



Article

# Insight into the Lifestyle of Amoeba *Willaertia magna* during Bioreactor Growth Using Transcriptomics and Proteomics

Issam Hasni <sup>1,2,3</sup>, Philippe Decloquement <sup>1</sup>, Sandrine Demaneche <sup>2</sup> , Rayane Mouh Mameri <sup>2</sup>, Olivier Abbe <sup>2</sup>, Philippe Colson <sup>1,3</sup> and Bernard La Scola <sup>1,3,\*</sup>

<sup>1</sup> Aix-Marseille University, Institut de Recherche pour le Développement IRD 198, Assistance Publique—Hôpitaux de Marseille (AP-HM), Microbes, Evolution, Phylogeny and Infection (MEΦI), UM63, 13005 Marseille, France; issemhasni@gmail.com (I.H.); Philippe.DECLOQUEMENT@univ-amu.fr (P.D.); philippe.colson@univ-amu.fr (P.C.)

<sup>2</sup> R&D Department, Amoéba, 69680 Chassieu, France; sandrine.demaneche@amoeba-biocide.com (S.D.); mameri77@yahoo.fr (R.M.M.); Olivier.Abbe@amoeba-biocide.com (O.A.)

<sup>3</sup> Institut Hospitalo-Universitaire (IHU)—Méditerranée Infection, 13005 Marseille, France

\* Correspondence: bernard.la-scola@univ-amu.fr; Tel.: +33-4-9132-4375; Fax: +33-4-9138-7772

Received: 5 May 2020; Accepted: 18 May 2020; Published: 21 May 2020



**Abstract:** *Willaertia magna* C2c maky is a thermophilic free-living amoeba strain that showed ability to eliminate *Legionella pneumophila*, a pathogenic bacterium living in the aquatic environment. The amoeba industry has proposed the use of *Willaertia magna* as a natural biocide to control *L. pneumophila* proliferation in cooling towers. Here, transcriptomic and proteomic studies were carried out in order to expand knowledge on *W. magna* produced in a bioreactor. Illumina RNA-seq generated 217 million raw reads. A total of 8790 transcripts were identified, of which 6179 and 5341 were assigned a function through comparisons with National Center of Biotechnology Information (NCBI) reference sequence and the Clusters of Orthologous Groups of proteins (COG) databases, respectively. To corroborate these transcriptomic data, we analyzed the *W. magna* proteome using LC-MS/MS. A total of 3561 proteins were identified. The results of transcriptome and proteome analyses were highly congruent. Metabolism study showed that *W. magna* preferentially consumed carbohydrates and fatty acids to grow. Finally, an in-depth analysis has shown that *W. magna* produces several enzymes that are probably involved in the metabolism of secondary metabolites. Overall, our multi-omic study of *W. magna* opens the way to a better understanding of the genetics and biology of this amoeba.

**Keywords:** *Willaertia magna*; amoeba; environmental; ecological; bioreactor; transcriptomics; proteomics; biocide

## 1. Introduction

Most free-living amoebas (FLAs) are classified in two suprankingdom-level groups: *Amoebozoa* including *Acanthamoeba* spp. or *Vermamoeba* spp. and *Excavata* comprising *Naegleria* spp., *Vahlkampfia* spp. or *Willaertia* spp. [1,2]. *Willaertia magna* is a FLA belonging to family *Vahlkampfiidae* [3]. This protist presents a three-stage life, switching between a trophozoite form, a flagellate form and a cystic form [4,5]. The *W. magna* trophozoite (50–100 µm in size) form corresponds to the main active stage during which this thermophilic amoeba feeds, moves and divides; it can grow at various temperatures ranging from 22 to 44 °C in xenic or axenic culture. Under unfavorable conditions, the amoeba adopts a cystic form (18–21 µm in size), which allows its survival until the conditions

become favorable for cell growth. The transformation to the flagellate stage occurs temporally, being triggered by various stressors [3].

FLAs graze on microorganisms to grow, and this regulates the microbial population in the environment [6,7]. However, some bacteria, such as *Legionella pneumophila*, have evolved to bypass the mechanism of elimination in order to survive and multiply within amoebae [8–10]. *L. pneumophila* is a human pathogen causing legionellosis, a disease transmitted by the inhalation of contaminated aerosol [11]. This Gram-negative bacterium is ubiquitous in natural and man-made fresh water environments such as cooling towers [12,13]. It has the ability to infect and invade a wide range of amoebae, including *Vermamoeba*, *Acanthamoeba* and *Naegleria* species [14]. Furthermore, the amoeba cysts serve as protection for *L. pneumophila* in adverse conditions making legionella management in cooling tower water using chemical biocides difficult [15,16]. Amoebae can adopt different behaviors with respect to *L. pneumophila*. If *Vermamoeba* and *Acanthamoeba* appear as highly susceptible to a wide range of *Legionella* spp., previous studies reported the elimination of pathogenic *L. pneumophila* strain Paris 1 (ATCC 33152) by *W. magna* C2c maky [17–20]. The Amoéba® company developed a natural biocide based on the property of *W. magna* C2c maky to control *L. pneumophila* proliferation and as an alternative to chemical biocides in cooling tower water [21]. To treat cooling tower water with this natural biocide, it is necessary to produce a large number of amoebae in a short time. However, traditional methods based on axenic amoebal cultivation in adhesion [4,22] do not provide high enough cell densities in a short time [23]. Therefore, high-throughput production of cells in bioreactor culture is a strategy developed to improve the productivity [24]. This process is usually used for the mass production of bacteria, fungi or microalgae of pharmaceutical interest [25–27].

In a previous study, the analysis of the *W. magna* genome indicated a size of 36 megabase pairs (Mbp) and a set of 18,561 genes [28]. The gene content survey included a detailed analysis of genes potentially involved in pathogenicity and of sequence transfers with pathogenic microorganisms. This study failed to detect virulence factors and showed the non-pathogenicity in vitro of *W. magna* [28]. Furthermore, we observed putative exchanges for 136 genes belonging to amoeba-resistant microorganisms [28].

To expand our knowledge on *W. magna* C2c maky, we explored the transcriptome as well as the proteome of this amoeba. RNA-seq is an effective method to understand and identify the level of expression of genes involved in the different metabolic pathways and biochemical processes [29]. It is accurately complemented by proteome analysis for these goals [30,31]. Furthermore, we compared our results with *Naegleria gruberi* data [32,33], a non-pathogenic amoeba that is phylogenetically close to *W. magna*, and with data on *Acanthamoeba castellanii*, which is the most studied FLA due to its interest regarding human health and its potential role in the ecosystem [34–36].

## 2. Materials and Methods

### 2.1. *W. magna* C2c maky Culture in Bioreactor

*W. magna* C2c maky cells (ATCC® PTA-7824) were precultured in SCGYEM medium [37] supplemented with 10% calf serum contained in 175 cm<sup>2</sup> culture flask at 37 °C. Exponentially growing cells were harvested and used to inoculate a 10-L bioreactor (GPC, La Rochelle, France; [38]) in SCGYEM medium without fetal calf serum (pH = 7) at a temperature varying between 30 and 45 °C under helix constant agitation (200 rpm) and oxygenation. The amoeba cells were harvested from a bioreactor at a volume of 50 mL and were centrifugated at 2000 × g for 10 min followed by three steps of washing using Page's modified Neff's Amoeba Saline medium (2 mM NaCl, 16 μM MgSO<sub>4</sub>, 27.2 μM CaCl<sub>2</sub>, 1 mM Na<sub>2</sub>HPO<sub>4</sub>, 1 mM KH<sub>2</sub>PO<sub>4</sub>). Amoeba quantification was performed using a KOVA® slide cell counting chamber (KOVA International, CA, USA).

### 2.2. Illumina Sequencing of the Transcriptome

The total RNA of each sample was extracted using a RNeasy Mini Kit (Qiagen, Hilden, France) according to the manufacturer's instructions. RNA was quantified and qualified using a QuantiFluor

RNA sample system (Promega, Charbonnières, France) and a nano RNA chip on a BioAnalyser 2100 (Agilent Technologies Inc., Santa Clara, CA, United States), respectively. Illumina sequencing was performed using ProfileXpert/Viroscan3D, Lyon. Briefly, libraries were prepared from total RNA using poly(A) enrichment of the mRNA to remove ribosomal RNA. NextFlex rapid directional RNAseq sample prep (Bio Scientific Corporation) was used to achieve the libraries. Quantification and validation of the libraries were performed using dsDNA HS Assay on a Quantus Fluorometer (Promega, Charbonnières, France) and on a BioAnalyser 2100 from Agilent using a HS DNA chip (Agilent Technologies Inc., Santa Clara, CA, United States), respectively. The library was sequenced in 75 base pair (bp) length paired end reads in NextSeq 500 Mid Output flow cell lines from Illumina (Illumina Inc, San Diego, CA, USA) generating 216.9 million raw read pairs.

### 2.3. Transcriptomic Analysis of *W. Magna* C2c Maky

The quality assessment of raw reads was checked using FastQC software (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) for reads with low quality ( $Q < 28$ ), and the adaptors were removed using Trimmomatic [39]. The identification of transcripts was performed as in the study conducted by Aherfi et al. [40]. Briefly, reads were mapped on the *W. magna* C2c maky assembled genome using Hisat2 with default parameters [41]. The expression of genes was quantified using Htseq [42], a software that processes data from high-throughput sequencing assays. This software identifies and counts the number of reads mapped to each gene. Predicted open reading frames (ORF) with a coverage of six reads or more were considered as transcribed [40]. This threshold was determined in order to obtain transcripts with high coverages (61 reads/ORFs).

### 2.4. Functional Annotation

To assign the biological functions of *W. magna* transcripts, we proceeded to homology searches for the transcripts in public protein databases. Predicted protein sequences were used as queries against the nr database using an e-value threshold of less than  $1e-03$ . Protein sequences were aligned to the Cluster Orthologous Groups of proteins (COG) database using EggNOG [43,44] with diamond as mapping mode. To improve understanding of the biological functions and metabolic pathways of the genes, the sequences were mapped on Kyoto Encyclopedia of Genes and Genomes Pathway (KEGG) [45] using BLASTp with an e-value cutoff of  $1e-3$ . Conserved protein domains were identified by mapping of proteins against the conserved domain database [46] and InterProScan database [47]. The nonribosomal peptide synthase and polyketide synthase domains were predicted using 2metDB [48] and NaPDoS [49] software. To predict the localization of proteins in the cell, we used TargetP 2.0 and Mitop2 [50,51]. To perform an analysis of genes related to anaerobic metabolism, we performed a BLASTp search against a non-redundant protein sequences database (nr), with an e-value cutoff of  $1e-03$ . Alignment of protein sequences was carried out using MUSCLE [52]. Phylogenetic construction was obtained using the maximum likelihood method with the Jones–Taylor–Thornton (JTT) model on MEGA 7.0.25 software [53]. Phylogenetic trees were analyzed using iTOL v3 online [54].

### 2.5. Proteomics

#### 2.5.1. Cell Lysis

Two samples of *W. magna* C2c maky obtained from two independent bioreactors were used in the proteomic study. *W. magna* pellets, containing  $5 \times 10^8$  cells, were lysed in 2 mL of buffer 1X phosphate buffered saline (pH 8.0) and 1% sodium dodecyl sulfate (SDS) by gram of pellet for 2 h at  $4^\circ\text{C}$  under agitation. Lysates were then centrifuged at  $3000 \times g$  for 15 min at  $4^\circ\text{C}$  to remove cellular debris, and the supernatants were stored at  $4^\circ\text{C}$ . Total protein concentration in lysate fractions was determined using the Pierce micro bicinchoninic acid (BCA) protein assay kit (Thermo Fisher Scientific, Illkirch, France) by using bovine serum albumin (BSA) as a standard.

### 2.5.2. SDS-Page and Western-Blot

Protein lysate samples were denatured and reduced in SDS/TCEP-loading buffer (Genentech, Inc., So. San Francisco, CA, US) and separated on a 4%–15% acrylamide gel (4%–15% Criterion<sup>®</sup> TGX Stain-Free<sup>™</sup> Gel, Bio-Rad, Hercules, CA, US). Migration was performed in Tris- Glycine-SDS buffer (Euromedex, Souffelweyersheim, France). The acrylamide gel was then activated under UV light with a ChemiDoc<sup>™</sup> MP system (Bio-Rad, Hercules, CA, US) for Stain-Free<sup>™</sup> detection. The SDS-PAGE gel was subsequently Coomassie-stained for 1 h with InstantBlue Ultrafast Protein Stain (Sigma-Aldrich Chimie S.a.r.l. Lyon, France) and proteins were visualized with a ChemiDoc<sup>™</sup> MP system (Bio-Rad, Hercules, CA, US).

### 2.5.3. Liquid Digestion

Following the protein precipitation step (using trichloroacetic acid 20% in volume, overnight at 4 °C), samples were washed in acetone twice and solubilized in 8 M urea. Then, samples were reduced (Tris(2-CarboxyEthyl)Phosphine, 5 mM, 57 °C, 1h), alkylated (iodoacetamide, 10 mM, RT, 45 min), and digested for 5 h at 37 °C with LysC and overnight at 37 °C with trypsin (1/100 ratio). Peptides digest was next fractionated on a high pH reversed phase fractionation spin column (Thermo Scientific) according to the manufacturer's instructions. The 8 fractions obtained were dried in a speed vacuum before nanoLC-MS/MS analysis and then suspended in 40 µL 0.1% HCOOH.

### 2.5.4. NanoLC-MS/MS Analysis

The fractions were analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose California) coupled online with a Q Exactive HF mass spectrometer via a nano-electrospray ionization source (Thermo Scientific, San Jose California).

2 µL of peptide mixtures were loaded on a C18 PepMap100 trap-column (75 µm i.d. × 2 cm, 5 µm, 100 Å, Thermo Fisher Scientific) for 3 min at 5 µL/min with 2% ACN, 0.05% TFA in H<sub>2</sub>O and then separated on a C18 Acclaim PepMap100 nano-column, 50 cm × 75 µm i.d, 2 µm, 100 Å (Thermo Scientific) with a 60 min linear gradient from 3.2% to 40% buffer B (A: 0.1% FA in H<sub>2</sub>O, B: 0.1% FA in ACN) and then from 40% to 90% of B in 2 min, held for 10 min and returned to the initial conditions in 1 min for 15 min. The total duration was set to 90 min at a flow rate of 300 nL/min. The oven temperature was kept constant at 40 °C.

The sample was analysed with the TOP20 HCD method: MS data were acquired in a data dependent strategy selecting the fragmentation events based on the 20 most abundant precursor ions in the survey scan (350–1600 Th). The resolution of the survey scan was 60,000 at m/z 200 Th. The Ion Target Value for the survey scans in the Orbitrap and the MS<sup>2</sup> mode were set to 3E6 and 1E5 respectively and the maximum injection time was set to 60 ms for both scan modes. Parameters for acquiring HCD MS/MS spectra were as follows: collision energy = 27 and an isolation width of 2 m/z. The precursors with unknown charge state or a charge state of 1 were excluded. Peptides selected for MS/MS acquisition were then placed on an exclusion list for 20 s using the dynamic exclusion mode to limit duplicate spectra.

## 2.6. Data Analysis

The data were converted to mgf format using RawConverter [55]. For each sample, the 8 fractions were merged into a single dataset and the resulting peak lists were searched against the protein sequences of *W. magna* C2c maky using Peaks Studio software (Bioinformatics Solutions Inc, Waterloo, Canada) [56]. The search was performed using the following settings, based on the *W. magna* genome database (18,519 sequences): peptide mass error tolerance, 25 ppm; fragment mass error tolerance, 0.1 Da; monoisotopic mass values, one missed cleavage, no non-specific cleavage, fixed modifications of carbamidomethylation; variable modifications of oxidation (M), deamidation (NQ), carbamylation

and oxidation or hydroxylation; 4 maximum variable post-translational modifications (PTMs) per peptide. The peptides identified were filtered based on an FDR (false discovery rate) cut-off of 0.5%.

### 3. Results

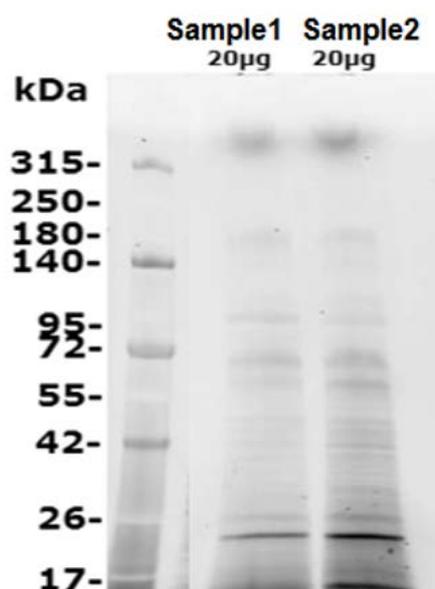
#### 3.1. Transcriptomic Analysis

RNA sequencing generated 217 million raw reads in total with 75 bp for both paired ends. After removal of adaptor sequences and improving the stringency quality, we obtained 14 gigabases of cleaned data, and 49,015,843 nucleotides (nt) were generated. To analyze gene expression, the reads were mapped against the *W. magna* C2c genome, resulting in 86.1% of reads mapped. The expression of the *W. magna* genes was examined and 8,790 *W. magna* C2c transcripts (47.4% of *W. magna* ORFs) were detected. Among the 8,790 transcripts, BLASTp analysis against the nr database found homologous sequences for 6,179 (70.3%) of the transcribed genes and identified 2,611 ORFans among the transcript genes (29.7%) (Table S1). Of the ten genes with the highest coverage (mean coverage of 4,316 reads/gene), two transcripts were assigned to a function (bactericidal permeability-increasing protein/lipopolysaccharide-binding protein and histidine triad motif (HIT) family protein), five were annotated as hypothetical proteins and two as ORFans (Table S1). To further analyze the most highly expressed genes, we studied the conserved domains of hypothetical proteins and ORFans. Among the five hypothetical proteins, we found a conserved domain for three predict genes, including a glutathione S-transferase domain, a WD40 domain and a flavoprotein domain (Table S1). For one of two ORFs with no match in the nr database, we identified one F-box domain (Table S1).

#### 3.2. Mean Proteomic Information

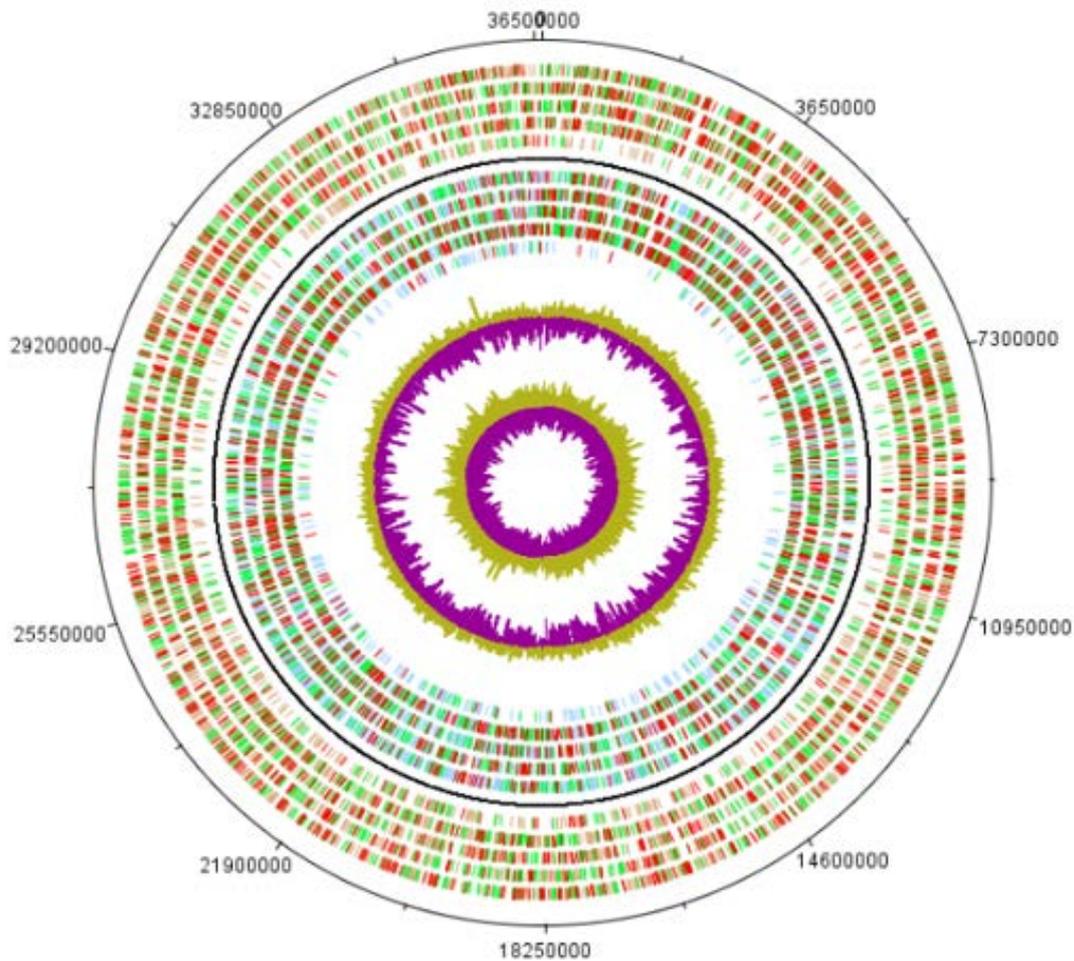
The RNA-seq analysis was supplemented with a proteomic analysis carried out on two biological replicates of *W. magna* grown in the bioreactor.

The lysate concentrations of replicates 1 and 2 were estimated using the micro BCA method at 11.6 and 15.6 mg/mL, respectively. A total of 20 µg of each sample were separated on SDS-PAGE gel and proteins were visualized using Stain-Free™ detection (Figure 1).



**Figure 1.** Detection of amoebal proteins using SDS-PAGE analysis. After separation on SDS-PAGE gel, proteins were visualized using stain-free™ detection sample 1 and sample 2: total proteins.

Expressed proteins were identified using the Peaks Studio program (Table S2) [57]. A total of 106,383 spectra along with 64,260 peptides were identified. In total, we detected 3561 proteins, which represented 19.2% of the predicted genes and 40% of the transcriptome (Figure 2 and Table S1).

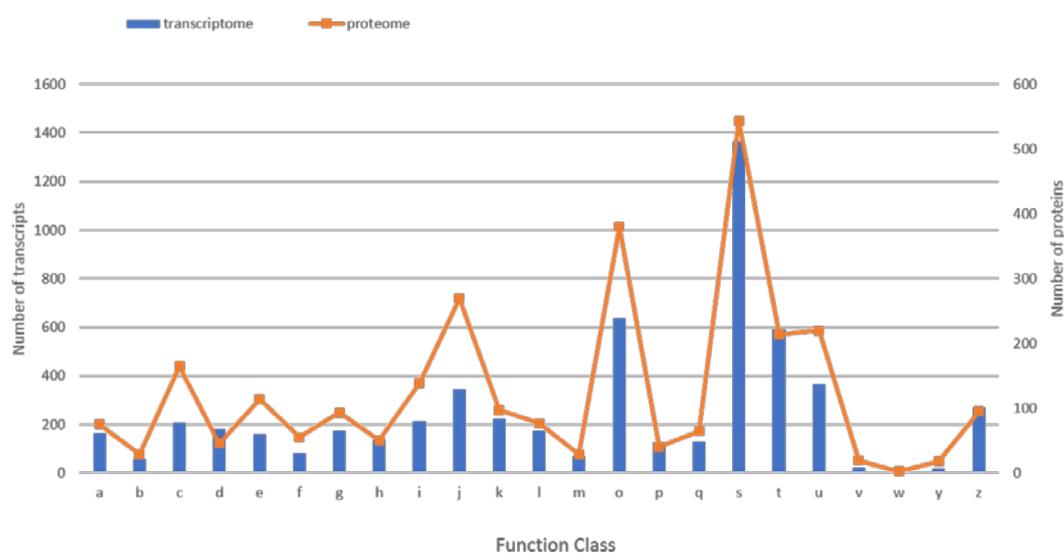


**Figure 2.** Circular map of the *W. magna* genome. Outside circle: genes on the forward strand colored by light brown are predicted genes; green bars are transcripts and red are proteins. Second to outside circle: genes on the reverse strand colored by light blue are predicted genes; green bars are transcripts and red bars are proteins. In the center: GC content and GC skew.

As we processed the proteomic analysis twice, we detected 2970 proteins from the first analysis and 3186 proteins from the second one (Table S2). The annotation using BLASTp against the National Center of Biotechnology Information (NCBI) protein sequence database assigned a putative function to 3376 proteins, among which 3013 had as best hits proteins belonging to *N. gruberi* and one best matched with a giant virus (*Moumouvirus* Monve) protein. The taxonomical distribution revealed a very high proportion of genes shared with eukaryotes (97.9%), followed by genes shared with bacteria (1.9%), archaea (0.15%) and viruses (0.03%) (Figure S1).

### 3.3. COG Enrichment Analysis

To further analyze the function of the transcripts and proteins, we conducted a comparison of the latter with the COG database (Figure 3).

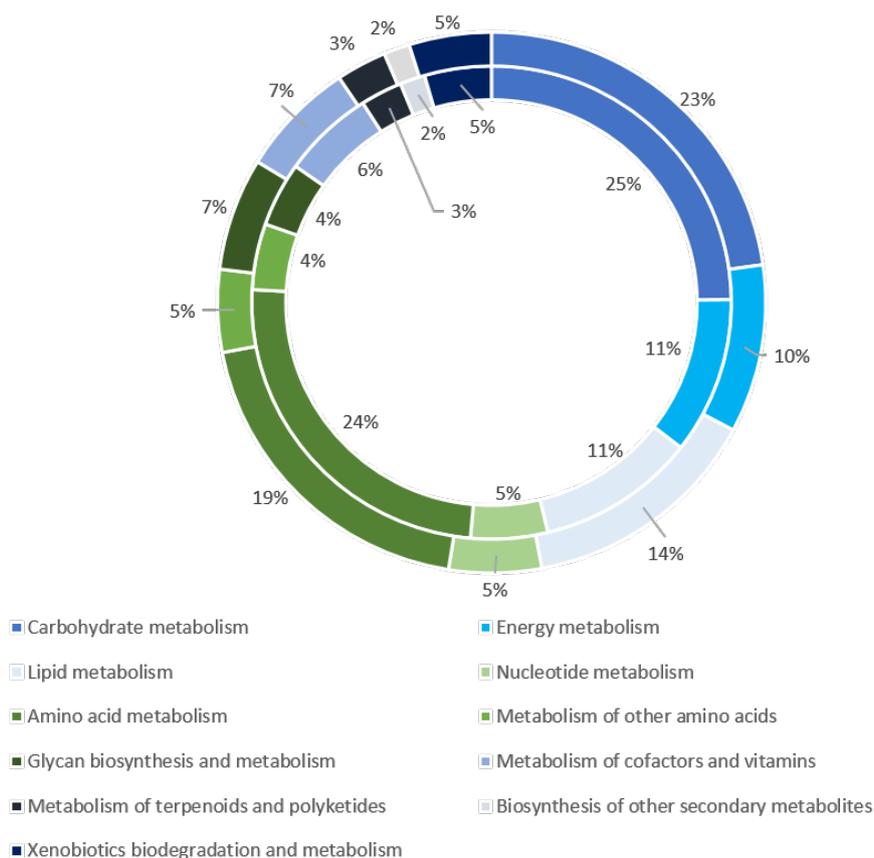


**Figure 3.** Representation of *W. magna* transcripts and proteins related to different functional categories in COG database. A: RNA processing and modification, B: Chromatin structure and dynamics, C: Energy production and conversion, D: Cell cycle control, cell division, chromosome partitioning, E: Amino acid transport and metabolism, F: Nucleotide transport and metabolism, G: Carbohydrate transport and metabolism, H: Coenzyme transport and metabolism, I: Lipid transport and metabolism, J: Translation, ribosomal structure and biogenesis, K: Transcription, L: Replication, recombination and repair, M: Cell wall/membrane/envelope biogenesis, O: Posttranslational modification, protein turnover, chaperones, P: Inorganic ion transport and metabolism, Q: Secondary metabolites biosynthesis, transport and catabolism, S: Function unknown, T: Signal transduction mechanisms, U: Intracellular trafficking, secretion and vesicular transport, V: Defense mechanisms, W: Extracellular structures, Y: Nuclear structure, Z: Cytoskeleton.

A total of 5341 (60.7%) and 2723 (76.5%) transcripts and proteins, respectively, were assigned to COG categories (Table S1). Overall, these transcripts and proteins were distributed in 23 COG categories. Regarding transcripts, the category “function unknown” (1366: 25%) was the most represented, followed by “post-translational modification, protein turnover, chaperones” (636: 12%), “signal transduction mechanisms” (593: 11%), “intracellular trafficking, secretion and vesicular transport” (366: 7%) and “translation, ribosomal structure and biogenesis” (346: 6%). Regarding proteins, the COG category in which the greatest number of proteins were assigned is the category “function unknown” (543: 20%) followed by “post-translational modification, protein turnover, chaperones” (380: 14%), “translation, ribosomal structure and biogenesis” (270: 10%), “intracellular trafficking, secretion and vesicular transport” (220: 8%) and “signal transduction mechanisms” (214: 8%). Categories “nuclear structure” (18: 0.3% transcripts and 18: 0.7% proteins) and “extracellular structure” (10: 0.2% and 3: 0.1% proteins) were weakly represented among *W. magna* transcripts and proteins. Thus, the COG analysis revealed a similar top five categories for transcripts and proteins.

### 3.4. KEGG Enrichment Analysis

An enrichment analysis with KEGG pathways was performed in order to get a better insight into the biological function and pathways of the expressed genes and proteins. Particularly, the KEGG study made it possible to identify the proteins involved in biochemical metabolism, the signal transduction pathway and genetic information processing. There were 3341 transcripts and 1999 proteins mapped into 368 and 362 KEGG pathways, respectively. Among them, 1173 transcripts and 846 proteins were assigned to metabolic pathways classified into 11 groups (Figure 4).



**Figure 4.** Representation of the *W. magna* transcripts and proteins related to metabolism. The transcripts and proteins are represented by the outer ring and inner ring, respectively.

The map with the greatest number of transcripts and proteins was the carbohydrate pathway (267 transcripts and 220 proteins), followed by amino acid metabolism (229 transcripts and 217 proteins), lipid metabolism (167 transcripts and 95 proteins) and energy metabolism (118 transcripts and 95 proteins). This study identified various numbers of transcripts and proteins involved in glycolysis, pyruvate metabolism, lipid degradation and amino acid degradation (Table S3). These metabolisms seemed to be the main pathways used by the amoeba to provide the pyruvate for the Krebs cycle and energy cells. Moreover, analysis of KEGG pathways revealed that 18 transcripts were classified into 10 sub-categories belonging to a secondary metabolite category, including genes involved in the streptomycin and penicillin biosynthesis. The latter were also found at the proteomic level. In addition, we found that 755 genes and 527 proteins matched to genetic information processing, involving transcription, translation, folding, sorting, degradation, replication and repair. Finally, KEGG analysis identified 1408 genes and 773 proteins classified into membrane transport, signal transduction and environmental adaptation.

### 3.5. *W. magna* Shape and Movement

The investigation on *W. magna* motility revealed the presence of 349 transcripts and 151 proteins related to the cytoskeleton (Table S4). As for in silico analyses for *N. gruberi* and *A. castellanii* previously published, the cytoskeleton of *W. magna* seems to be mainly composed of microtubules and actin filaments [32,58]. Among the genes related to actin, we identified actin proteins that are the globular components of the microfilament (Table S4). In addition, we found actin-related proteins and myosin, which regulates actin microfilaments through its interaction with actin proteins [58] (Table S4). We detected components involved in microfilament organization including a Rho-family protein (Ras homolog family member) (Table S4). RhoA activates diaphanous-related formins, thereby contributing

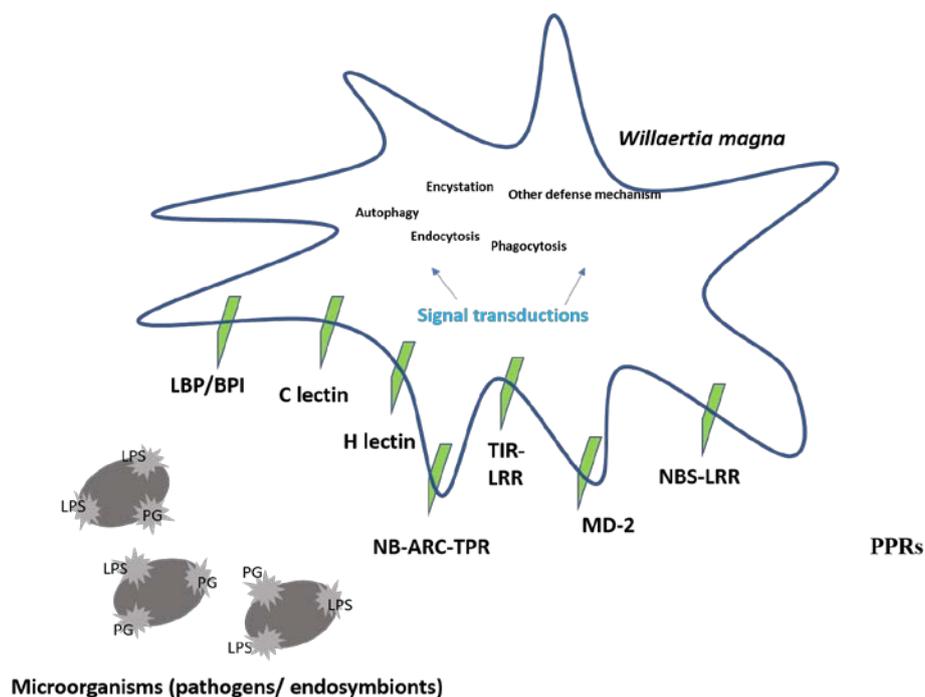
to the stimulation of actin polymerization [59]. The cell motility and cellular transport of *W. magna* are regulated by microtubules. Coherently, we found the presence of several transcripts (132) and proteins (31) related to microtubules, such as tubulins or motor proteins (dynein and kinesin). The association of protein motors and microtubules plays a critical role in intracellular organelle transport and is involved in the maintenance of cell shape [58]. We also detected microtubule proteins that make up the internal structure of the flagella (Table S4). In addition to genes related to cytoskeleton, some specific genes related to flagellar formation were identified (cilia- and flagella-associated protein 43, cilia- and flagella-associated protein 74 and long flagella protein lf4). This might suggest that *W. magna* is temporarily present in its flagellate form during suspension culture, but such form could not be observed using microscopy (unpublished data). Among the genes related to movement, we found that the genes encoding long flagella protein lf4, zinc finger, regulator of chromosome condensation, tubulin and dynein were the most expressed (Table S4).

### 3.6. Metabolism

The analysis of *W. magna* metabolism enabled a set of enzymes involved in glycolysis to be identified (Table S5). The presence of glucokinase, ribokinase and fructokinase revealed that the *W. magna* is able to use glucose and also several other monosaccharides for this carbohydrate requirement. *W. magna* shares some similarities with *N. gruberi* regarding the glycolysis pathway [33]. Indeed, *W. magna* does not have a hexokinase, an enzyme involved in the first step of glycolysis for the phosphorylation of all hexoses. Nevertheless, we reported the expression of a hexokinase homolog (glucokinase), which is a glucose-specific enzyme. Furthermore, the phosphorylation of fructose 6-phosphate in the second step of glycolysis is not carried out by classical ATP-dependent phosphofructokinase (PFK) but is catalyzed by a pyrophosphate-fructose 6-phosphate 1-phosphotransferase (ppi-PFK). The hexose-monophosphate pathway is another metabolic pathway used by *W. magna*, which degrades the glucose in pyruvate. Moreover, we detected all the enzymes involved in the consumption of glycerol that may serve as an energy substrate (Table S5). Finally, we identified components involved in the Krebs cycle and mitochondrial respiratory chain. *W. magna* is an aerobic organism that uses oxygen as terminal acceptor. However, we reported the presence of a nitrate reductase, which is an enzyme involved in anaerobic metabolism (Table S5) [32,33]. Furthermore, our analysis revealed the presence of FeFe-hydrogenase and one associated maturase (FeFe-H<sub>2</sub>ase maturase proteins HydG). These enzymes are involved in energy metabolism for microaerophilic or anaerobic microorganisms producing molecular hydrogen [35,60]. Of these three proteins, two seemed to have N-terminal mitochondrial transit peptides (Table S5). These three proteins produced by *W. magna* shared best hits with *Naegleria gruberi* proteins. The phylogenetic reconstructions based on these genes related to anaerobic respiration showed a clustering with their homologs in *Naegleria* organisms (Figures S2–S4). All these genes related to anaerobic respiration were weakly expressed (Table S5).

### 3.7. Defense Mechanisms

To characterize the defense mechanisms of *W. magna*, we searched for the presence of predicted pattern-recognition receptors (PRRs) (Figure 5 and Table S6).



**Figure 5.** Mechanism of pattern recognition and pattern-recognition receptors (PRRs) identified within *Willaertia magna*: Lipopolysaccharide (LPS), Proteoglycan (PG), Lipopolysaccharide binding protein/Bactericidal Permeability-Increasing protein (LPS/BPI), C-lectin, H-lectin, Nucleotide-Binding Adaptor R-gene and Tetratricopeptide Repeat (NB-ARC-TPR), Toll/Interleukin-1 Receptor and Leucine-Rich Repeat (TIR-LRR domain), Nucleotide Binding Site and Leucine Rich Repeat (NBS-LRR domain) and MD-2.

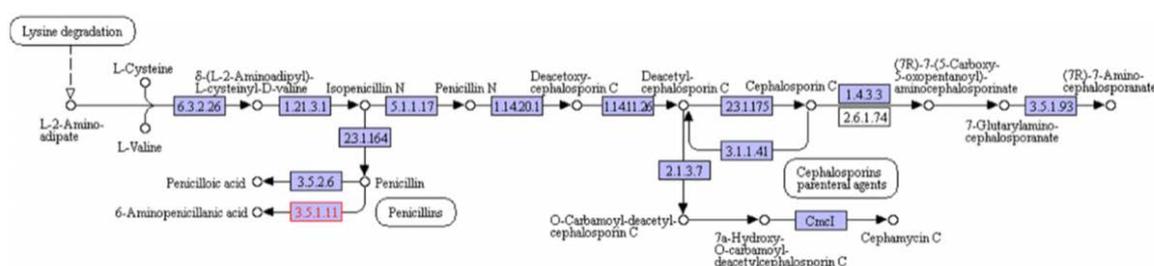
*W. magna* contains three members of the bactericidal permeability-increasing protein/lipopolysaccharide-binding protein family and a MD-2-related protein. All these four receptors play an important role in triggering defensive responses through their interactions with lipopolysaccharide of bacteria [61]. This gene shared a best hit in the nr database with a *N. gruberi* sequence. Furthermore, we reported the presence of C- and H-lectin domains (three) that belong to a large family of receptors that bind carbohydrates and induce endocytic, phagocytic and antimicrobial responses [35] (Table S6). Four mannose-binding proteins were revealed within the *W. magna* transcriptome. Among them, two best matched with *N. gruberi*, one with *Tetrahymena thermophila* and one with *Rathayibacter tanacetii*. These glycoproteins are involved in the endocytosis of *L. pneumophila* by *A. castellanii* [34]. In *A. castellanii* and *Dictyostelium discoideum*, leucine-rich repeat (LRR) containing nucleotide-binding adaptor R-gene (NB-ARC) tetratricopeptide repeat (TPR) or Toll/interleukin-1 receptor (TIR) domains are assumed to be involved in the immune responses [35,62]. *W. magna* was predicted here to possess 18 LRR-containing proteins according to its transcriptome; 17 have a transmembrane domain, two are homologs of a NB-ARC-TPR, one is homologous to a nucleotide binding site and leucine rich repeat (NBS-LRR) domain and one is homologous to a TIR domain, such as those found in Toll-like receptors. Although *W. magna* is produced in high-throughput axenic culture, we identified the genes involved in the bacterial degradation. Indeed, we identified a homologous of Nox2, an important protein of the NADPH-oxidase membrane-bound enzyme complex that is implicated in the oxidative burst of phagocytic cells and plays a role in bacterial killing by free radicals [63]. This gene was highly transcribed, but we had not identified this protein within proteome data. The *W. magna* also displayed a large range of enzymes involved in the hydrolysis of carbohydrates, such as chitinases, lysozymes,  $\beta$ -hexosaminidases and amylase. In addition, we reported the presence of different isoforms of cathepsins, a group of lysosomal hydrolases involved in protein degradation [64,65]. Finally, we found transcripts and proteins related to the

phago-lysosomal system (phospholipases and lipases), which is consistent with the capability of *W. magna* to digest bacteria (Table S6).

*W. magna* possesses the essential weapons to deal with microorganism attacks. However, all the genes identified in bacterial defense and destruction are either transcribed at a low level (such as amylase) or they could have another function, including a role in metabolic pathways, molecular signaling or transport. These results suggested that these genes may be transcribed for another reason, other than bacterial defense in an axenic medium (Table S6).

### 3.8. Ecological Context

Amoebae cope with a variety of microorganisms in the environment. To survive and defend themselves, they produce a repertoire of several classes of proteins. We identified transcripts and proteins involved in various metabolic pathways related to the biosynthesis of secondary metabolites (Figure S5 and Table S7). Among them was a penicillin amidase transcribed at low level. This enzyme serves for the biosynthesis of penicillin derivatives (Figure 6).



**Figure 6.** Representation of metabolic pathway performed on KEGG website. The map shows the pathway of secondary metabolite biosynthesis. In red color are represented the enzymes found in transcriptome and proteome of *W. magna*.

We also identified two genes with multidomains related to polyketide synthases producing natural metabolites with antimicrobial properties [66,67] (Figure S5). Among them, we found a gene encoding polyketide synthase containing ketide synthase, acyltransferase, thioesterase, enoyl reductase and ketoreductase domains (Figure S5). This gene, sharing the best hit with a *N. gruberi* gene, was expressed at a high level. Interestingly, several transcripts ( $n = 33$ ) and proteins (20) were related to xenobiotic biodegradation and metabolism, encompassing 2-haloacid dehalogenase (K01560) and alkane 1-monooxygenase (K00496), which are enzymes specifically involved in chloroalkane and chloroalkene degradation and in caprolactam degradation, respectively (Table S2). These enzymes have crucial functions in the degradation of persistent organic pollutants, which cause substantial problems for the environment [68]. Moreover, we identified the presence of ATP-binding cassette (ABC) transporters that constitute a protein family involved in the import and export of a large variety of substrates and in the regulation of several cellular processes, such as chemotaxis [69]. These transporters could have played the role of detoxifiers by exporting xenobiotics and endogenous secondary metabolites [70]. Furthermore, we found several enzymes involved in the metabolism of terpenoid compounds, such as terpene and squalene, many of which have pharmaceutical or biological properties [71] (Table S7).

## 4. Discussion

This study analyzed for the first time the transcriptome and proteome of *W. magna*. It provides unique data on the molecular mechanisms and metabolic pathways used by *W. magna* C2c maky during axenic mass culture. Firstly, high throughput RNA-sequencing technology was carried out to obtain a full set of transcripts of the amoeba in bioreactor culture condition. A similar dataset was generated for the transcriptome of *Naegleria fowleri* [72] performed with Illumina HiSeq 2000 technology. Among the 8,790 transcripts, 70.3% matched with known proteins in the nr database. The absence of matches in the

nr database for the remaining transcripts was probably due to the currently limited set of sequenced amoebal genomes. Indeed, only the genomes of *A. castellanii*, *Acanthamoeba triangularis*, *Vermamoeba vermiformis*, *Balamuthia mandrillaris* and four genomes of the *Vahlkampfiidae* family (*W. magna*, *N. gruberi*, *N. fowleri* and *Naegleria lovaniensis*) were sequenced and analyzed [28,32,72–77].

Based on and completing genomic and transcriptomic data, the proteomic analysis provided a better description and understanding of the molecular and physiological processes of the amoeba. The proteome was performed from two independent bioreactor samples each divided into eight fractions. The analysis allowed 3561 proteins to be identified of which a very large majority (3376; 95%) had an assigned function in nr. To obtain a detailed and complete characterization into gene function, we explored the different classification systems. The COG and KEGG classification uncovered that the transcripts and proteins are involved in various metabolic pathways and diverse molecular mechanisms.

For transcripts and proteins, the top five enriched COG categories were similar, although they presented in a different order. Furthermore, we observed that the proportion of proteins assigned to the top five COG categories (“unknown function”, “post-translational modification, protein turnover, chaperones”, “signal transduction mechanisms”, “intracellular trafficking secretion and vesicular transport” and “translation, ribosomal structure and biogenesis”) was higher than the proportion for the transcripts. Furthermore, KEGG-based classification reported a large proportion of enzymes involved in the metabolism of carbohydrates, amino acids and lipids. No apparent differences were observed between metabolic pathways identified through the analyses of the transcriptome and the proteome. This suggests that the proteomic data mainly confirmed here the transcriptomic data. Although the data of the transcriptome and the proteome are congruent, we noted some differences such as the high expression of the *Nox2* gene, whereas the protein was not identified within the proteomic data. These differences between the transcriptomics and proteomics data have already been widely described in the literature. Indeed, these differences can be explained due to several reasons, as the detection of RNA-seq is a much more sensitive method than the determination of the proteome [78]. Moreover, the regulatory processes which can occur after transcription of mRNA, including post-transcriptional, translational, post-translational and protein degradation regulation mechanisms, as well as the half-life of RNA and of the corresponding proteins, could be another reason for the difference between the transcriptomics and proteomics results [31,79,80]. The cytoskeleton allows cellular “scaffolding”, which is essential for cell motility, intracellular transport and architectural structure [81]. Generally, amoebae adhere to a support on which they adopt a particular morphology with pseudopodia presence and move by an amoeboid movement [58,82]. Our transcriptomic and proteomic data revealed multiple transcripts (349) and proteins (151) related to cytoskeleton. In a previous study, the analysis of the *W. magna* genome revealed 625 genes related to cytoskeleton, which represented 3.4% of the gene repertoire [28]. This proportion of genes related to cytoskeleton is greater than for *Naegleria* species, which could be related to the high mobility of *W. magna*. In our COG analysis, we found that only 1.5% of genes related to cytoskeleton were expressed, and 0.5% genes were translated. These low proportions could suggest that the amoebae cultivated in a bioreactor (suspension mode) require a structural organization of the cytoskeleton different from that of amoebae grown on a support (adherence mode). Microtubules and microfilaments were found in the present work to be essential structural components of the *W. magna* cytoskeleton. Under specific conditions, *W. magna* has a flagellate stage, as is for the case of *Naegleria* species [5]. In our study, the identification of microtubule-related proteins involved in flagellate formation indicates that *W. magna* could temporarily move in its flagellated form during bioreactor growth.

In a large-scale laboratory fermenter, *W. magna* uses carbohydrates, lipids and amino acids through the Krebs cycle and a branched respiratory chain with oxygen as final electron acceptor. Although considered to be fully aerobic, the identification of transcripts and proteins involved in anaerobic respiration shows that this amoeba has the ability to adapt to low oxygen concentrations. This parallels previous findings for *A. castellanii*, *N. gruberi* and other organisms regarding the versatility of metabolism and the possibility of a facultative, anaerobic metabolism [32,33,35,83]. For some protists

living under low-oxygen condition, the anaerobic respiration takes place in mitochondrion-related organelles named hydrogenosomes [84]. In our study, we identified N-terminal mitochondrial transit peptides for two proteins related to anaerobic respiration. The same result was predicted by bioinformatic for *N. gruberi* anaerobic protein, with which *W. magna* shared best hits [32]. However, an in vitro study showed that FeFe-hydrogenase and its associated maturase is located in cytosol [85]. These results suggest that the FeFe-hydrogenase and its associated maturase in *W. magna* are likely located in mitochondria or in cytosol.

The use of such facultative anaerobic metabolism could indicate that some cells were not oxygenated enough. Therefore, the detection of proteins involved in *W. magna* metabolism versatility may provide us with an indication of parameters that could be modified in order to improve growing conditions. Free-living amoebae have a predatory role in the control of bacterial populations in soils and aquatic environments [6,7,86,87]. To feed on bacteria, amoebae have a set of genes that set up a variety of molecular mechanisms to destroy bacteria [33,88]. Several years ago, we cultivated *W. magna* C2c maky in axenic medium, in which the amoebae lived without the presence of bacteria and only by the absorption of dissolved nutrients. However, the identification of transcripts and proteins related to the degradation and phagocytosis of bacteria suggests that *W. magna* still has the ability to feed on bacteria. This hypothesis was verified by plating *W. magna* C2c maky cultivated on non-nutrient agar plates seeded with *E. coli* (data not shown). This is an important feature for this amoeba, which is produced in bioreactors with the objective to decontaminate cooling towers from legionella. The expression of these proteins may be necessary for the degradation of nutrients in the growth medium. Similar results were obtained in a *N. gruberi* metabolism study, in which Opperdoes et al. reported that the *N. gruberi* strain NEG-M, cultivated for many years in an axenic medium, had retained all the genes involved in bacterial degradation within the genome [33]. Moreover, we found the sequences related to the bacterial degradation in the transcriptome of *N. gruberi* strain NEG-M [72]. However, all these genes related to bacterial defense mechanisms have another function, which could explain the reason for the expression of these genes in the absence of bacteria. In the environment, amoebae live in communities with other microorganisms and face many complex ecological challenges [7]. Indeed, they are in constant competition with fungi, bacteria, protists and other amoebae for food and multiplication. Furthermore, amoebae must deal with predation and toxins [70]. To survive, *W. magna* seems to harbor an ancient weapon in the defense against the growth of other microorganisms. Analysis reported the presence of some enzymes involved in the biosynthesis of secondary metabolites such as a penicillin amidase or polyketides synthase, which may allow it to defend itself against the microbial communities in the environment [7,70]. However, penicillin amidase is also found in many different microorganisms that appear to use this enzyme for other purposes such as the assimilation of carbon source [89]. *A. castellanii* has demonstrated potent bactericidal properties against Methicillin-resistant *Staphylococcus aureus* (MRSA) [90]. A large repertoire of putative polyketide synthases has been identified in the *D. discoideum* genome [91]. Therefore, we can hypothesize that amoebae could be potential candidates as a new source of antimicrobial compounds. Furthermore, we have identified transcripts and proteins of *W. magna* that support a potential involvement in the degradation of xenophobic compounds. These man-made compounds cause considerable problems to the environment because they are often refractory to degradation [68,92]. As for some bacteria, *W. magna* could be tested for the control of the degradation of persistent organic chemicals [93].

RNA sequencing has a broad range of applications allowing several scientific questions to be responded to such as the composition of transcriptome or comparison of gene expression profiling between different samples [94]. Biological replicates are essential for differential expression analyses in interpretation of the quantification of the level of gene expression [95–97]. To analyze the composition of transcriptome, they are less important as they are only used to overcome natural biological variability. Association with proteome analysis, done in duplicate in the present study, overcomes this problem. Thus, we could identify the proteins for nearly all transcripts detected, even those with a limited number of reads. This is likely because our RNA-seq study included a high number of reads and

coverage [98]. Indeed, we obtained 217 million reads from RNA sequencing, which is efficient for a transcriptomic study [72,95].

## 5. Conclusions

The study provides new insight in the understudied amoebic field. Overall, our transcriptomic and proteomic survey expands the still limited knowledge on FLAs, which are common in the environment. These analyses provide new insights into the metabolic pathways and biological processes of *W. magna* cultivated using untraditional methods. Moreover, the identification of enzymes associated with secondary metabolite pathways suggests that *W. magna* could have the capacity to produce and export compounds with antimicrobial activity.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2076-2607/8/5/771/s1>, Figure S1: Representation of taxonomic distribution of protein assigned to a function in nr database, Figure S2: Phylogenetic tree for *W. magna* [FeFe] hydrogenase maturation protein, Figure S3: Phylogenetic tree for *W. magna* [FeFe] hydrogenase maturation protein, Figure S4: Phylogenetic tree for [FeFe]-hydrogenase maturation protein HydG and Figure S5: Representation of conserved domain of beta-ketoacyl synthase; Table S1: Transcriptomic and proteomic annotation, Table S2: Proteomic data, Table S3: KEGG annotation, Table S4: Motility study, Table S5: Metabolism study, Table S6: Mechanism defense study and Table S7: Secondary metabolism study.

**Availability of Data:** The transcriptome data are deposited in EBI-EMBL under bioproject number PRJEB30797. The proteome generated during the current study are deposited in ProteomeXchange via the PRIDE database under the project accession number PXD016724.

**Author Contributions:** I.H. performed genome analysis and wrote the manuscript, P.D. performed proteomic analysis, S.D. reviewed the manuscript, R.M.M. and O.A. prepared amoeba, P.C. reviewed genome analysis, B.L.S. conceived and supervised the study and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by a grant from the French State managed by the National Research Agency under the “Investissements d’avenir (Investments for the Future)” program with the reference ANR-10-IAHU-03 (Méditerranée Infection) and by the Région Provence-Alpes-Côte-d’Azur and the European funding FEDER PRIMI.

**Acknowledgments:** The authors thank Anass Jawhari and Sebastien Igonet of Calixar company (Lyon, France) for their contribution to the work of total *Willaertia magna* proteins extraction. The authors acknowledge the financial support from ITMO Cancer AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé, National Alliance for Life Sciences and Health) within the framework of the cancer plan for Orbitrap mass spectrometer founding and Adeline Page and Frédéric Delolme for performing the mass spectrometry analysis (Protein Science Facility, SFR BioSciences CNRS UMS3444, Inserm US8, UCBL, ENS de Lyon, 50 Avenue Tony Garnier, 69007 Lyon, France).

**Conflicts of Interest:** I. Hasni had a CIFRE grant supported by the Amoeba society, Mameri is a former employee of the Amoeba society and S. Demanche is employed by the Amoeba society.

## References

1. Adl, S.M.; Bass, D.; Lane, C.E.; Lukeš, J.; Schoch, C.L.; Smirnov, A.; Agatha, S.; Berney, C.; Brown, M.W.; Burki, F.; et al. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J. Eukaryot. Microbiol.* **2018**, jeu.12691. [[CrossRef](#)] [[PubMed](#)]
2. Bertelli, C.; Greub, G. Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. *Front. Cell. Infect. Microbiol.* **2012**, *2*. [[CrossRef](#)]
3. Pánek, T.; Čepička, I. Diversity of Heterolobosea. *Genet. Divers. Microorg.* **2012**. [[CrossRef](#)]
4. de Jonckheere, J.F.; Dive, D.G.; Pussard, M.; Vickerman, K. *Willaertia Magna* gen. nov., sp. nov. (Vahlkampfiidae), a Thermophilic Amoeba Found in Different Habitats. Available online: <https://eurekamag.com/research/001/281/001281223.php> (accessed on 12 August 2018).
5. Robinson, B.S.; Christy, P.E.; De Jonckheere, J.F. A temporary flagellate (mastigote) stage in the vahlkampfiid amoeba *Willaertia magna* and its possible evolutionary significance. *Biosystems* **1989**, *23*, 75–86. [[CrossRef](#)]
6. Clarholm, M. Protozoan grazing of bacteria in soil—impact and importance. *Microb. Ecol.* **1981**, *7*, 343–350. [[CrossRef](#)] [[PubMed](#)]
7. Rodríguez-Zaragoza, S. Ecology of Free-Living Amoebae. *Crit. Rev. Microbiol.* **1994**, *20*, 225–241. [[CrossRef](#)]
8. Greub, G.; Raoult, D. Microorganisms Resistant to Free-Living Amoebae. *Clin. Microbiol. Rev.* **2004**, *17*, 413–433. [[CrossRef](#)]

9. Rowbotham, T.J. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J. Clin. Pathol.* **1980**, *33*, 1179–1183. [CrossRef]
10. Rowbotham, T.J. Isolation of *Legionella pneumophila* from clinical specimens via amoebae, and the interaction of those and other isolates with amoebae. *J. Clin. Pathol.* **1983**, *36*, 978–986. [CrossRef]
11. Stout, J.E.; Yu, V.L. Legionellosis. *N. Engl. J. Med.* **1997**, *337*, 682–687. [CrossRef]
12. Barbaree, J.M.; Fields, B.S.; Feeley, J.C.; Gorman, G.W.; Martin, W.T. Isolation of protozoa from water associated with a legionellosis outbreak and demonstration of intracellular multiplication of *Legionella pneumophila*. *Appl. Environ. Microbiol.* **1986**, *51*, 422–424. [CrossRef] [PubMed]
13. Scheikl, U.; Sommer, R.; Kirschner, A.; Rameder, A.; Schrammel, B.; Zweimüller, I.; Wesner, W.; Hinker, M.; Walochnik, J. Free-living amoebae (FLA) co-occurring with legionellae in industrial waters. *Eur. J. Protistol.* **2014**, *50*, 422–429. [CrossRef] [PubMed]
14. Dupuy, M.; Binet, M.; Bouteleux, C.; Herbelin, P.; Soreau, S.; Héchar, Y. Permissiveness of freshly isolated environmental strains of amoebae for growth of *Legionella pneumophila*. *FEMS Microbiol. Lett.* **2016**, *363*, fnw022. [CrossRef] [PubMed]
15. Kilvington, S.; Stevens, C.; Ebert, F.; Michel, R.; Beeching, J.R. A comparative study of *Williaertia magna* (free-living amoeba) from different geographic areas using whole-cell and small-subunit rRNA restriction fragment length polymorphisms. *J. Protozool. Res.* **1995**, *5*, 97–107.
16. Linder, J.W.-K. Ewert Free-living Amoebae Protecting *Legionella* in Water: The Tip of an Iceberg? *Scand. J. Infect. Dis.* **1999**, *31*, 383–385. [CrossRef]
17. Cirillo, J.D.; Falkow, S.; Tompkins, L.S. Growth of *Legionella pneumophila* in *Acanthamoeba castellanii* enhances invasion. *Infect. Immun.* **1994**, *62*, 3254–3261. [CrossRef]
18. Delafont, V.; Rodier, M.-H.; Maisonneuve, E.; Cateau, E. *Vermamoeba vermiformis*: A Free-Living Amoeba of Interest. *Microb. Ecol.* **2018**, *76*, 991–1001. [CrossRef]
19. Dey, R.; Bodennec, J.; Mameri, M.O.; Permin, P. Free-living freshwater amoebae differ in their susceptibility to the pathogenic bacterium *Legionella pneumophila*. *FEMS Microbiol. Lett.* **2009**, *290*, 10–17. [CrossRef]
20. Hasni, I.; Jarry, A.; Quelard, B.; Carlino, A.; Eberst, J.-B.; Abbe, O.; Demanèche, S. Intracellular Behaviour of Three *Legionella pneumophila* Strains within Three Amoeba Strains, Including *Williaertia magna* C2c Maky. *Pathogens* **2020**, *9*, 105. [CrossRef]
21. Amoeba|Biocide by Nature. Available online: <http://www.amoeba-biocide.com/fr> (accessed on 7 January 2020).
22. Neff, R.J. Purification, Axenic Cultivation, and Description of a Soil Amoeba, *Acanthamoeba* sp. *J. Protozool.* **1957**, *4*, 176–182. [CrossRef]
23. Weekers, P.H.H.; Vogels, G.D. Axenic cultivation of the free-living amoebae, *Acanthamoeba castellanii* and *Hartmannella vermiformis* in a chemostat. *J. Microbiol. Methods* **1994**, *19*, 13–18. [CrossRef]
24. Beshay, U.; Friehs, K.; Azzam, A.-E.-M.; Flaschel, E. Analysis of the behaviour of *Dictyostelium discoideum* in immobilised state by means of continuous cultivation. *Bioprocess. Biosyst. Eng.* **2003**, *26*, 117–122. [CrossRef] [PubMed]
25. Mimouni, V.; Ulmann, L.; Pasquet, V.; Mathieu, M.; Picot, L.; Bougaran, G.; Cadoret, J.-P.; Morant-Manceau, A.; Schoefs, B. The potential of microalgae for the production of bioactive molecules of pharmaceutical interest. *Curr. Pharm. Biotechnol.* **2012**, *13*, 2733–2750. [CrossRef] [PubMed]
26. Newman, J.D.; Marshall, J.; Chang, M.; Nowroozi, F.; Paradise, E.; Pitera, D.; Newman, K.L.; Keasling, J.D. High-level production of amorpha-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered *Escherichia coli*. *Biotechnol. Bioeng.* **2006**, *95*, 684–691. [CrossRef] [PubMed]
27. Papaspyridi, L.-M.; Aligiannis, N.; Christakopoulos, P.; Skaltsounis, A.-L.; Fokialakis, N. Production of bioactive metabolites with pharmaceutical and nutraceutical interest by submerged fermentation of *Pleurotus ostreatus* in a batch stirred tank bioreactor. *Procedia Food Sci.* **2011**, *1*, 1746–1752. [CrossRef]
28. Hasni, I.; Chelkha, N.; Baptiste, E.; Mameri, M.R.; Lachuer, J.; Plasson, F.; Colson, P.; Scola, B.L. Investigation of potential pathogenicity of *Williaertia magna* by investigating the transfer of bacteria pathogenicity genes into its genome. *Sci. Rep.* **2019**, *9*, 1–12. [CrossRef]
29. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [CrossRef]
30. Lomsadze, A.; Burns, P.D.; Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **2014**, *42*, e119. [CrossRef]

31. Vogel, C.; Marcotte, E.M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **2012**, *13*, 227–232. [[CrossRef](#)]
32. Fritz-Laylin, L.K.; Prochnik, S.E.; Ginger, M.L.; Dacks, J.B.; Carpenter, M.L.; Field, M.C.; Kuo, A.; Paredez, A.; Chapman, J.; Pham, J.; et al. The Genome of *Naegleria gruberi* Illuminates Early Eukaryotic Versatility. *Cell* **2010**, *140*, 631–642. [[CrossRef](#)]
33. Opperdoes, F.R.; De Jonckheere, J.F.; Tielens, A.G.M. *Naegleria gruberi* metabolism. *Int. J. Parasitol.* **2011**, *41*, 915–924. [[CrossRef](#)] [[PubMed](#)]
34. Alsam, S.; Sissons, J.; Dudley, R.; Khan, N.A. Mechanisms associated with *Acanthamoeba castellanii* (T4) phagocytosis. *Parasitol. Res.* **2005**, *96*, 402–409. [[CrossRef](#)]
35. Clarke, M.; Lohan, A.J.; Liu, B.; Lagkouravdos, I.; Roy, S.; Zafar, N.; Bertelli, C.; Schilde, C.; Kianianmomeni, A.; Bürglin, T.R.; et al. Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol.* **2013**, *14*, R11. [[CrossRef](#)] [[PubMed](#)]
36. Khan, N.A. *Acanthamoeba*: Biology and increasing importance in human health. *FEMS Microbiol. Rev.* **2006**, *30*, 564–595. [[CrossRef](#)] [[PubMed](#)]
37. De Jonckheere, J. Use of an axenic medium for differentiation between pathogenic and nonpathogenic *Naegleria fowleri* isolates. *Appl. Environ. Microbiol.* **1977**, *33*, 751–757. [[CrossRef](#)]
38. Bioréacteur Pilote GPC. Available online: <http://www.gpcbio.com/bioracteurpilote.html> (accessed on 21 March 2020).
39. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
40. Aherfi, S.; Andreani, J.; Baptiste, E.; Oumessoum, A.; Dornas, F.P.; Andrade, A.C.; Chabriere, E.; Abrahao, J.; Levasseur, A.; Raoult, D.; et al. A Large Open Pangenome and a Small Core Genome for Giant Pandoraviruses. *Front. Microbiol.* **2018**, *9*. [[CrossRef](#)]
41. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [[CrossRef](#)]
42. Anders, S.; Pyl, P.T.; Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169. [[CrossRef](#)]
43. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [[CrossRef](#)]
44. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [[CrossRef](#)] [[PubMed](#)]
45. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]
46. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I.; et al. CDD: NCBI’s conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222–D226. [[CrossRef](#)] [[PubMed](#)]
47. Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L.; et al. InterPro: The integrative protein signature database. *Nucleic Acids Res.* **2009**, *37*, D211–D215. [[CrossRef](#)] [[PubMed](#)]
48. Bachmann, B.O.; Ravel, J. Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enzymol.* **2009**, *458*, 181–217. [[CrossRef](#)]
49. Ziemert, N.; Podell, S.; Penn, K.; Badger, J.H.; Allen, E.; Jensen, P.R. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **2012**, *7*, e34064. [[CrossRef](#)]
50. Almagro Armenteros, J.J.; Salvatore, M.; Emanuelsson, O.; Winther, O.; von Heijne, G.; Elofsson, A.; Nielsen, H. Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* **2019**, *2*. [[CrossRef](#)]
51. Prokisch, H.; Andreoli, C.; Ahting, U.; Heiss, K.; Ruepp, A.; Scharfe, C.; Meitinger, T. MitoP2: The mitochondrial proteome database—now including mouse data. *Nucleic Acids Res.* **2006**, *34*, D705–D711. [[CrossRef](#)]
52. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]

53. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
54. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [[CrossRef](#)] [[PubMed](#)]
55. He, L.; Diedrich, J.; Chu, Y.-Y.; Yates, J.R. Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Anal. Chem.* **2015**, *87*, 11361–11367. [[CrossRef](#)] [[PubMed](#)]
56. Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2003**, *17*, 2337–2342. [[CrossRef](#)] [[PubMed](#)]
57. McCafferty, J.; Reid, R.; Spencer, M.; Hamp, T.; Fodor, A. Peak Studio: A tool for the visualization and analysis of fragment analysis files. *Environ. Microbiol. Rep.* **2012**, *4*, 556–561. [[CrossRef](#)] [[PubMed](#)]
58. Khan, N.A. *Acanthamoeba: Biology and Pathogenesis*, 2nd ed.; Caister Academic Press: Norfolk, UK, 2018.
59. Otomo, T.; Otomo, C.; Tomchick, D.R.; Machius, M.; Rosen, M.K. Structural Basis of Rho GTPase-Mediated Activation of the Formin mDia1. *Mol. Cell* **2005**, *18*, 273–281. [[CrossRef](#)] [[PubMed](#)]
60. Hug, L.A.; Stechmann, A.; Roger, A.J. Phylogenetic Distributions and Histories of Proteins Involved in Anaerobic Pyruvate Metabolism in Eukaryotes. *Mol. Biol. Evol.* **2010**, *27*, 311–324. [[CrossRef](#)]
61. Elsbach, P. The bactericidal/permeability-increasing protein (BPI) in antibacterial host defense. *J. Leukoc. Biol.* **1998**, *64*, 14–18. [[CrossRef](#)]
62. Cosson, P.; Lima, W.C. Intracellular killing of bacteria: Is *D. dictyostelium* a model macrophage or an alien?: Intracellular bacterial killing in *Dictyostelium*. *Cell. Microbiol.* **2014**, *16*, 816–823. [[CrossRef](#)]
63. Winterbourn, C.C.; Kettle, A.J. Redox Reactions and Microbial Killing in the Neutrophil Phagosome. *Antioxid. Redox Signal.* **2012**, *18*, 642–660. [[CrossRef](#)]
64. Hong, Y.; Kang, J.-M.; Joo, S.-Y.; Song, S.-M.; Lê, H.G.; Thái, T.L.; Lee, J.; Goo, Y.-K.; Chung, D.-I.; Sohn, W.-M.; et al. Molecular and Biochemical Properties of a Cysteine Protease of *Acanthamoeba castellanii*. *Korean J. Parasitol.* **2018**, *56*, 409–418. [[CrossRef](#)]
65. Turk, V.; Stoka, V.; Vasiljeva, O.; Renko, M.; Sun, T.; Turk, B.; Turk, D. Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim. Biophys. Acta* **2012**, *1824*, 68–88. [[CrossRef](#)] [[PubMed](#)]
66. Miyanaga, A. Structure and function of polyketide biosynthetic enzymes: Various strategies for production of structurally diverse polyketides. *Biosci. Biotechnol. Biochem.* **2017**, *81*, 2227–2236. [[CrossRef](#)]
67. Robinson, J.A. Polyketide synthase complexes: Their structure and function in antibiotic biosynthesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **1991**, *332*, 107–114. [[CrossRef](#)] [[PubMed](#)]
68. Singh, D.R. BIODEGRADATION OF XENOBIOTICS- A WAY FOR ENVIRONMENTAL DETOXIFICATION. *Int. J. Dev. Res.* **2017**, *7*, 14082–14087.
69. Miranda, E.R.; Nam, E.A.; Kuspa, A.; Shaulsky, G. The ABC transporter, AbcB3, mediates cAMP export in *D. discoideum* development. *Dev. Biol.* **2015**, *397*, 203–211. [[CrossRef](#)] [[PubMed](#)]
70. Eichinger, L.; Pachebat, J.A.; Glöckner, G.; Rajandream, M.-A.; Suggang, R.; Berriman, M.; Song, J.; Olsen, R.; Szafranski, K.; Xu, Q.; et al. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **2005**, *435*, 43–57. [[CrossRef](#)]
71. Brown, S.; Clastre, M.; Courdavault, V.; O'Connor, S.E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3205–3210. [[CrossRef](#)]
72. Zysset-Burri, D.C.; Müller, N.; Beuret, C.; Heller, M.; Schürch, N.; Gottstein, B.; Wittwer, M. Genome-wide identification of pathogenicity factors of the free-living amoeba *Naegleria fowleri*. *BMC Genomics* **2014**, *15*. [[CrossRef](#)]
73. Liechti, N.; Schürch, N.; Bruggmann, R.; Wittwer, M. The genome of *Naegleria lovaniensis*, the basis for a comparative approach to unravel pathogenicity factors of the human pathogenic amoeba *N. fowleri*. *BMC Genomics* **2018**, *19*. [[CrossRef](#)]
74. Hasni, I.; Andréani, J.; Colson, P.; La Scola, B. Description of Virulent Factors and Horizontal Gene Transfers of Keratitis-Associated Amoeba *Acanthamoeba Triangularis* by Genome Analysis. *Pathogens* **2020**, *9*, 217. [[CrossRef](#)]
75. Chelkha, N.; Hasni, I.; Louazani, A.C.; Levasseur, A.; La Scola, B.; Colson, P. Vermamoeba vermiformis CDC-19 draft genome sequence reveals considerable gene trafficking including with candidate phyla radiation and giant viruses. *Sci. Rep.* **2020**, *10*. [[CrossRef](#)] [[PubMed](#)]

76. Detering, H.; Aebischer, T.; Dabrowski, P.W.; Radonić, A.; Nitsche, A.; Renard, B.Y.; Kiderlen, A.F. First Draft Genome Sequence of *Balamuthia mandrillaris*, the Causative Agent of Amoebic Encephalitis. *Genome Announc.* **2015**, *3*. [[CrossRef](#)] [[PubMed](#)]
77. Greninger, A.L.; Messacar, K.; Dunnebacke, T.; Naccache, S.N.; Federman, S.; Bouquet, J.; Mirsky, D.; Nomura, Y.; Yagi, S.; Glaser, C.; et al. Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: The continuing case for reference genome sequencing. *Genome Med.* **2015**, *7*. [[CrossRef](#)] [[PubMed](#)]
78. Fan, M.; Sun, X.; Xu, N.; Liao, Z.; Li, Y.; Wang, J.; Fan, Y.; Cui, D.; Li, P.; Miao, Z. Integration of deep transcriptome and proteome analyses of salicylic acid regulation high temperature stress in *Ulva prolifera*. *Sci. Rep.* **2017**, *7*, 1–19. [[CrossRef](#)]
79. Lee, M.V.; Topper, S.E.; Hubler, S.L.; Hose, J.; Wenger, C.D.; Coon, J.J.; Gasch, A.P. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol. Syst. Biol.* **2011**, *7*, 514. [[CrossRef](#)]
80. de Sousa Abreu, R.; Penalva, L.O.; Marcotte, E.M.; Vogel, C. Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* **2009**, *5*, 1512–1526. [[CrossRef](#)]
81. Fletcher, D.A.; Mullins, R.D. Cell mechanics and the cytoskeleton. *Nature* **2010**, *463*, 485–492. [[CrossRef](#)]
82. Tekle Yonas, I.; Williams Jessica, R. Cytoskeletal architecture and its evolutionary significance in amoeboid eukaryotes and their mode of locomotion. *R. Soc. Open Sci.* **2016**, *3*, 160283. [[CrossRef](#)]
83. Atteia, A.; van Lis, R.; Gelius-Dietrich, G.; Adrait, A.; Garin, J.; Joyard, J.; Rolland, N.; Martin, W. Pyruvate formate-lyase and a novel route of eukaryotic ATP synthesis in *Chlamydomonas mitochondria*. *J. Biol. Chem.* **2006**, *281*, 9909–9918. [[CrossRef](#)]
84. Barberà, M.J.; Ruiz-Trillo, I.; Tufts, J.Y.A.; Bery, A.; Silberman, J.D.; Roger, A.J. *Sawyeria marylandensis* (Heterolobosea) has a hydrogenosome with novel metabolic properties. *Eukaryot. Cell* **2010**, *9*, 1913–1924. [[CrossRef](#)]
85. Tsaousis, A.D.; Nývltová, E.; Šuták, R.; Hrdý, I.; Tachezy, J. A Nonmitochondrial Hydrogen Production in *Naegleria gruberi*. *Genome Biol. Evol.* **2014**, *6*, 792–799. [[CrossRef](#)] [[PubMed](#)]
86. Siddiqui, R.; Khan, N.A. War of the microbial worlds: Who is the beneficiary in *Acanthamoeba*–bacterial interactions? *Exp. Parasitol.* **2012**, *130*, 311–313. [[CrossRef](#)] [[PubMed](#)]
87. Denet, E.; Coupat-Goutaland, B.; Nazaret, S.; Pélandakis, M.; Favre-Bonté, S. Diversity of free-living amoebae in soils and their associated human opportunistic bacteria. *Parasitol. Res.* **2017**, *116*, 3151–3162. [[CrossRef](#)] [[PubMed](#)]
88. Akya, A.; Pointon, A.; Thomas, C. Mechanism involved in phagocytosis and killing of *Listeria monocytogenes* by *Acanthamoeba polyphaga*. *Parasitol. Res.* **2009**, *105*, 1375–1383. [[CrossRef](#)]
89. Tishkov, V.I.; Savin, S.S.; Yasnaya, A.S. Protein Engineering of Penicillin Acylase. *Acta Nat.* **2010**, *2*, 47–61. [[CrossRef](#)]
90. Iqbal, J.; Siddiqui, R.; Khan, N.A. *Acanthamoeba* and bacteria produce antimicrobials to target their counterpart. *Parasit. Vectors* **2014**, *7*, 56. [[CrossRef](#)]
91. Zucko, J.; Skunca, N.; Curk, T.; Zupan, B.; Long, P.F.; Cullum, J.; Kessin, R.H.; Hranueli, D. Polyketide synthase genes and the natural products potential of *Dictyostelium discoideum*. *Bioinforma. Oxf. Engl.* **2007**, *23*, 2543–2549. [[CrossRef](#)]
92. Rozgaj, R. [Microbial degradation of xenobiotics in the environment]. *Arh. Hig. Rada Toksikol.* **1994**, *45*, 189–198.
93. Janssen, D.B.; Dinkla, I.J.T.; Poelarends, G.J.; Terpstra, P. Bacterial degradation of xenobiotic compounds: Evolution and distribution of novel enzyme activities. *Environ. Microbiol.* **2005**, *7*, 1868–1882. [[CrossRef](#)]
94. Li, W.V.; Li, J.J. Modeling and analysis of RNA-seq data: A review from a statistical perspective. *Quant. Biol. Beijing China* **2018**, *6*, 195–209. [[CrossRef](#)]
95. Manga, P.; Klingeman, D.M.; Lu, T.-Y.S.; Mehlhorn, T.L.; Pelletier, D.A.; Hauser, L.J.; Wilson, C.M.; Brown, S.D. Replicates, Read Numbers, and Other Important Experimental Design Considerations for Microbial RNA-seq Identified Using *Bacillus thuringiensis* Datasets. *Front. Microbiol.* **2016**, *7*. [[CrossRef](#)] [[PubMed](#)]
96. Schurch, N.J.; Schofield, P.; Gierliński, M.; Cole, C.; Sherstnev, A.; Singh, V.; Wrobel, N.; Gharbi, K.; Simpson, G.G.; Owen-Hughes, T.; et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* **2016**, *22*, 839–851. [[CrossRef](#)] [[PubMed](#)]

97. Rapaport, F.; Khanin, R.; Liang, Y.; Pirun, M.; Krek, A.; Zumbo, P.; Mason, C.E.; Socci, N.D.; Betel, D. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.* **2013**, *14*, 3158. [[CrossRef](#)] [[PubMed](#)]
98. Liu, Y.; Zhou, J.; White, K.P. RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics* **2014**, *30*, 301–304. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).