

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1JS5R>

Title: Filogenômica e macroecologia em Orchidaceae: integrando a citogenética e a ecologia no estudo da evolução de orquídeas neotropicais

Creator: Ana Moraes - **ORCID:** [0000-0002-9878-3925](https://orcid.org/0000-0002-9878-3925)

Affiliation: Universidade Federal do ABC (ufabc.edu.br)

Contributor: Magdalena Vaio, Juliano S. Cabral, Eve Lucas, Mark W. Chase, Fábio Pinheiro

Funder: São Paulo Research Foundation (fapesp.br)

Funding opportunity number: 2022/05890-3

Template: Digital Curation Centre (português)

Project abstract:

A existência de associações entre os caracteres citogenéticos, em especial a ploidia, e as características ecológicas, como a colonização de ambientes extremos, foi inicialmente proposta a quase um século, mas ainda hoje são poucos os estudos que avaliam se tais associações representam tendências gerais em plantas ou são resultados de casos pontuais. Tais análises macroecológicas, integrando dados citogenéticos e ecológicos, dependem obrigatoriamente de um *background* evolutivo que só pode ser oferecido por hipóteses filogenéticas robustas, como a filogenômica empregando bibliotecas enriquecidas. Soma-se a essa abordagem, as vantagens oferecidas pela metodologia de *Hyb-Seq*, permitindo explorar, a partir do mesmo sequenciamento, tanto a fração não-repetitiva do genoma (enriquecida pelas sondas), como também a repetitiva. Assim é possível identificar comparativamente as sequências repetitivas presentes nas espécies estudadas na filogenia, além de mapear tais sequências nos cromossomos, possibilitando integrar a citogenética molecular às análises macroecológicas. Por fim, é necessário um grupo vegetal com alta diversidade nos caracteres citogenéticos e ecológicos, como Orchidaceae, uma família amplamente distribuída, com grande variabilidade de habitats e com ampla variação de número cromossômico e de tamanho de genoma. Desta forma, o presente projeto visa investigar o uso potencial das sequências repetitivas na sistemática e evolução de Orchidaceae, além de testar se há associação entre os caracteres citogenéticos e o ecológicos empregando a maior subfamília de Orchidaceae, Epidendroideae. Assim, pretende-se revisitar as hipóteses sugeridas a quase um século, porém com análise mais robustas e consistentes que permitam identificar padrões gerais que possam ser aplicados a plantas.

Start date: 01-01-2023

End date: 12-31-2024

Last modified: 01-18-2024

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

Filogenômica e macroecologia em Orchidaceae: integrando a citogenética e a ecologia no estudo da evolução de orquídeas neotropicais

Neste projeto serão criados uma grande diversidade de dados físicos e digitais. Os dados físicos referem-se aos espécimens que serão coletados em trabalho de campo, enquanto os demais dados serão todos digitais.

Os dados digitais brutos a serem gerados são: (a) Imagens digitais das plantas coletadas; (b) Imagens digitais das preparações cromossômicas; (c) Dados de tamanho de genoma; (d) Dados de sequência de genoma; (e) Levantamentos de caracteres citogenéticos (números cromossômicos e tamanho de genoma); (f) Levantamento de hábito; (g) Dados de reconstrução da filogenia; (h) Dados de anotação do genoma. Adicionalmente serão gerados scripts para as análises em ambiente R.

Os dados físicos serão coletados em viagens de campo, junto com as imagens digitais das plantas coletadas.

As imagens digitais das preparações cromossômicas serão obtidas em microscópio de fluorescência Olympus BX51 com câmera Olympus CCD XC10 acoplada e capturadas usando o software Olympus Stream.

Os dados de tamanho de genoma serão obtidos com citômetro de fluxo da marca BD, modelo FACSCanto II com auxílio do software Cell Quest. A avaliação estatística será realizada no programa Flowing Software 2.5.1.

Dados de sequência de genoma serão obtidos em equipamento Illumina MiSeq (Illumina, San Diego, CA, United States). As sequências serão *pair end* com 150pb.

O levantamento de dados de caracteres citogenéticos e ecológicos serão obtidos via consulta aos bancos de dados C-value, CCDB e revisão da literatura via plataforma Web of Science e Google Academics.

Os dados de reconstrução da filogenia serão obtidos via análise bioinformática com softwares específicos, enquanto a anotação da fração repetitiva do genoma será realizada em plataforma Repeat Explorer2.

Adicionalmente serão gerados scripts para as análises em ambiente R que ficarão disponíveis via GitHub do grupo de pesquisa (github.com/PlantCytogenomics).

Os metadados serão codificados segundo o padrão DarwinCore (DwC; disponível em <http://rs.tdwg.org/dwc.htm>) sempre acompanhando o nome da agência financiadora e número de processo. Este padrão básico de metadados é bem estabelecido entre os padrões disponíveis e amplamente aceito na área de Biodiversidade, nas disciplinas de botânica, ecologia e genética contempladas neste projeto.

A equipe deste projeto completará os metadados para cada dataset em arquivos "readme.txt" que ficarão disponíveis, junto aos dados tabelas e resumos disponíveis no repositório dataverse, ligado ao metabuscador da FAPESP (<https://metabuscador.uspdigital.usp.br/>). Isso potencializará as possibilidades de re-uso dos dados. Os arquivos de metadados refletirão as atualizações tanto dos dados como dos próprios metadados.

Sobre coleta de amostras à campo: Por se tratar de amostras de Orchidaceae, protegidas por lei, as localizações das

coletas de material vivo só poderão ser disponibilizadas sob solicitação ao pesquisador responsável. Serão solicitadas licenças aos órgãos competentes para realização das coletas.

Sobre diretrizes para uso dos dados de pesquisa: No caso do uso de metadados pelo presente projeto, entraremos em contato com o pesquisador(a) responsável pelos dados que serão acessados. Considera-se uma questão de ética profissional reconhecer o trabalho de pesquisadores que resultou em dados e metadados que poderão ser usados. Assim, o uso de dados será sempre acompanhado pela citação dos autores originais e agradecimentos apropriados.

Todos os dados serão feitos disponíveis sob licença CC-BY. Os metadados derivados deste projeto serão disponibilizados após, ou concomitante, à publicação dos artigos científicos, ficando públicos no repositório de dados dataverse da UFABC ligado ao metabuscador da FAPESP.

Aos pesquisadores que utilizarem esses dados e metadados serão solicitados fazer referência aos autores do presente projeto, via citação incluindo o identificador de objeto digital (DOI) e a fonte de financiamento da pesquisa.

Os dados de imagem de espécimens coletados à campo e de células serão armazenados em arquivos .jpg e/ou .tiff; os dados de citometria de fluxo em arquivos .fsc e dados de genoma serão armazenadas em arquivos .fasta. Por fim, os dados de levantamento de dados da bibliografia serão armazenadas em planilhas .csv.

Os dados brutos de imagem de células serão armazenados em backups físicos (HD externo ligado aos computadores acoplados à câmera do microscópio). Periodicamente será feito backup na nuvem, utilizando conta institucional do GoogleDrive e backup adicional no repositório zenodo (<https://zenodo.org/>). Todos os demais arquivos brutos das demais técnicas também serão armazenadas nos backup virtuais (GoogleDrive fornecido pela instituição UFABC, repositório zenodo e GitHub).

Os backups virtuais serão de responsabilidade da pesquisadora responsável pelo projeto. de todos

As pastas com os dados brutos depositados nos repositórios escolhidos serão de acesso exclusivo da pesquisadora responsável pelo projeto. Os acessos aos dados será protegido por senha.

Os dados de sequência de genoma serão armazenadas a longo prazo nos repositórios virtuais GenBank e ENA. A manutenção desses dados é dos repositórios.

Os demais dados serão mantidos a longo prazo no repositório Dataverse da UFABC. Para tal, serão selecionadas as melhores imagens de células e plantas, assim como os melhores histogramas que embasem as publicações. Estes dados brutos serão armazenadas a longo prazo no repositório, enquanto os dados completos serão armazenadas pela pesquisadora responsável do projeto (backup físico e na nuvem em conta pessoal).

As tabelas de levantamento de dados também serão armazenadas no repositório Dataverse.

Os dados brutos poderão ser compartilhadas sob solicitação e os metadados e dados selecionados para as publicações serão disponibilizados livremente no repositório Dataverse.

Serão priorizados repositórios providos pela instituição sede e repositórios livres (ie, zenodo e GitHub) viabilizando

o armazenamento a longo prazo dos dados e metadados.

Os dados serão compartilhados via palestras, TCC/dissertações/teses, relatórios para agência de fomento e, por fim, via artigos científicos. Os metadados serão disponibilizados como material suplementar nestes artigos e armazenadas em repositórios virtuais de livre acesso. Os dados brutos poderão ser obtidos via solicitação à pesquisadora responsável.

Não existem restrições ao compartilhamento de dados.

Neste projeto a obtenção de dados e produção dos metadados serão responsabilidade de alunos associados ao projeto e dos pesquisadores associados, cada um participando de um ou dois subprojetos. A pesquisadora responsável participará da obtenção de dados de todos os subprojetos.

O controle de qualidade do material produzido será da pesquisadora responsável e dos pesquisadores associados.

O armazenamento de dados e backup, assim como o compartilhamento de dados, será de responsabilidade exclusiva da pesquisadora responsável.
