Molecular & Biochemical Parasitology 178 (2011) 7–14

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

Analysis of the transcriptome of the root lesion nematode *Pratylenchus coffeae* generated by 454 sequencing technology

Annelies Haegeman, Soumi Joseph, Godelieve Gheysen*

Ghent University, Dpt. Molecular Biotechnology, Coupure links 653, B-9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 25 January 2011
Received in revised form 30 March 2011
Accepted 4 April 2011
Available online 12 April 2011

Keywords:
Root lesion nematode
Pratylenchus
Expressed sequence tags
Transcriptome
454 sequencing

ABSTRACT

To study interactions between plants and plant-parasitic nematodes, several omics studies have nowadays become extremely useful. Since most data available so far is derived from sedentary nematodes, we decided to improve the knowledge on migratory nematodes by studying the transcriptome of the nematode *Pratylenchus coffeae* through generating expressed sequence tags (ESTs) on a 454 sequencing platform.

In this manuscript we present the generation, assembly and annotation of over 325,000 reads from *P. coffeae*. After assembling these reads, 56,325 contigs and singletons with an average length of 353 bp were selected for further analyses. Homology searches revealed that 25% of these sequences had significant matches to the Swiss-prot/trEMBL database and 29% had significant matches in nematode ESTs. Over 10,000 sequences were successfully annotated, corresponding to over 6000 unique Gene Ontology identifiers and 5000 KEGG orthologues. Different approaches led to the identification of different sequences putatively involved in the parasitism process. Several plant cell wall modifying enzymes were identified, including an arabinogalactan galactosidase, so far identified in cyst nematodes only. Additionally, some new putative cell wall modifying enzymes are present belonging to GHF5 and GHF16, although further functional studies are needed to determine the true role of these proteins. Furthermore, a homologue to a chorismate mutase was found, suggesting that this parasitism gene has a wider occurrence in plant-parasitic nematodes than previously assumed. Finally, the dataset was searched for orthologues against the Meloidogyne genomes and genes involved in the RNAi pathway.

In conclusion, the generated transcriptome data of *P. coffeae* will be very useful in the future for several projects: (1) evolutionary studies of specific gene families, such as the plant cell wall modifying enzymes, (2) the identification and functional analysis of candidate effector genes, (3) the development of new control strategies, e.g. by finding new targets for RNAi and (4) the annotation of the upcoming genome sequence.

1. Introduction

The root lesion nematode *Pratylenchus coffeae* is a parasite of banana causing lesions, necrosis and toppling of the plants. It has a worldwide distribution, probably due to spreading of infected banana planting material. It also causes damage to various other crops such as yam, ginger, turmeric, abaca and coffee. It is the most important nematode causing damage to banana in South East Asia, Central and South America, and the Pacific. In Africa, its distribution is more localized, and it generally occurs in mixed populations with other nematodes such as *Radopholus similis*, *Helicotylenchus multicinctus* and Meloidogyne species [1].

In recent years, several plant-parasitic nematodes have been subjected to molecular analyses, especially transcriptome analyses by means of expressed sequence tags (ESTs). Currently, approximately 175,000 ESTs from plant-parasitic nematodes have been submitted to the NCBI database, all derived from traditional Sanger sequencing (December 2010). The sedentary root-knot nematodes (Meloidogynidae) have been extensively studied with the generation of over 70,000 ESTs and the sequencing of two complete genomes of *Meloidogyne incognita* [2] and *Meloidogyne hapla* [3]. The Pratylenchidae, a family of migratory nematodes, is the family most closely related to the Meloidogynidae [4]. It currently accounts for approximately 15,000 ESTs, of which less than 8000 are derived from the genus Pratylenchus. A molecular comparison of members of the Meloidogynidae and Pratylenchidae could provide insights on the differences and similarities between sedentary and migratory nematodes. Keeping this in mind, we decided to characterize the transcriptome of a mixed-stage *P. coffeae* population by 454 pyrosequencing. The latter technique has become a relatively rapid and cost-effective method for high-throughput sequencing of ESTs of non-model organisms. So far, 454 sequencing

* Corresponding author. Tel.: +32 92645888; fax: +32 92646219.
E-mail address: Godelieve.Gheysen@UGent.be (G. Gheysen).

0166-6851/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2011.04.001

has only been used to generate ESTs from animal-parasitic nematodes [5–7], for plant-parasitic nematodes no reports on 454 sequencing studies are available yet. One of the goals of this project was to identify putative effector genes by several approaches. More specifically, we wanted to look if *P. coffeae* possesses a similar arsenal of plant cell wall modifying enzymes as the Meloidogyne species. The latter enzymes are necessary for plant-parasitic nematodes to penetrate the rigid plant cell wall. Numerous enzymes have been identified in different families of plant-parasitic nematodes, such as endo-1,4-beta-glucanase, xylanase, pectate lyase, polygalacturonase, arabinogalactan galactosidase, and arabinase [8]. These genes may have been acquired by horizontal gene transfer from bacteria and fungi [9]. In Pratylenchus species, so far only an endo-1,4-beta-glucanase has been identified [10], although there is also EST evidence for an expansin, a protein known to loosen the cell wall non-enzymatically [11].

2. Materials and methods

2.1. RNA extraction, cDNA synthesis and sequencing

P. coffeae was cultured on carrot discs at a constant temperature of 25 °C. RNA was isolated from mixed stages using the TRI reagent (Sigma) according to the manufacturer's instructions. First strand cDNA synthesis was done with the Super SMART PCR cDNA synthesis kit (Clontech, CA, USA) including an amplification step of 20 cycles as described in the manual. Subsequently, the amplified cDNA was purified using the Qiaquick PCR purification kit (Qiagen, Germany) and normalized using the TRIMMER kit (Evrogen). The normalized cDNA sample was sent to LGC Genomics (Berlin, Germany), where it was sequenced in two separate runs of ¼ of a picotiter plate on a 454 FLX Titanium platform (Roche, Branford, CT, USA) by a shotgun approach. The data was submitted to the NCBI Sequence Read Archive (SRA) with accession number SRA028814.

2.2. Cleaning and assembly

The resulting reads were processed with the CLC Genomics Workbench 4.0.2 software. SMART adapter sequences and 454 sequencing primers were trimmed from all reads. Additionally, low quality reads (<99.5% accuracy) and short reads (<50 bp) were discarded. The assembly was done using standard settings.

2.3. Homology searches

All contigs and singletons longer than 150 bp were blasted locally (blastx) against Swiss-prot and trEMBL (October 2010) with a bit score cut off of 50. Additionally, all nematode ESTs were downloaded from the EST division of Genbank, and split into three different datasets according to the nematode's lifestyle: animal-parasitic nematodes, plant-parasitic nematodes and free-living nematodes. A local tblastx search (bit score > 50) looked for homologues for all sequences in these datasets. Since a lot of plant cell wall degrading enzymes in nematodes are thought to originate through horizontal gene transfer, we tried to identify HGT candidates. Therefore we did blast searches (bit score > 50) against different datasets of plant, nematode and bacterial sequences. We downloaded the coding sequences of the following genomes from the RefSeq database of NCBI: *Caenorhabditis elegans* (NC_003279–NC_003284), *Brugia malayi* (NZ_AAQA00000000), *Arabidopsis thaliana* (NC_003070–NC_003071, NC_003074–NC_003076) and all completed genomes of plant pathogenic bacteria and an oomycete (*Pectobacterium atrosepticum*, NC_004547; *Ralstonia solanacearum*, NC_003295; *Xylella fastidosa* 9a5c, NC_002488; *Agrobacterium tumefaciens*, NC_003062; *Xanthomonas campestris* pv. campestris, NC_003902; *Xanthomonas axonopodis* pv. citri, NC_003919; *Pseudomonas syringae* pv. syringae, NC_007005; *Xylella fastidosa* Temecula 1, NC_004556; *Pseudomonas syringae* pv. tomato, NC_004578; *Leifsonia xyli* subsp. xyli, NC_006087; *Pseudomonas syringae* pv. phaseolicola, NC_005773; *Xanthomonas campestris* pv. campestris, NC_007086; *Xanthomonas campestris* pv. vesicatoria, NC_007508; Aster yellows witches broom Phytoplasma, NC_007716; *Clavibacter michiganensis* subsp. sepedonicus, NC_010407; *Candidatus Phytoplasma* mai, NC_011047; *Dickeya dadantii*, NC_012880; *Dickeya zeae*, NC_012912; *Phytophthora infestans* T30-4, NZ_AATU00000000).

All putative proteins from the genomes of the root-knot nematodes *M. incognita* and *M. hapla* were downloaded from the projects' websites [2,3]. To look for orthologues, a reciprocal blast strategy was used: the Pratylenchus sequences longer than 150 bp were blasted (blastx, bit score > 50) against the Meloidogyne proteins as well as the opposite strategy (tblastn, bit score > 50). Only when both pairs of blast hits were the same, they were considered as true orthologues.

2.4. Annotation

All Pratylenchus sequences longer than 150 bp were annotated based on the blastx results against Swiss-prot and trEMBL. A sequence was annotated based on the top hit information, only if the bit score > 50 and if the description of the top hit did not contain any terms that would suggest it is a hypothetical or unknown protein ("unknown", "putative", "uncharacterized", "hypothetical", "similar", "predicted", "probable"). Gene Ontology terms were retrieved for all unique protein identifiers from annotated sequences using QuickGO from the EBI website (http://www.ebi.ac.uk/QuickGO/GAnnotation). KEGG orthologues were identified using the KEGG Automated Annotation Server (KAAS) with default parameters [12]. Subsequently, KEGG BRITE mapping was applied to find the most common classifications.

2.5. Translation into putative proteins

To predict putative proteins, the sequences longer than 150 bp were translated using OrfPredictor [13]. To look for putative parasitism genes, the presence of signal peptides was predicted with SignalP 3.0 [14] and transmembrane domains were predicted using TMHMM 2.0 [15].

2.6. Searching for specific genes

As described above, three sequence sets putatively related to parasitism were retained: the first dataset was derived from homology to plant pathogenic bacteria and/or plants only, the second one was derived from homology to parasitic nematode ESTs exclusively, and the third one was derived from putative proteins with a signal peptide. The blastx hits of these datasets were retrieved and manually searched for the presence of putative plant cell wall modifying enzymes. The sequences that showed similarity to these genes were locally blasted (tblastx, bit score > 50) against all Pratylenchus sequences longer than 150 bp to identify any additional family members.

The following putative effector genes were retrieved from Genbank and used for homology searches: 10A06, 14-3-3b, 16D10, 19C07, 7E12, acid phosphatase, annexin, calreticulin, chitinase, chorismate mutase, CLE peptide, ERp99, galectin, glutathione peroxidase, glutathione-S-transferase, map-1, nodL factor, peroxiredoxin, SPRYSEC RBP-1, RING-H2 zinc finger protein, fatty acid and retinol binding protein or SEC-2, SKP1-like protein, SXP/RAL-2, transthyretin-like protein, ubiquitin extension protein and venom allergen protein. Accession numbers and references can be found