Draft template for Applications in Plant Sciences - Genomic Resources Article
Short Title for Running Head: Surname of the first author, followed, as appropriate, with the surname of a sole co-author, or et al. (if there are three or more authors) - 2 to 4 descriptive words [e.g., Smith et al. - Root staining in Fabaceae]

A transcriptome resource for the cleistogamous herb, *Impatiens capensis* (Balsaminaceae)[1]

[no more than 125 characters; after a species name, include family name in parentheses]

Shannon E. Smiley[2], William Y. Brown[2], and James D. Bryant[2,3,4]

[2]Rancho Santa Ana Botanical Garden, 1500 N. College Ave., Claremont, California 91711 USA

[3]Department of Biological Sciences, University of Pittsburgh, 4249 Fifth Ave., Pittsburgh, Pennsylvania 15260 USA

Email addresses: SES: smileSE@test.edu

WUB: brownwy@test.edu

JDB: jdbryant@test.edu

Number of words: YYYY *[6000 or fewer, consisting of Introduction, Methods & Results, Discussion, and Conclusion sections]*.

[1] Manuscript received _____; revision accepted _____.

[4] Author for correspondence: jdbryant@test.edu

Acknowledgments should be limited to no more than 75 words.

ABSTRACT

The abstract should capture the interest of the general plant sciences community as well as specialists within the area. The abstract is 200 words or less, written in a structured format:

- Premise of the study (why the genomic resource is necessary)
- Methods (how the genomic resource was developed and tested)
- Results (results of the study, how the new resource compares to other genomes, etc.)
- Discussion (how the resource is applicable to the plant sciences)

Avoid references; if essential, cite parenthetically with journal name, volume number, pages, and year.

Here is a sample abstract:

- *Premise of the study:* Cleistogamy is a reproductive system in which plants produce open, chasmogamous flowers and closed, cleistogamous flowers. Although classic reproductive investigations focused on the model system *Impatiens capensis*, gene expression patterns associated with the two floral types remain unknown due to lack of a genomic resource.
- *Methods:* Using the 454 sequencing platform, nearly 1.1 million sequences were collected from multiple cDNA libraries representing different floral tissues and assembled into 46,342 contigs with a combined length of 14,299,355 bp.
- *Results:* Known genes were identified from these contigs and compared to *Arabidopsis* and to GenBank sequences, which resulted in annotation of 11,788 contigs. Gene ontology was conducted to characterize the types of genes in this resource and their functional class.
- *Discussion:* The resource described here for *I. capensis* will enable investigations of gene expression in floral development and may serve as a reference for other cleistogamous species.

Key words: Balsaminaceae; chasmogamy; cleistogamy; *Impatiens capensis*; transcriptome.

[Please list 3 to 6 key words here in alphabetical order, separated by semicolons.]

INTRODUCTION

This section should consist of no more than five paragraphs outlining the reasons behind

the development of the new genomic resource, including a convincing argument for why it is necessary and also its importance relative to other resources currently available. Authors must also provide readers with any background information necessary to understand the applicability of the resource to the plant sciences.

METHODS

The **Methods** section will consist of no more than seven paragraphs in which the development of the genomic resource (transcriptome, whole plant genome, etc.) should be adequately characterized. If any chemicals or supplies are mentioned, the location of the supplying company must be provided in parentheses (e.g. "Li-Cor, Lincoln, Nebraska, USA"). If plant samples were used in the development or testing of the protocol, the number and geographic origin of specimens analyzed (using GPS decimal degrees or to the nearest second) must also be included; any voucher specimens must also be given here, in a table footnote, or in an appendix (in the case of multiple voucher specimens).

The information provided in this section may vary with the type of genomic resource presented, but for transcriptome studies there must be a clear description of the sequencing procedure, method of contig and scaffold assembly, and metrics on the number, sizes, and distribution of contigs/scaffolds (including estimated genome coverage of the combined scaffolds). At a minimum, the following run and assembly statistics should be included: number of reads, number of contigs, coverage depth statistics, and N50 length. Data must also be deposited into an appropriate online repository prior to submission of the manuscript. Authors must ensure that all

raw and edited data is readily accessible to readers upon publication by deposition to a publically-accessible database archive (such as the NCBI Transcriptome Shotgun Assembly [TSA] sequence database); the corresponding accession number(s) must be provided in the manuscript.

For whole plant genomes, papers must include an indication that coverage was appropriate and a complete description of the assembly (e.g., N50, L50, maximum scaffold/contig size, total number of bases in scaffolds). In addition, gene prediction and functional annotation must be provided, based on either transcriptome resources (reported in the manuscript or published elsewhere) or *ab initio* gene prediction models. Assembly accuracy should be addressed by at least one method. For next-generation-sequencing–based studies, authors must submit raw data to the NCBI Sequence Read Archive (or equivalent databases for EMBL or DDBJ) and include the corresponding accession numbers in the manuscript. Genome assemblies and related resources (scaffolds; protein predictions) must be made publically available by web server.

RESULTS

The **Results** section will consist of no more than six paragraphs. The authors must clearly demonstrate the quality and utility of the resource by comparing the resource to other genomes (which can be shown graphically in terms of divergence with other taxa, for example). If appropriate, genes should be parsed into known functional groups and examined in terms of gene ontology, especially if sampling was conducted in multiple tissues or different environments. Results of gene prediction annotation much also be presented. Tables, figures, or multimedia content associated with these analyses are recommended. For whole plant genomes, other elements

could include sections with re-sequencing for genetic diversity, non-coding RNAs, plastid sequence assembly, and analysis of genome duplication (in the case of polyploids. Manuscripts that simply present a new resource without any other supporting information or that lack a rigorous examination of the data will be returned to the authors without review.

DISCUSSION

In this section, the author(s) should clearly articulate the importance of this genomic resource and state in several paragraphs the main conclusions that have been reached, focusing on the effectiveness and applicability of the genomic resource in comparison to other existing resources. In addition, the authors may address potential advances that would be gained through the application of this new resource.

LITERATURE CITED *[no more than 60]*

GOUDET, J. 1995. FSTAT: A computer program to calculate *F* statistics, version 1.2. *Journal of Heredity* 86: 485–486.

STEBBINS, G. L. 1974. Flowering plants: Evolution above the species level. Belknap Press, Cambridge, Massachusetts, USA.

STEVENS, P. F. 2001 onward. Angiosperm phylogeny website, version 8, June 2007 [more or less continuously updated]. Website http://www.mobot.org/MOBOT/research/APweb/ [accessed 00 Month Year].

TURNER, B. L., AND R. M. KING. 1977. Chromosome numbers in the Compositae. VIII. Mexican

and Central American species. *Southwestern Naturalist* 9: 27–39.

WHITE, T. J., T. D. BRUNS, S. B. LEE, AND J. W. TAYLOR. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White [eds.], PCR protocols: A guide to methods and applications, 315–322. Academic Press, San Diego, California, USA.

Tables

Tables must follow the general format of all APPS manuscripts in which each table has a brief legend, with footnotes explaining any abbreviations within the table itself.

Figure and Legend

Figures are strongly encouraged for Genomic Resources manuscripts to enhance key concepts. As just one example, a transcriptome report could include a histogram of annotated contig length distributions; phenograms, Venn diagrams, or scatterplots for comparisons with other taxa or expression levels in different tissues; and pie graphs or histograms detailing gene ontology categories. Figures should be uploaded as separate files with the legend included in the text file.

Appendix

If necessary, the appendix can be used for supplementary information. If there are multiple voucher specimens, they should be listed in the appendix, using the following format:

APPENDIX 1. Species, population voucher, municipality/state, country, and GPS coordinates of all samples used in this study. Abbreviations: AM = Amazonas; BA = Bahia; MG = Minas Gerais; MS = Mato Grosso do Sul; MT = Mato Grosso; PR = Paraná; SC = Santa Catarina; SP = São Paulo.

Utricularia gibba—UG3: Mogi das Cruzes/SP, Brazil (−23.532917, −46.143972); UG4: Mogi das Cruzes/SP, Brazil (−23.557844, −46.137386); UG5: Itararé/SP, Brazil (−24.084417, −49.201639); UG6: Guaratuba/PR, Brazil (−26.023625, −48.770411); UG7: São Bento do Sul/SC, Brazil (−26.361697, −49.388964); UG8: Corumbá/MS, Brazil (−19.008889, −57.652778); UG9: Presidente Figueiredo/AM, Brazil (−60.020556, −2.052489).

Utricularia neottioides—UN1: Chapada dos Guimarães/MT, Brazil (15.383333, −55,833333); UN2: Piatã/BA, Brazil (−13.151356, −41.758842); UN3: Santa Bárbara/MG, Brazil (−19.958889, −43.415000); UN4: Raudal Caldero/Amazonas, Venezuela (4.766667, −66.683333).

Utricularia reniformis—UR1: Salesópolis/SP, Brazil, (−23.649222, −45.677833); UR2: Biritiba-Mirim/SP, Brazil (−23.658306, −46.034556); UR3: Bananal/SP, Brazil (−22.798722, −44.377917); UR5: Itararé/SP, Brazil (−24.115472, −49.363611); UR6: Mogi das Cruzes/SP, Brazil (−23.751353, −46.126506); UR7: Campina Grande do Sul/PR, Brazil (−25.245278, −48.834167); UR8: Corupá/SC, Brazil (−26.393211, −49.354878); UR9: Morretes/PR, Brazil (−25.127778, −48.820278).

Utricularia subulata—US1: Salesópolis/SP, Brazil (−23.556856, −46.137842); US2: Mogi das Cruzes/SP, Brazil (−23.534294, −46.144850); US3: Jaguariaíva/PR, Brazil (−24.250833, −49.705833).

Appendix 2. Voucher information for *Armeria* species used in this study.

Species	Voucher specimen accession no.[a]	Collection locality[b]	Geographic coordinates	No. of individuals
A. caespitosa	Acp-003-AG	Cabeza de Hierro, Madrid	40°47′57.14″N, 3°57′3.21″W	20
A. caespitosa	Acp-015-AG	Pico del Lobo, Guadalajara	41°11′0.23″N, 3°27′58.91″W	20
A. bigerrensis	Abg-002-AG	Morezón, Ávila	40°14′56.13″N, 5°16′11.33″W	5
A. cantabrica	Act-001-AG	Torrecerredo, Asturias	43°12′3.26″N, 4°50′53.19″W	5
A. maritima	Amt-001-AG	Cabo Mayor, Santander	43°29′26.94″N, 3°47′26.27″W	5

Note: AG = Alfredo García, collector.

[a] Vouchers deposited at Universidad Rey Juan Carlos, Departamento de Biología y Geología, Germplasm bank.

[b] Locality and Spanish province.