
Plan Overview

A Data Management Plan created using DMPTool

Title: Inference, analysis and computational resources of coexpression networks in bioenergy crops

Creator: Jorge Muñoz

Affiliation: Universidade de São Paulo (www5.usp.br)

Funder: São Paulo Research Foundation (fapesp.br)

Funding opportunity number: 2020/13102-0

Template: USP Template - Minimum

Project abstract:

Life is organized through networks, e.g. metabolic pathways, gene regulatory circuits, protein-protein interactions. In which parts interact among themselves. It is from these interactions that complex phenomena originate. Our understanding of biological networks is still very limited, even more so for important commercial crops such as Sugarcane. Among these biological networks, the gene co-expression networks have gained increased attention, mostly due to studies relating genes to clinical traits, but also in plant biology where it was shown the potential to identify genes related to agronomic traits. Recently the pan-transcriptome concept referred to the groups of transcripts present in all individuals of a species (core), the elements present in some individuals (accessory) and the elements found only in certain individuals (exclusive) has gained popularity in omics analysis in eukaryotic organisms due the possibility of avoiding the reference bias.

We aim to infer pan-transcriptome co-expression networks for Sugarcane and Sorghum by exploiting publicly available data, and in such a way identify network components that are conserved in these crops; Our second goal in this project is to develop and deploy a web-accessible, user-friendly OMICS integrative database, called CaneMine, envisioned as a one-stop for all OMICS information in sugarcane and sorghum, and relevant data from closely related species. The information generated within this project will aid in the understanding of the establishment and evolution of complex plant phenotypes that could be transferred into the existing breeding programs through collaborations with scientific partners.

Start date: 01-01-2021

End date: 01-01-2025

Last modified: 05-20-2022

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Inference, analysis and computational resources of coexpression networks in bioenergy crops - Description of Data and Metadata produced by the project

Data Creation and Collection

What data will be collected or created?

Os principais dados usados neste projeto serão dados de sequenciamento de transcriptoma e genoma de diversos genótipos de cana-de-açúcar e do sorgo, chamados esses dados de primários. Só serão usados dados que estão no domínio público em repositórios públicos do National Center for Biotechnology Information (NCBI: <https://www.ncbi.nlm.nih.gov/>) ou do Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>).

Esses dados vêm em dois tipos de formatos:

fastq que incluem a sequência de uma leitura de sequenciamento e a qualidade de cada base sequenciada, (https://en.wikipedia.org/wiki/FASTQ_format) e são usados para armazenar as leituras de sequenciamento, i.e., são dados brutos, e fasta que só incluem a sequência das bases, (https://en.wikipedia.org/wiki/FASTA_format) e são usadas para armazenar as montagens de transcriptomas ou genomas, i.e., são então dados processados.

How data will be collected or created

Serão pesquisados os dados de RNAseq (specific strand e paired end reads) publicados em artigos científicos para diferentes genótipos de Cana e Sorgo os metadados serão coletados manualmente.

Planned Research Outputs

Software - "CaneMine"

Como parte deste projeto, a partir desses dados públicos serão gerados novos dados: 1) pan- transcriptoma da cana-de-açúcar e do sorgo, no formato fasta 2) valores de expressão dos diferentes transcritos dessas culturas, na forma de matrizes 3) grupos de genes co-expressados, na forma de tabelas e 4) grupos de genes ortólogos na forma de tabelas

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
CaneMine	Software	Unspecified	Open	None specified		None specified	None specified	No	No