BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

BIOMEDICAL INFORMATICS

www.bioinformation.net

Volume 12(4)

Database

A searchable database for the genome of *Phomopsis longicolla* (isolate MSPL 10-6)

Omar Darwish[1,$], Shuxian Li[2,$,*], Zane May[1], Benjamin Matthews[3], Nadim W. Alkharouf[1+]

[1]Department of Computer and Information Sciences, Towson University, MD 21252, USA; [2]United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Crop Genetics Research Unit, Stoneville, MS 38776, USA; [3]USDA-ARS, Beltsville Agriculture Research Center, Beltsville, MD 21075, USA; Shuxian Li – E-mail: shuxian.li@ars.usda.gov; [$]Equal contribution, *Corresponding author.

Received May 9, 2016; Revised May 25, 206; Accepted May 26, 206; Published July 26, 2016

Abstract:
Phomopsis longicolla (syn. *Diaporthe longicolla*) is an important seed-borne fungal pathogen that primarily causes *Phomopsis* seed decay (PSD) in most soybean production areas worldwide. This disease severely decreases soybean seed quality by reducing seed viability and oil quality, altering seed composition, and increasing frequencies of moldy and/or split beans. To facilitate investigation of the genetic base of fungal virulence factors and understand the mechanism of disease development, we designed and developed a database for *P. longicolla* isolate MSPL 10-6 that contains information about the genome assemblies (contigs), gene models, gene descriptions and GO functional ontologies. A web-based front end to the database was built using ASP.NET, which allows researchers to search and mine the genome of this important fungus. This database represents the first reported genome database for a seed borne fungal pathogen in the *Diaporthe–Phomopsis* complex. The database will also be a valuable resource for research and agricultural communities. It will aid in the development of new control strategies for this pathogen.

Availability: http://bioinformatics.towson.edu/*Phomopsis_longicolla*/HomePage.aspx

Key words: *Phomopsis longicolla*, MSPL 10-6, database, annotations and genomic sequence.

Background:
Phomopsis longicolla (syn. *Diaporthe longicolla*) is an important seed-borne fungal pathogen that primarily causes *Phomopsis* seed decay (PSD) in most soybean production areas worldwide **[1, 2].** This disease severely decreases soybean seed quality by reducing seed viability and oil quality, altering seed composition, and increasing frequencies of moldy and/or split beans **[3-6].** Research on analysis of the internal transcribed spacer (ITS) region **[7],** the small subunit of the mitochondrial ribosomal RNA gene **[8],** and other genes/regions of *P. longicolla* have been reported. Recently, the genome of a *P. longicolla* isolate MSPL 10-6 which was isolated from field-grown soybean seed in Mississippi, USA was sequenced **[9].** Development of a database for *P. longicolla* isolate MSPL 10-6 that contains information about the genome assemblies (contigs), gene models, gene descriptions and GO functional ontologies will allow researchers to search and mine the genome of this important fungus. The database will be a valuable resource for research and agricultural communities, and facilitate investigation of the genetic base of fungal virulence factors and an understanding of the mechanism of disease development. To our knowledge, this database represents the first reported genome database for a seed borne fungal pathogen in the *Diaporthe–Phomopsis* complex.

Methodology of Development:
The database was designed, implemented and hosted using Microsoft SQL Server 2008 Enterprise Edition. Microsoft Visual Studio 2013 was used to design and implement the web pages, which were programmed using ASP.NET framework 4.0 with C# programming language. Both the database and the website are on the same server at Towson University in Baltimore, MD, USA. This

ISSN 0973-2063 (online) 0973-8894 (print)

server is running Microsoft Windows Server 2012 and Internet Information Services (IIS V7.0). The database stores the assembly of the *P. longicolla* MSPL 10-6 genome (108 scaffolds) **[9]** and their annotations. In addition to the sequences, the database also houses information on gene function and gene ontology distributions.

Utility to the biological community:

The database contains the genome sequence of *P. longicolla* MSPL 10-6 and the 16,597 genes that were annotated. The annotation includes GO ontologies that have been assigned to most genes (process, molecular function and cellular component). The database's web-accessible interface **(Figure 1)** provides an easy way to search, browse and download the sequences and functional annotation data stored in the database. The following are the main functions the website provides:

[1] **Search:**
Users can search by GO ontology terms, or by sequence description **(Figure 2).** Partial characters can be used if one is not sure of the full GO term or gene name. Both the search by GO ontologies and search by description return their results in a nice tabular format that allows the user to select any record of the returned search results to see details about that specific sequence\gene. The information includes sequence name, sequence description, sequence length, blast e-value, gene ontolgy, InterProScan results and the actual sequence in FASTA.

[2] **Statistics and Graphs:**
The web site provides static pages that display the annotation statistics (lengths of coding regions, number of exons…etc.) along with bar graphs depicting the GO ontologies distributions.

[3] **Download:**
The web site allows user to download the complete assembled genome (FASTA format) and the annotations in both FASTA and GF3 formats. Raw sequences can be found from the SRA database, located at: http://www.ncbi.nlm.nih.gov/nuccore/ AYRD00000000/

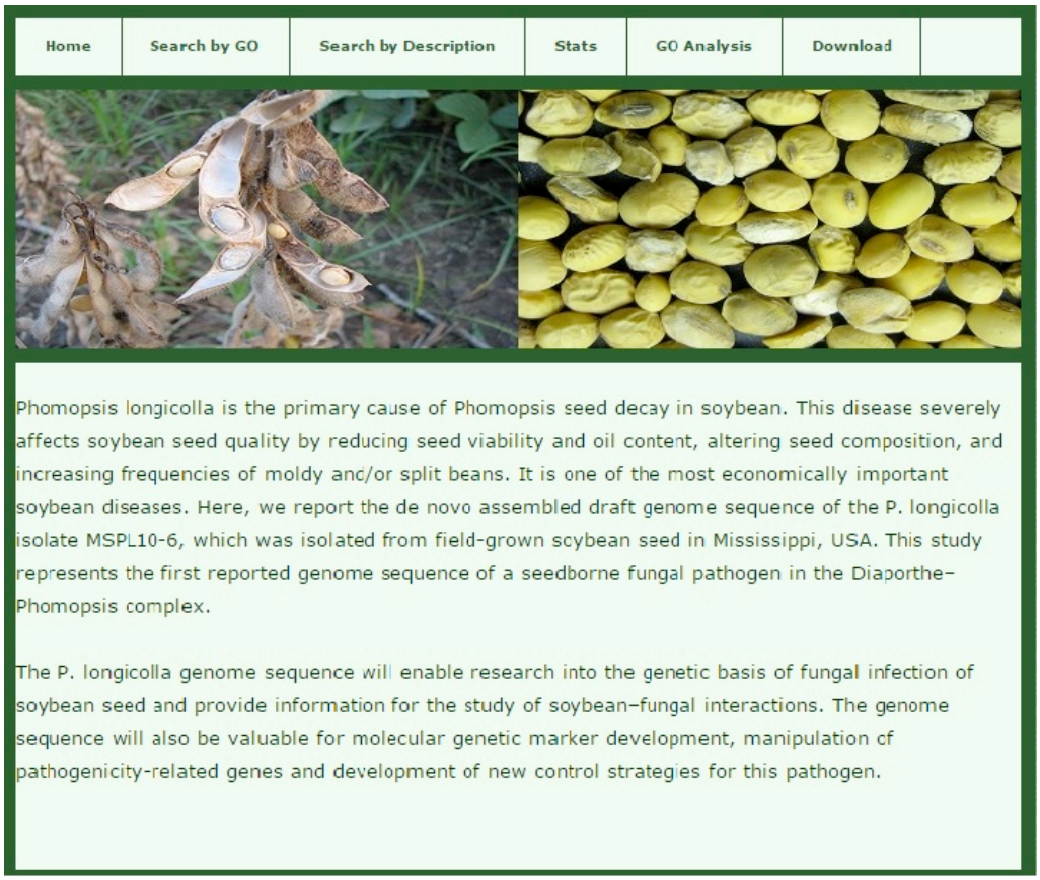

Figure 1: A snapshot of the database main web page showing a quick summary of the project and its functions in the website.

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

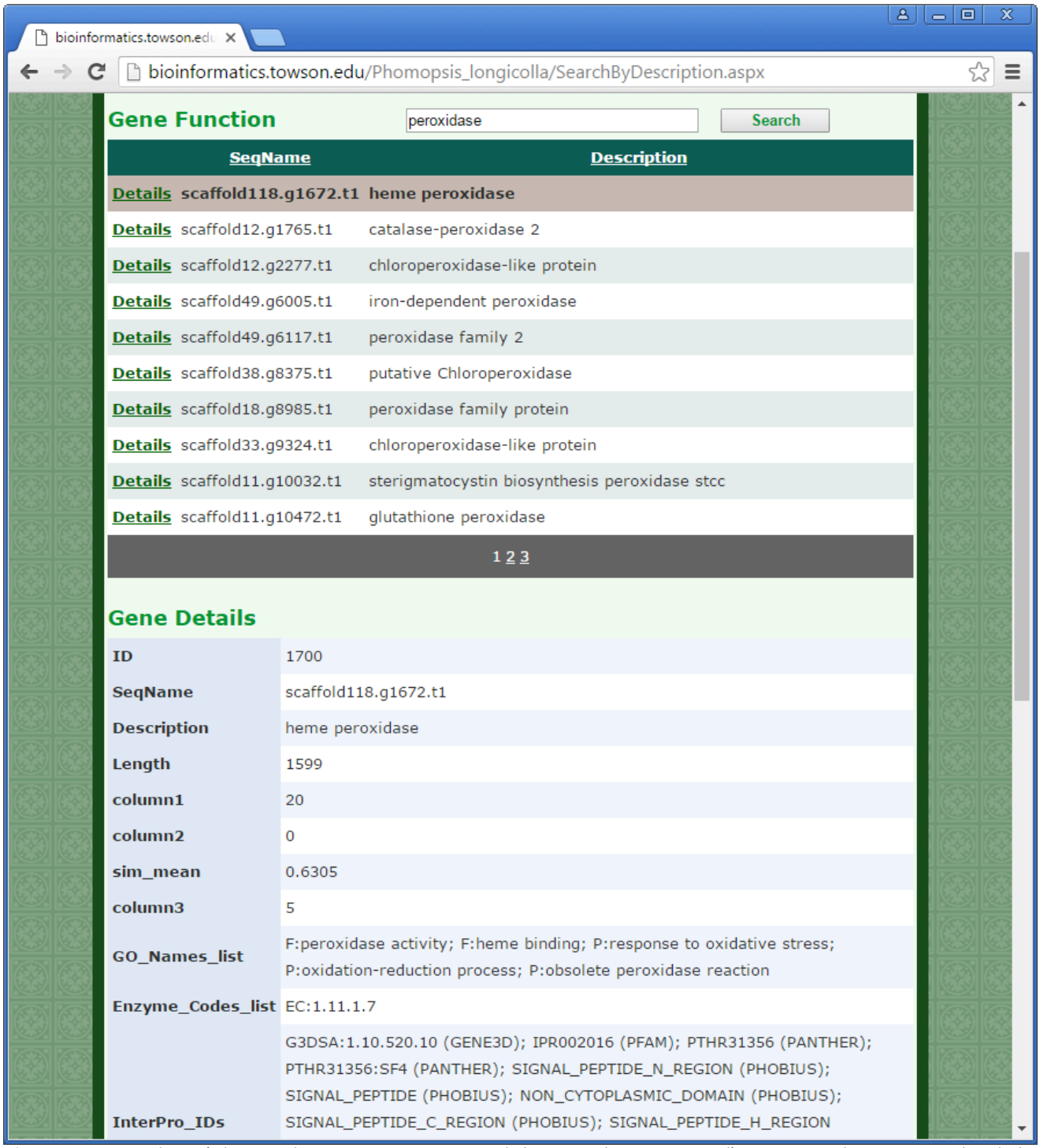

Figure 2: A snapshot of the search pages. Users can search by gene description and\or GO ontologies. More detailed information of a record (i.e. sequence) can be obtained by clicking on the "Details" link next to the sequence ID.

ISSN 0973-2063 (online) 0973-8894 (print)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

Caveats:
The assembly and genome annotation *P. longicolla* cannot be considered a complete reference for the species, as only one strain (MSPL 10-6) was sequenced.

Future Development:
Other strains of *P. longicolla* will be included on this database\site once they have been sequenced and annotated.

Authors' contributions:
Omar Darwish and Zane May, designed and developed the database and user interface under the guidance of Nadim Alkharouf at Towson University. Shuxian Li at the USDA-ARS led and coordinated the project and was in charge of fungal culture and DNA preparation for sequencing as well as the overall design of the experiments. Benjamin Matthews acted as a scientific consultant. All authors contributed to the writing of the manuscript.

Acknowledgments:
This work was partially supported by the USDA-ARS projects 6066-21220-012-00D. We are grateful to Phillip SanMiguel at Purdue Genomics Core Facility for sequencing. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture. USDA is an equal opportunity provider and employer.

References:
[1] Hobbs TW *et al*. *Mycologia* 1985 **77:** 535
[2] Li S *et al*. *Plant Dis*. 2010 **94:** 1035
[3] Hepperly PR & Sinclair JB, *Phytopathology* 1978 **68:** 1684
[4] Sinclair JB, *Plant Dis*. 1993 **77:** 329
[5] Li S, *Phomopsis seed decay of soybean,* In Soybean: Molecular Aspects of Breeding. Intech Publisher, Vienna Austria 2011 p277-292.
[6] Li S *et al*. *Phomopsis Seed Decay,* In Compendium of Soybean Diseases and Pets, Fifth Edition. APS Press. Minnesota, USA. 2015 p47-48.
[7] Zhang AW *et al*. *Phytopathology* 1998 **88:** 1306
[8] Li S *et al*. *Plant Dis*. 2001 **85:** 1031
[9] Li S *et al*. *Genome Data* 2014 **3**: 55 [PMID: 26484148]

Edited by P Kangueane
Citation: **Darwish,** Bioinformation 12(4): 233-236 (2016)