

Bioinformatic tools help molecular characterization of *Perkinsus olseni* differentially expressed genes

Rita M. T. Ascenso

School of Technology and Management, Polytechnic Institute of Leiria, Morro do Lena, Alto do Vieiro, Leiria, Portugal, rita.ascenso@ipleiria.pt

Abstract

In the 80ies, in Southern Europe and in particular in Ria Formosa there was an episode of heavy mortality of the economically relevant clam *Ruditapes (R.) decussatus* associated with a debilitating disease (Perkinsosis) caused by *Perkinsus olseni*. This protozoan parasite was poorly known concerning its' differential transcriptome in response to its host, *R. decussatus*. This laboratory available protozoan system was used to identify parasite genes related to host interaction. Beyond the application of molecular biology technologies and methodologies, only the help of Bioinformatics tools allowed to analyze the results of the study. The strategy started with SSH technique, allowing the identification of parasite up-regulated genes in response to its natural host, then a macroarray was constructed and hybridized to characterize the parasite genes expression when exposed to bivalves hemolymph from permissive host (*R. decussatus*), resistant host (*R. philippinarum*) and non permissive bivalve (*Donax trunculus*) that cohabit in the same or adjacent habitats in Southern Portugal. Genes and respective peptides full molecular characterization depended on several Bioinformatic tools application. Also a new Bioinformatic tool was developed.

1 Introduction

Perkinsus (P.) is a protozoan parasite affecting economically relevant mollusks worldwide, and was associated to clam heavy mortalities caused by Perkinsosis [1] [2] in Huelva coast [3] in South of Spain, being speculated that the higher water temperatures in the southern regions may impact on disease severity [4]. This disease had, also, negative impact in Asia, namely in Korea [5], China [6] and Japan [7]. Also, in USA, there was an oyster susceptibility to parasitism by *P. marinus* [8]. *P. olseni* [1] [2][9] was found to be the causative agent of Perkinsosis in Southern Portugal coastal region, where nearly 90 percent of the national carpet-shell clam production occurs [10].

This parasite was firstly described in Australia, infecting commercial relevant abalone species [11]. This situation led to an increase of interest to further characterize *P. olseni* response to its host, facilitated thanks to a continuous in vitro clonal cell culture of trophozoites derived from parasites infecting the carpet-shell clam, *R. decussatus* [12][13] (deposited in the American Type Cell Culture, ATCC 50984).

To characterize molecularly the parasite response to bivalves and identify differentially expressed genes from an organism from which the genome is unknown, it was used a Suppression Subtractive Hybridization (SSH). This involved the construction of a cDNA subtractive library [14], based in the Diatchenko methodology [15]. In parallel, a strategy of enrichment, Mirror Orientation Selection (MOS) was used to decrease background [16]. In order to obtain the

largest diversity of up-regulated genes, two libraries (Dfsl and efMOSI) were constructed and screened [14]. Duplicates of membranes with clones from the forward cDNA libraries were spotted and hybridized against forward and reverse libraries PCR products from the cDNA library. The autoradiography films with the impressed signals were analyzed using Bioinformatic tools.. Each bacteria clone spotted represent an EST (Expressed Sequence Tag), i.e. a cDNA fragment representative of parasite gene expressed.

Among the numerous bivalves in South Portugal, *R. decussatus* and *R. philippinarum* are both susceptible bivalves to *P. olsenii* infection but Perkinsosis does not significantly affect the wedge-shell clam *Donax trunculus* (Linnaeus, 1767), the blue mussel *Mytilus galloprovincialis* (Lamarck, 1919), the Portuguese oyster *Crassostrea angulata* (Lamarck, 1819) or the Japanese oyster, *C. gigas* (Thunberg, 1873), which also constitute important fishing resources. These bivalves have different habitats within the same coast, being *D. trunculus* found in the open, wave exposed sites, while the other species inhabit more sheltered areas.

To detect parasite relevant genes involved in bivalves' response and characterize its' expression a cDNA macroarray was constructed and hybridized [17]. Since this parasite was yet not very well described, being a protozoan, it could be relevant to compare its' transcriptome with other parasites from Apicomplexa [18], group such as *Plasmodium* sp. and *Toxoplasma* sp., responsible for Malaria and Toxoplasmosis.

Based on relative gene expression, it were selected the most represented EST and the functionally relevant gene to further characterization. Using Bioinformatic tools was possible to interpret and treat full cDNAs and full gene sequence obtained using Molecular Biology methodologies [19]. The prediction and characterization of peptides was entirely obtained from Bioinformatic analysis. Several needs in macroarray data analysis promoted the collaboration with Department of Electronics and Informatics Engineering to develop a new Bioinformatic tool.

Although the original work was centered in Molecular Biology [20], in this paper we pretend to give an overview of the Bioinformatic tools used during the characterization of this poorly known parasite, starting from the analysis of its up-regulated transcriptome until the molecular characterization of genes and respective proteins.

2 Uncovering Parasite differential transcriptome

Transcriptome is the set of genes active in certain circumstances; *P. olsenii* transcriptome studied in this work was a set of up-regulated genes in response to its host hemolymph, the clam *R. decussatus*. The differentially expressed genes were obtained using two approaches involving subtractive cDNA libraries. A cDNA subtractive library construction results in a forward and a reverse library, allowing the identification of down-regulated and up-regulated genes. In order to search for sober-expressed genes, the forward library was spotted and hybridized with forward and reverse library [15]. Two different forward libraries were constructed: (1) The Direct forward subtractive library (Dfsl), which consists on direct cloning of the forward cDNA subtractive library, followed by plate lifts screening to reveal up-regulated genes, and (2) the MOS enrichment methodology [16], used after completion of the SSH step and followed by *in situ* differential screening by bacteria dot blot (efMOSI) [14].

2.1 Libraries screening

Differential screening [21] was performed by double filter hybridization of each set of membranes for each approach, Dfsl and efMOSI, according to SSH protocol [15]. Numerous membranes were prepared with bacteria clones and each spot was analyzed after hybridization with forward and reverse library. After autoradiography of two replicates, the impressed signals were quantified by densitometry (Quantity One, BioRad, Richmond, USA). Quantity One is a powerful, flexible software package for imaging and analyzing 1-D electrophoresis gels, dot blots and other arrays. Available software was used to quantify gene expression, allowing the selection of genes of interest for a macroarray construction (Table 1).

For efMOSI, the scanned images were easily treated since the distribution of clones in the array and the software used a 96-well standard grid; facilitating signal detection and signal density quantification. In Dfsl, signal position was manually detected followed by signal quantification. Both data analysis involved the background quantification and subtractions from each foreground signal, performed in excel. The gene expression was obtained from the ratio between data from forward and data from reverse hybridization. Naked eye analysis was also performed to confirm information provided by the software analysis. Spots corresponding to clones showing more than two fold up-expression were further processed for sequencing. 96 and 204 clones were analyzed from Dfsl and efMOSI respectively [14].

2.2 Clones sequencing and analysis

The gene identification was possible based on sequencing at Macrogen (South Korea), which used a 3730XL DNA Sequencer and Monitoring tool, LIMS (Laboratory Information Management system). The sequences were observed one by one to check sequencing quality, using Chromas Lite, by chromatogram files observation. This permitted to ensure the sequence and the adaptors used. The adaptors sequences were removed using VecScreen¹ and stored in the laboratory database, Vector NTI (Invitrogen, Carlsbad, USA), which allows to manage and analyze sequences.

2.3 Sequence identification

The major objective was to characterize the parasite response to its host, so molecular function was determined using GOBlet server, which consists of a gene ontology (GO) search server based on similarity searches against known protein databases [22] (Table 1). For *P. olseni* cDNA sequences it was used a cut-off for E-value of 0.1 and 26 different GO subclasses were identified (Table 1), 25 from Dfsl and 6 from efMOSI. From BLASTx analysis the larger diversity obtained in the Dfsl library was confirmed with 82 different hits compared to 27 hits in the efMOSI [20][17]. To improve the description of the poorly characterized parasite, the sequences were submitted to BLASTx [23] taking into account the organism hit, the majority of clones corresponded to Protist kingdom. For Dfsl, 32 percent of total clones corresponded to Protist Kingdom and for efMOSI, 48 percent hits also belonged to Protozoan's, as expected [20]. In fact *Perkinsus* genus fit in Protist Kingdom [2] although there are numerous genes, in minority, belonging to all other kingdoms.

¹<http://www.ncbi.nlm.nih.gov/projects/VecScreen/>

Table 1: Work-flow followed for *P. olseni* transcriptome and molecular characterization, showing the Molecular Biology strategy aside with Bioinformatic tools used.

MOLECULAR BIOLOGY		BIOINFORMATICS
Parasite cells	Parasite cells exposed to host	
cDNA Subtractive Library Two approaches		
Dfsl library	efMOSI library	
Differential gene expression		<i>in silico</i> data analysis
Clones screening Scanning image analysis Signal quantification Gene expression		Quantity One Quantity One Excel
Differential Transcriptome		<i>in silico</i> sequence analysis
EST sequence read and quality check Sequence storage and treatment Gene Molecular function Gene identification		Cromas Lite Vector NTI GOblet BLASTx
Gene expression pattern		Macroarray analysis
Macroarray construction and hybridization Gene expression quantification		Quantity One Developed tool: Automatic Macroarray Analysis
Gene and Protein Molecular Characterization		<i>in silico</i> analysis
Gene and Protein identification		BLASTn and BLASTx
Primers design Sequence assemble ORF and peptide prediction Intron-Exon boundaries		Primer Premier Vector NTI ORF Finder B12seq
Gene promoter characterization		MatInspector
Protein characterization Protein domain's motifs Protein functional family		SignalP, MotifScan, TMprep Clustalx and WebLOGO Clustalx and MEGA4

3 Transcriptome characterization

To study genes pattern of expression, macroarrays were hybridized against transcripts from parasite exposed to permissive, resistant and non permissive bivalves [20]. A macroarray was constructed with 98 selected clones, 72 from Dfsl and 26 from efMOSI, namely those with molecular function related to cell adhesion, binding, chaperone, signal transduction, transport, cellular process, transcription, translation, structural, catalysis and basal metabolism, and clones matching hypothetical parasite genes, all over-expressed in response to host hemolymph [14].

The identification of genes and/or proteins was based on comparison with annotated databases (Table 1). BLAST was performed against the Apicomplexa Database² [18]. A few genes confirmed its previously described function; others remained unknown or revealed no significant hit, below 0.07 E-value cut-off. A more generalized search was performed against non-redundant database from NCBI³ [23]; still, some ESTs gave no hit, which points out for an

²<http://apidb.org/apidb/>

³<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

atypical organism, poorly known and needing further molecular characterization. Databases updating may contribute for future comparison and improved results, for example the recently updated Apidb, now Eukaryotic Pathogen Data Base Resource or the recent ESTs described for *P. marinus* [24].

3.1 Parasite gene expression pattern

Two macroarrays were constructed and hybridized twice. From each there were various autoradiography degrees of exposure to get films with non saturated and saturated signal.

Gene expression was again based on image scanning, analysis and signal quantification from each experiment and replicates of macroarrays, using Quantity One (Table 1). The number of genes up-regulated one hundred fold or more in response to each bivalve hemolymph was: 58 genes up-regulated in response to *R. philippinarum* hemolymph and 84 genes up-regulated in response to *D. trunculus* hemolymph [14]. The differentially expressed genes were considered relevant in the host-parasite interaction.

3.2 Development of macroarray analysis tool

Having numerous autoradiography films of diverse saturation levels and with arrays of variable sizes from cDNA library screening and from macroarray; for quick macroarray data analysis, from image analysis to gene expression, it was developed a Automatic Macroarray Analysis Tool (Table 1), in cooperation with the Department of Electronics and Informatics Engineering of Faculty of Science and Technology of University of Algarve [25]. Lately this Bioinformatic tool was updated to allow more than quantification of a set of experiments with macroarrays, analyzing gene expression data. It was improved to be able for other densitometric analysis involving circular dish plates such as petri dish plates, or multiple cell plates [26]. This tool is proposed to be useful in the measurement of cell mineralization analyzing calcium quantification based on Von Kossa staining.

4 Characterization of most represented gene

From the 300 EST resulting from both approaches of SSH, we chose to focus on some of the highly represented. Through a cDNA subtractive library were identified those parasite genes up-regulated when challenged with host hemolymph. From these, the gene represented in more copies (15 ESTs) was obtained from the efMOS1 cDNA subtracted library. First identified as *Moorella* sp. cytosine deaminase, at the time [20], following BLASTx analysis [23], was considered related to pyrimidine salvage pathway. Later a higher homology was obtained with proteins from the chlorohydrolase family. So, it was named PoClhl, Chlorohydrolase like.

4.1 PoClhl gene identification and promoter characterization

The chlorohydrolase was shown to be firmly linked with a major amidohydrolase protein superfamily [27]. Homologous proteins catalyzing different reactions are being discovered at an

increasing rate with functional genomics focusing attention on the interplay between molecular sequence and function. This amidohydrolase superfamily comprehend some of the enzymes responsible for nitrogen recycle, essential metabolites for DNA and nucleic acid synthesis [27], including 30 percent of the steps in four intermediate metabolic pathways [28]. The *P. olsenii* EST was identified, through *in silico* search analysis, BLAST [23], as being a chlorohydrolase protein family gene, with significant hits for TrzA, AtzA and CD.

The promoter analysis to search for putative transcription factors binding sites was used MatInspector at Genomatix Portal⁴. The corroboration of gene organization was obtained comparing *P. olsenii* and *P. marinus* genes [29] to full characterize the PoClhl gene, but also the PoAdh and PoNHE genes. The promoter was carefully analyzed *in silico* and its characteristics compared to other *Perkinsus* spp. promoters. PoClhl and PmClhl had no canonical TATA box, but a T-rich region was identified. A comparable T-rich region was identified in other *Perkinsus* gene promoters such as the PmSOD1 gene, which contains a proximal TATA (TATTTTA) 34 bases from the transcription start point, a position typical of many Eukaryotic promoters [30], while PmNramp has a putative TATA box at -162 position [29]. The confirmation was performed by multiple alignment [31] and WebLOGO visualization [32].

4.2 PoClhl protein characterization

From *in silico* analysis, the PoClhl polypeptide appears as a non-secretory protein as expected, a result confirmed by SignalP analysis since no signal peptide was identified. Motif Scan predicted various putative phosphorylation sites, one cyclin recognition site and one putative tyrosine sulphating site⁵. The comparison of the presently described polypeptides with those from amidohydrolase family demonstrated a higher similarity with the Proteobacteria, *N. oceani*, amidohydrolase, described as a cytosine deaminase and related metal-dependent hydrolase according to predicted function. The *in silico* identification of PoClhl polypeptide as a member of the amidohydrolase family was tested by protein expression studies. The absence of an effective transfection assay for *P. olsenii* limited further studies [20].

In order to search for functional domains described for metal-dependent protein family, some selected proteins were aligned with PoClhl and PmClhl. The expected motif HxH, present in the proteins suggested them to be coordinated by divalent-metals bound to specific residues, as for the other family proteins. It also contains the motif HxHxxE, relevant in substrate coordination at the active site. The metal ligands, four histidines and one aspartic acid residue, are strictly conserved in the three enzyme families, and define a subtle but sharp sequence signature of a metal-dependent hydrolases superfamily [28]. Multiple alignment using ClustalX [31] followed by WebLOGO⁶ [32] permitted the visual identification of conserved residues related to protein function, namely according to described examples [28]. This analysis confirmed the predicted function of PoClhl.

⁴<http://www.genomatix.de>

⁵http://myhits.isb-sib.ch/cgi-bin/motif_scan

⁶<http://weblogo.berkeley.edu/logo.cgi>

5 Characterization of functional relevant genes

From macroarray gene expression and identification, two genes were further analyzed; adhesion related (PoAdh) and transporter, Na⁺/H⁺ antiporter (PoNHE) [20] were molecularly characterized once their function appear to be related to host response. DNA sequence, ORF, introns and exons and predicted peptide were obtained by Molecular Biology methodologies [19] and use of several Bioinformatics (Table 1). These two genes were submitted to an exhaustive molecular characterization which results won't be presented, but only the strategy followed, as for PoClhl.

5.1 Genes molecular characterization

Full cDNA sequence was obtained based on Rapid Amplification of both 5' and 3' cDNA Ends (RACE) using Marathon cDNA library (Clontech, Palo Alto, USA) and primers designed based on the clones of each gene from the subtractive library. Primer Premier Software was used for all primers design (PREMIER Biosoft International, Palo Alto, USA) [20]. Numerous clones obtained with diverse sequence size, for 5' and 3' of the cDNA; these were assemble using the assemble tool of Vector NTI Software. The full consensus cDNA sequence was analyzed, identifying the start codon ATG and stop codon using ORF Finder at NCBI⁷. ORF Finder tool allowed to predict the peptides coded. Having access to species *P. marinus* genome sequence project and ESTs [24], a local search was performed to confirm the genes presence.

Full gene sequencing was easily obtained based on *P. olsenii* cDNA sequence aligned with *P. marinus* gene sequencing to predict exon region to design primers. Gene sequencing including the search of promoter region behind 5' UTR was obtained using a Universal Genome Walker library to get upstream and downstream regions of the gene [20]. Two libraries were prepared with two different endonucleases enzymes, after several amplifications, cloning and sequencing was obtained several clones with fragments of full genes sequence. Having several sequences, these were assembled to complete each full gene consensus, after cleaning the sequence from sequencing bad quality ends and primer sequence. The assembling was performed using the Vector NTI Software. In parallel to molecular biology techniques, Bioinformatic tools allow the results analysis and presentation, namely full genes sequence and molecular organization.

5.2 Proteins characterization

Protein analysis was performed based on the predicted peptides sequences for each ORF sequence. The secondary structure prediction was performed using ExPASy available tools⁸, namely TMprep for transmembrane regions identification [33]. More relevant than hydrophathy profile, the identification of protein domains and motifs are very useful to predict the protein function compared to other proteins of the same family. In order to comprehend post-translational modifications, SignalP [34] was used to detect secreted proteins and MotifScan [35] for protein domains prediction. To screen the protein path, MITOPROT and PATS were used to predict if one of this predicted peptides were targeted to mitochondria or apicoplast, respectively. Also the targeting to chloroplast was checked using CloroP. The knowledge about

⁷<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>

⁸<http://expasy.org/>

Cation Proton Antiporter 1 super-family and NHE protein family [36] [37], allowed the comparison among 46 proteins using multiple alignments and phylogeny analysis helped to ensure protein function and cellular localization. ClustalX [31] and MEGA4 [38] analysis allowed to infer that NHE is close related to Endosomal-TNG or Plant vacuolar, part of the intracellular clade, suggesting that PoNHE may be responsible for adjustment to different environmental conditions, possibly responding to an osmotic shock related with the interface parasite-host [20]. The characterization of this parasite wouldn't be possible without making use of Bioinformatic tools.

6 Conclusion

This paper intend to present the workflow of a molecular characterization of differential transcriptome of *P. olsenii* parasite, in response to its host. The Molecular Biology technology and methodologies demand a multiple experiment design, with replicates and numerous sequencing data. The amount of data analysis was only possible using Bioinformatic tools. For detection and identification of parasite up-regulated genes and to clarify the full cDNA, gene and protein characterization of three genes were used diverse Bioinformatic tools. The presented strategy was used along the development of PhD work and resulted also in the development of a new Bioinformatic tool, Automatic Macroarray Analysis Tool [25]. Future work aims optimizing this instrument in order to help screening differential expressed genes based in macroarrays, for poorly characterized systems. This paper may help starting Molecular Biologists during a molecular characterization using, mainly, free web available Bioinformatic tools.

Acknowledgements

Acknowledgments to M.Sc. Ana Filipa Nogueira, PhD Adriano Benvenho and M.Sc. Mário Antunes for technical support in Latex utilization.

References

- [1] C. Azevedo. *Perkinsus atlanticus* nova espécie protoparasita da ameijoia *Ruditapes decussatus* (mollusca, bivalvia). *Pesca e Navegação*, pages 8–10, 1989.
- [2] C. Azevedo. Fine structure of *Perkinsus atlanticus* n. sp. (apicomplexa, perkinsea) parasite of the clam *Ruditapes decussatus* from portugal. *Journal of Parasitology*, 75(4):627–635, 1989.
- [3] J. I. Navas, M.C. Castillo, P. Vera, and M. Ruiz-Rico. Principal parasites observed in clams, *Ruditapes decussatus* (l.), *Ruditapes philippinarum* (adams et reeve), *Venerupis pullastra* (montagu) and *Venerupis aureus* (gmelin), from the huelva coast (s.w. spain). *Aquaculture*, 107:193–199, 1992.
- [4] A. Villalba and S.M. Casas. Effect of perkinsosis on the energetic physiology of the clam *Ruditapes decussatus*. *Journal of Shellfish Research*, 20:559–560, 2001.

- [5] K. I. Park, K. S. Choi, and J. W. Choi. Epizootiology of *Perkinsus* sp. found in the manila clam, *Ruditapes philippinarum* in komsoe bay, korea. *Journal of Shellfish Research*, 32(3):303–309, 1999.
- [6] Y. B. Liang, X. C. Zhang, L. J. Wang, B. Yang, Y. Zhang, and C. L. Cai. Prevalence of *Perkinsus* sp. in the manila clam *Ruditapes philippinarum* along northern coast of yellow sea in china. *Oceanologia et Limnologia Sinica*, 32:502–511, 2001.
- [7] M. Hamaguchi, N. Suzuki, H. Usuki, and H. Ishioka. *Perkinsus* protozoan infection in short-necked clam *Tapes (=Ruditapes) philippinarum* in japan. *Fish Pathology*, 33:473–480, 1998.
- [8] D. Bushek, J. Scarpa, and S.E Laramore. Susceptibility of the caribbean oyster *Crassostrea rhizophorae* to *Perkinsus marinus*. *Journal of Shellfish Research*, 21:371–372, 2002.
- [9] Ricardo B. Leite, Ricardo Afonso, and M. Leonor Cancela. *Perkinsus* sp. infestation in carpet-shell clams, *Ruditapes decussatus* (l), along the portuguese coast. results from a 2-year survey. *Aquaculture*, 240:39–53, 2004.
- [10] IPIMAR. Relatório de actividades do ipimar. *ipimar-iniap.ipimar.pt*, 2003.
- [11] R. J. G. Lester and G. H. G. Davis. A new *Perkinsus* species (apicomplexa, perkinsea) from the abalone *Haliotis ruber*. *Journal of Invertebrate Pathology*, 37(2):181–187, 1981.
- [12] José A. F. Robledo, Patrícia A. Nunes, M. Leonor Cancela, and Gerardo R. Vasta. Development of an in vitro clonal culture and characterization of the rna gene cluster of *Perkinsus atlanticus*, a protistan parasite of the clam *Tapes decussatus*. *Journal of Eukaryotic Microbiology*, 49(5):414–422, 2002.
- [13] S. M. Casas, J.F. Peyre, K.S. Reece, C. Azevedo, and A. Villalba. Continuous in vitro culture of the carpet shell clam *Tapes decussatus* protozoan parasite *Perkinsus atlanticus*. *Diseases of Aquatic Organisms*, 52:217–231, 2002.
- [14] Rita M. T. Ascenso, Ricardo B. Leite, Ricardo Afonso, and Maria Leonor Cancela. Suppression-subtractive hybridization: A rapid and inexpensive detection methodology for up-regulated *Perkinsus olseni* genes. *African Journal Biochemistry Research*, 1(3):24–28, 2007.
- [15] Luda Diatchenko, Yun-Fai Chis Lau, Aaron. Campbell, Alex Chenchick, Fauzia Moqadam, Betty Huang, Sergey Lukyanov, Konstantine Lukyanov, Nadya Gurskaya, Eugene D. Sverdlov, and Paul D. Siebert. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *PNAS*, 93:6025–6030, 1996.
- [16] Denis V. Rebrikov, Olga V. Britanova, Nadya G. Gurskaya, Konstantine A. Lukyanov, Victor S. Tarabykin, and Sergey A. Lukyanov. Mirror orientation selection (mos): a method for eliminating false positive clones from libraries generated by suppression subtractive hybridization. *Nucleic Acids Research*, 28(20):1–4, 2000.

- [17] R.M.T. Ascenso, R.B. Leite, R. Afonso, and M.L. Cancela. Expression pattern of *Perkinsus olseni* genes in response to bivalves with different susceptibility to perkinsosis. *Journal of Fish Disease*, 32(7):633–636, 2009.
- [18] C Aurrecochea, M Heiges, H Wang, Z Wang, S Fischer, P Rhodes, J Miller, E Kraemer, C J Jr Stoeckert, D S Roos, and J C Kissinger. Apidb: integrated resources for the apicomplexan bioinformatics resource center. *Nucleic Acids Research*, 35, 2007.
- [19] J. Sambrook, E.F. Fritsch, and T. Maniatis. Molecular cloning: A laboratory manual. *Cold Spring Harbor Press, Cold Spring Harbor, New York, USA*, (2nd Ed), 1989.
- [20] Rita M. T. Ascenso. *Study of Perkinsus olseni infection mechanisms: Identification and regulation of parasite genes differentially expressed in response to host and environmental stress*. PhD thesis, 2008.
- [21] Vera G. Fonseca, Asuncion Lago-Leston, Vincent Laizé, and M. Leonor Cancela. Rapid identification of differentially expressed genes by in situ screening of bacteria. *Molecular Biotechnology*, 30:163–165, 2005.
- [22] Detlef Groth, Hans Lehrach, and Steffen Hennig. Goblet: a platform for gene ontology annotation of anonymous sequence data. *Nucleic Acids Research*, 32:313–317, 2004.
- [23] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Scherffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402, 1997.
- [24] Sandeep J Joseph, José A Fernandez-Robledo, Malcolm J Gardner, Najib M El-Sayed, Chih-Horng Kuo, Eric J Schott, Haiming Wang, Jessica C Kissinger, and Gerardo R Vasta. The alveolate *Perkinsus marinus*: Biological insights from est gene discovery. *BMC Genomics*, 11(228), 2010.
- [25] António Anjos, Hamid Shahbazkia, and Rita M. T. Ascenso. Automatic macroarray analysis - tool and methods. ISCA 21st International Conference on Computer Applications in Industry and Engineering, 23 a 25 de March 2008.
- [26] António Anjos, Ricardo Leite, M. L. Cancela, and Hamid Reza Shahbazkia. Maq - a bioinformatics tool for automatic macroarray analysis. *International Journal of Computer Applications*, 7, 2010.
- [27] Jennifer L. Seffernick, Mervyn L. De Souza, Michael J. Sadowsky, and Lawrence P. Wackett. Melamine deaminase and atrazine chlorohydrolase: 98 percent identical but functionally different. *Journal of Bacteriology*, 183(8):2405–2410, 2001.
- [28] Liisa Holm and Chris Sander. An evolutionary treasure: Unification of a broad set of amidohydrolases related to urease. *Proteins: structure, function and genetics*, 28:72–82, 1997.
- [29] José António F. Robledo, Pascal Courville, Mathieu F. M. Cellier, and Gerardo R. Vasta. Gene organization and expression of the divalent cation transporter nramp in the protistan parasite *Perkinsus marinus*. *Journal of Parasitology*, 90(5):1004–1014, 2004.

- [30] Eric J. Schott, José A. F. Robledo, Anita C. Wright, Abelardo M. Silva, and Gerardo R. Vasta. Gene organization and homology modeling of two iron superoxide dismutases of the early branching protist *Perkinsus marinus*. *Gene*, 309:1–9, 2003.
- [31] J. D. Thompson, T. J. Gibson, and F. Plewniak. The clustal-x windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 25(24):4876–4882, 1997.
- [32] Gavin E. Crooks, Gary Hon, John-Marc Chandonia, and Steven E. Brenne. Weblogo: A sequence logo generator. *Genome research*, 14(6):1188–1190, 2004.
- [33] K. Hofmann and W. Stoffel. Tmbase - a database of membrane spanning proteins segments. *Biological Chemistry Hoppe-Seyler*, 47(166), 1993.
- [34] Emanuelsson Olof, Brunak Soren, Heijne Gunnar von, and Nielsen Henrik. Locating proteins in the cell using targetp, signalp, and related tools. *Nature Protocols*, 2, 2007.
- [35] M Pagni, V Ioannidis, L Cerutti, M Zahn-Zabal, CV Jongeneel, J Hau, O Martin, D Kuznetsov, and L Falquet. Myhits: improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Research*, 35, 2007.
- [36] L. K. Putney, S. P. Denker, and D. L. Barber. The changing face of the na⁺/h⁺ exchanger, nhe1: Structure, regulation, and cellular actions. *Annual Review of Pharmacology and Toxicology*, 42:527–552, 2002.
- [37] Christopher L. Brett, Mark Donowitz, and Rajini Rao. Evolutionary origins of eukaryotic sodium/proton exchangers. *American Journal of Cell Physiology*, 288:C223–C239, 2005.
- [38] K. Tamura, J. Dudley, M. Nei, and S. Kumar. Mega4: Molecular evolutionary genetics analysis (mega) software version 4.0. *Molecular Biology and Evolution*, 24:1596–1599, 2007.