

**Prediction of protein coding genes.** Gene predictions were performed using EuGene[14], optimized for *M. incognita* models and tested on a data set of 230 nonredundant, full-length cDNAs. Translation starts and splice sites were predicted by SpliceMachine[41]. Available *M. incognita* ESTs were aligned on the genome using GenomeThreader[42]. Similarities to *C. elegans* and other species' protein, genome and EST sequences were identified using BLAST[43]. Repetitive sequences were masked using RepeatMasker (http://repeatmasker.org/, **Supplementary Methods**, section 8.8; **Supplementary Fig. 15** online).

**Automatic functional annotation.** Protein domains were searched with InterproScan[44]. We also submitted proteins from seven additional species to the same InterproScan search. We included three other nematodes (*C. elegans*, *C. briggsae* and *B. malayi*), the fruitfly (*D. melanogaster*) and three fungi (*Magnaporthe grisea, Gibberella zea* and *Neurospora crassa*). To identify clusters of orthologous genes between *M. incognita* and the seven additional species, we used OrthoMCL[15] (**Supplementary Methods**, section 8.9).

**Expert functional annotation.** The collection of predicted protein coding genes was manually annotated by a consortium of laboratories. Each laboratory focused on a particular process or gene family relevant to the different aspects of *M. incognita* biology. Patterns of presence and/or absence and expansion and/or reduction in comparison to *C. elegans*, and other species were examined. The quality of predicted genes was manually checked and a functional annotation was proposed accordingly (**Supplementary Methods**, sections 8.10–8.20). A genome browser and additional information on the project are available from http://meloidogyne.toulouse.inra.fr/.

**Accession codes.** The 9,538 contigs resulting from the *Meloidogyne incognita* genome assembly and annotation were deposited in the EMBL/Genbank/DDBJ databases under accession numbers CABB01000001–CABB01009538.

*Note: Supplementary information is available on the Nature Biotechnology website.*

### ACKNOWLEDGMENTS

SCRI laboratory (V.C.B. and J.T.J.) received funding from the Scottish Government. This work benefited from links funded via COST Action 872. G.V.M. and V.L. are supported by ARC, CNRS, EMBO, MENRT and Region Rhone-Alpes. G.V.M., M.R.-R. and V.L. are also funded by the EU Cascade Network of Excellence and the integrated project Crescendo. M.-C.C. is supported by MENRT. We thank Philippe Lecomte for critical reading of the manuscript and all our collaborators from the "Plant-Nematode interaction" team of INRA Sophia Antipolis for technical help and support.

### AUTHOR CONTRIBUTIONS

P.A. and J.G. contributed equally as first authors. J.-M.A., P.C.-S., E.G.J.D., E.D. and L.P.-B. contributed equally as second authors. T.J.B., M.B., T.B-Z., E.L.D., J.J.E., B.F., E.G., B.H., J.T.J., V.L., A.G.M., H.Q., M.-N.R., T.S., G.S., J.W. and P.W. contributed equally as senior authors. P.A., M.B., P.C.-S. and E.G.J.D. wrote the manuscript with input from J.T.J. and A.G.M. For biological material,

contributions were as follows. F.D., M.M. and L.P.-B. for strain growth, control and selection and DNA extraction. P.A., M.-C.C., F.D., E.D., B.F., M.-N.R. and L.P.-B. for cDNA libraries and EST data. For genome sequencing and assembly, contributions were as follows. B.S., E.U., J.P., V.A. for sequencing. C.J. for assembly. C.D. for cDNA clustering and library analyses. J.-M.A., O.J., C.J., F.A. for bioinformatics of allelism characterization. J.W. and P.W. supervision and coordination of the sequencing. For genome structure and organization, contributions were as follows. P.C.-S., T.F., H.Q. and D.S. for repetitive and transposable elements. J.G., E.S. for rRNAs, tRNAs, miRNAs. M.B. for operonic structures. M.-N.R., E.S. and C.V.G. for splice leaders (SL). For *in-silico* global genome analysis, contributions were as follows. E.D., J.G. and T.S. for gene predictions, automatic functional annotation, databases and bioinformatics. E.D. and B.F. for global protein set comparative analysis. Proteome expert annotation was as follows: P.M.C., E.G.J.D. and B.H., for Carbohydrate-Active enZymes. P.C.-S. and E.G. for proteases. M.-C.C., E.L.D., M.E., B.F., E.G.J.D., E.D., E.G., J.T.J., N.H., L.P.-B., G.S. and T.T. for candidate nematode parasitism and pioneer genes. P.A., T.B.-Z., E.G.J.D., E.D., J.J.E., J.V.G., G.P. and M.-N.R. for protection against plant defenses and immune system. V.L., G.V.M. and M.R.-R. for nuclear receptors. T.J.B., T.H. and T.R.M. for the kinome. E.G.J.D. and L.P.-B. for GPCRs. T.B.-Z., F.D.L., P.L. and P.V. for collagen. A.G.M. and P.M.V. for neuropeptides. J.T.J. for sex determination. V.C.B., E.G.J.D. and L.P.-B. for RNAi pathway and lethal RNAi phenotypes.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/





1. Trudgill, D.L. & Blok, V.C. Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. *Annu. Rev. Phytopathol.* **39**, 53–77 (2001).
2. Blaxter, M.L. Nematoda: genes, genomes and the evolution of parasitism. *Adv. Parasitol.* **54**, 101–195 (2003).
3. Caillaud, M.C. *et al.* MAP65–3 Microtubule-associated protein is essential for nematode-induced giant cell ontogenesis in *Arabidopsis*. *Plant Cell* **20**, 423–437 (2008).
4. Caillaud, M.C. *et al.* Root-knot nematodes manipulate plant cell functions during a compatible interaction. *J. Plant Physiol.* **165**, 104–113 (2008).
5. The C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
6. Stein, L.D. *et al.* The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* **1**, E45 (2003).
7. Ghedin, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **317**, 1756–1760 (2007).
8. Jaffe, D.B. *et al.* Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13**, 91–96 (2003).
9. Leroy, S., Duperray, C. & Morand, S. Flow cytometry for parasite nematode genome size measurement. *Mol. Biochem. Parasitol.* **128**, 91–93 (2003).
10. Triantaphyllou, A.C. in *An Advance Treatise on Meloidogyne* vol. 1 (eds. Sasser, J.N. & Carter, C.C.) 113–126, (North Carolina State University Graphics, Raleigh, USA, 1985).
11. Castagnone-Sereno, P. Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. *Heredity* **96**, 282–289 (2006).
12. Mark Welch, D.B., Cummings, M.P., Hillis, D.M. & Meselson, M. Divergent gene copies in the asexual class Bdelloidea (Rotifera) separated before the bdelloid radiation or within bdelloid families. *Proc. Natl. Acad. Sci. USA* **101**, 1622–1625 (2004).
13. Piotte, C., Castagnone-Sereno, P., Bongiovanni, M., Dalmasso, A. & Abad, P. Cloning and characterization of two satellite DNAs in the low-C-value genome of the nematode *Meloidogyne* spp. *Gene* **138**, 175–180 (1994).
14. Foissac, S. & Schiex, T. Integrating alternative splicing detection into gene prediction. *BMC Bioinformatics* **6**, 25 (2005).
15. Li, L., Stoeckert, C.J. Jr. & Roos, D.S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
16. Davis, E.L., Hussey, R.S. & Baum, T.J. Getting to the roots of parasitism by nematodes. *Trends Parasitol.* **20**, 134–141 (2004).
17. Wei, Y.D. *et al.* Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry longicorn beetle, *Apriona germari*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **145**, 220–229 (2006).
18. Qin, L. *et al.* Plant degradation: a nematode expansin acting on plants. *Nature* **427**, 30 (2004).
19. Lambert, K.N., Allen, K.D. & Sussex, I.M. Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*. *Mol. Plant Microbe Interact.* **12**, 328–336 (1999).
20. Hotson, A. & Mudgett, M.B. Cysteine proteases in phytopathogenic bacteria: identification of plant targets and activation of innate immunity. *Curr. Opin. Plant Biol.* **7**, 384–390 (2004).
21. Tang, X., Xiao, Y. & Zhou, J.M. Regulation of the type III secretion system in phytopathogenic bacteria. *Mol. Plant Microbe Interact.* **19**, 1159–1166 (2006).
22. Huang, G. *et al.* A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode *Meloidogyne incognita*. *Mol. Plant Microbe Interact.* **16**, 376–381 (2003).
23. Lindblom, T.H. & Dodd, A.K. Xenobiotic detoxification in the nematode *Caenorhabditis elegans*. *J. Exp. Zoolog. A Comp. Exp. Biol.* **305**, 720–730 (2006).
24. Menzel, R., Bogaert, T. & Achazi, R. A systematic gene expression screen of *Caenorhabditis elegans* cytochrome P450 genes reveals CYP35 as strongly xenobiotic inducible. *Arch. Biochem. Biophys.* **395**, 158–168 (2001).
25. Ewbank, J.J. Signaling in the immune response. *WormBook* doi/10.1895/wormbook.1.83.1, <http://www.wormbook.org/> (2006).
26. Alegado, R.A. & Tan, M.W. Resistance to antimicrobial peptides contributes to persistence of Salmonella typhimurium in the *C. elegans* intestine. *Cell Microbiol.* **10**, 1259–1273(2008).
27. Paschinger, K., Gutternigg, M., Rendic, D. & Wilson, I.B. The N-glycosylation pattern of *Caenorhabditis elegans*. *Carbohydr. Res.* **343**, 2041–2049 (2007).
28. Bertrand, S. *et al.* Evolutionary genomics of nuclear receptors: from 25 ancestral genes to derived endocrine systems. *Mol. Biol. Evol.* **21**, 1923–1937 (2004).
29. Robinson-Rechavi, M., Maina, C.V., Gissendanner, C.R., Laudet, V. & Sluder, A. Explosive lineage-specific expansion of the orphan nuclear receptor HNF4 in nematodes. *J. Mol. Evol.* **60**, 577–586 (2005).
30. Plowman, G.D., Sudarsanam, S., Bingham, J., Whyte, D. & Hunter, T. The protein kinases of *Caenorhabditis elegans*: a model for signal transduction in multicellular organisms. *Proc. Natl. Acad. Sci. USA* **96**, 13603–13610 (1999).
31. Robertson, H.M. & Thomas, J.H. The putative chemoreceptor families of *C. elegans*. *WormBook* doi/10.1895/wormbook.1.66.1, <http://www.wormbook.org/> (2006).
32. Marks, N.J. & Maule, A.G. in *Neuropeptide Systems as Targets for Parasite and Pest Control* (eds. Geary, T.G. & Maule, A.G.) (Landes Bioscience/Eurekah.com, Austin, TX, USA, 2008).
33. Zarkower, D. Somatic sex determination. *WormBook* doi/10.1895/wormbook.1.84.1, <http://www.wormbook.org/> (2006).
34. Papadopoulou, J. & Triantaphyllou, A.C. Sex-determinantion in *Meloidogyne incognita* and anatomical evidence of sexual reversal. *J. Nematol.* **14**, 549–566 (1982).
35. Rosso, M.N., Dubrana, M.P., Cimbolini, N., Jaubert, S. & Abad, P. Application of RNA interference to root-knot nematode genes encoding esophageal gland proteins. *Mol. Plant Microbe Interact.* **18**, 615–620 (2005).
36. Huang, G., Allen, R., Davis, E.L., Baum, T.J. & Hussey, R.S. Engineering broad root-knot resistance in transgenic plants by RNAi silencing of a conserved and essential root-knot nematode parasitism gene. *Proc. Natl. Acad. Sci. USA* **103**, 14302–14306 (2006).
37. Zawadzki, J.L., Presidente, P.J., Meeusen, E.N. & De Veer, M.J. RNAi in Haemonchus contortus: a potential method for target validation. *Trends Parasitol.* **22**, 495–499 (2006).
38. Birky, C.W. Jr. Bdelloid rotifers revisited. *Proc. Natl. Acad. Sci. USA* **101**, 2651–2652 (2004).
39. Mark Welch, D. & Meselson, M. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**, 1211–1215 (2000).
40. Mark Welch, D.B., Mark Welch, J.L. & Meselson, M. Evidence for degenerate tetraploidy in bdelloid rotifers. *Proc. Natl. Acad. Sci. USA* **105**, 5145–5149 (2008).
41. Degroeve, S., Saeys, Y., De Baets, B., Rouze, P. & Van de Peer, Y. SpliceMachine: predicting splice sites from high-dimensional local context representations. *Bioinformatics* **21**, 1332–1338 (2005).
42. Gremme, G., Brendel, V., Sparks, M.E. & Kurtz, S. Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* **47**, 965–978 (2005).
43. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
44. Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116–W120 (2005).
45. Hillier, L.W. *et al.* Genomics in *C. elegans*: so many genes, such a little worm. *Genome Res.* **15**, 1651–1660 (2005).