

# Genes found in partial sequencing of *Phytophthora cinnamomi* genome

## Genes descobertos na sequenciação parcial do genoma de *Phytophthora cinnamomi*

Luís Santos<sup>1\*</sup>, Ivone M. Martins<sup>2,4</sup>, Vera Maia<sup>3,4</sup>, Ángel Domínguez<sup>3</sup> e Altino Choupina<sup>4</sup>

<sup>1</sup> Agricultural College of Bragança, Polytechnic Institute of Bragança, Campus de Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal;

<sup>2</sup>CEB - Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal;

<sup>3</sup>Department of Microbiology and Genetics, CIETUS-IBSAL, University of Salamanca/CSIC, Plaza de los Drs. de la Reina s/n, 37007 Salamanca, Spain;

<sup>4</sup> CIMO-Mountain Research Center, Agricultural College of Bragança, Polytechnic Institute of Bragança, Campus de Santa Apolónia, Apartado 1172, 5300-253 Bragança, Portugal.

(\*E-mail: lsantos@ipb.pt)

<http://dx.doi.org/10.19084/RCA16083>

Recebido/received: 2015.06.17

Aceite/accepted: 2016.07.07

### ABSTRACT

Members of the oomycete cause extensive losses in agriculture and widespread degradation in natural plant communities, being responsible for the death of thousands of trees every year. Two of the representative species are *Phytophthora infestans*, which causes late blight of potato, and *Phytophthora cinnamomi*, which causes chestnut ink disease, responsible for losses on sweet chestnut production in Europe. Genome sequencing efforts have been focused on the study of three species: *P. infestans*, *P. sojae* and *P. ramorum*. *Phytophthora infestans* has been developed as the model specie for the genus, possessing excellent genetic and genomics resources including genetic maps, BAC libraries, and EST sequences. Our research team is trying to sequence the genome of *P. cinnamomi* in order to gain a better understanding of this oomycete, to study changes in plant-pathogen relationships including those resulting from climate change and trying to decrease the pathogen's impact on crops and plants in natural ecosystems worldwide. We present here a preliminary report of partially sequenced genomic DNA from *P. cinnamomi* encoding putative protein-coding sequences and tRNAs. Database analysis reveals the presence of genes conserved in oomycetes.

**Keywords:** DNA library, oomycetes, *Phytophthora cinnamomi*, tRNA.

### RESUMO

Os oomicetas causam enormes perdas na agricultura e degradação generalizada em áreas de crescimento natural, sendo responsáveis pela morte de milhares de árvores por ano. Duas das espécies representativas de oomicetas são *Phytophthora infestans*, que provoca o míldio da batateira, e *Phytophthora cinnamomi*, que causa a doença da tinta do castanheiro, responsável por perdas de produção de castanha na Europa. Os esforços de sequenciação de genomas de oomicetas têm estado focados no estudo de três espécies: *P. infestans*, *P. sojae* e *P. ramorum*. *Phytophthora infestans* foi considerada como espécie modelo para o género, pois possui excelentes recursos genéticos e de genómica, incluindo mapas genéticos, bibliotecas BAC, e sequências de EST. Esta equipa de investigadores procura sequenciar o genoma de *P. cinnamomi*, a fim de obter uma melhor compreensão deste oomiceta, para estudar mudanças nas relações planta-patogénio, incluindo as resultantes das alterações climáticas, e tentando diminuir o impacto do patogénio nas culturas e plantas em ecossistemas naturais em todo o mundo. É apresentada uma abordagem preliminar do DNA genómico parcialmente sequenciado a partir de *P. cinnamomi*, o qual codifica possíveis sequências de proteínas e tRNAs. As análises em bases de dados revelam a presença de genes conservados em oomicetas.

**Palavras-chave:** Biblioteca DNA, oomicetas, *Phytophthora cinnamomi*, tRNA.

## INTRODUCTION

Plant-pathogenic oomycetes belongs to the eukaryotic Stramenopile lineage and cause devastating diseases in numerous crops as well as ornamental and native plants. They include more than 60 species of the genus *Phytophthora*. The best known potato late blight disease, which is caused by the fungus *P. infestans*, results in considerable loss of potato crop yield worldwide (Judelson *et al.*, 2009). *P. cinnamomi*, a destructive and widespread soil-borne pathogen that infects woody plant hosts, has been estimated to directly affect over 2000 of the 9000 native plant species in the south west of Western Australia. In the north east of Portugal and in the Spanish adjoin region, *P. cinnamomi* has been reported to affect chestnut-tree (*Castanea sativa*) and cherry-tree (*Prunus avium*) (Martins *et al.*, 2014b). Fundamental insights into the pathogenesis of fungal surface infections and the strategy of fungal invasion of a host can be achieved by identifying the genes of *P. infestans* and *P. cinnamomi* that are expressed during standard lab conditions and during infection. This would help to establish the genes involved in the pathogenic processes.

The environmental changes associated with invasion and spread in the host demand specific adaptations. Indeed, it has been shown that *P. infestans* and *P. cinnamomi* express specific genes at certain stages of infection. It is unlikely that a single pathogenicity program would be sufficient to account for the range of infections that can be caused by these pathogens. A permanent interplay exists between the host and the pathogen that regulates the transcriptional profile of *P. infestans* and *P. cinnamomi*; in this scenario, gene expression of *P. infestans* and *P. cinnamomi* is regulated by dynamic changes in relation to the environment.

Determining which genes are expressed during the different stages and types of infections is of essential importance in understanding the biological and physiological responses of those fungal pathogens to their environment (Choupina *et al.*, 2014). While many data exist about the *P. infestans* and *P. sojae* genome, those related to *P. cinnamomi* are scanty. Genomic DNA libraries are a standard procedure which has been used to sequencing genes in several organisms (Kanjawanawattawong *et al.*, 2014; Tan *et al.*, 2014). By digestion of genomic DNA from *P. cinnamomi* we have constructed a DNA library

with an insert size average of 2-12 kb in *Escherichia coli*. Six clones comprising 30000-40000 bp were sequenced, finding 12 putative protein-coding sequences and tree tRNAs.

Database analysis reveals the presence of conserved genes in oomycetes. Our group, using asymmetric PCR, have identified and characterized the structural proteins Actin1, Actin2 and Tubulin1 (GenBank: AM412175.1, AM412176.1, AM412177.1) (Martins *et al.*, 2013), and some other proteins involved in mechanisms of infection by *P. cinnamomi* namely endo-1,3-beta-glucanase (GenBank: AM259651.1) (complete cds) (Meirinho *et al.*, 2010); endo-glucanase (GenBank: AJ964942.1) (partial cds) (unpublished results); glucanase inhibitor protein (GIP) (GenBank: AM259384.1) (complete cds) (Martins *et al.*, 2014a); necrosis-inducing *Phytophthora* protein 1 (NPP1) (GenBank: AM403130.1) (complete cds) (unpublished results) (Gijzen and Nurnberger, 2006) and transglutaminase (GenBank: AM403129.1) (unpublished results) (Martins *et al.*, 2014b; Yu *et al.*, 2008). By standard PCR with DNA of *P. cinnamomi* and primers designed by homologous sequences from different *Phytophthora* species, our group also sequenced ribosomal RNA subunits (GenBank: AJ421265.2 and AJ457835.2) (unpublished results).

## MATERIAL AND METHODS

### *Biological material*

*Phytophthora cinnamomi* PR120 strain was grown in potato dextrose agar (PDA) medium with cellophane at 28°C for 1 week. Total DNA was obtained from mycelial mats, mycelia was ground with a mortar and pestle in liquid nitrogen. Total DNA was extracted by standard DNA extraction methods (Reinoso *et al.*, 2005).

### *Gene library construction*

The genomic DNA library was prepared by digestion with *Hind*III, which generated fragments in the range of 2-12 kb. The fragments were inserted into the corresponding site of the pBluescript KS+ vector. The recombinant plasmids were used to transform the *E. coli* XL1 Blue strain. Transformants were selected on LB media supplemented with 100 µg/ml ampicillin and six clones were taken randomly for further sequencing (Sambrook *et al.*, 1989). Sequencing reactions were conducted using universal and reverse primers with the ABI 377

**Table 1** - Genes found after partial sequencing of *Phytophthora cinnamomi* genome

Clone No.	Gene	Length (bp)	Homology to... (%)
1	<i>PcRBP14</i>	187	<i>P. infestans</i> ribosomal protein L14 (95.0) <i>P. infestans</i> ribosomal protein L14 (93.0) <i>Saprolegnia feraz</i> ribossomal protein L14 (80.0) <i>Methylococcus capsulatus</i> ribossomal protein L14 (55.0)
1	<i>PcRBP5</i>	534	<i>P. infestans</i> ribosomal protein L5 (89.0) <i>P. infestans</i> ribosomal protein L5 (89.0) <i>Thalassiosira pseudonana</i> ribossomal protein L5 (35.0) <i>Malawimonas jakobiformis</i> ribossomal protein L5 (35.0)
2	<i>PcPLP1</i>	5354	<i>P. infestans</i> pol polyprotein (51.0) <i>P. infestans</i> putative polyprotein (56.0) <i>P. infestans</i> putative polyprotein (40.0) <i>Hordeum vulgare</i> polyprotein (34.0)
2	<i>PcHP1</i>	1388	<i>Rhodobacter sphaeroides</i> chromosome segregation protein SMC (26.0) <i>Erythrobacter litorales</i> hypothetical protein (24.0) <i>Oceanicola granulosis</i> hypothetical protein (30.0)
3	<i>PcPLP2</i>	6473	<i>Strongylocentrotus purpuratus</i> similar to Zn-finger, CCHC type and RNA directed DNA polymerase and integrase, catalytic domain protein (29.0-30.0) <i>S. purpuratus</i> similar to retrovirus-related pol polyprotein from transposon (29.0) <i>Schistosoma mansoni</i> TPA pol polyprotein (28.0)
4	<i>PcHP2</i>	666	<i>Aspergillus nidulans</i> hypothetical protein AN5649.2 (25.0) <i>Canis familiaris</i> cardiac titin isoform N2BA (26.0) <i>Caenorhabditis elegans</i> hypothetical protein C25A1.10b (21.0)
4	<i>PcUPL1</i>	1812	<i>Anopheles gambiae</i> ENSANGP00000025653 (36.0) <i>Arabidopsis thaliana</i> UPL1 (ubiquitin-protein ligase 1) (34.0) <i>A. thaliana</i> UPL1 (34.0) <i>Mus musculus</i> Smad ubiquitination regulatory factor 1 (34.0)
5	<i>PcPLP3</i>	594	<i>P. infestans</i> putative polyprotein (36.0) <i>P. infestans</i> pol protein (25.0) <i>C. familiaris</i> similar to serine/arginine repetitive matrix isoform 2 (34.0) <i>Homo sapiens</i> serine/arginine repetitive matrix 1 (35.0)
5	<i>PcHP3</i>	1062	<i>Pan troglodytes</i> hypothetical protein XP_531346 (31.0) <i>H. sapiens</i> RNA binding protein (23.0) <i>Streptomyces aureofaciens</i> unknown (26.0) <i>C. elegans</i> hypothetical protein H39E23.3 (26.0)
6	<i>PcGLP</i>	942	<i>Theileria parva</i> hypothetical protein TP01_0256 (32.0) <i>Gibberella zeae</i> hypothetical protein FG10731.1 (32.0) <i>Cryptosporidium parvum</i> 60 kDa glycoprotein (29.0) <i>C. hominis</i> glycoprotein (27.0) <i>C. parvum</i> 60 kDa glycoprotein (27.0)
6	<i>PcPGP</i>	1284	<i>Listeria monocytogenes</i> putative peptidoglycan bound protein (LPXTG motif) (32.0) <i>L. monocytogenes</i> probable peptidoglycan bound protein (32.0) <i>Bacillus clausii</i> hypothetical protein (35.0) <i>Jannaschia</i> sp mucin-associated surface protein (39.0)
6	<i>PcHP4</i>	726	<i>Cryptococcus neoformans</i> conserved hypothetical protein (22.0) <i>Bos taurus</i> similar to Rb-inducible coiled coil protein 1 (22.0) <i>Plasmodium berghei</i> hypothetical protein PB405340.00.0 (26.0) <i>C. parvum</i> hypothetical protein cgd4_4490 (36.0)

sequencer (Applied Biosystems, Inc.), using the Taq DyeDeoxy™ Terminator Cycle Sequencing Kit as supplied by the manufacturer. Junctions were sequenced with walking primers using the entire plasmid. The quality of the final sequence was checked by visual inspection of the sequencing profiles at each position on each DNA strand. The sequence was considered correct only when an unambiguous reading of each nucleotide on each strand was achieved.

#### Computer-assisted sequence analysis

Assembly of the sequences was done with the SeqMan program of the DNASTAR program package (DNASTAR Ltd.). The open reading frames (ORFs) were predicted using the DNA strider software (Figure 1).

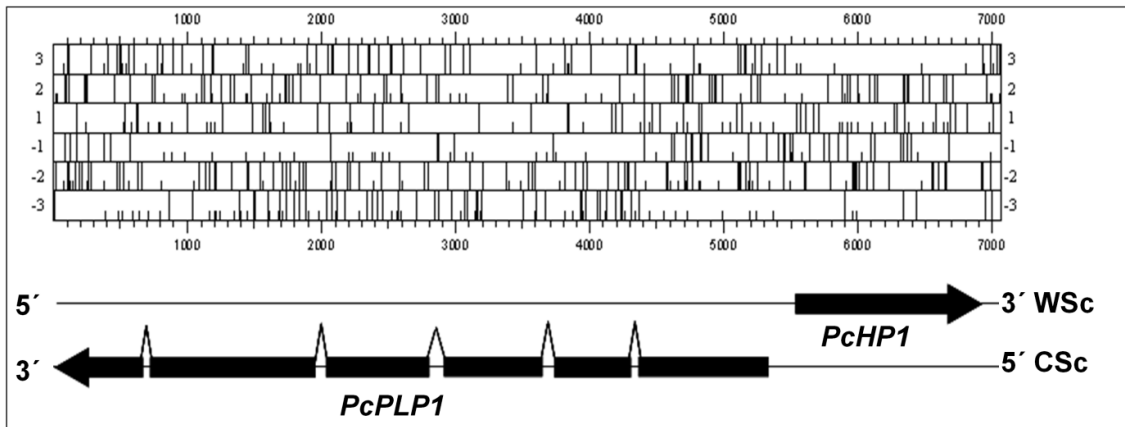


Figure 1 - Genes found after partial sequencing of *Phytophthora cinnamomi* genome.

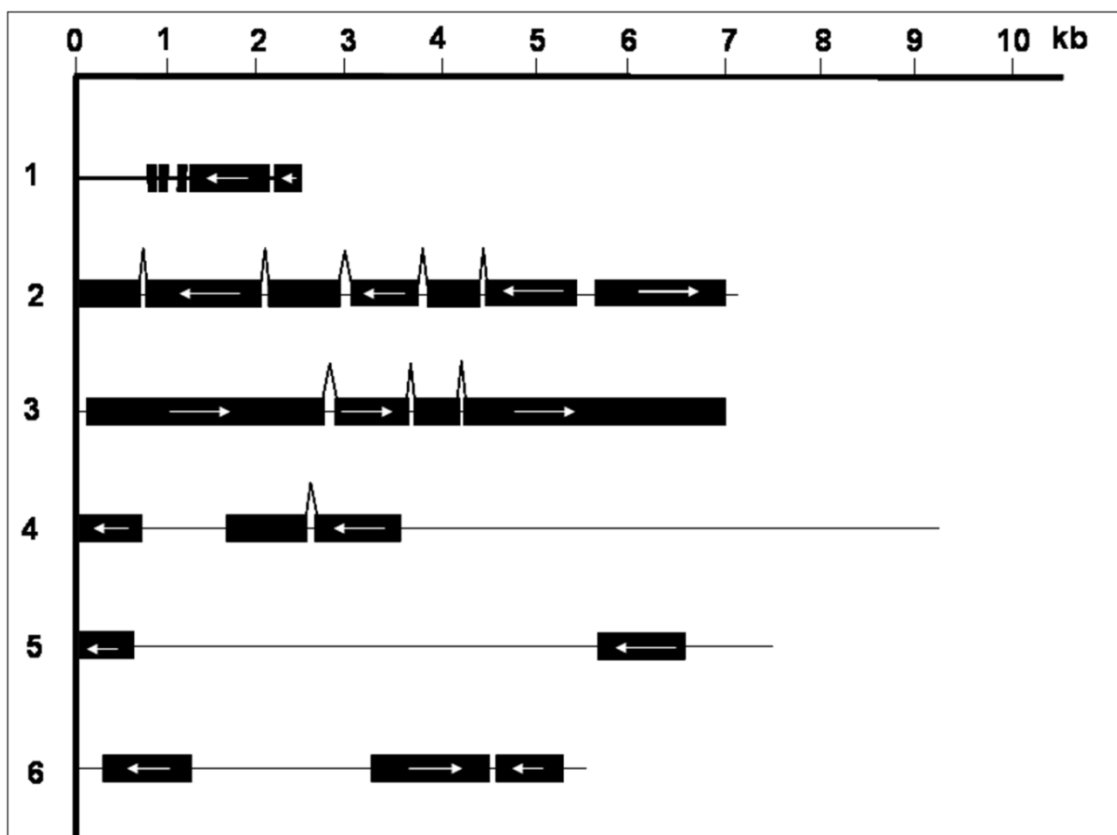


Figure 2 - Genes found after partial sequencing of *Phytophthora cinnamomi* genome.

The gene structure prediction was performed, using the Twinscan (<http://genes.cs.wustl.edu/>) and FGENESH (<http://sun1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind>). ORFs were named with the prefix Pc (*P. cinnamomi*). Searches for homologies were done using the BLAST or FASTA programs. Multiple-sequence alignments were obtained

using the CLUSTAL program. In addition, a Blast analysis was performed on PhytophthoraDB database available on <http://phytophthora.vbi.vt.edu/browse/index.html> and in the Phytophthora Functional Genomics Database available at <http://www.pfgd.org/>. Sequences of predicted proteins were also scanned for Motif Scan.

## RESULTS AND DISCUSSION

### Sequence analysis

Plasmids 1 to 6 contained inserts between 2246 (clone No.1) and 9218 bp (clone No.4), which after sequencing, covered 38,146 bp altogether.

**Clone No.1** has 2246 bp which 750 bp correspond to two possible ORFs, both have identity with known proteins; A+T content is 79.61% (Figure 1, Table 1). In addition, two Glycine tRNA and one Tyrosine tRNA were detected using the software tRNAscan-SE 1.21 (<http://lowelab.ucsc.edu/cgi-bin/tRNAscan-SE>). They have a size of 71, 72 and 83 nucleotides, respectively.

**Clone No.2** is a 7060 bp sequence, with 6742 bp corresponding to two possible ORFs. The first ORF (*PcPLP1*) has 5 introns and has identity with known proteins. The A+T content is 42.58% (Figure 1, Table 1).

**Clone No.3** is a 6758 bp sequence, with 6483 bp corresponding to one possible ORF (presenting 3 introns) with homology to known proteins; A+T content is 44.48% (Figure 1, Table 1).

**Clone No.4** has 9218 bp which 2921 bp correspond to two possible ORFs. The first one is partial and the second has 2 introns with identity to known proteins; A+T content is 44.37% (Figure 1, Table 1).

**Clone No.5** has 7422 bp which 1735 bp correspond to two possible ORFs. The first is partial with identity to known proteins; A+T content is 42.29% (Figure 1, Table 1).

**Clone No.6** has 5442 bp which 2949 bp correspond to three possible ORFs, the two first have identity with known proteins; A+T content is 43.68% (Figure 1, Table 1).

## CONCLUSIONS

Although sequencing is partial, 0.05% of the total genome, it may be indicative of some aspects of the genomic organization of *P. cinnamomi*, which allow to provide the following conclusions: the first sequence (Clone No.1) has a 20,39% GC content and presents two coding sequences from mitochondrial DNA (one of them partial) and tree tRNAs. The other five sequences have a 56.52% GC content and present 10 putative coding sequences (three of them partial). The 12 coding sequences (3 of them containing at least one intron) cover 47% of the total sequence. Eight sequences have similarity with known genes of some bacteria of Stramenopiles lineage and fungal. The number of introns varies from 1 to 5. Intron lengths are quite variable from 41 to 146 nucleotides and conserved splicing sequence elements were found. Intron boundaries shows the canonical splice site 5'GT ... AG3' (Kamoun, 2003).

## ACKNOWLEDGMENTS

The Project COMBATINTA/SP2.P11/02 Interreg IIIA – Cross-Border Cooperation Spain-Portugal, financed by The European Regional Development Found, supported this work.

## REFERENCES

- Choupina, A.; Estevinho, L. & Martins, I. (2014) - Scientifically advanced solutions for chestnut ink disease. *Applied Microbiology and Biotechnology*, vol. 98, n. 9, p. 3905-3909. <http://dx.doi.org/10.1007/s00253-014-5654-2>
- Gijzen, M. & Nurnberger, T. (2006) - Nep1-like proteins from plant pathogens: recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry*, vol. 67, n. 16, p. 1800-1807. <http://dx.doi.org/10.1016/j.phytochem.2005.12.008>
- Judelson, H.; Narayan, R.; Ah-Fong, A. & Kim, K. (2009) - Gene expression changes during asexual sporulation by the late blight agent *Phytophthora infestans* occur in discrete temporal stages. *Molecular Genetics and Genomics*, vol. 281, n. 2, p. 193–206. <http://dx.doi.org/10.1007/s00438-008-0407-5>
- Kanjanawattanawong, S.; Tangphatsornruang, S.; Triwitayakorn, K.; Ruang-areerate, P.; Sangsrakru, D.; Poopear, S.; Somyong, S. & Narangajavana, J. (2014) - Characterization of rubber tree microRNA in phytohormone response using large genomic DNA libraries, promoter sequence and gene expression analysis. *Molecular Genetics and Genomics*, vol. 289, n. 5, p. 921-933. <http://dx.doi.org/10.1007/s00438-014-0862-0>



- Kamoun, S. (2003) - Molecular genetics of pathogenic oomycetes. *Eukaryotic Cell*, vol. 2, n. 2, p. 191-199. <http://dx.doi.org/10.1128/EC.2.2.191-199.2003>
- Martins, I.; López, M.; Dominguez, A. & Choupina, A. (2013) - Isolation and sequencing of Actin1, Actin2 and Tubulin1 genes involved in cytoskeleton formation in *Phytophthora cinnamomi*. *Journal of Plant Pathology & Microbiology*, vol. 4, a. 194. <http://dx.doi.org/10.4172/2157-7471.1000194>
- Martins, I.; Martins, F.; Belo, H.; Vaz, M.; Carvalho, M.; Cravador, A. & Choupina, A. (2014a) - Cloning, characterization and in vitro and in planta expression of a glucanase inhibitor protein (GIP) of *Phytophthora cinnamomi*. *Molecular Biology Reporter*, vol. 41, n. 4, p. 2453-2462. <http://dx.doi.org/10.1007/s11033-014-3101-1>
- Martins, I.; Matos, M.; Costa, R.; Silva, F.; Pascoal, A.; Estevinho, L. & Choupina, A. (2014b) - Transglutaminases: recent achievements and new sources. *Applied Microbiology and Biotechnology*, vol. 98, n. 16, p. 6957-6964. <http://dx.doi.org/10.1007/s00253-014-5894-1>
- Meirinho, S.; Carvalho, M.; Domínguez, A. & Choupina, A. (2010) - Isolation and characterization by asymmetric PCR of the ENDO1 gene for glucan endo-1,3-B-D-glucosidase in *Phytophthora cinnamomi* associated with the ink disease of *Castanea sativa* Mill. *Brazilian Archives of Biology and Technology*, vol. 53, n. 3, p. 513-518. <http://dx.doi.org/10.1590/S1516-89132010000300003>
- Reinoso, C.; Nino-Vega, G.; San-Blast, G. & Dominguez, A. (2005) - Random sequencing of *Paracoccidioides brasiliensis* genes. *Medical Mycology*, vol. 43, n. 8, p. 681-689. <http://dx.doi.org/10.1080/13693780500129673>
- Sambrook, J.; Fritsch, E. & Maniatis, T. (1989) - *Molecular cloning: A laboratory manual*. 2. ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Tan, C.; Wu, Y.; Taliaferro, C.; Bell, G.; Martin, D. & Smith, M. (2014) - Development and characterization of genomic SSR markers in *Cynodon transvaalensis* Burt-Davy. *Molecular Genetics and Genomics*, vol. 289, n. 4, p. 523-531. <http://dx.doi.org/10.1007/s00438-014-0829-1>
- Yu, Y.; Wu, S.; Chan, H.; Chen, Y.; Chen, Z. & Yang, M. (2008). Overproduction of soluble recombinant transglutaminase from *Streptomyces netropsis* in *Escherichia coli*. *Applied Microbiology and Biotechnology*, vol. 81, n. 3, p. 523-532. <http://dx.doi.org/10.1007/s00253-008-1688-7>