ARN y análisis transcriptómica para biología organísmica

Carrie M Tribble, PhD
University of Hawai`i at Mānoa
School of Life Sciences

@tribbletweets
ctribble09@gmail.com

ANTECEDENTES

ANTECEDENTES

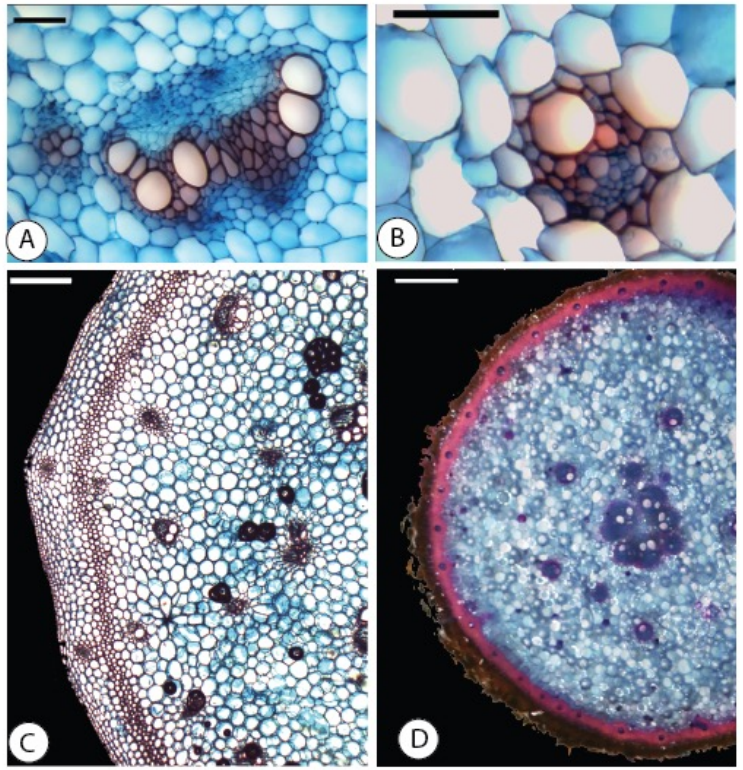
ANTECEDENTES

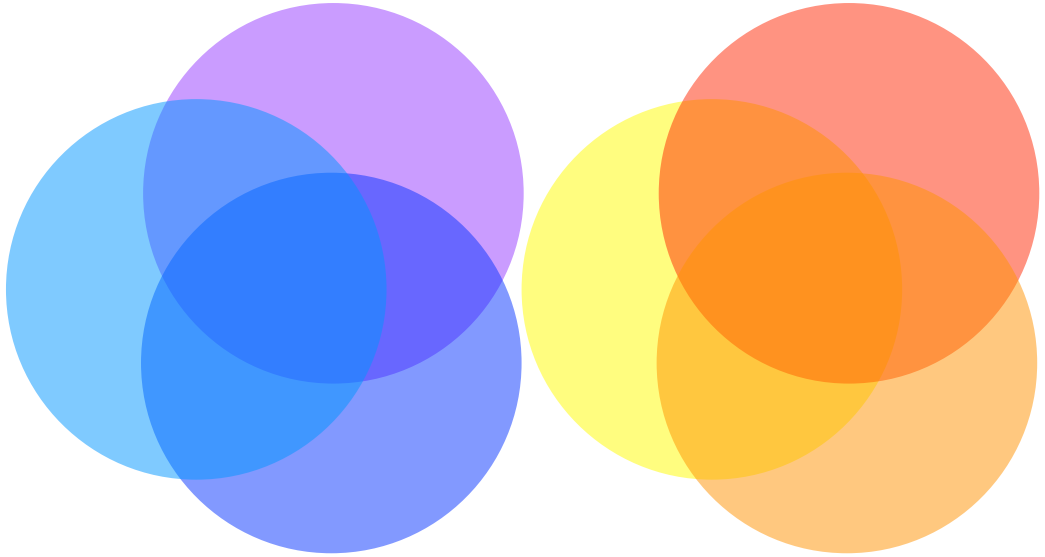
ANTECEDENTES

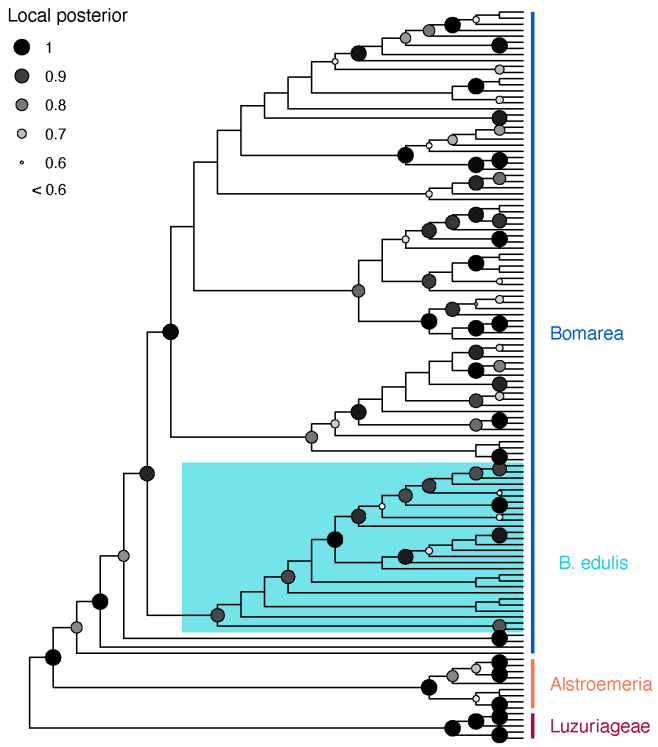
~4x

ANTECEDENTES

Are repeatedly-evolved phenotypes the result of the same underlying molecular and anatomical mechanisms?





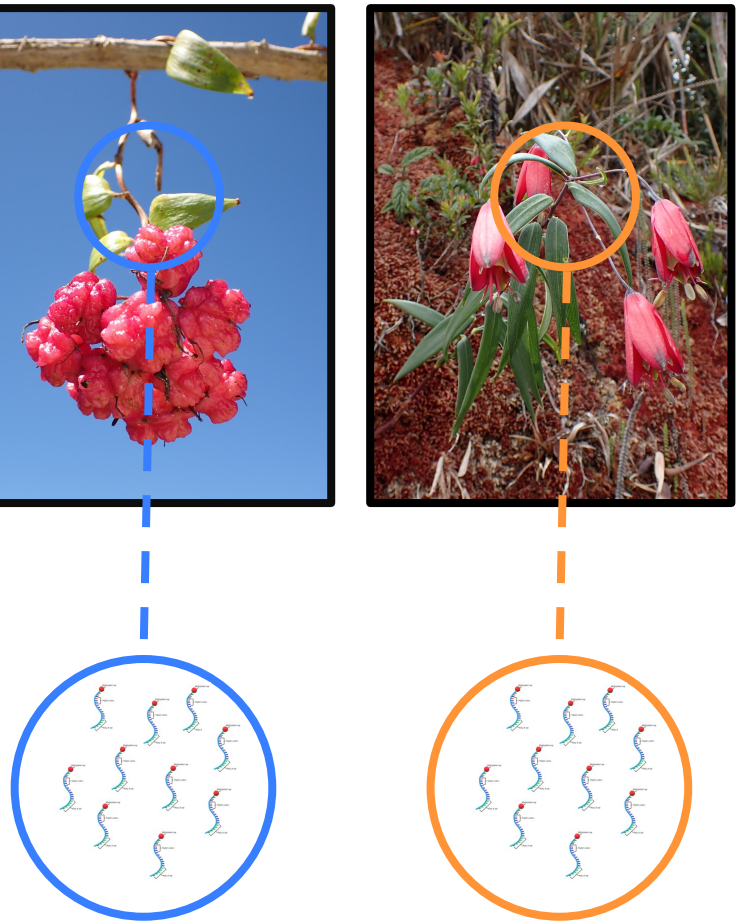


Anatomía

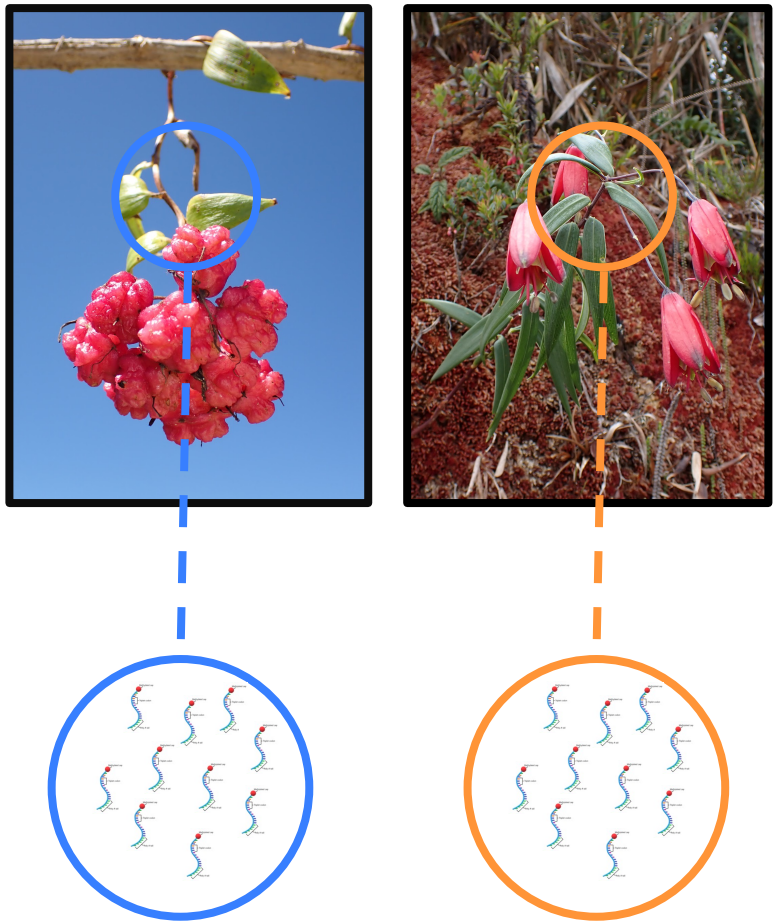
Genoma

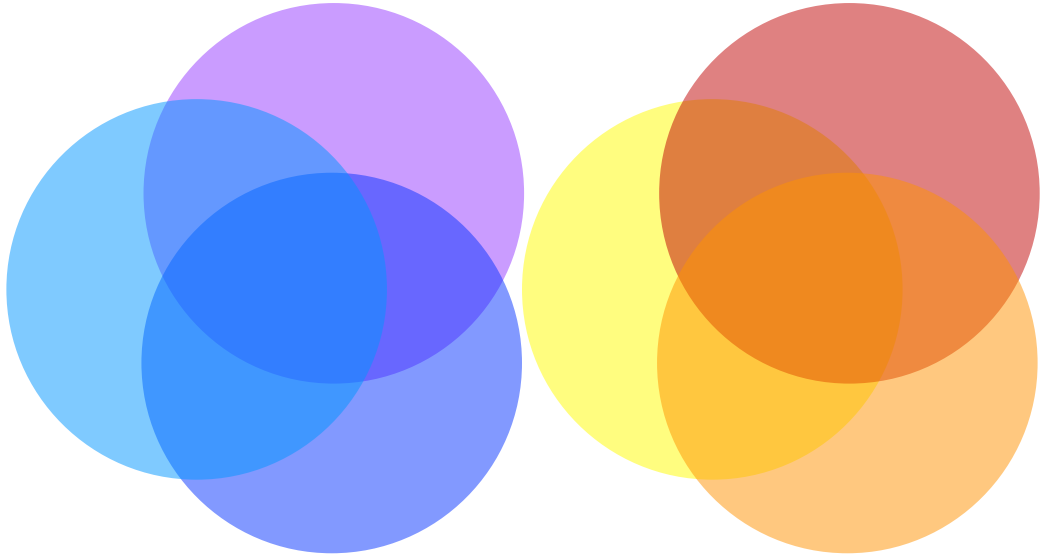
Macroevolución

Identificar genes candidatos

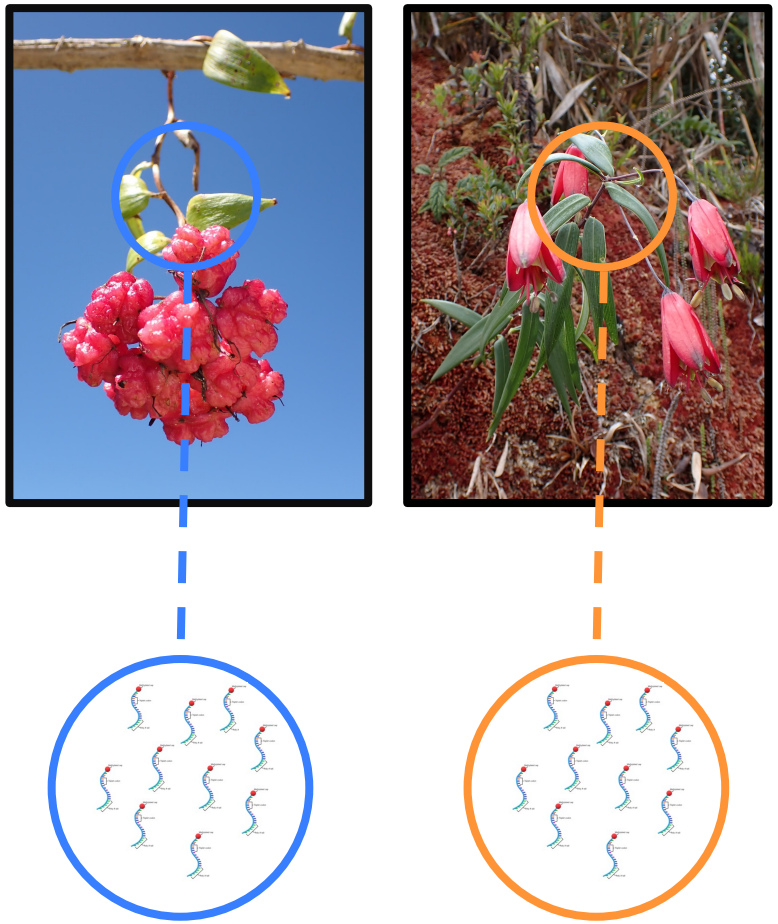


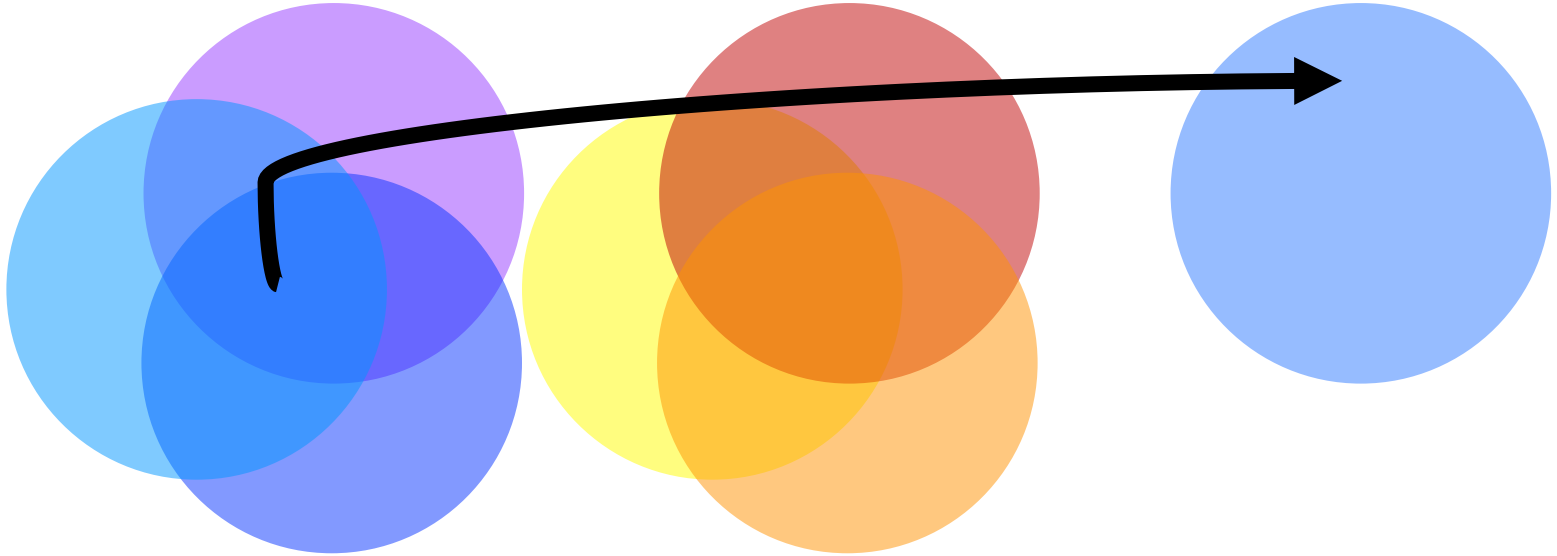
Identificar genes candidatos



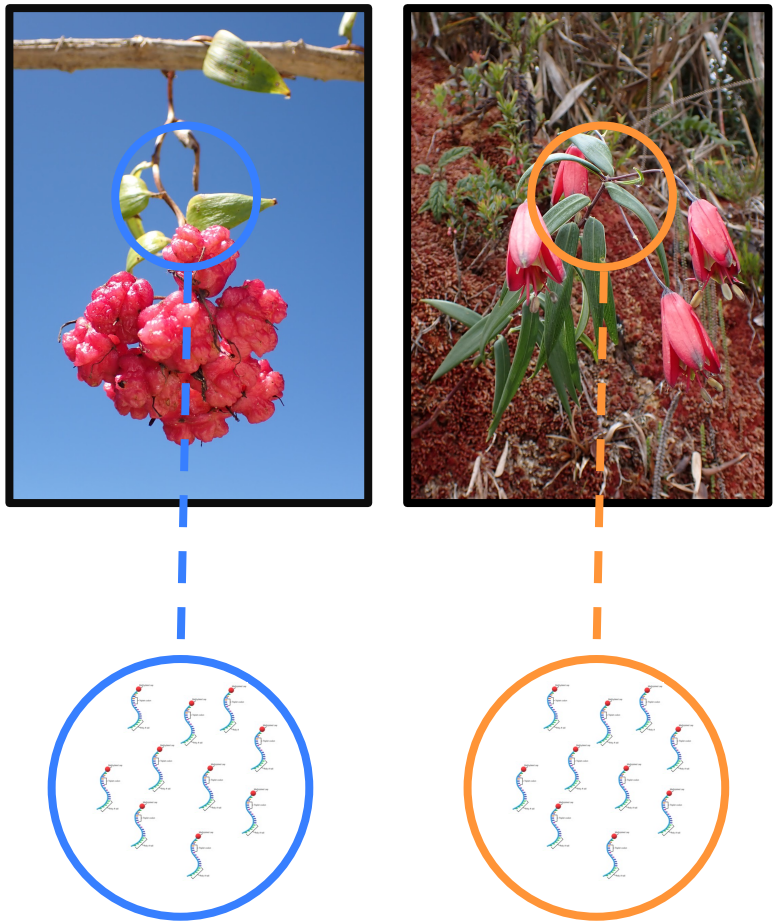


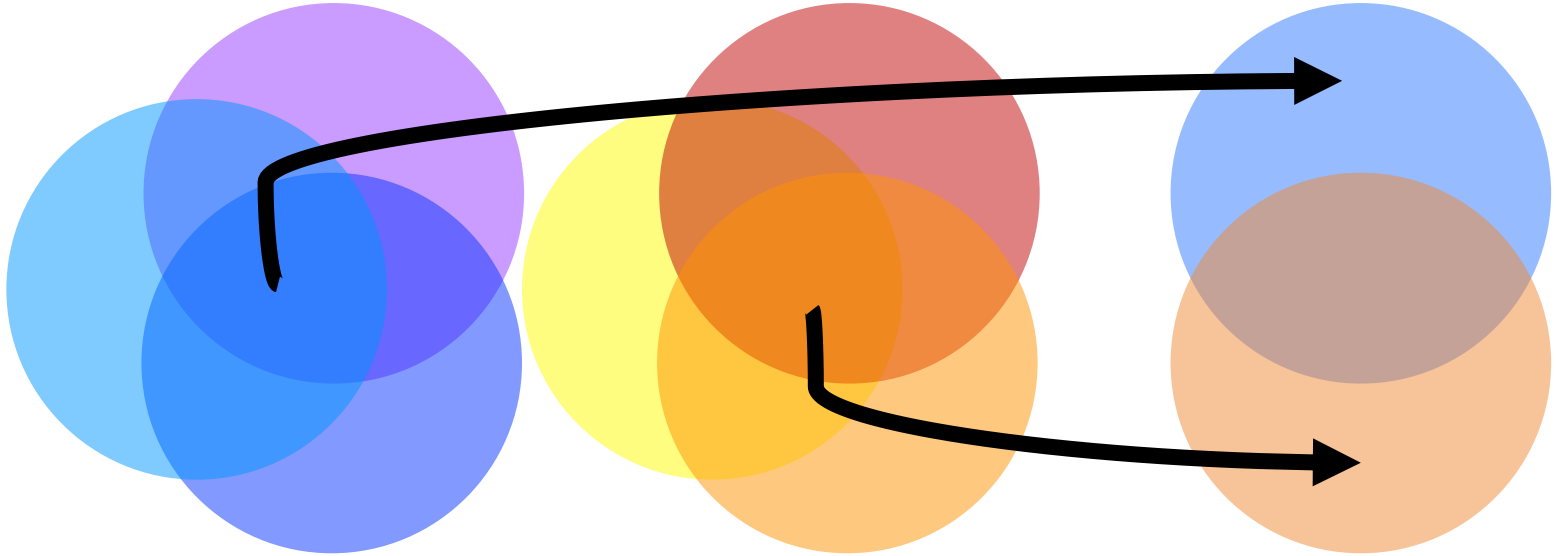
Identificar genes candidatos



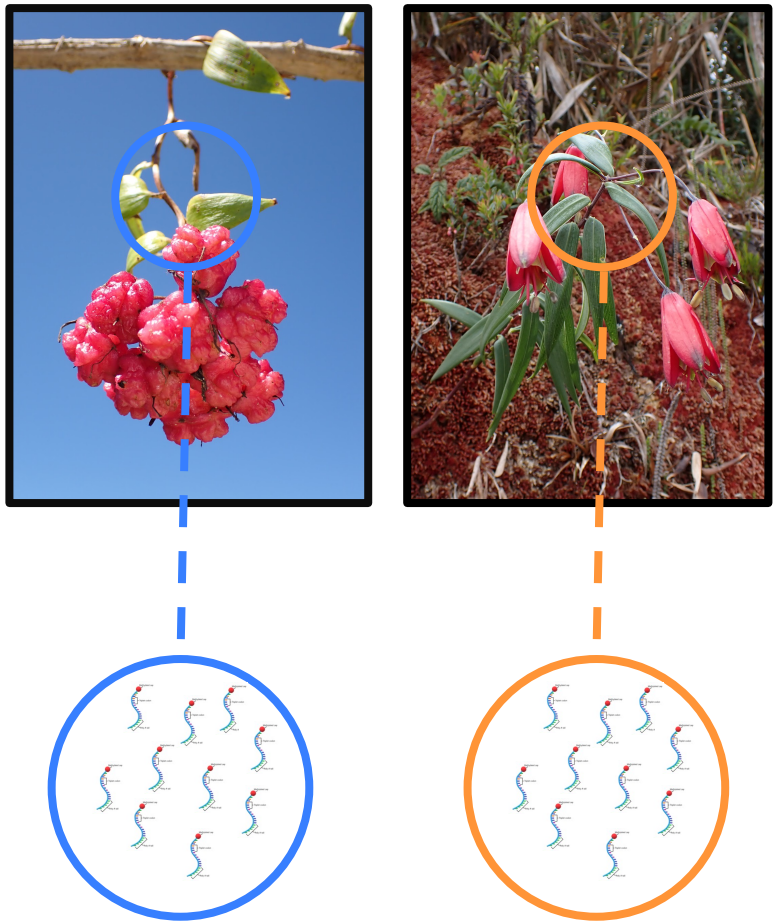


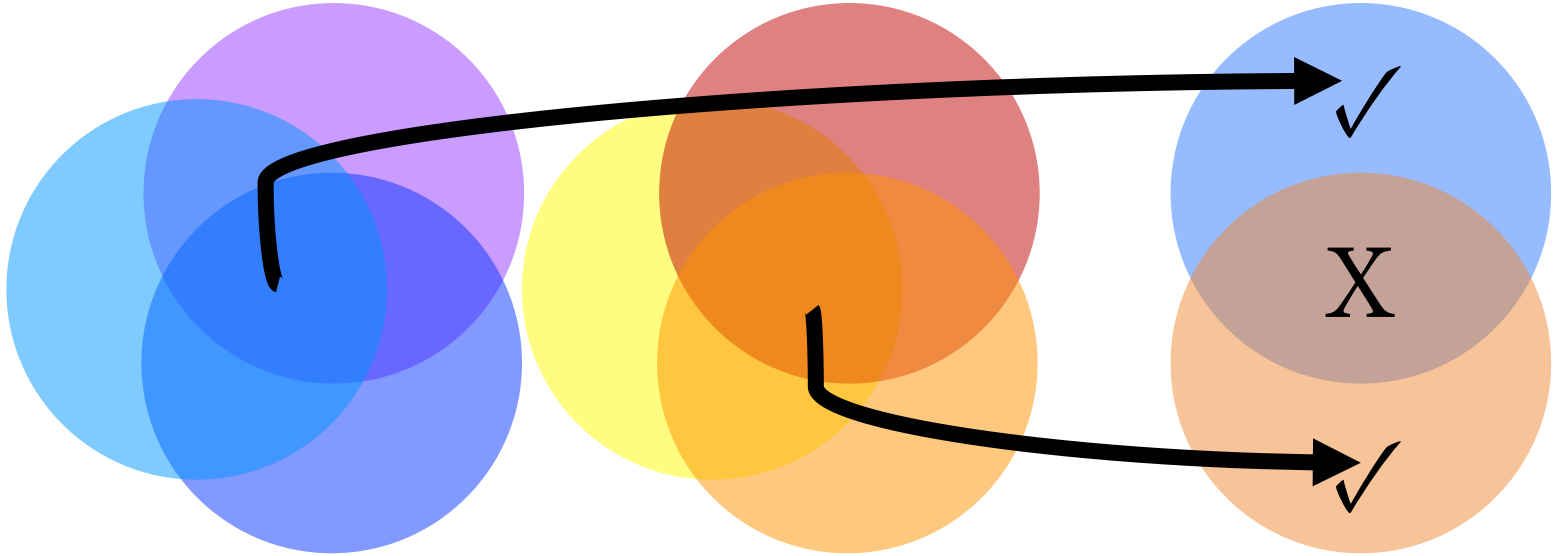
Identificar genes candidatos





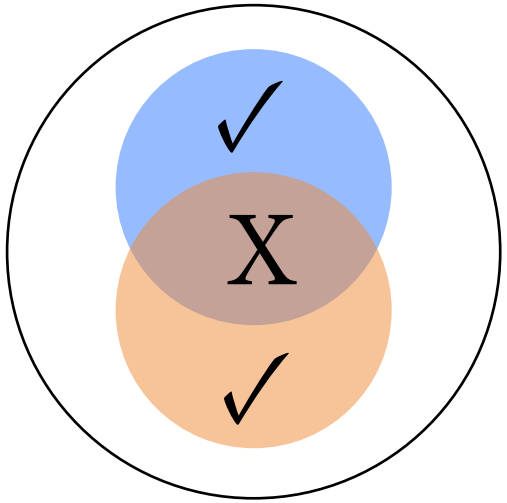
Identificar genes candidatos

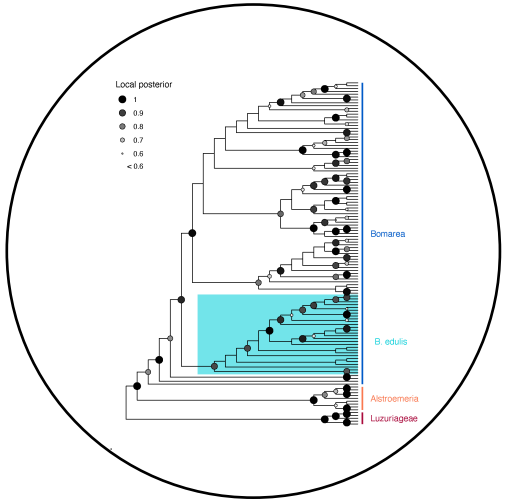


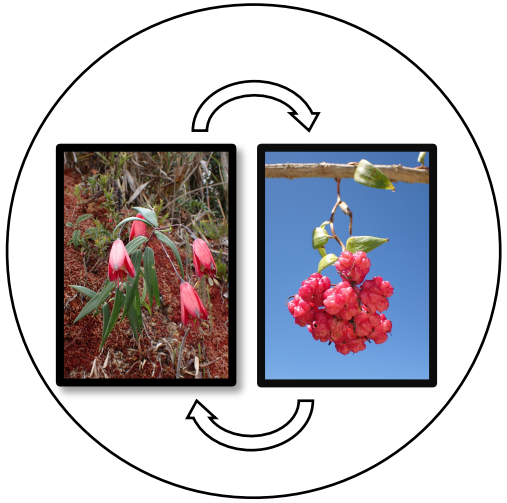


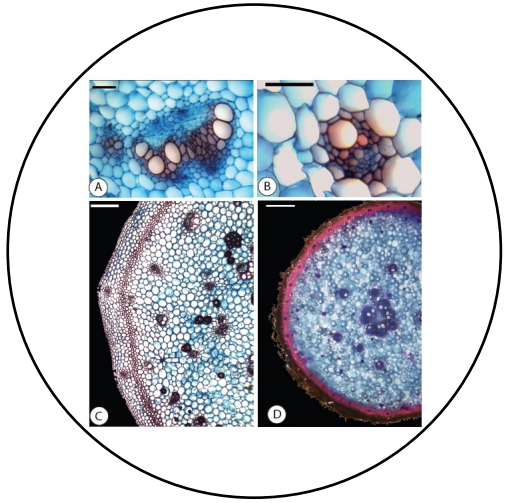
Escandentes

Erectas

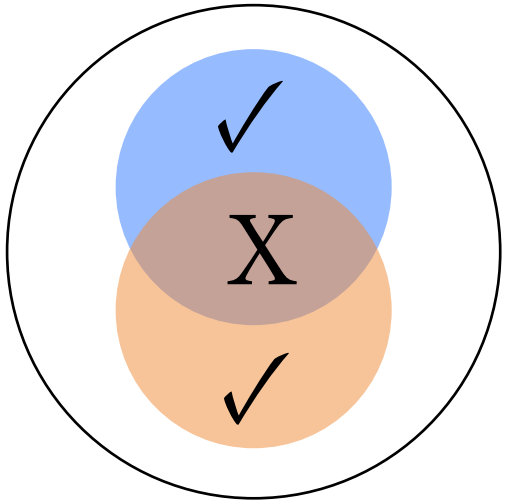


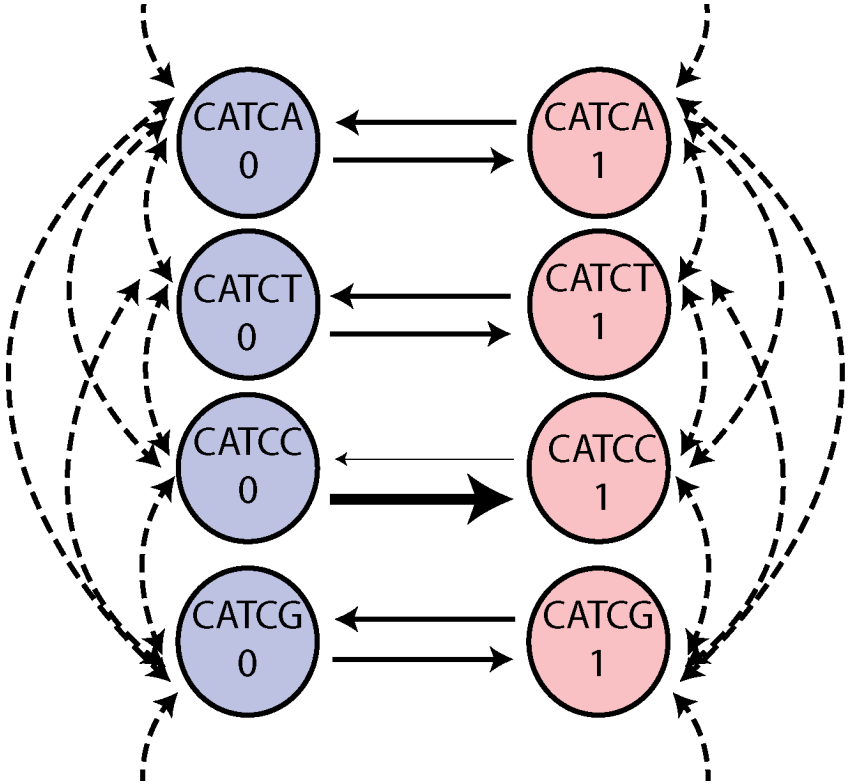


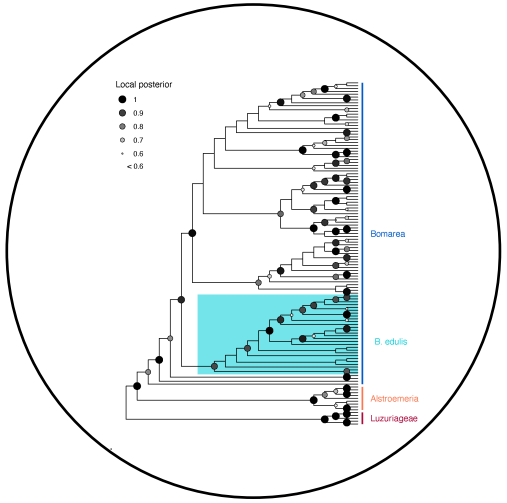


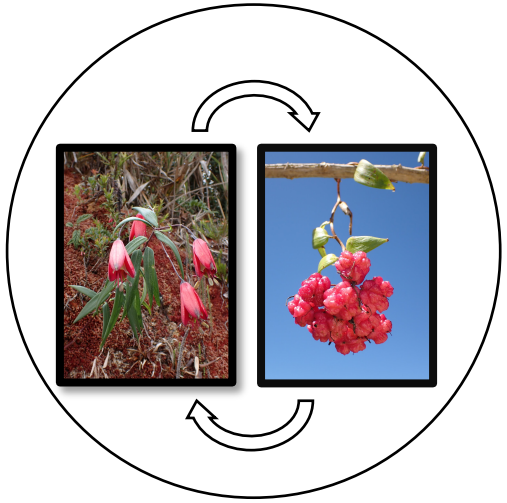


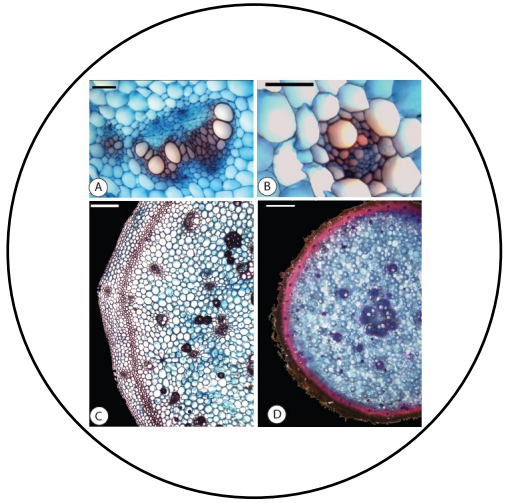
ANTECEDENTES











ANTECEDENTES

https://diytranscriptomics.com/

https://datacarpentry.org/lessons/#genomics-workshop

https://biovcnet.github.io/

https://training.galaxyproject.org/

https://diytranscriptomics.com/

https://datacarpentry.org/lessons/#genomics-workshop

https://biovcnet.github.io/

https://training.galaxyproject.org/ [opciones en español]

Descripción general

Preguntas:

- ¿Qué genes se expresan diferencialmente entre las células G1E y los megacariocitos?
- ¿Cómo podemos generar un transcriptoma de novo a partir de datos de secuenciación de ARN?

Objetivos:

- Análisis de datos de secuenciación de ARN usando un genoma de referencia
- Reconstrucción de transcritos sin transcriptoma de referencia (de novo)
- Análisis de genes expresados diferencialmente

Requisitos:

- Introducción a los análisis de galaxias
- Análisis de secuencia
  - Control de calidad: diapositivas - las manos en
  - Cartografía: diapositivas - las manos en

Tiempo estimado: 6 horas

Materiales de apoyo:

conjuntos de datos Flujos de trabajo preguntas frecuentes Grabaciones Disponible en estos Galaxy

Última modificación: 28/11/2022

Licencia: El contenido del tutorial tiene licencia de Creative Commons Attribution 4.0 International License El Marco GTN tiene licencia de MIT

1) Teoría de transcriptomas (1 hora)

2) Proceso de realizar un experimento (0.5 hora)

3) Analizar datos en R (2 horas)

1) Teoría de transcriptomas (1 hora)

2) Proceso de realizar un experimento (0.5 hora)

3) Analizar datos en R (2 horas)

1) Qué es el ARN?

2) Qué es el transcriptoma?

3) A cuáles preguntas podemos responder utilizando ARN?

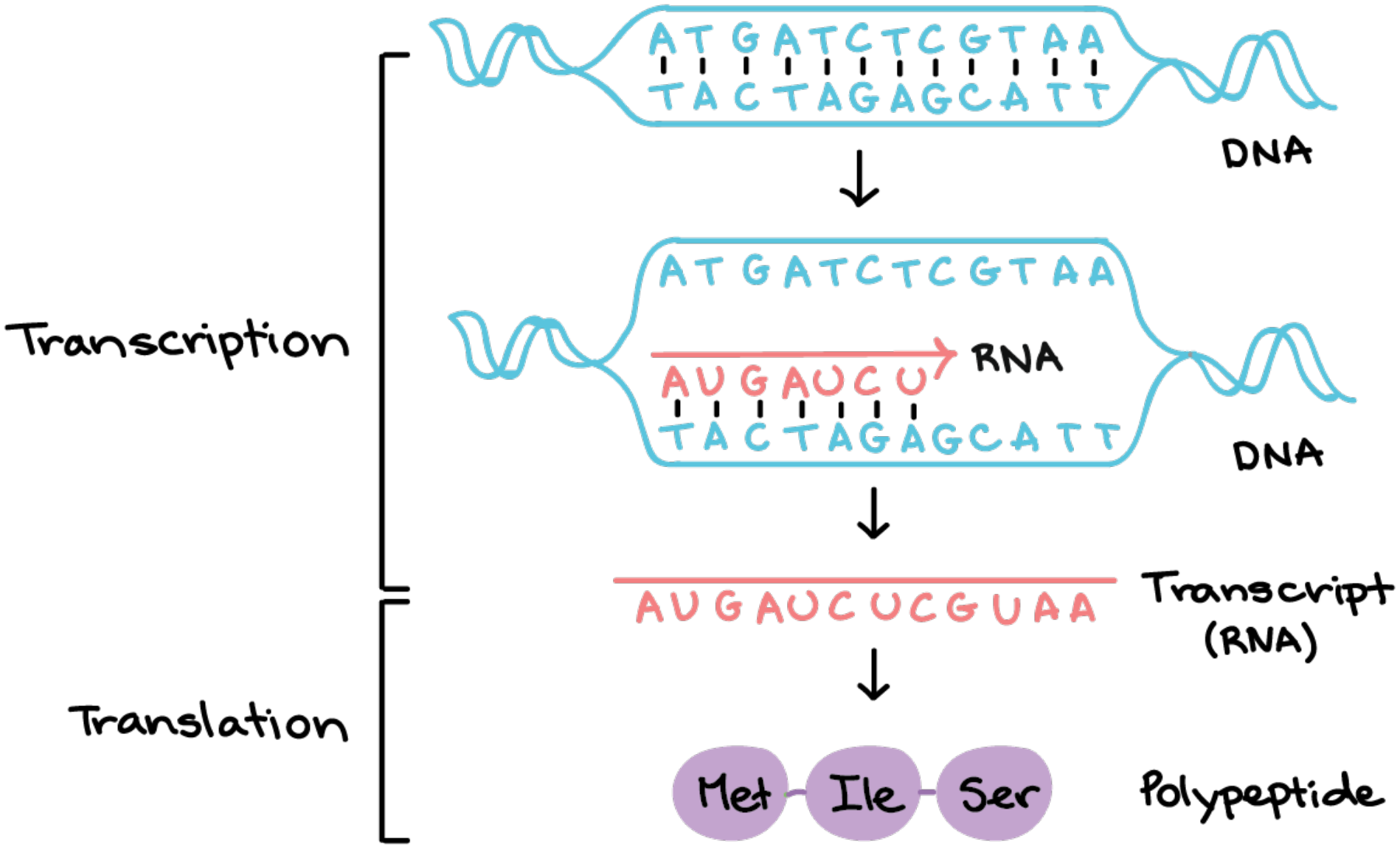
4) Cómo diseñar un experimento?

1) Qué es el ARN?

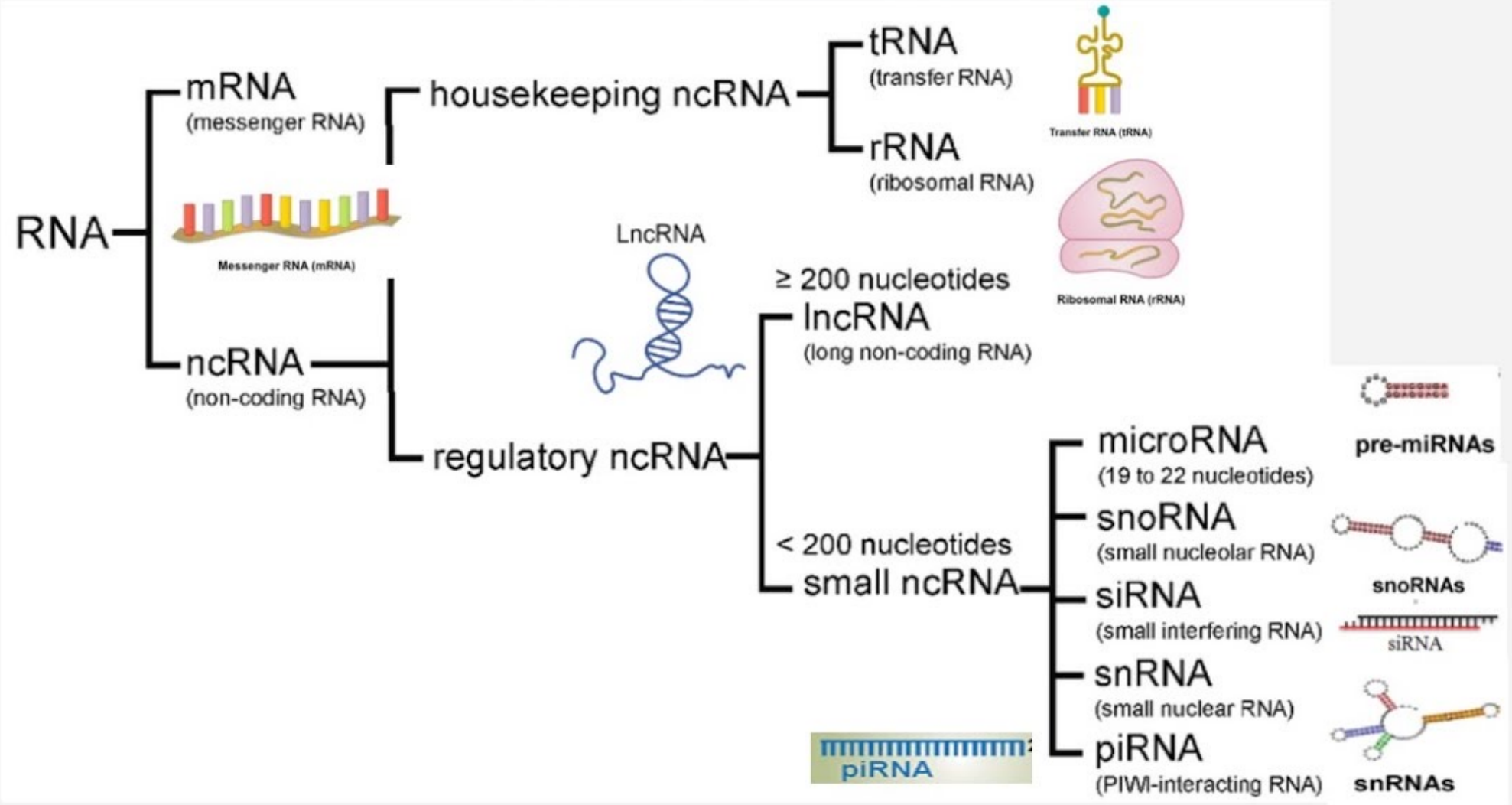
2) Qué es el transcriptoma?

3) A cuáles preguntas podemos responder utilizando ARN?

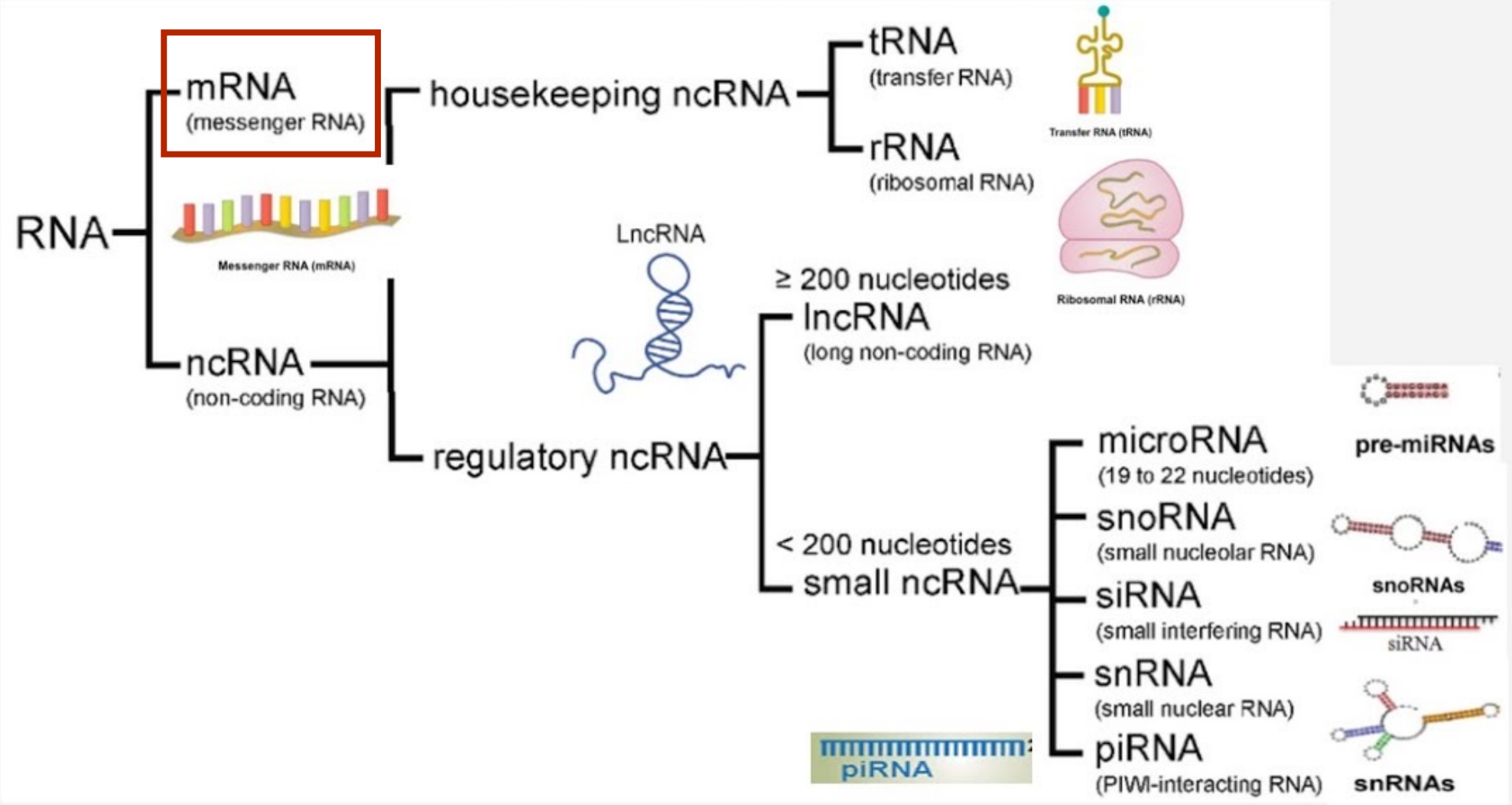
4) Cómo diseñar un experimento?

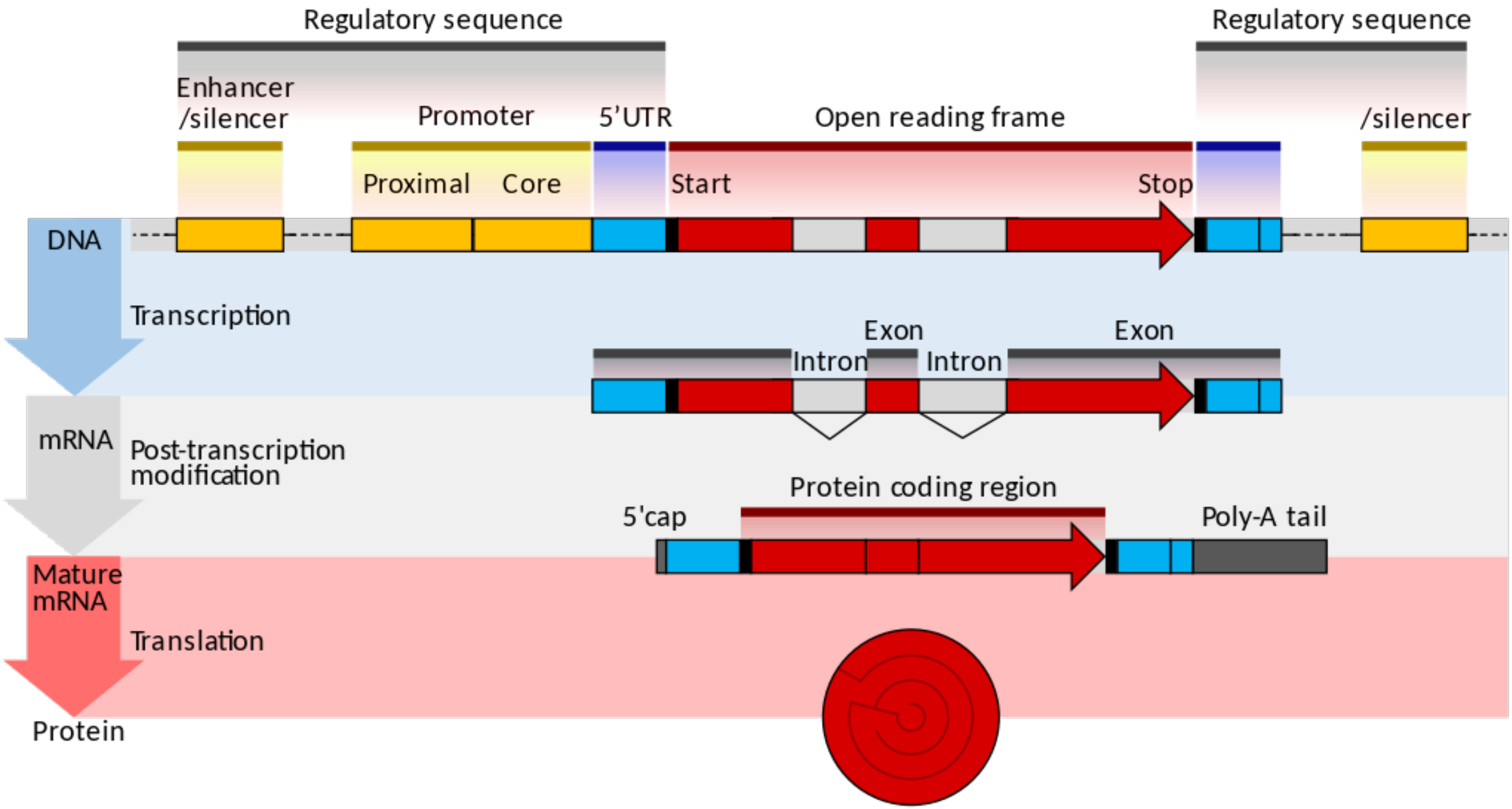

ATGATCTCGTAA
TACTAGAGCATT
DNA
ATGATCTCGTAA
RNA
AUGAUCU
TACTAGAGCATT
DNA
AUGAUCUCGUAA
Transcript (RNA)
Met
Ile
Ser
Polypeptide
Transcription
Translation

TYPES OF RNA


RNA
mRNA
(messenger RNA)
Messenger RNA (mRNA)
ncRNA
(non-coding RNA)
housekeeping ncRNA
tRNA
(transfer RNA)
Transfer RNA (tRNA)
rRNA
(ribosomal RNA)
Ribosomal RNA (rRNA)
regulatory ncRNA
LncRNA
≥ 200 nucleotides
lncRNA
(long non-coding RNA)
< 200 nucleotides
small ncRNA
microRNA
(19 to 22 nucleotides)
pre-miRNAs
snoRNA
(small nucleolar RNA)
snoRNAs
siRNA
(small interfering RNA)
siRNA
snRNA
(small nuclear RNA)
snRNAs
piRNA
(PIWI-interacting RNA)
piRNA

TYPES OF RNA


RNA
mRNA
(messenger RNA)
Messenger RNA (mRNA)
ncRNA
(non-coding RNA)
housekeeping ncRNA
tRNA
(transfer RNA)
Transfer RNA (tRNA)
rRNA
(ribosomal RNA)
Ribosomal RNA (rRNA)
regulatory ncRNA
LncRNA
≥ 200 nucleotides
lncRNA
(long non-coding RNA)
< 200 nucleotides
small ncRNA
microRNA
(19 to 22 nucleotides)
pre-miRNAs
snoRNA
(small nucleolar RNA)
snoRNAs
siRNA
(small interfering RNA)
siRNA
snRNA
(small nuclear RNA)
snRNAs
piRNA
(PIWI-interacting RNA)
piRNA


Regulatory sequence
Regulatory sequence
Enhancer
/silencer
Promoter
5'UTR
Open reading frame
/silencer
Proximal
Core
Start
Stop
DNA
Transcription
Exon
Exon
Intron
Intron
mRNA
Post-transcription
modification
Protein coding region
5'cap
Poly-A tail
Mature
mRNA
Translation
Protein

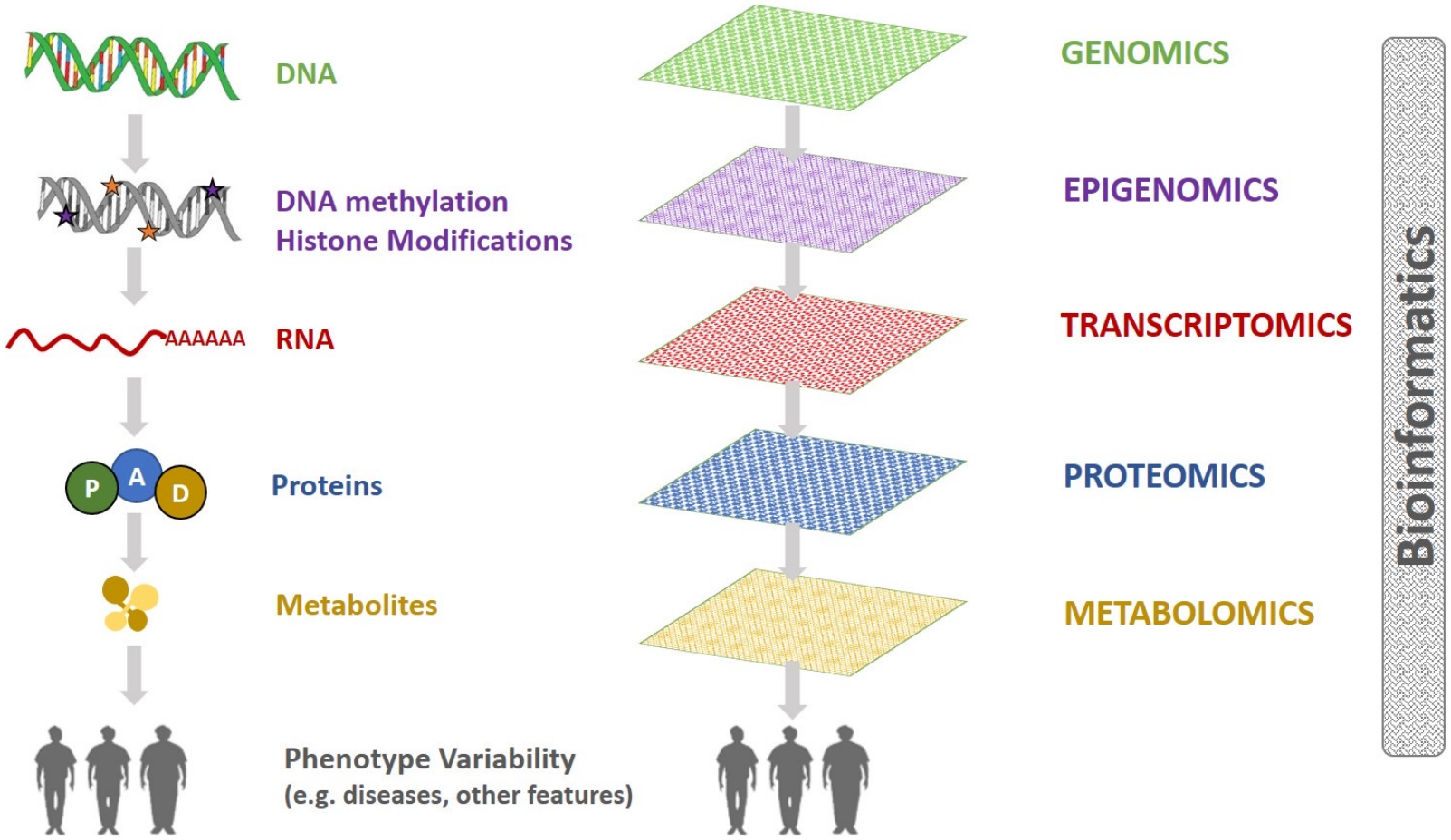
1) Qué es el ARN?

2) Qué es el transcriptoma?

3) A cuáles preguntas podemos responder utilizando ARN?

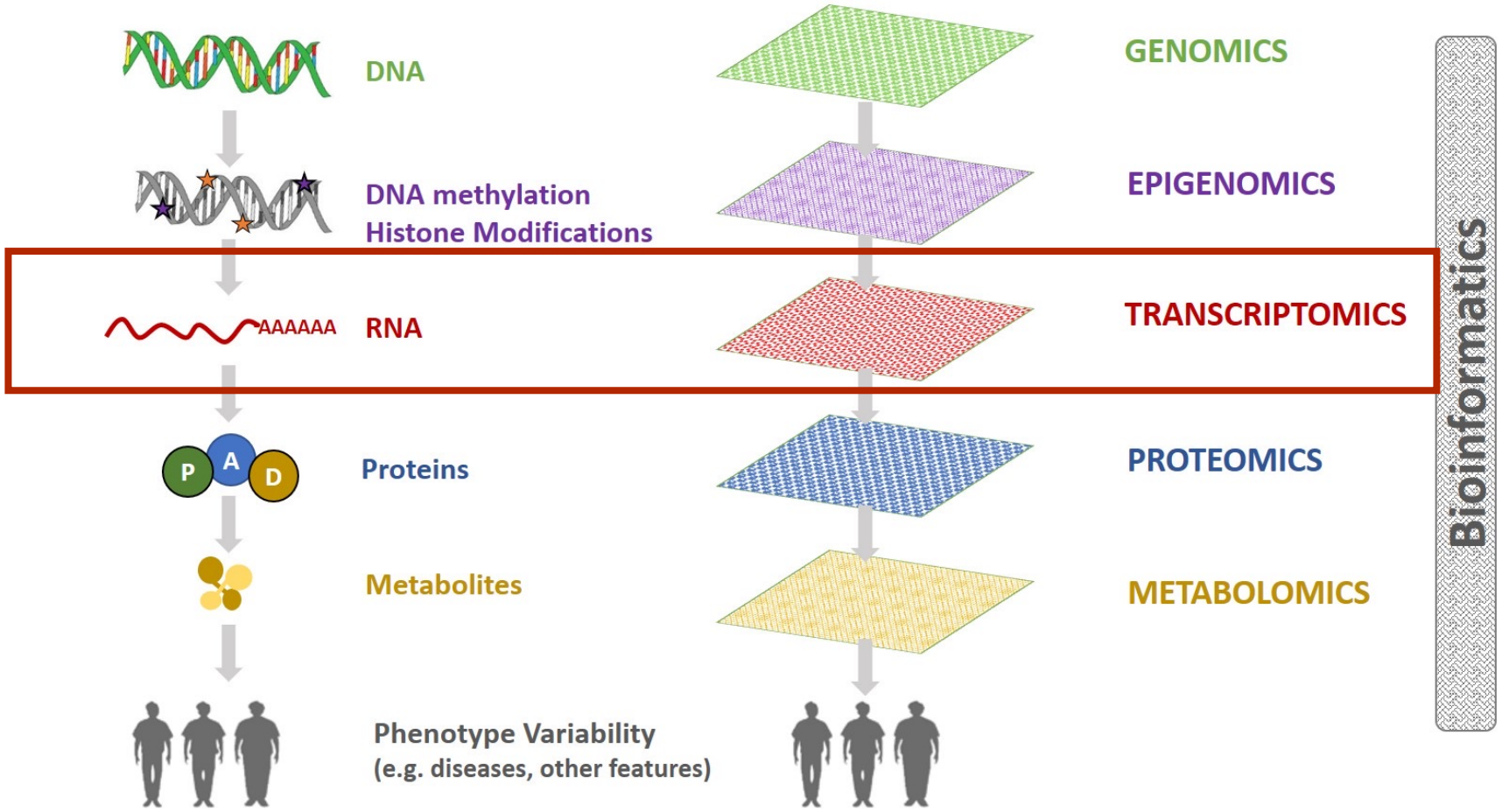
4) Cómo diseñar un experimento?

Integration of Multi-Omics Data



TRANSCRIPTOMA

Integration of Multi-Omics Data



TRANSCRIPTOMA

Transcriptoma = todos los genes expresados del tejido/ de la cédula **en el momento de colectar**

Cuáles son las ventajas de usar transcriptomas en vez que genomas?

1) Qué es el ARN?

2) Qué es el transcriptoma?

3) A cuáles preguntas podemos responder utilizando ARN?

4) Cómo diseñar un experimento?

Gallego *et al. BMC Genomics* (2021) 22:551
https://doi.org/10.1186/s12864-021-07871-0

BMC Genomics

RESEARCH ARTICLE Open Access

Transcriptomic analyses of cacao flavonoids produced in photobioreactors

Adriana M. Gallego[1], Luisa F. Rojas[2], Wilmar G. Valencia[3], Lucía Atehortúa[1], Aura I. Urrea[1], Andrew S. Fister[4,5,6], Mark J. Guiltinan[4,5], Siela N. Maximova[4,5*] and Natalia Pabón-Mora[7*]

Abstract

Background: *Theobroma cacao* is a major source of flavonoids such as catechins and their monomers proanthocyanidins (PAs), widely studied for their potential benefits in cardiovascular diseases. Light has been shown to promote plant secondary metabolite production in vitro. In this study, cacao cells cultured in 7.5 L stirred tank photobioreactors (STPs) were exposed to a change of white to blue LED lights for 28 days (d).

Results: Transcriptomic analyses were performed in three time points comparing changing expression patterns, after cell exposure to white light (d0-VS-d14), after a shift from white to blue light (d14-VS-d15), and after an extended period of blue light for the following 15 days (d15-VS-d28). Under white light, there was enrichment in metabolic pathways associated with cell growth (carbon, glycolysis, and amino acid biosynthesis) accompanied by a significant increase in the PAs content. In the shift to blue light, further increase in PAs content was observed concomitantly with the significant expression of *TWO-COMPONENT RESPONSE REGULATOR* genes involved in the early stress responses via circadian clock and hormone pathways. Under blue light exposure, we observed a depletion of PAs content associated with ROS-mediated stress pathways.

Conclusions: Light effects on large-scale cell cultures in photobioreactors are complex and pleiotropic; however, we have been able to identify key regulatory players upstream cacao flavonoid biosynthesis in STPs, including *TWO-COMPONENT SYSTEM* and ROS-signaling genes. The crosstalk between flavonoid biosynthesis and regulatory networks led to understand the dynamics of flavonoid production and degradation in response to light-driven ROS signals. This can be used to optimize the time, and the yield of in vitro targeted metabolites in large-scale culture systems.

Keywords: Flavonoids, Gene expression, Oxidative stress, Photobioreactors, *Theobroma cacao*

Dr. Natalia Pabón Mora

“Cómo afecta el tipo de luz a la producción de metabolitos secundarios en cacao?”

PREGUNTAS

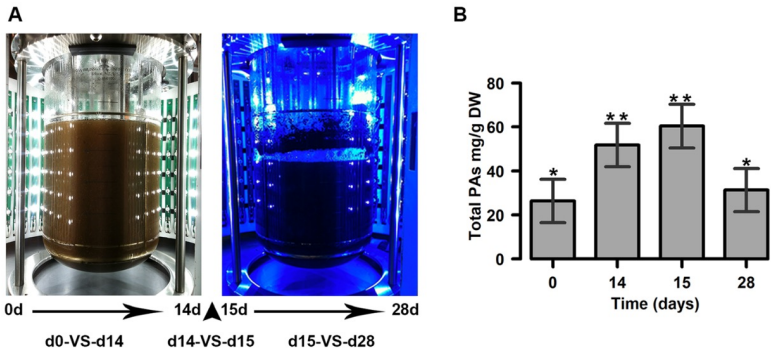


Fig. 1 Experimental design and PAs accumulation in cacao cell cultures grown in stirred tank photobioreactors (STPs). **a** Bioreactors under white (Day 0-14) and blue (Day 15-28) light conditions. Fresh media was added at day 14 (arrowhead). Three comparisons were performed: d0-VS-d14, d14-VS-d15, and d15-VS-d28. **b** Total proanthocyanidins (PAs) content measured in cacao cell cultures grown in STPs. Treatments showing a different number of asterisks are considered significant with *p*-value < 0.05. Bars represent probability intervals. DW: dry weight

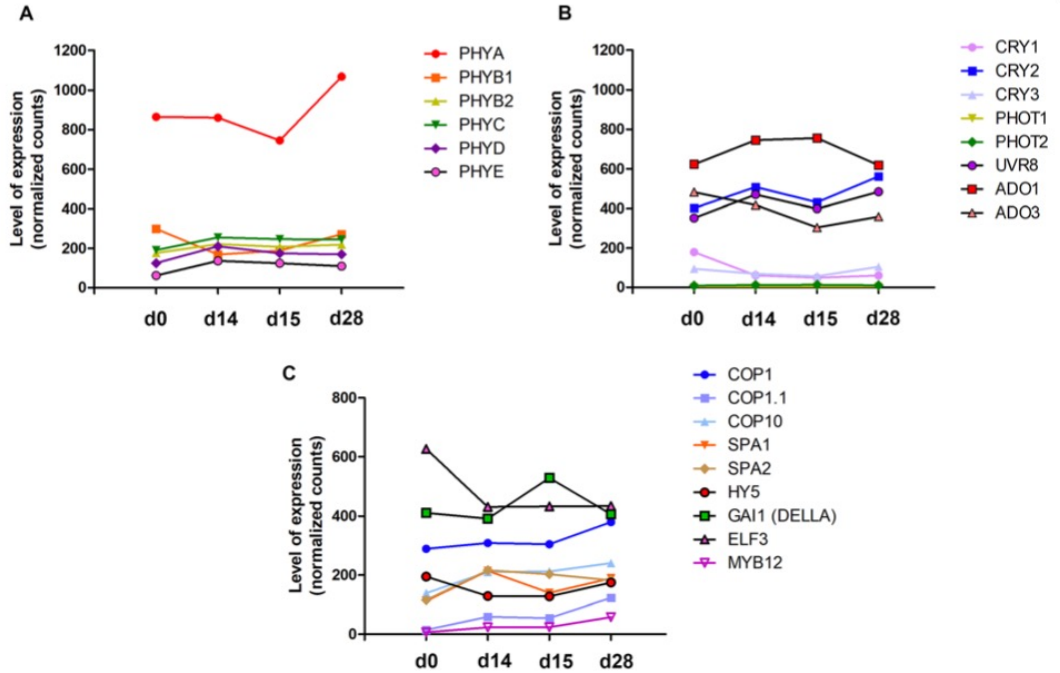


Fig. 3 Gene expression patterns of light-signaling genes identified in cacao cells grown in STPs. **a** Red/Far-red responsive photoreceptors. **b** Blue light and UV responsive photoreceptors. **c** Downstream light signaling genes. The data corresponds to the mean of three biological replicates. d, day. Note that *PHYE* and *COP1.1* were found to be differentially expressed under d0-VS-d14 and d15-VS-d28 respectively

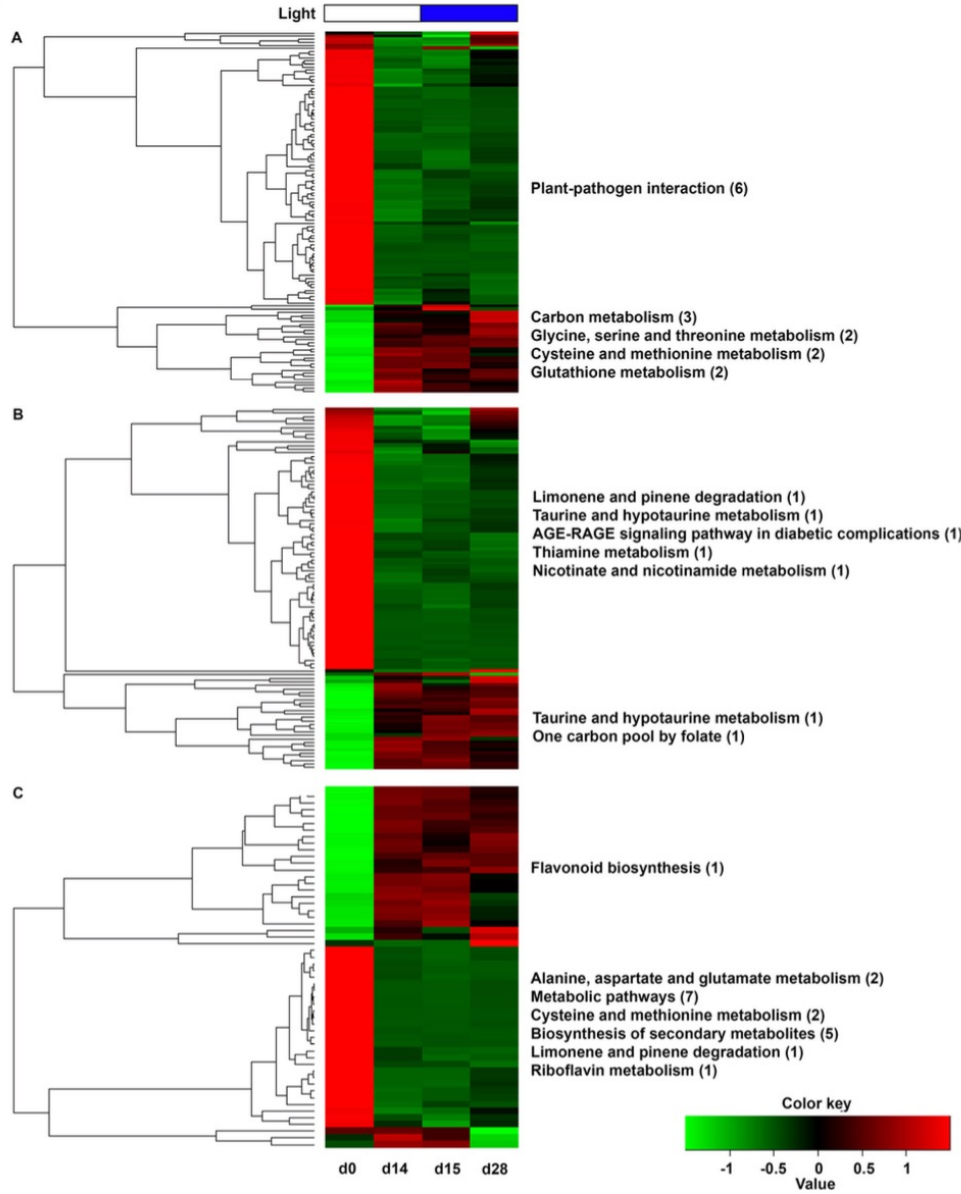


Fig. 4 Cluster analysis showing the dynamics of ROS markers gene expression under white-blue treatments of cacao cells grown in STPs. DEG genes were subjected to metabolic pathway KEGG enrichment analysis for each type of ROS. Only significant enriched categories are shown (*p*-value < 0.05). Upregulated and downregulated genes are colored in red and green respectively. **a** Singlet oxygen ($^1O_2$). **b** Hydrogen peroxide (H_2O_2). **c** Superoxide (O_2^-). d, day

PREGUNTAS

tiempo: d0, d14, d15, d28

luz: blanco (d0 & d14) vs. azul (d15 & d28)

réplicas: 3

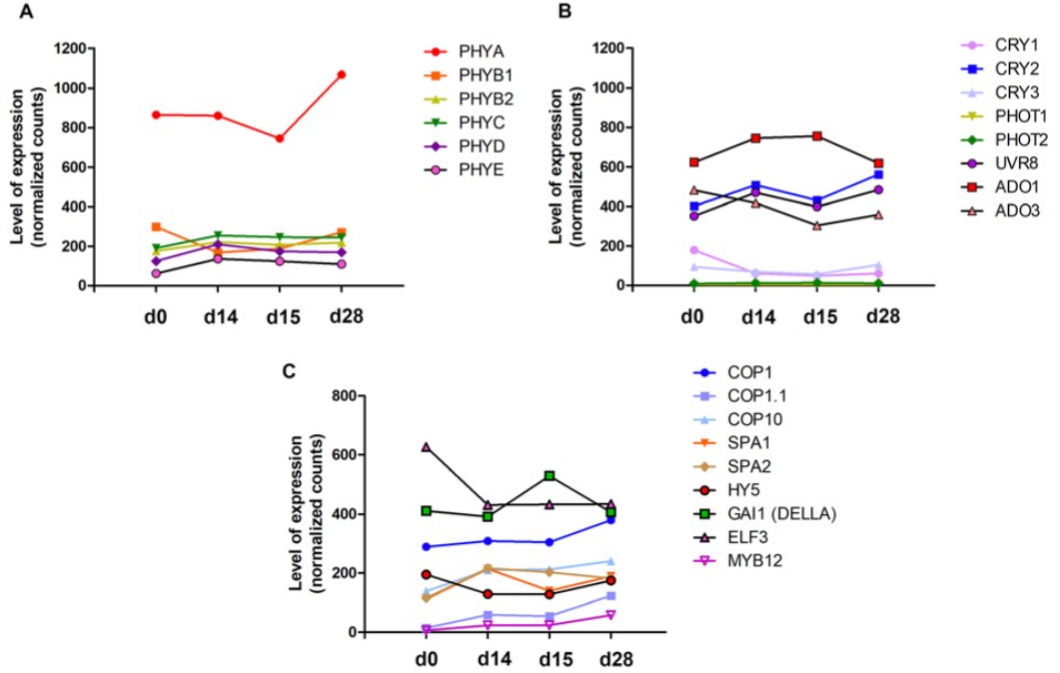


Fig. 3 Gene expression patterns of light-signaling genes identified in cacao cells grown in STPs. **a** Red/Far-red responsive photoreceptors. **b** Blue light and UV responsive photoreceptors. **c** Downstream light signaling genes. The data corresponds to the mean of three biological replicates. d, day. Note that *PHYE* and *COP1.1* were found to be differentially expressed under d0-VS-d14 and d15-VS-d28 respectively

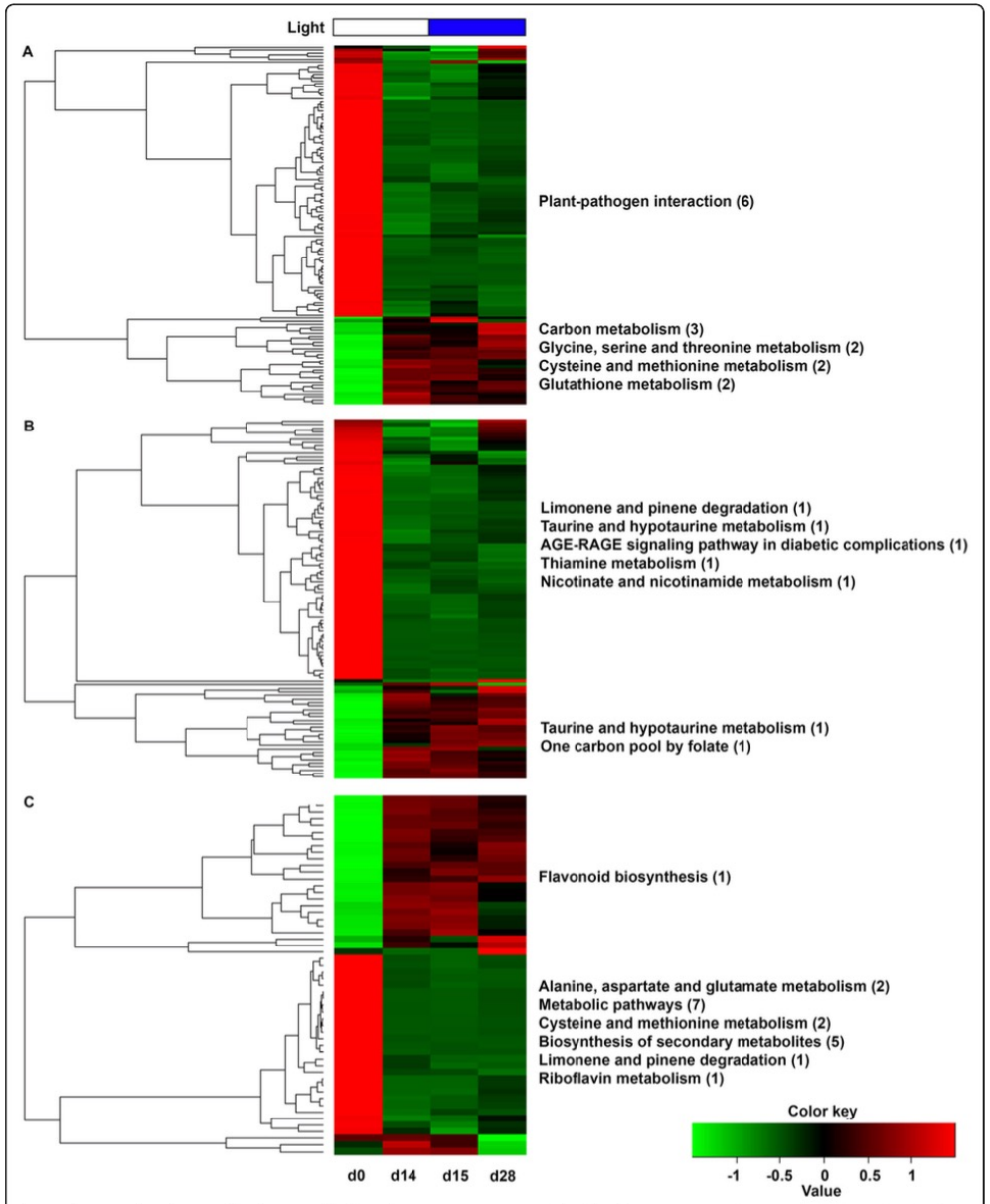


Fig. 4 Cluster analysis showing the dynamics of ROS markers gene expression under white-blue treatments of cacao cells grown in STPs. DEG genes were subjected to metabolic pathway KEGG enrichment analysis for each type of ROS. Only significant enriched categories are shown (p-value < 0.05). Upregulated and downregulated genes are colored in red and green respectively. **a** Singlet oxygen ($^1O_2$). **b** Hydrogen peroxide (H_2O_2). **c** Superoxide (O_2^-). d, day

PREGUNTAS

SOCIETY FOR EXPERIMENTAL BIOLOGY

Journal of Experimental Botany, Vol. 72, No. 11 pp. 4161–4179, 2021
doi:10.1093/jxb/erab051 Advance Access Publication 17 February 2021

Journal of Experimental Botany

RESEARCH PAPER

Natural ultraviolet radiation exposure alters photosynthetic biology and improves recovery from desiccation in a desert moss

Jenna T.B. Ekwealor[1,*], Theresa A. Clark[2], Oliver Dautermann[3], Alexander Russell[2], Sotodeh Ebrahimi[2], Lloyd R. Stark[2], Krishna K. Niyogi[3,4,5] and Brent D. Mishler[1]

[1] Department of Integrative Biology, and University and Jepson Herbaria, University of California, Berkeley, CA, USA
[2] School of Life Sciences, University of Nevada, Las Vegas, NV, USA
[3] Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA
[4] Howard Hughes Medical Institute, University of California, Berkeley, CA, USA
[5] Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

* Correspondence: jtbe@berkeley.edu

Received 1 December 2020; Editorial decision 26 January 2021; Accepted 1 February 2021

Editor: Karl-Josef Dietz, Bielefeld University, Germany

Abstract

Plants in dryland ecosystems experience extreme daily and seasonal fluctuations in light, temperature, and water availability. We used an *in situ* field experiment to uncover the effects of natural and reduced levels of ultraviolet radiation (UV) on maximum PSII quantum efficiency (F_v/F_m), relative abundance of photosynthetic pigments and antioxidants, and the transcriptome in the desiccation-tolerant desert moss *Syntrichia caninervis*. We tested the hypotheses that: (i) *S. caninervis* plants undergo sustained thermal quenching of light [non-photochemical quenching (NPQ)] while desiccated and after rehydration; (ii) a reduction of UV will result in improved recovery of F_v/F_m; but (iii) 1 year of UV removal will de-harden plants and increase vulnerability to UV damage, indicated by a reduction in F_v/F_m. All field-collected plants had extremely low F_v/F_m after initial rehydration but recovered over 8 d in lab-simulated winter conditions. UV-filtered plants had lower F_v/F_m during recovery, higher concentrations of photoprotective pigments and antioxidants such as zeaxanthin and tocopherols, and lower concentrations of neoxanthin and Chl *b* than plants exposed to near natural UV levels. Field-grown *S. caninervis* underwent sustained NPQ that took days to relax and for efficient photosynthesis to resume. Reduction of solar UV radiation adversely affected recovery of F_v/F_m following rehydration.

Keywords: Biological soil crust, desiccation tolerance, maximum PSII quantum efficiency (F_v/F_m), Mojave Desert, moss, non-photochemical quenching (NPQ), photosynthetic efficiency, photosynthetic pigments, *Syntrichia caninervis*, UV tolerance.

Dr. Jenna Ekwealor

“Cuál es el efecto de la luz ultravioleta en musgos de desierto?”

PREGUNTAS

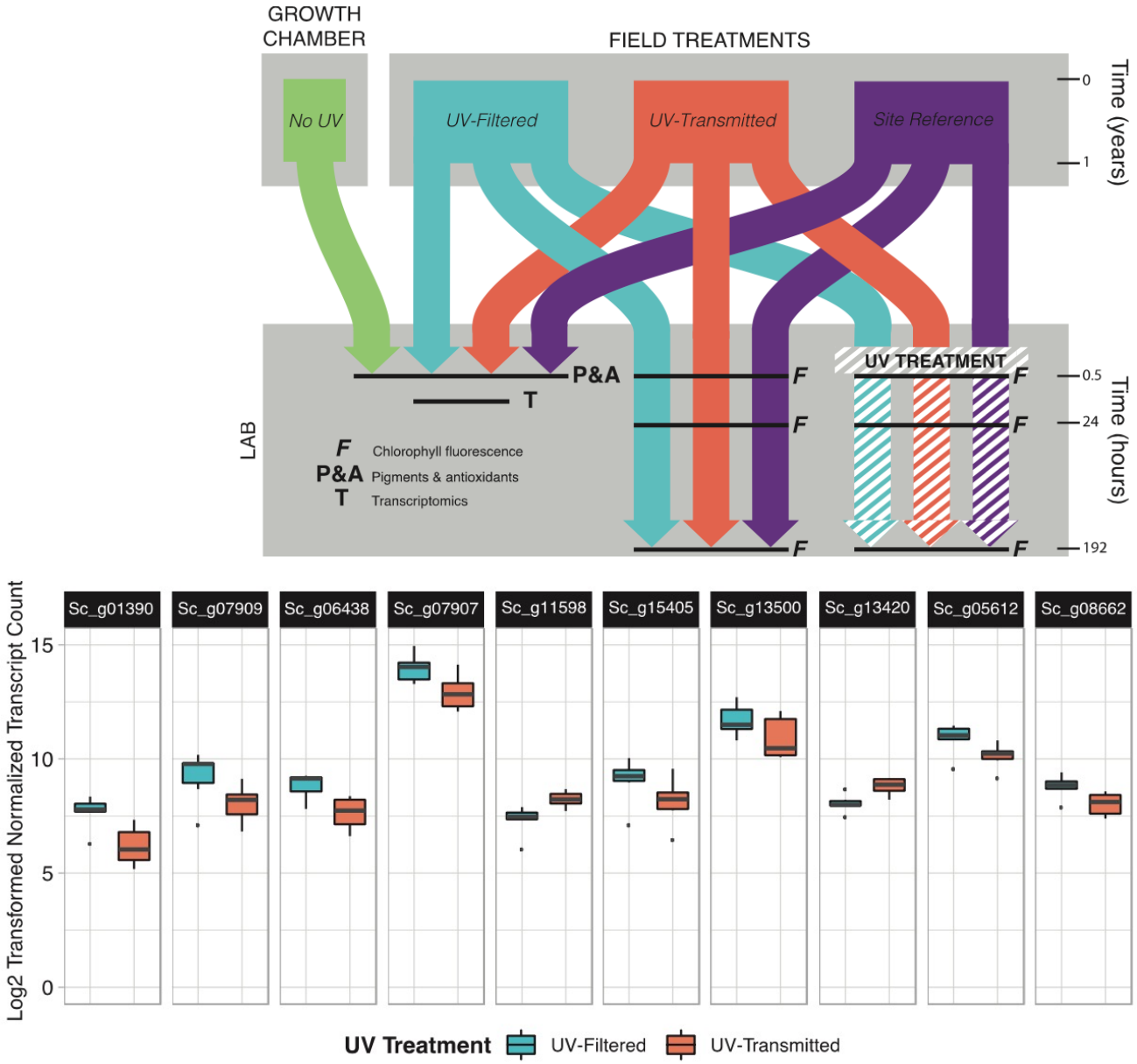
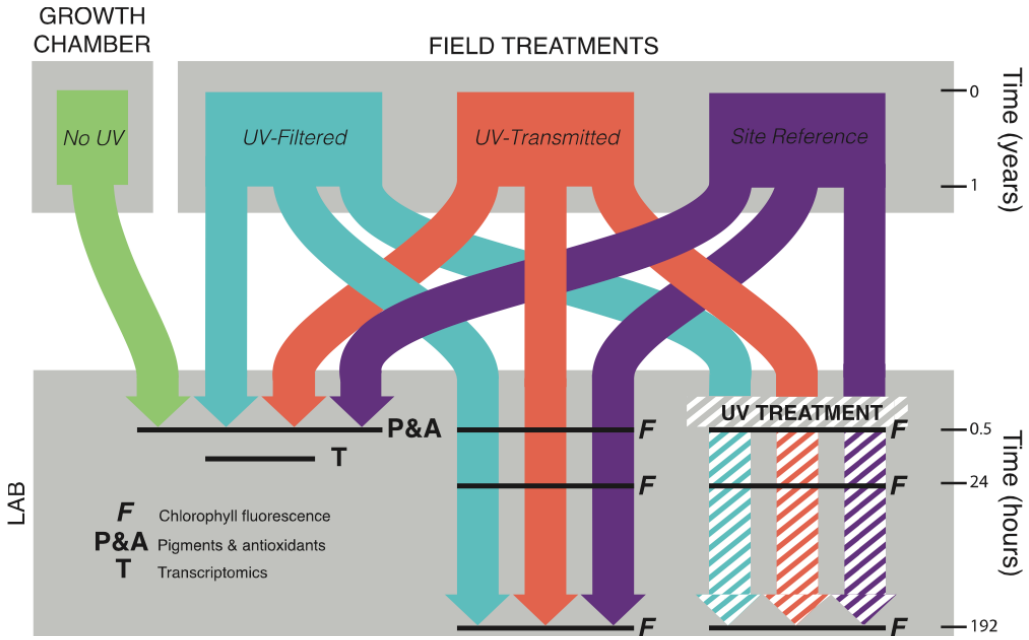


Fig. 8. Normalized transcript counts of the top 10 most differentially abundant ($P<0.005$) transcripts with UV filtering in field-treated *Syntrichia caninervis*.

PREGUNTAS

UV vs. no UV
6 réplicas



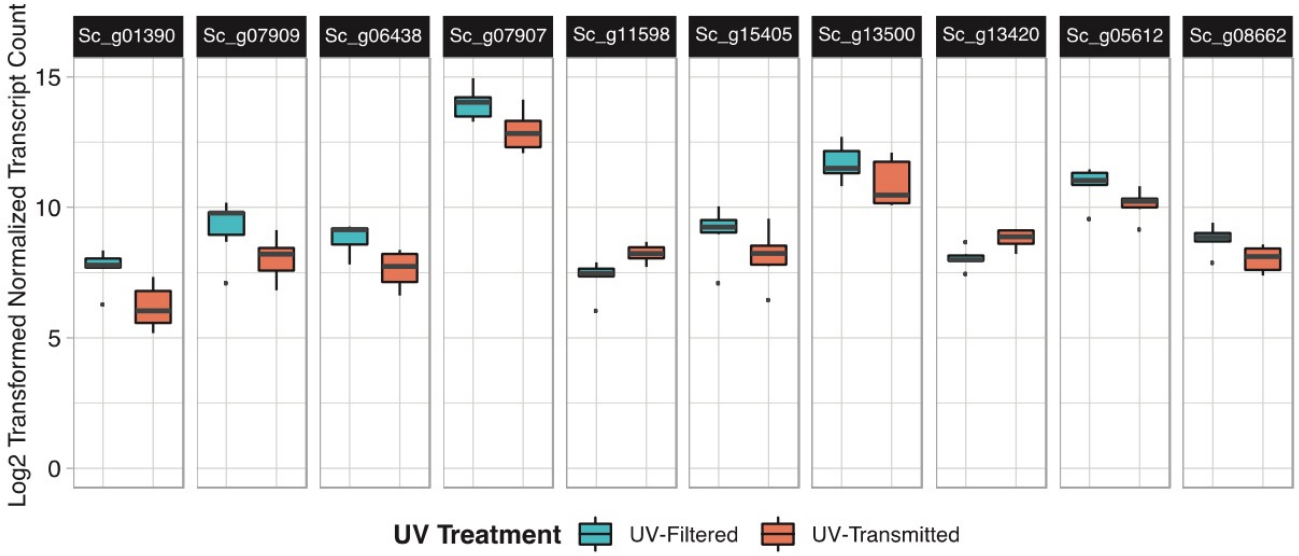


Fig. 8. Normalized transcript counts of the top 10 most differentially abundant ($P<0.005$) transcripts with UV filtering in field-treated *Syntrichia caninervis*.

PREGUNTAS

PNAS RESEARCH ARTICLE | EVOLUTION

Layered evolution of gene expression in “superfast” muscles for courtship

James B. Pease[a,1,2], Robert J. Driver[b,1], David A. de la Cerda[a,c], Lainy B. Day[d], Willow R. Lindsay[d,e], Barney A. Schlinger[f,g,h], Eric R. Schuppe[a], Christopher N. Balakrishnan[b], and Matthew J. Fuxjager[i,2]

Edited by Gene Robinson, University of Illinois at Urbana–Champaign Institute for Genomic Biology, Urbana, IL; received November 4, 2021; accepted February 11, 2022

Identifying the molecular process of complex trait evolution is a core goal of biology. However, pinpointing the specific context and timing of trait-associated changes within the molecular evolutionary history of an organism remains an elusive goal. We study this topic by exploring the molecular basis of elaborate courtship evolution, which represents an extraordinary example of trait innovation. Within the behaviorally diverse radiation of Central and South American manakin birds, species from two separate lineages beat their wings together using specialized “superfast” muscles to generate a “snap” that helps attract mates. Here, we develop an empirical approach to analyze phylogenetic lineage-specific shifts in gene expression in the key snap-performing muscle and then integrate these findings with comparative transcriptomic sequence analysis. We find that rapid wing displays are associated with changes to a wide range of molecular processes that underlie extreme muscle performance, including changes to calcium trafficking, myocyte homeostasis and metabolism, and hormone action. We furthermore show that these changes occur gradually in a layered manner across the species history, wherein which ancestral genetic changes to many of these molecular systems are built upon by later species-specific shifts that likely finalized the process of display performance adaptation. Our study demonstrates the potential for combining phylogenetic modeling of tissue-specific gene expression shifts with phylogenetic analysis of lineage-specific sequence changes to reveal holistic evolutionary histories of complex traits.

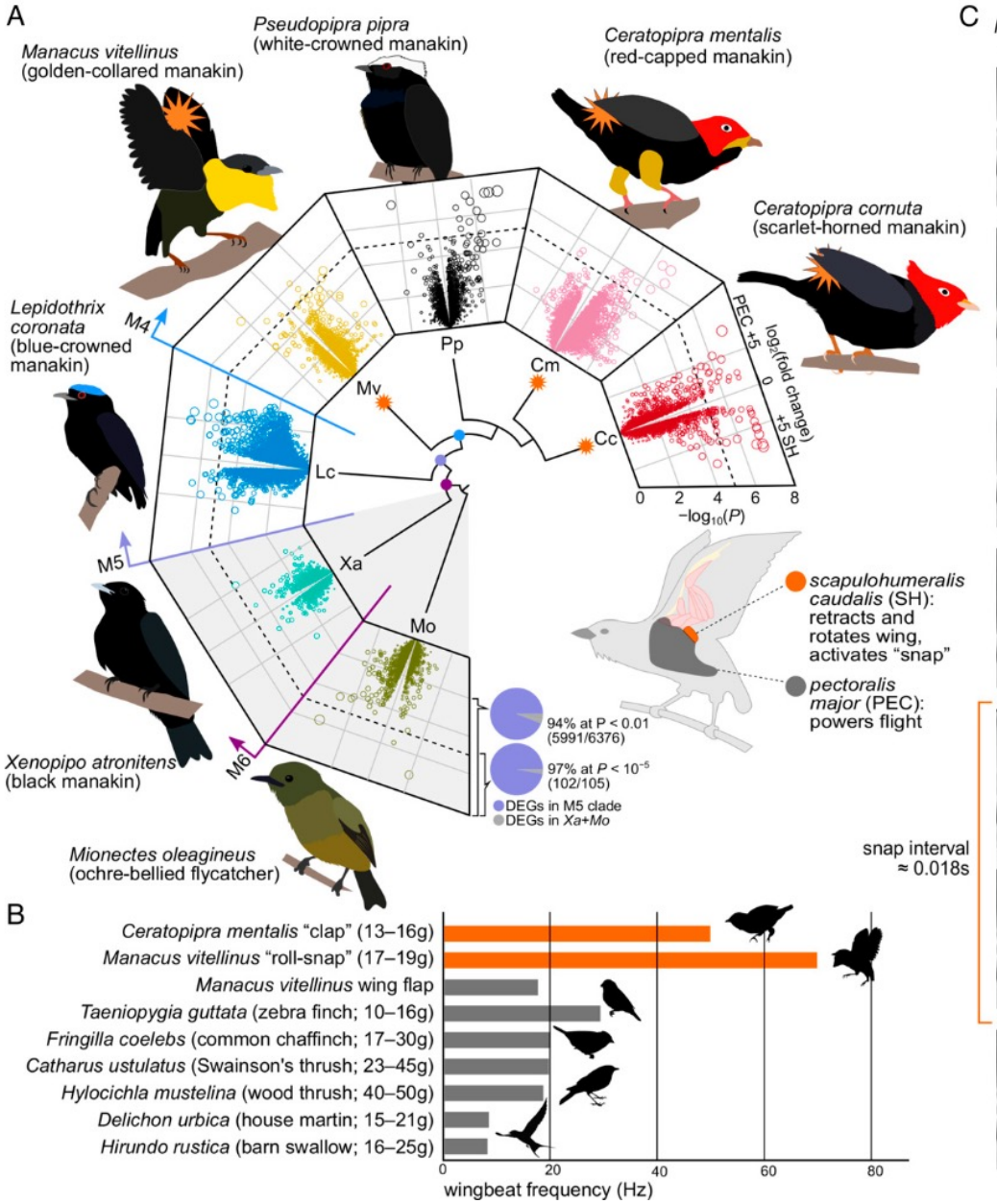
Significance

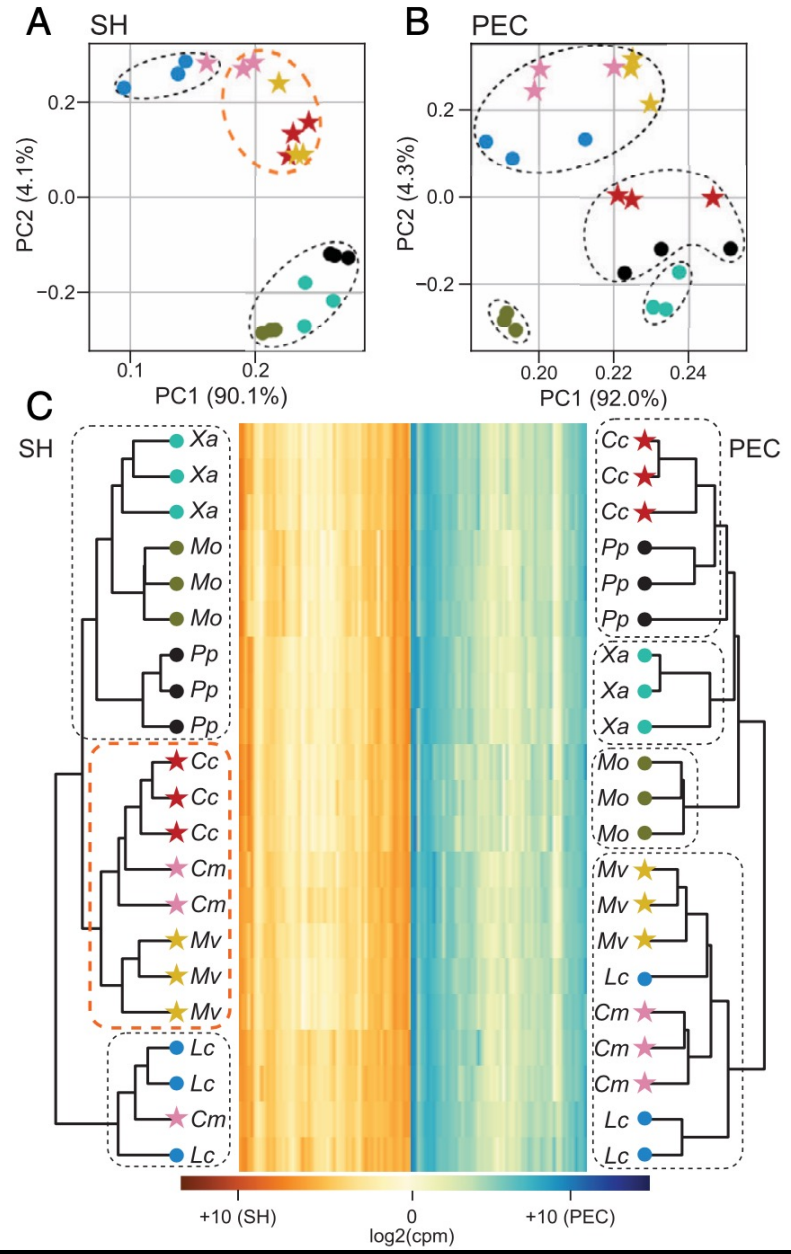
The process by which new complex traits evolve has been a persistent conundrum throughout the history of evolutionary inquiry. How multiple physiological changes at the organism level and genetic changes at the molecular level combine is still unclear for many traits. Here, we studied the displays of manakins, who beat their wings together at nearly twice the speed of other songbirds to produce a loud “snap” that attracts mates. We simultaneously analyzed evolution of gene

Dr. James Pease

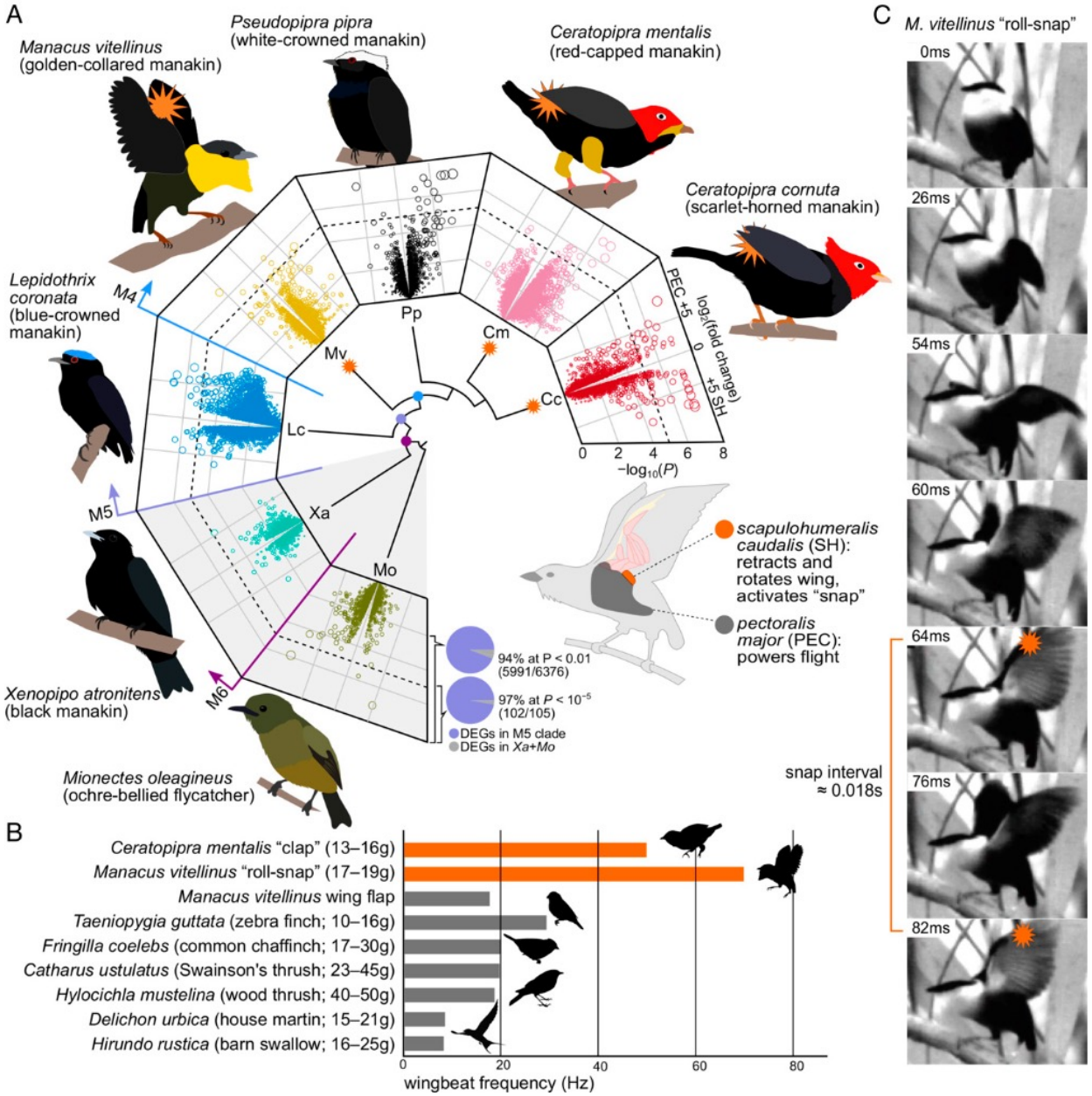
“Cuales son los genes involucrados en los músculos que producen movimientos rápidos en las alas de aves?”

PREGUNTAS

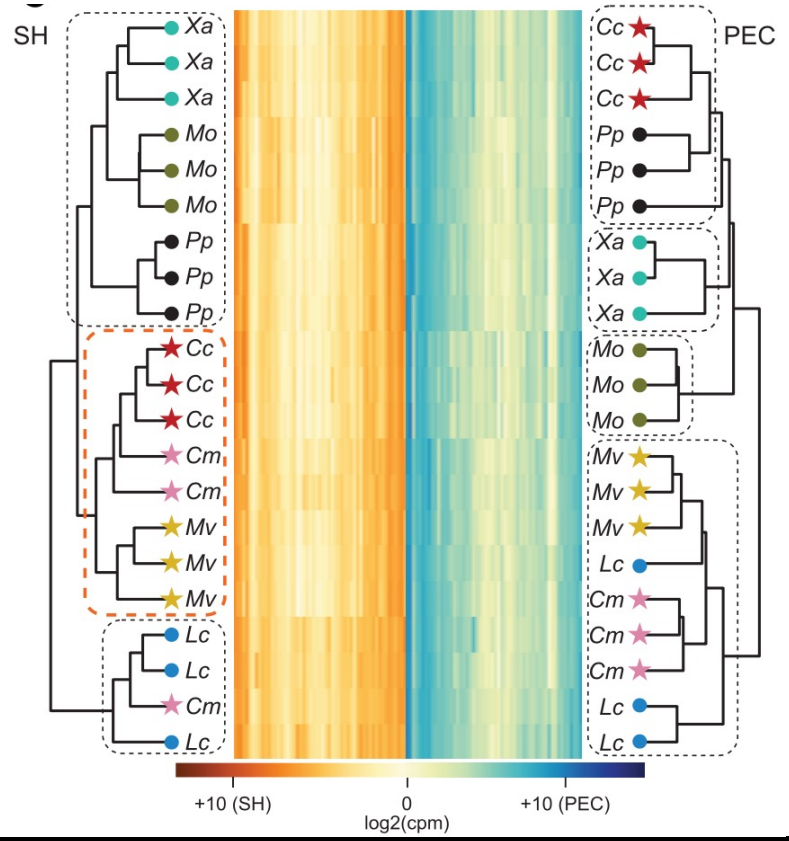

A
Manacus vitellinus
(golden-collared manakin)
Pseudopipra pipra
(white-crowned manakin)
Ceratopipra mentalis
(red-capped manakin)
Ceratopipra cornuta
(scarlet-horned manakin)
Lepidothrix coronata
(blue-crowned manakin)
Xenopipo atronitens
(black manakin)
Mionectes oleagineus
(ochre-bellied flycatcher)
M4
M5
M6
Pp
Cm
Mv
Cc
Lc
Xa
Mo
PEC +5
log2(fold change)
0
+5 SH
0 2 4 6 8
−log10(P)
scapulohumeralis caudalis (SH): retracts and rotates wing, activates "snap"
pectoralis major (PEC): powers flight
94% at P < 0.01 (5991/6376)
97% at P < 10−5 (102/105)
DEGs in M5 clade
DEGs in Xa+Mo
B
Ceratopipra mentalis "clap" (13–16g)
Manacus vitellinus "roll-snap" (17–19g)
Manacus vitellinus wing flap
Taeniopygia guttata (zebra finch; 10–16g)
Fringilla coelebs (common chaffinch; 17–30g)
Catharus ustulatus (Swainson's thrush; 23–45g)
Hylocichla mustelina (wood thrush; 40–50g)
Delichon urbica (house martin; 15–21g)
Hirundo rustica (barn swallow; 16–25g)
0 20 40 60 80
wingbeat frequency (Hz)
C
M. vitellinus "roll-snap"
0ms
26ms
54ms
60ms
64ms
76ms
82ms
snap interval ≈ 0.018s


A
SH
PC2 (4.1%)
0.2
0.0
−0.2
0.1
0.2
PC1 (90.1%)
B
PEC
PC2 (4.3%)
0.2
0.0
−0.2
0.20
0.22
0.24
PC1 (92.0%)
C
SH
PEC
Xa
Xa
Xa
Mo
Mo
Mo
Pp
Pp
Pp
Cc
Cc
Cc
Cm
Cm
Mv
Mv
Mv
Lc
Lc
Cm
Lc
Cc
Cc
Cc
Pp
Pp
Pp
Xa
Xa
Xa
Mo
Mo
Mo
Mv
Mv
Mv
Lc
Cm
Cm
Cm
Lc
Lc
+10 (SH)
0
+10 (PEC)
log2(cpm)

PREGUNTAS



6 especies
3 réplicas



PREGUNTAS

Annals of Botany **130**: 671–685, 2022
https://doi.org/10.1093/aob/mcac117, available online at www.academic.oup.com/aob

ANNALS OF BOTANY Founded 1887

Phylotranscriptomics reveal the spatio-temporal distribution and morphological evolution of *Macrozamia*, an Australian endemic genus of Cycadales

Sadaf Habib[1,2], Yiqing Gong[1], Shanshan Dong[1], Anders Lindstrom[3], Dennis William Stevenson[4], Yang Liu[1], Hong Wu[2] and Shouzhou Zhang[1,*]

[1]Key Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen, 518004, China, [2]College of Life Sciences, South China Agricultural University, Guangzhou, 510642, China, [3]Global Biodiversity Conservancy 144/124 Moo 3, Soi Bua Thong, Bangsalae, Sattahip, Chonburi 20250, Thailand, and [4]New York Botanical Garden, Bronx, NY 10458, USA
** For correspondence. E-mail shouzhouz@szbg.ac.cn*

Received: 12 May 2022 Returned for revision: 4 September 2022 Editorial decision: 8 September 2022 Accepted: 14 September 2022
Electronically published: 16 September 2022

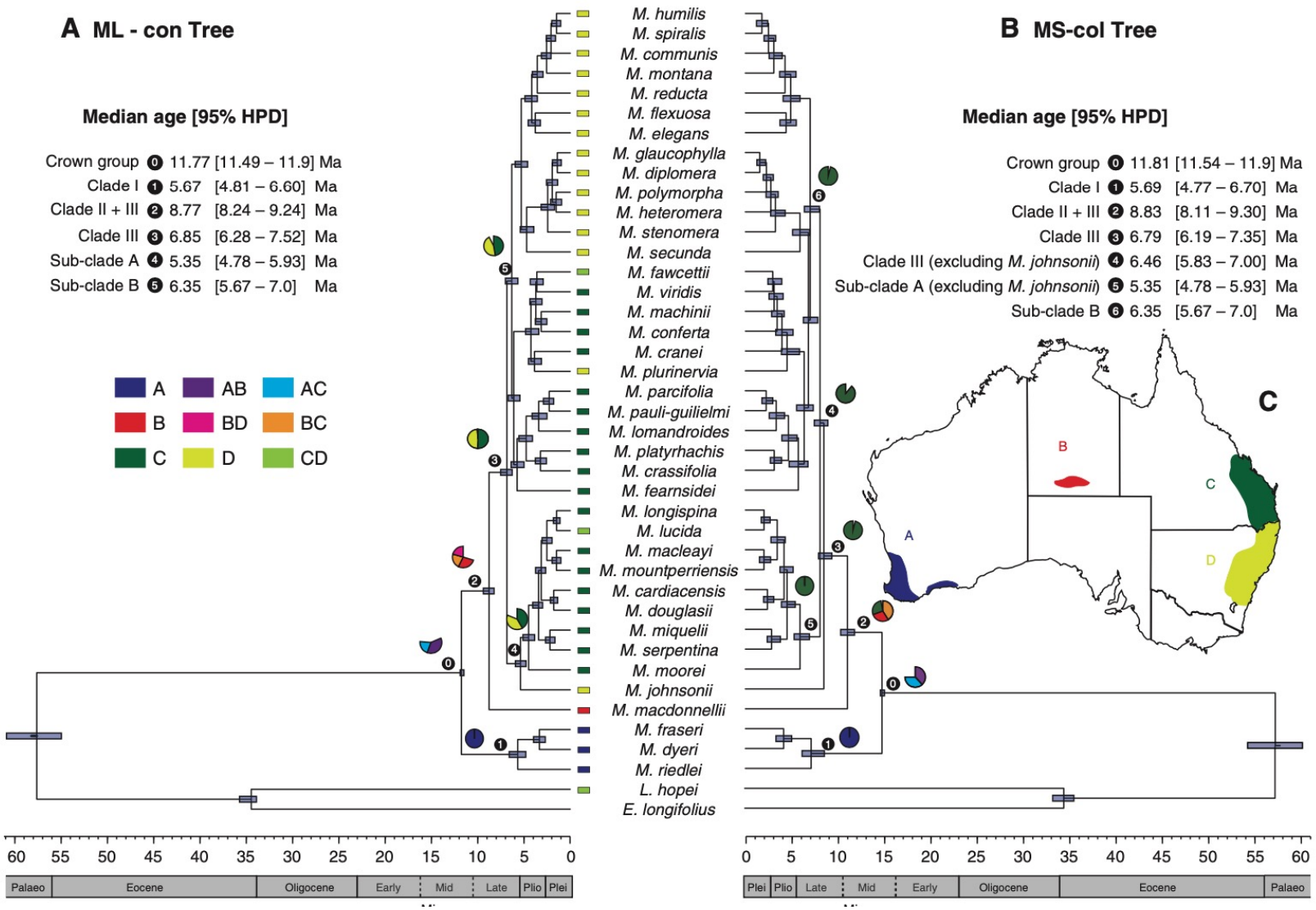
• **Background and Aims** Cycads are regarded as an ancient lineage of living seed plants, and hold important clues to understand the early evolutionary trends of seed plants. The molecular phylogeny and spatio-temporal diversification of one of the species-rich genera of cycads, *Macrozamia*, have not been well reconstructed.
• **Methods** We analysed a transcriptome dataset of 4740 single-copy nuclear genes (SCGs) of 39 *Macrozamia* species and two outgroup taxa. Based on concatenated (maximum parsimony, maximum likelihood) and multispecies coalescent analyses, we first establish a well-resolved phylogenetic tree of *Macrozamia*. To identify cyto-nuclear incongruence, the plastid protein coding genes (PCGs) from transcriptome data are extracted using the software HybPiper. Furthermore, we explore the biogeographical history of the genus and shed light on the pattern of floristic exchange between three distinct areas of Australia. Six key diagnostic characters are traced on the phylogenetic framework using two comparative methods, and infra-generic classification is investigated.
• **Key Results** The tree topologies of concatenated and multi-species coalescent analyses of SCGs are mostly congruent with a few conflicting nodes, while those from plastid PCGs show poorly supported relationships. The genus contains three major clades that correspond to their distinct distributional areas in Australia. The crown group of *Macrozamia* is estimated to around 11.80 Ma, with a major expansion in the last 5–6 Myr. Six morphological characters show homoplasy, and the traditional phenetic sectional division of the genus is inconsistent with this current phylogeny.
• **Conclusions** This first detailed phylogenetic investigation of *Macrozamia* demonstrates promising prospects of SCGs in resolving phylogenetic relationships within cycads. Our study suggests that *Macrozamia*, once widely distributed in Australia, underwent major extinctions because of fluctuating climatic conditions such as cooling and mesic biome disappearance in the past. The current close placement of morphologically distinct species in the phylogenetic tree may be related to neotenic events that occurred in the genus.

Key words: Cycadales, *Macrozamia*, biogeography, phylotranscriptomics, character evolution, gymnosperms.

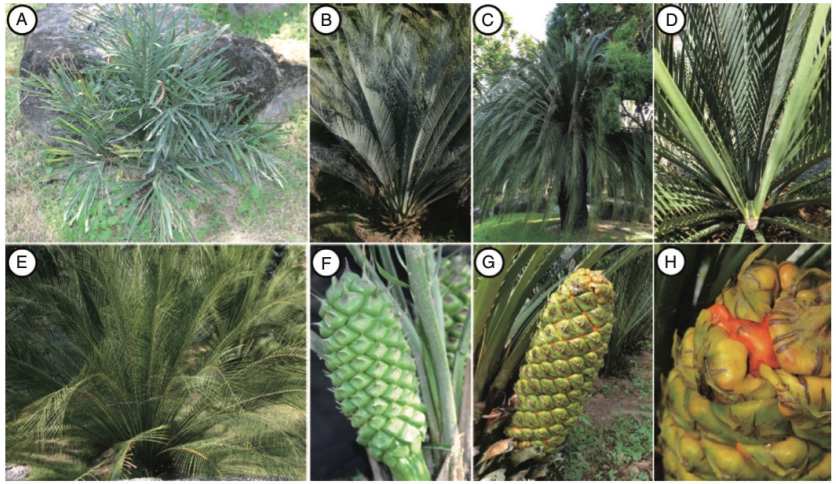
Dr. Sadaf Habib

“Cuales son las relaciones evolutivas entre especies de Macrozamia?”

PREGUNTAS



41 especies
1 réplica



PREGUNTAS

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-022-31003-1 OPEN

Check for updates

A single-cell transcriptomic atlas characterizes the silk-producing organ in the silkworm

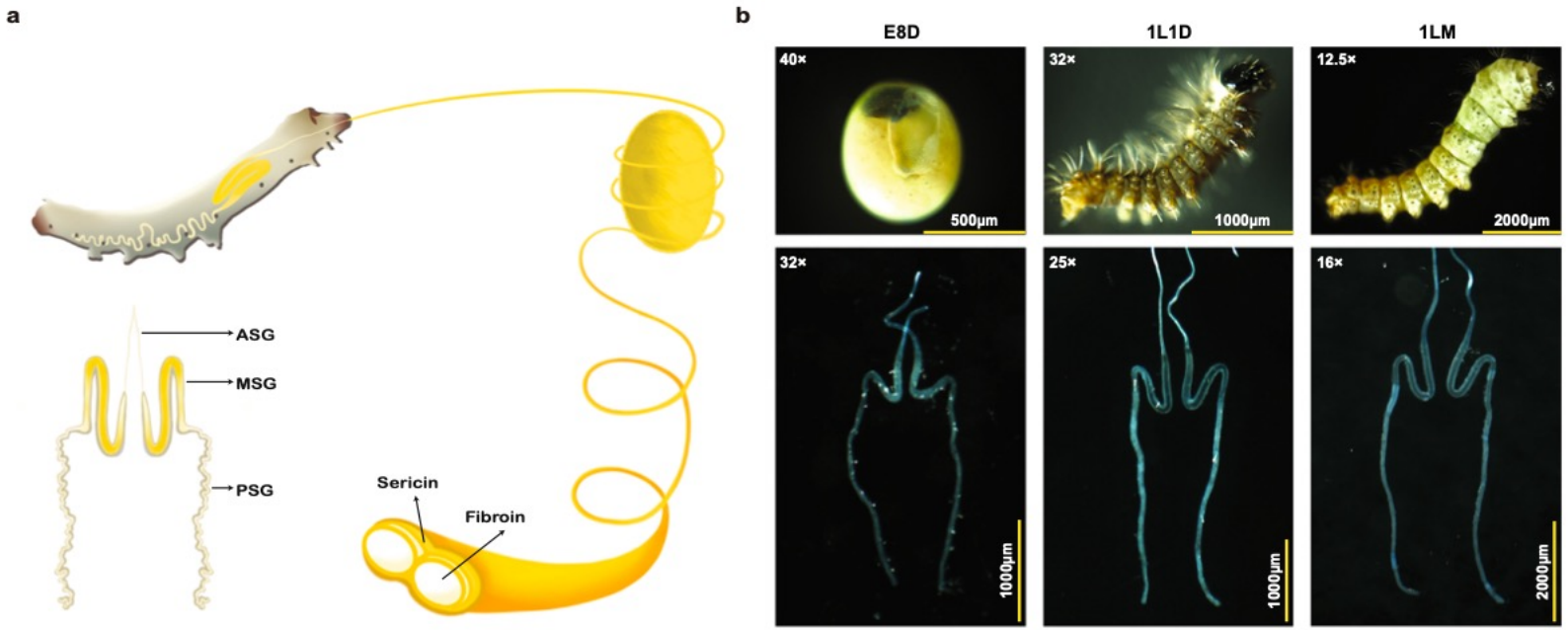
Yan Ma[1,3], Wenhui Zeng[1,3], Yongbing Ba[2,3], Qin Luo[1], Yao Ou[1], Rongpeng Liu[1], Jingwen Ma[1], Yiyun Tang[1], Jie Hu[1], Haomiao Wang[1], Xuan Tang[2], Yuanyuan Mu[1], Qingjun Li[1], Yuqin Chen[1], Yiting Ran[1], Zhonghuai Xiang[1] & Hanfu Xu[1]✉

The silk gland of the domesticated silkworm *Bombyx mori*, is a remarkable organ that produces vast amounts of silk with exceptional properties. Little is known about which silk gland cells execute silk protein synthesis and its precise spatiotemporal control. Here, we use single-cell RNA sequencing to build a comprehensive cell atlas of the silkworm silk gland, consisting of 14,972 high-quality cells representing 10 distinct cell types, in three early developmental stages. We annotate all 10 cell types and determine their distributions in each region of the silk gland. Additionally, we decode the developmental trajectory and gene expression status of silk gland cells. Finally, we discover marker genes involved in the regulation of silk gland development and silk protein synthesis. Altogether, this work reveals the heterogeneity of silkworm silk gland cells and their gene expression dynamics, affording a deeper understanding of silk-producing organs at the single-cell level.

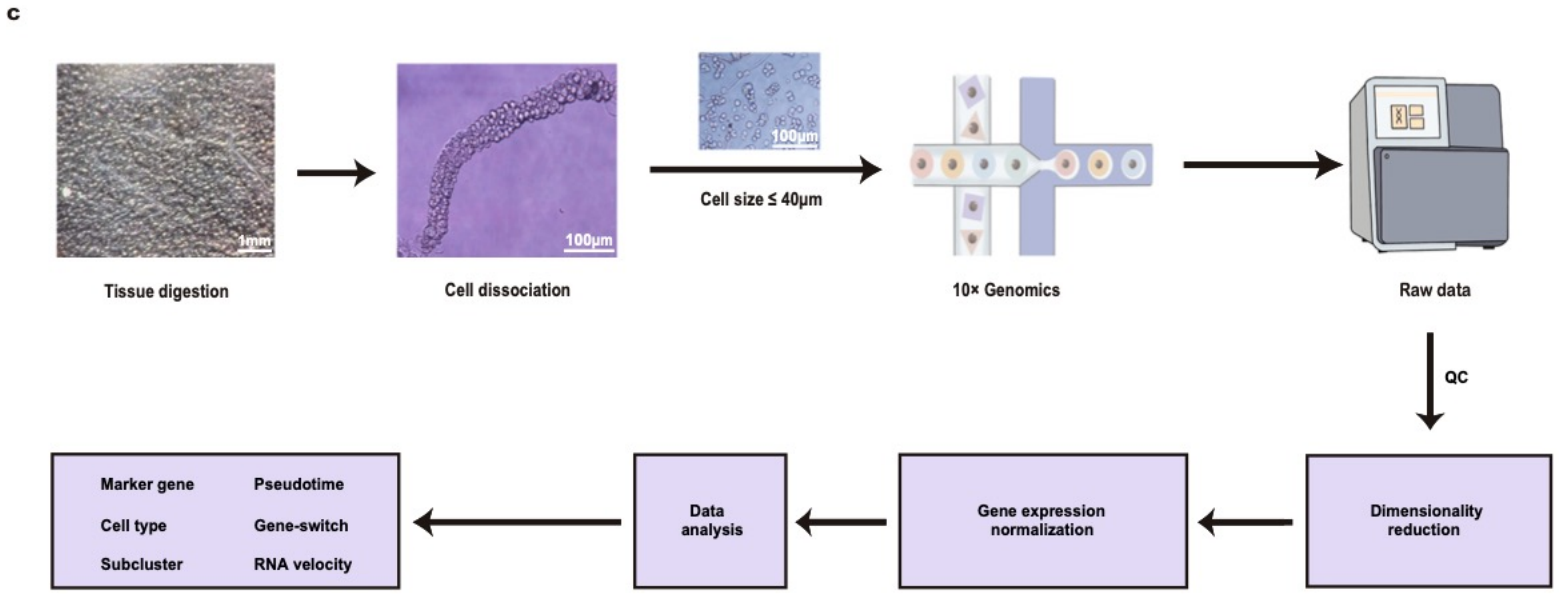
Dr. Hanfu Xu

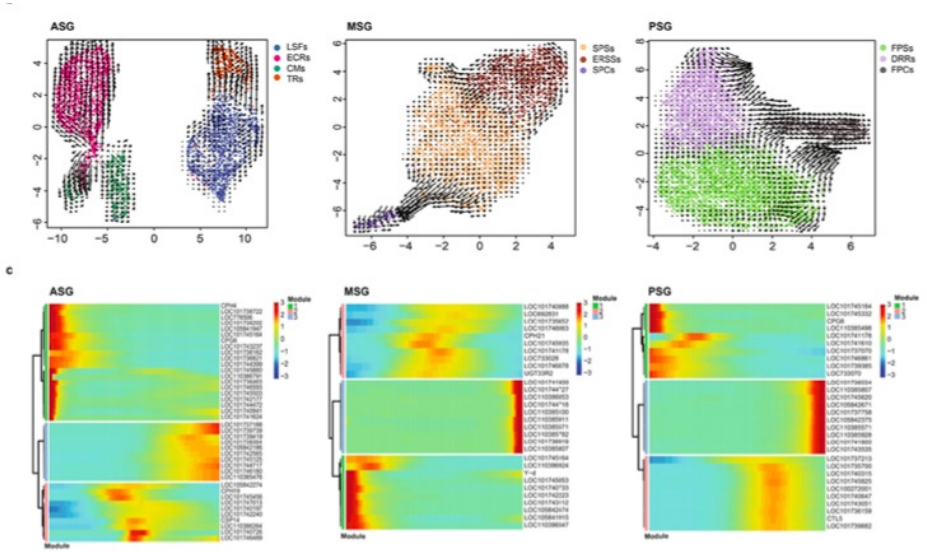
“Cuales son las características moleculares de diferentes tipos de cédulas en el órgano que produce seda (de los gusanos de seda) a largo del desarollo?”

PREGUNTAS



14972 cédulas en total
10 tipos de cédulas
3 etapas de desarollo





PREGUNTAS

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
PLANT MORPHOGENESIS
BODY PATTERNING
NATURAL VARIATION
PLANT EVOLUTION

Received
11 June 2013

Accepted
29 July 2014

Published
29 August 2014

Correspondence and requests for materials should be addressed to A.M.R.A. (ana_almeida@berkeley.edu; anagaivota@gmail.com)

Co-option of the polarity gene network shapes filament morphology in angiosperms

Ana Maria Rocha de Almeida[1], Roxana Yockteng[1,2], James Schnable[1], Elena R. Alvarez-Buylla[1,3], Michael Freeling[1] & Chelsea D. Specht[1,4]

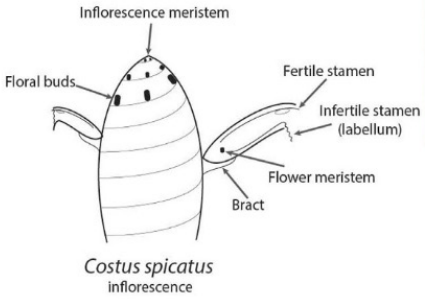
[1]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720 USA, [2]Institut de Systématique, Evolution et Biodiversité (UMR 7205 CNRS, Muséum National d'Histoire Naturelle, CP39, 16 rue Buffon, 75231 Paris Cedex 05, France, [3]Laboratorio de Genética, Epigenética, Desarrollo y Evolución de Plantas, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad Universitaria, 3er Circuito Exterior Junto a Jardín Botánico, Coyoacán, México DF 04510, [4]Department of Integrative Biology and The University and Jepson Herbaria, University of California, Berkeley, CA 94720 USA.

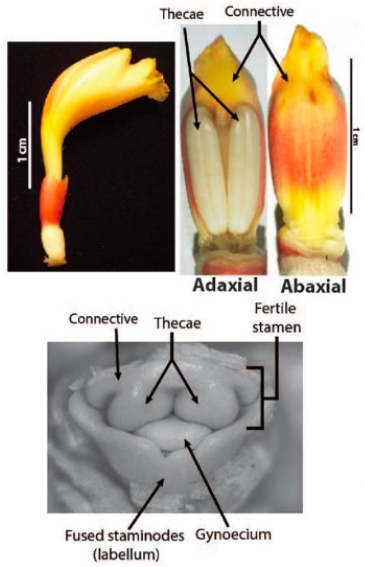
The molecular genetic mechanisms underlying abaxial-adaxial polarity in plants have been studied as a property of lateral and flattened organs, such as leaves. In leaves, laminar expansion occurs as a result of balanced abaxial-adaxial gene expression. Over- or under- expression of either abaxializing or adaxializing genes inhibits laminar growth, resulting in a mutant radialized phenotype. Here, we show that co-option of the abaxial-adaxial polarity gene network plays a role in the evolution of stamen filament morphology in angiosperms. RNA-Seq data from species bearing laminar (flattened) or radial (cylindrical) filaments demonstrates that species with laminar filaments exhibit balanced expression of abaxial-adaxial (ab-ad) genes, while overexpression of a *YABBY* gene is found in species with radial filaments. This result suggests that unbalanced expression of ab-ad genes results in inhibition of laminar outgrowth, leading to a radially symmetric structure as found in many angiosperm filaments. We anticipate that co-option of the polarity gene network is a fundamental mechanism shaping many aspects of plant morphology during angiosperm evolution.

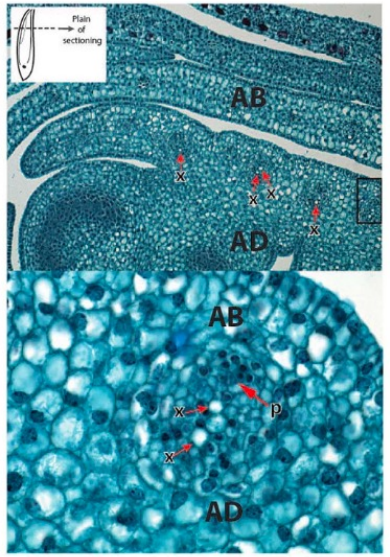
Dr. Ana Maria Almeida

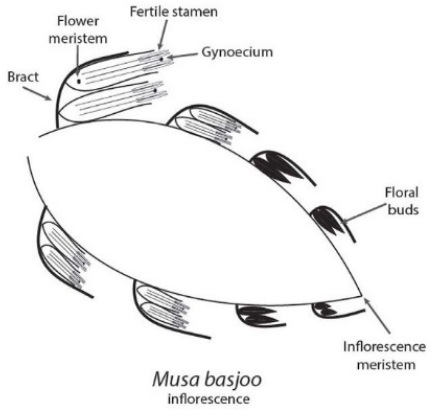
"Cuales son los genes responsables por la morfología de filamentos de flores?"

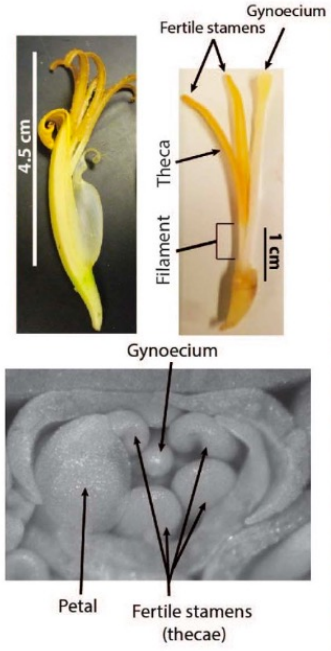
PREGUNTAS

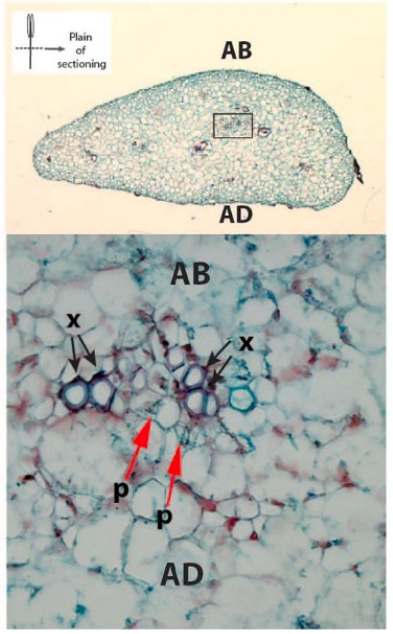

Inflorescence meristem
Floral buds
Fertile stamen
Infertile stamen (labellum)
Flower meristem
Bract
Costus spicatus
inflorescence

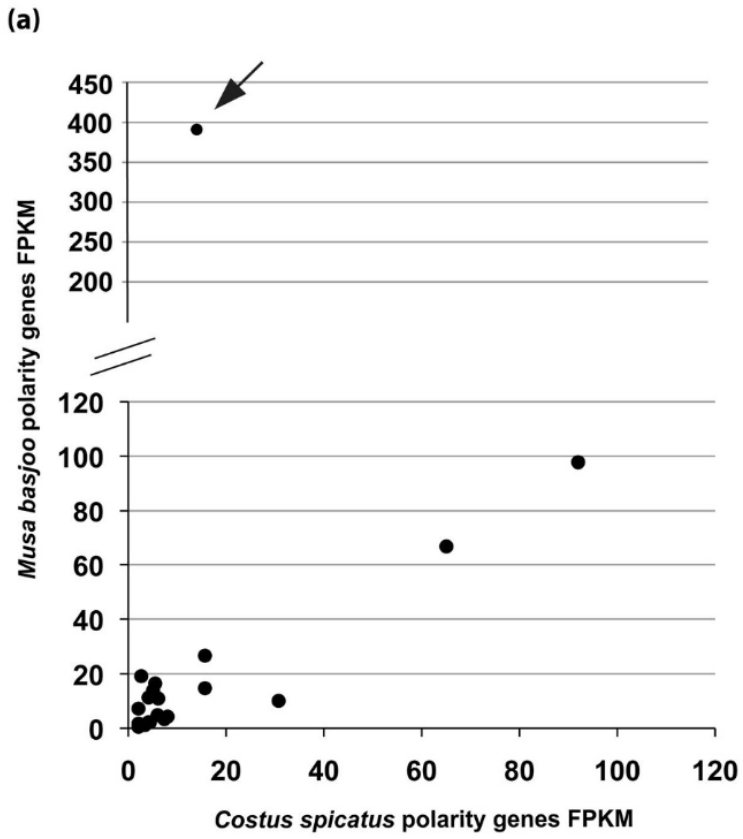

Thecae
Connective
1 cm
Adaxial
Abaxial
Connective
Thecae
Fertile stamen
Fused staminodes (labellum)
Gynoecium

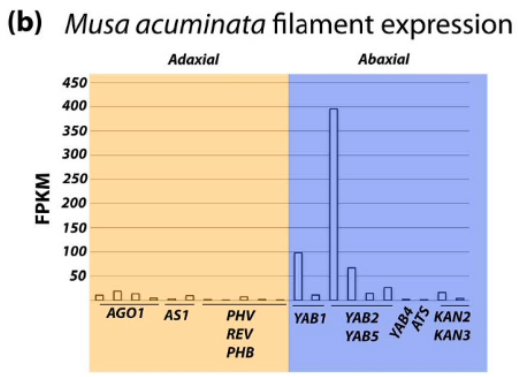

Plain of sectioning
AB
AD
x
p

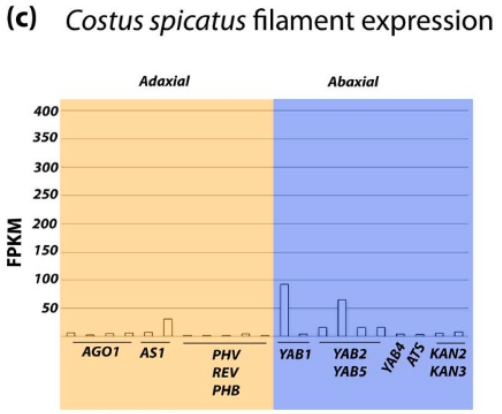

Flower meristem
Fertile stamen
Gynoecium
Bract
Floral buds
Inflorescence meristem
Musa basjoo
inflorescence

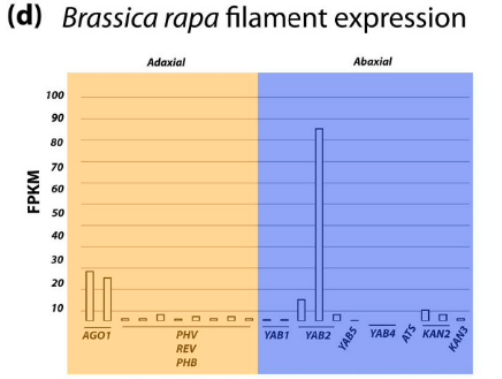

4.5 cm
Gynoecium
Fertile stamens
Theca
Filament
1 cm
Gynoecium
Petal
Fertile stamens (thecae)


Plain of sectioning
AB
AD
x
p

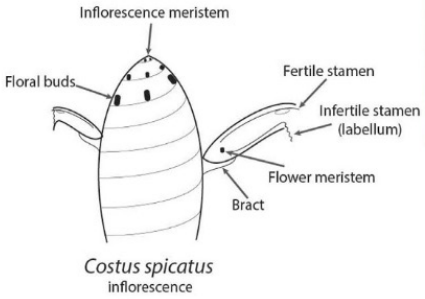

(a)
Musa basjoo polarity genes FPKM
Costus spicatus polarity genes FPKM
0
20
40
60
80
100
120
200
250
300
350
400
450

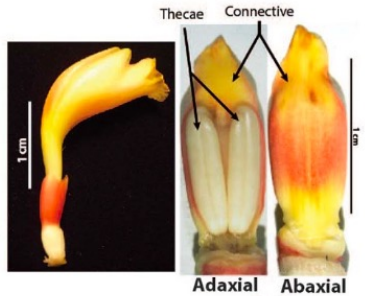

(b) Musa acuminata filament expression
Adaxial
Abaxial
FPKM
AGO1
AS1
PHV
REV
PHB
YAB1
YAB2
YAB5
YAB4
ATS
KAN2
KAN3

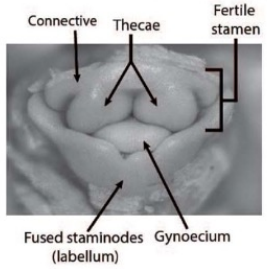

(c) Costus spicatus filament expression
Adaxial
Abaxial
FPKM
AGO1
AS1
PHV
REV
PHB
YAB1
YAB2
YAB5
YAB4
ATS
KAN2
KAN3

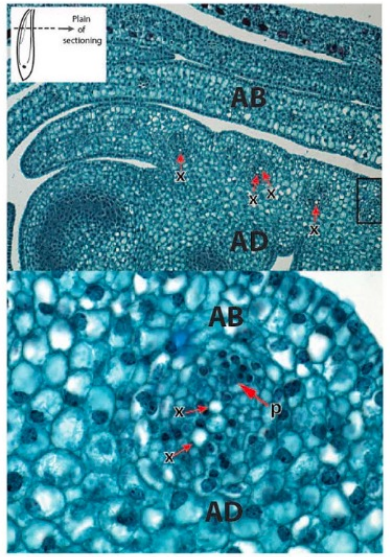

(d) Brassica rapa filament expression
Adaxial
Abaxial
FPKM
AGO1
PHV
REV
PHB
YAB1
YAB2
YAB5
YAB4
ATS
KAN2
KAN3

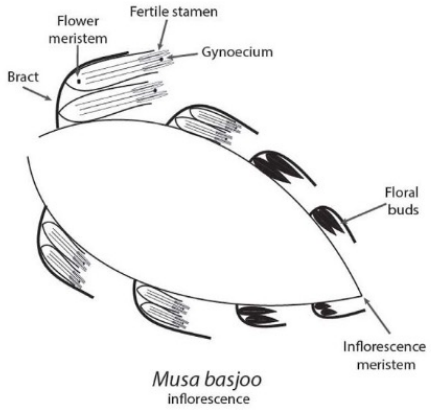
PREGUNTAS

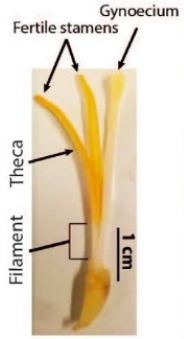


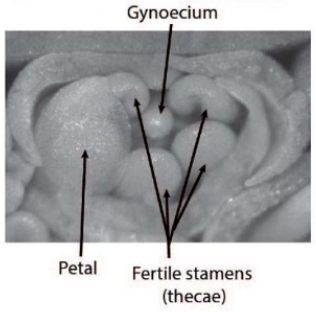


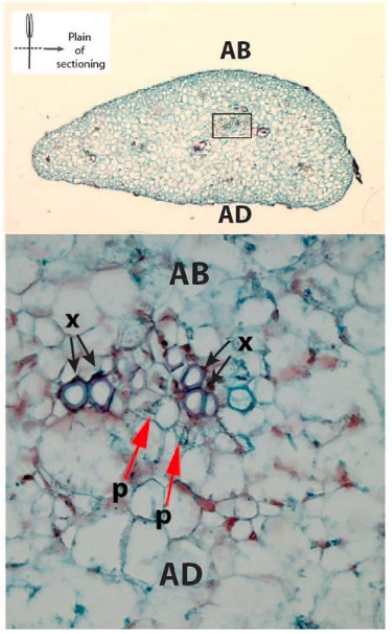


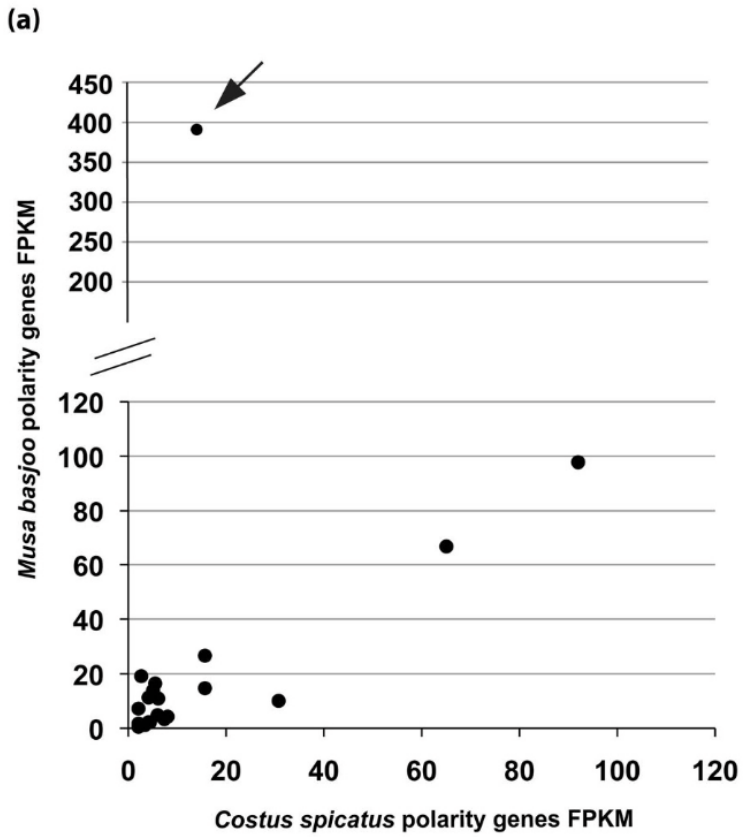


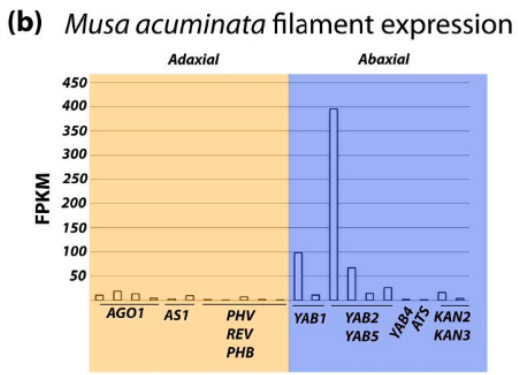


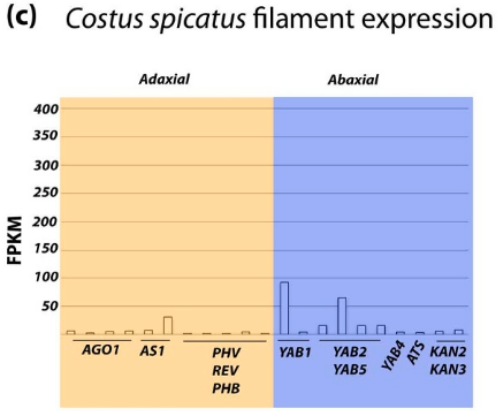


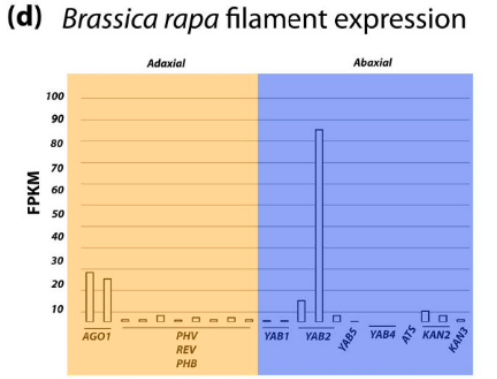












2 especies
2 órganos florales
2 réplicas

PREGUNTAS

RESEARCH

WILEY

Comparative transcriptomics of a monocotyledonous geophyte reveals shared molecular mechanisms of underground storage organ formation

Carrie M. Tribble[1] | **Jesús Martínez-Gómez**[1,2] | **Fernando Alzate-Guarín**[3] | **Carl J. Rothfels**[1] | **Chelsea D. Specht**[2]

[1]Department of Integrative Biology and University Herbarium, University of California, Berkeley, California, USA

[2]School of Integrative Plant Sciences, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, New York, USA

[3]Grupo de Estudios Botánicos (GEOBOTA) and Herbario Universidad de Antioquia (HUA), Facultad de Ciencias Exactas y Naturales, Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

Correspondence
Carrie M. Tribble, Department of Integrative Biology and University Herbarium, University of California, Berkeley, Berkeley, CA 94720, United States.
Email: ctribble09@gmail.com

Funding information
UC Berkeley College of Natural Resources, Grant/Award Number: NA; University and Jepson Herbaria, Grant/Award Number: NA

Abstract

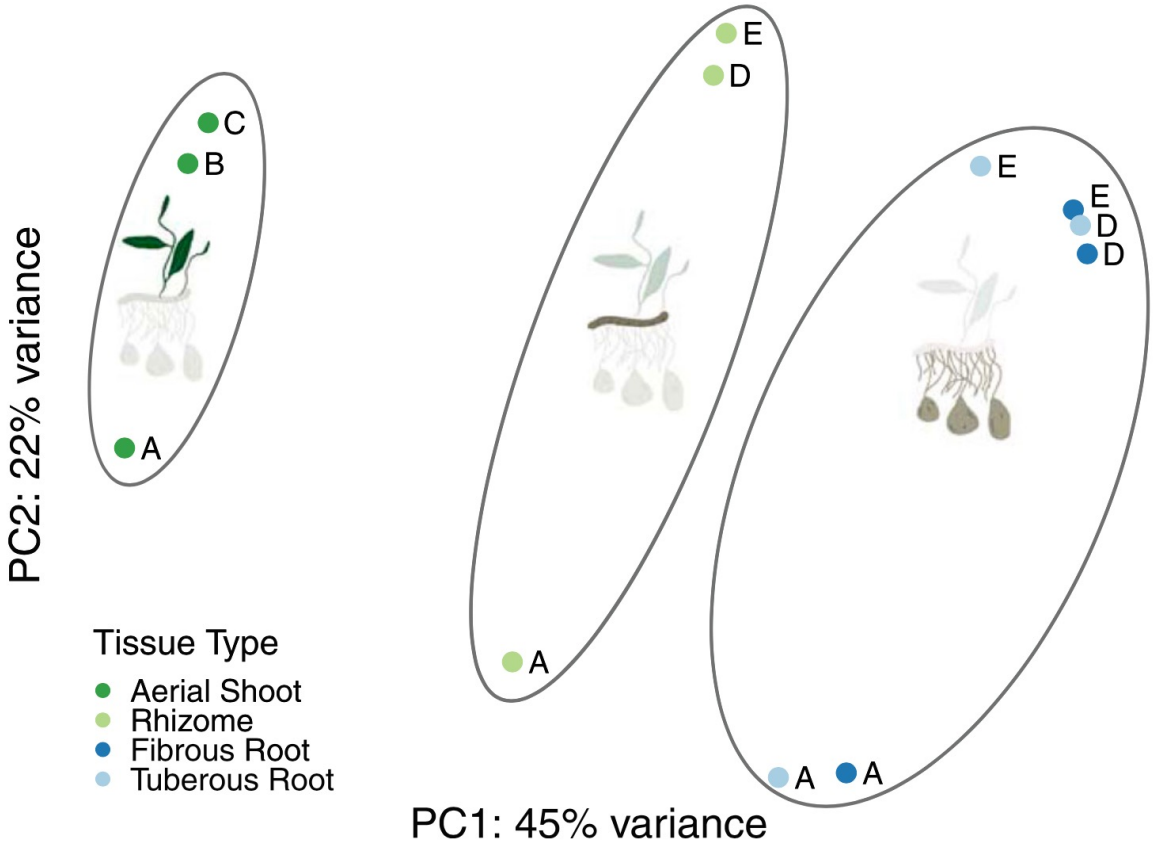
Many species from across the vascular plant tree-of-life have modified standard plant tissues into tubers, bulbs, corms, and other underground storage organs (USOs), unique innovations which allow these plants to retreat underground. Our ability to understand the developmental and evolutionary forces that shape these morphologies is limited by a lack of studies on certain USOs and plant clades. We take a comparative transcriptomics approach to characterizing the molecular mechanisms of tuberous root formation in *Bomarea multiflora* (Alstroemeriaceae) and compare these mechanisms to those identified in other USOs across diverse plant lineages; *B. multiflora* fills a key gap in our understanding of USO molecular development as the first monocot with tuberous roots to be the focus of this kind of research. We sequenced transcriptomes from the growing tip of four tissue types (aerial shoot, rhizome, fibrous root, and root tuber) of three individuals of *B. multiflora*. We identified differentially expressed isoforms between tuberous and non-tuberous roots and tested the expression of a priori candidate genes implicated in underground storage in other taxa. We identify 271 genes that are differentially expressed in root tubers versus non-tuberous roots, including genes implicated in cell wall modification, defense response, and starch biosynthesis. We also identify a phosphatidylethanolamine-binding protein, which has been implicated in tuberization signalling in other taxa and, through gene-tree analysis, place this copy in a phylogenetic context. These findings suggest that some similar molecular processes underlie the formation of USOs across flowering plants despite the long evolutionary distances among taxa and non-homologous morphologies (e.g., bulbs vs. tubers). (Plant development, tuberous roots, comparative transcriptomics, geophytes).

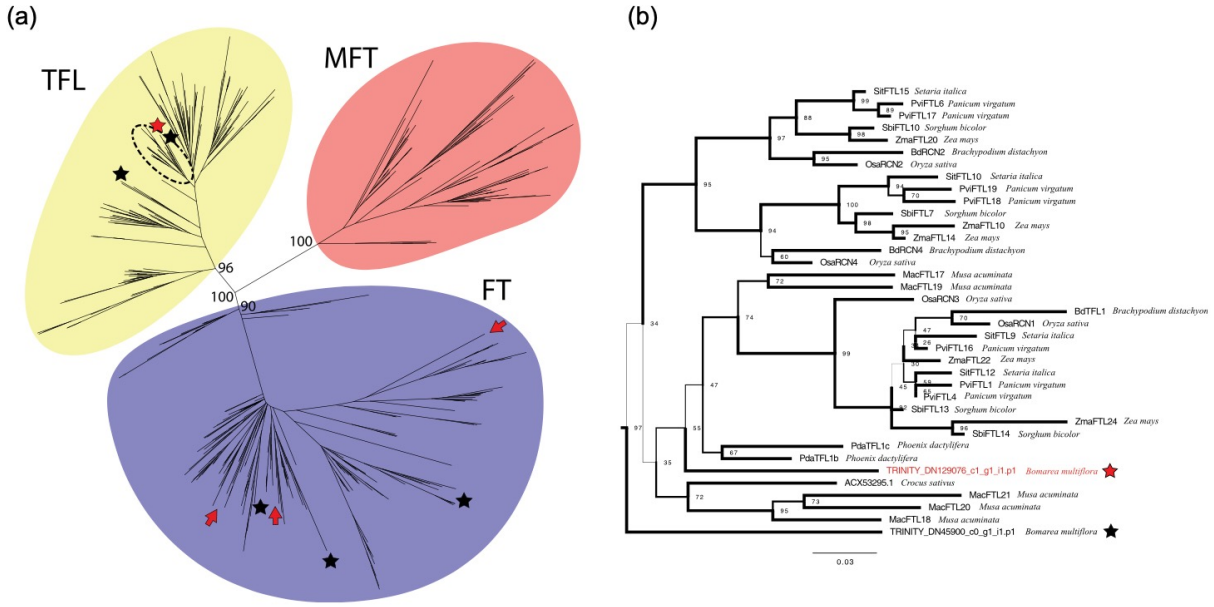
KEYWORDS
comparative transcriptomics, geophytes, tuberous roots

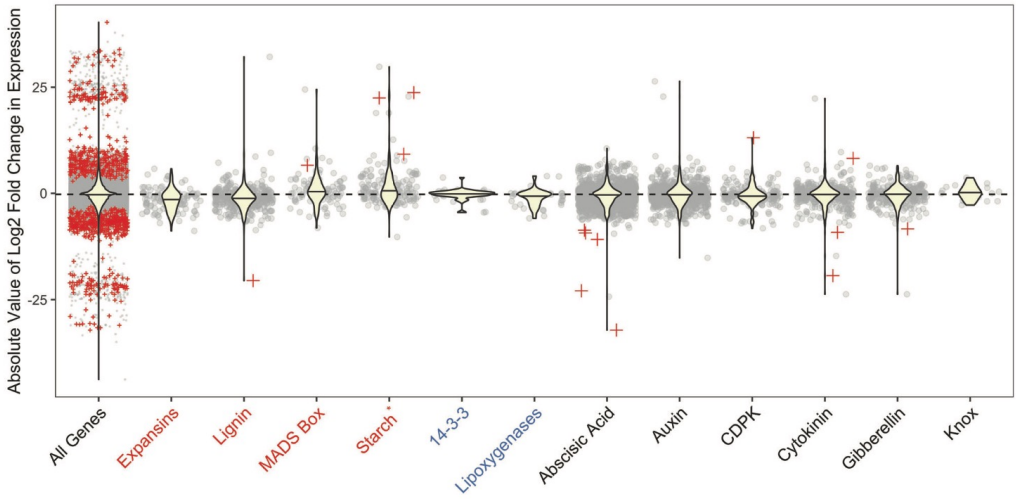
Dr. Carrie Tribble

“Cuales son los genes responsables por el desorollo de túberculos de raíz?”

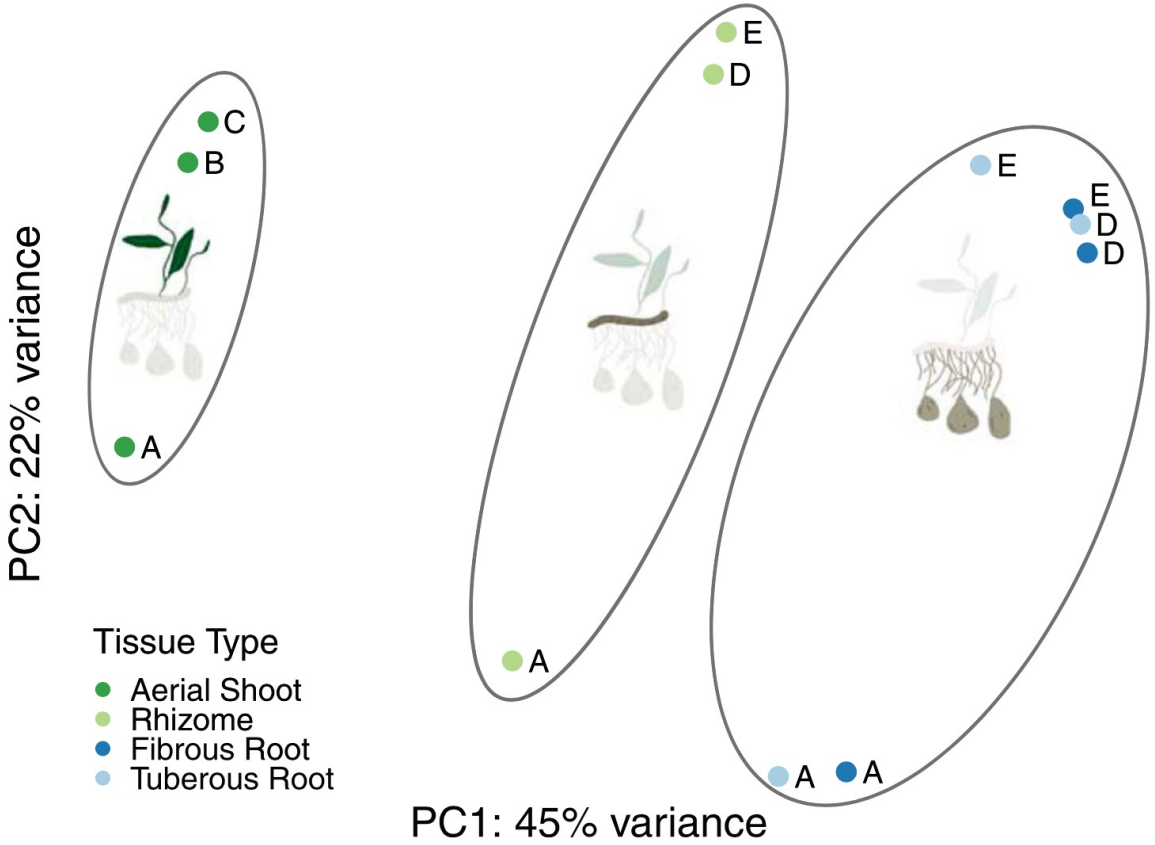
PREGUNTAS

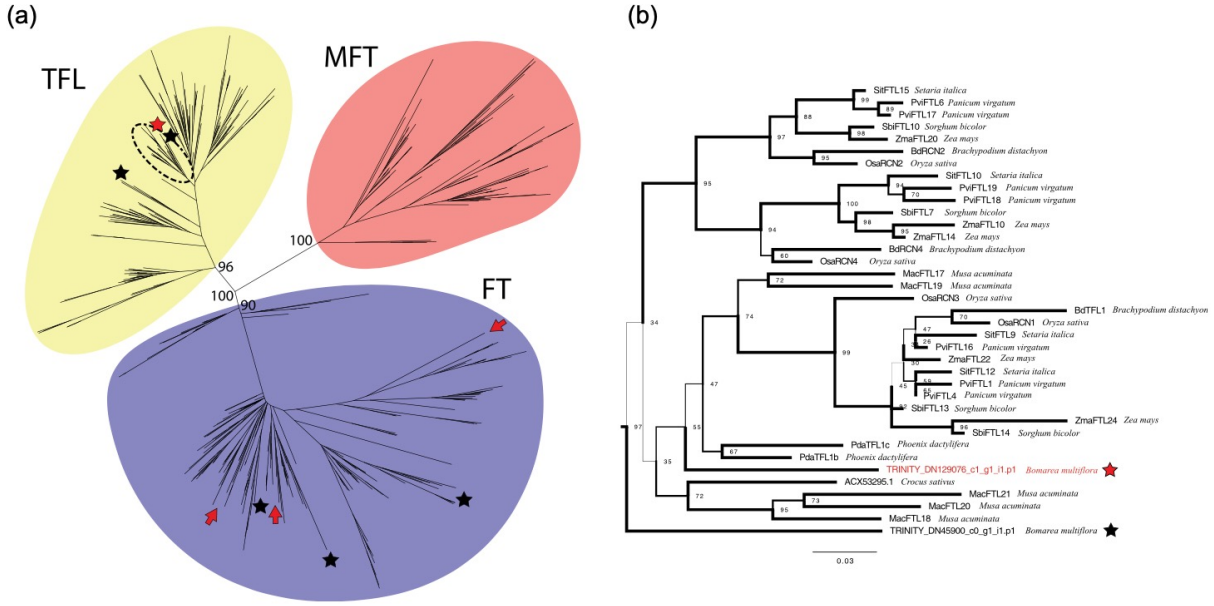

PC2: 22% variance
PC1: 45% variance
Tissue Type
Aerial Shoot
Rhizome
Fibrous Root
Tuberous Root
A
B
C
D
E

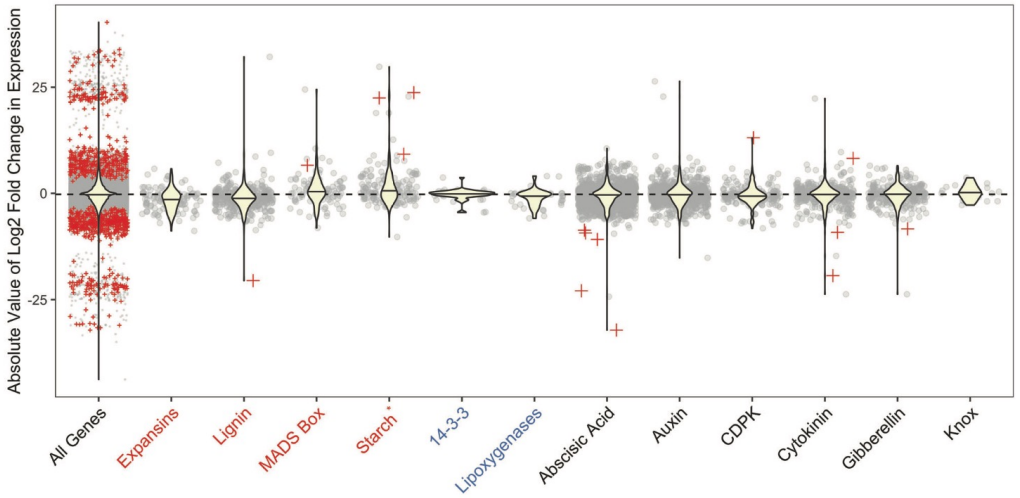

(a)
TFL
MFT
FT
(b)


Absolute Value of Log2 Fold Change in Expression
All Genes
Expansins
Lignin
MADS Box
Starch
14-3-3
Lipoxygenases
Abscisic Acid
Auxin
CDPK
Cytokinin
Gibberellin
Knox

PREGUNTAS







2(4) órganos
3 replicas

PREGUNTAS

A cuáles preguntas de investigación les interesa aplicar estos métodos?

1) Qué es el ARN?

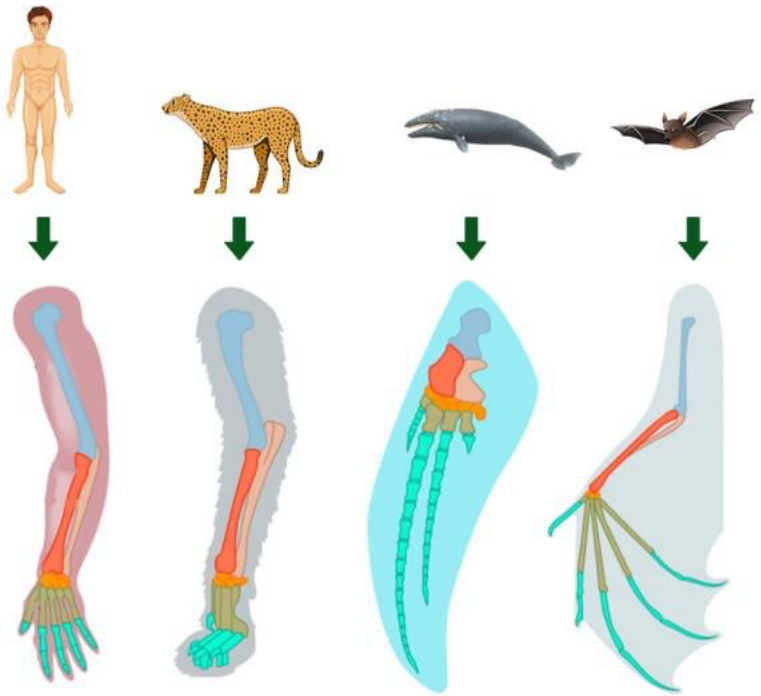
2) Qué es el transcriptoma?

3) A cuáles preguntas podemos responder utilizando ARN?

4) Cómo diseñar un experimento?

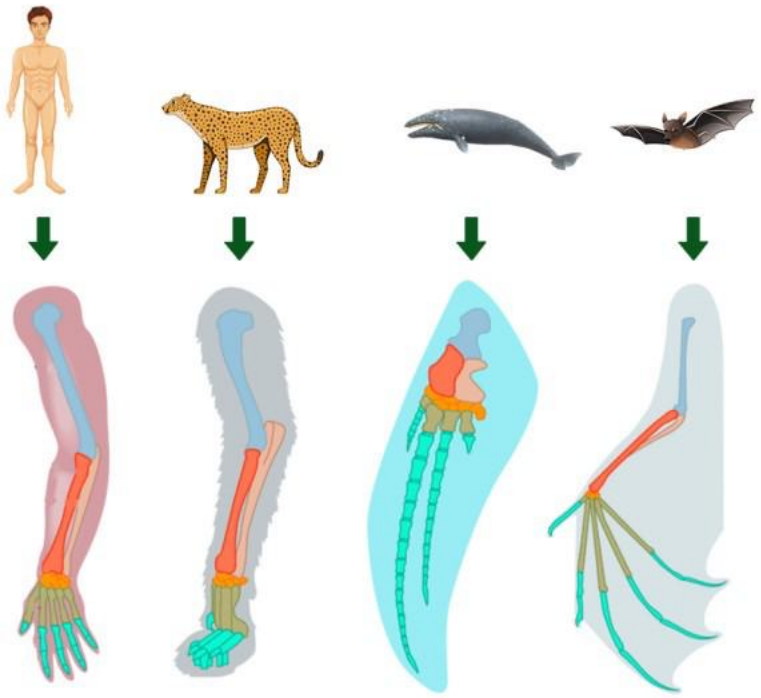
Que se está comparando?

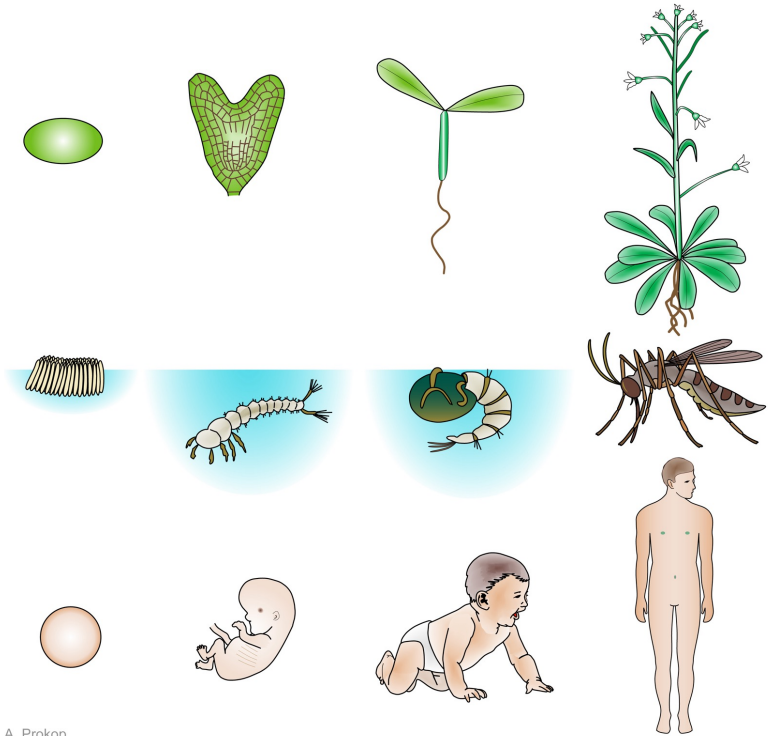
Que se está comparando?



homología

Que se está comparando?

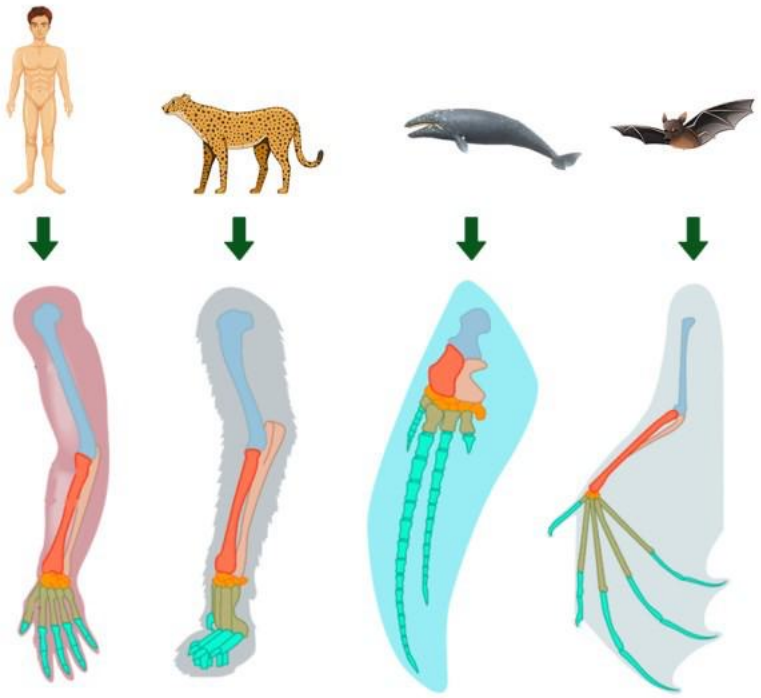




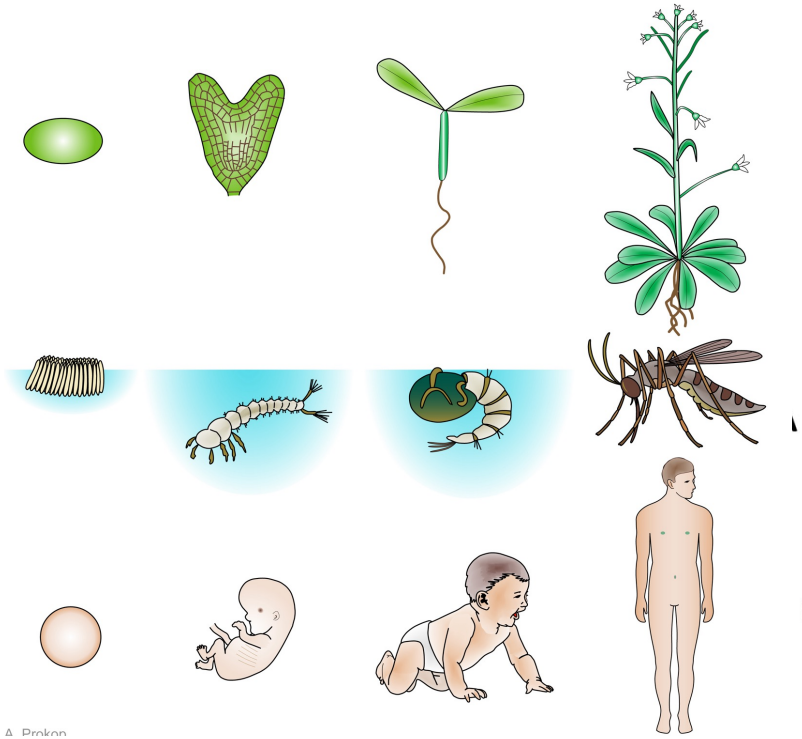
homología

desarollo

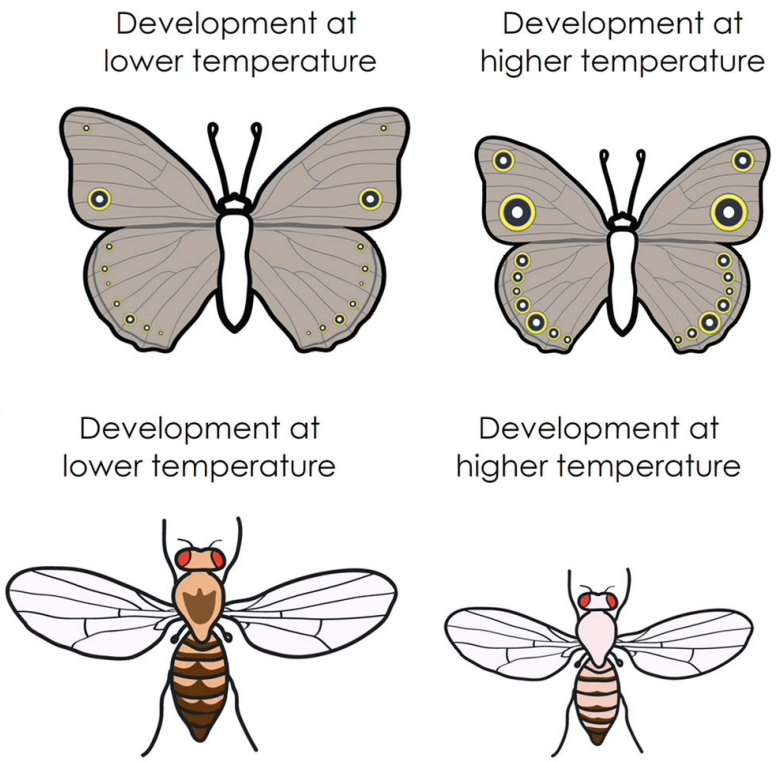
Que se está comparando?



homología



desarollo



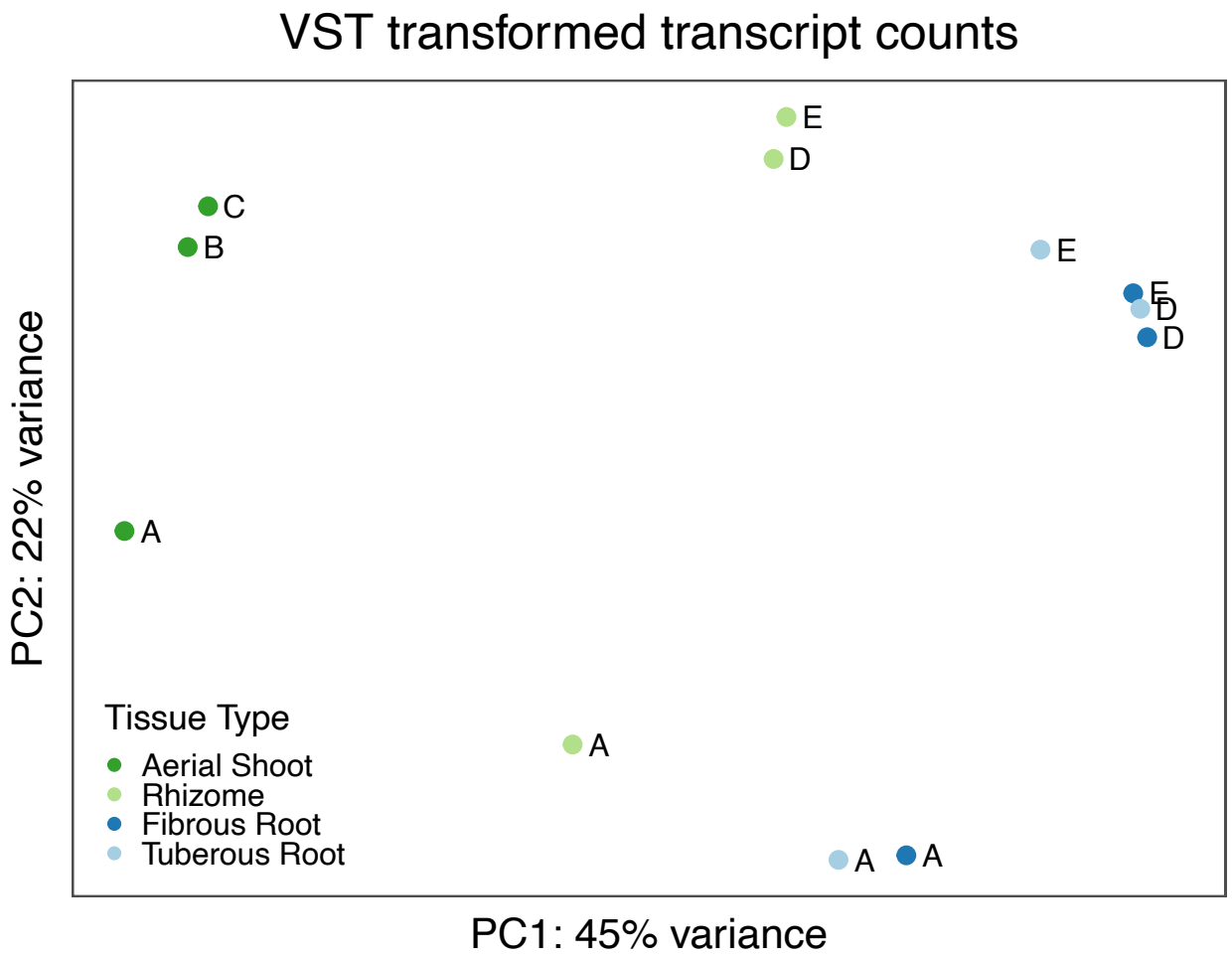
ambiente

Cuántas réplicas se necesita?

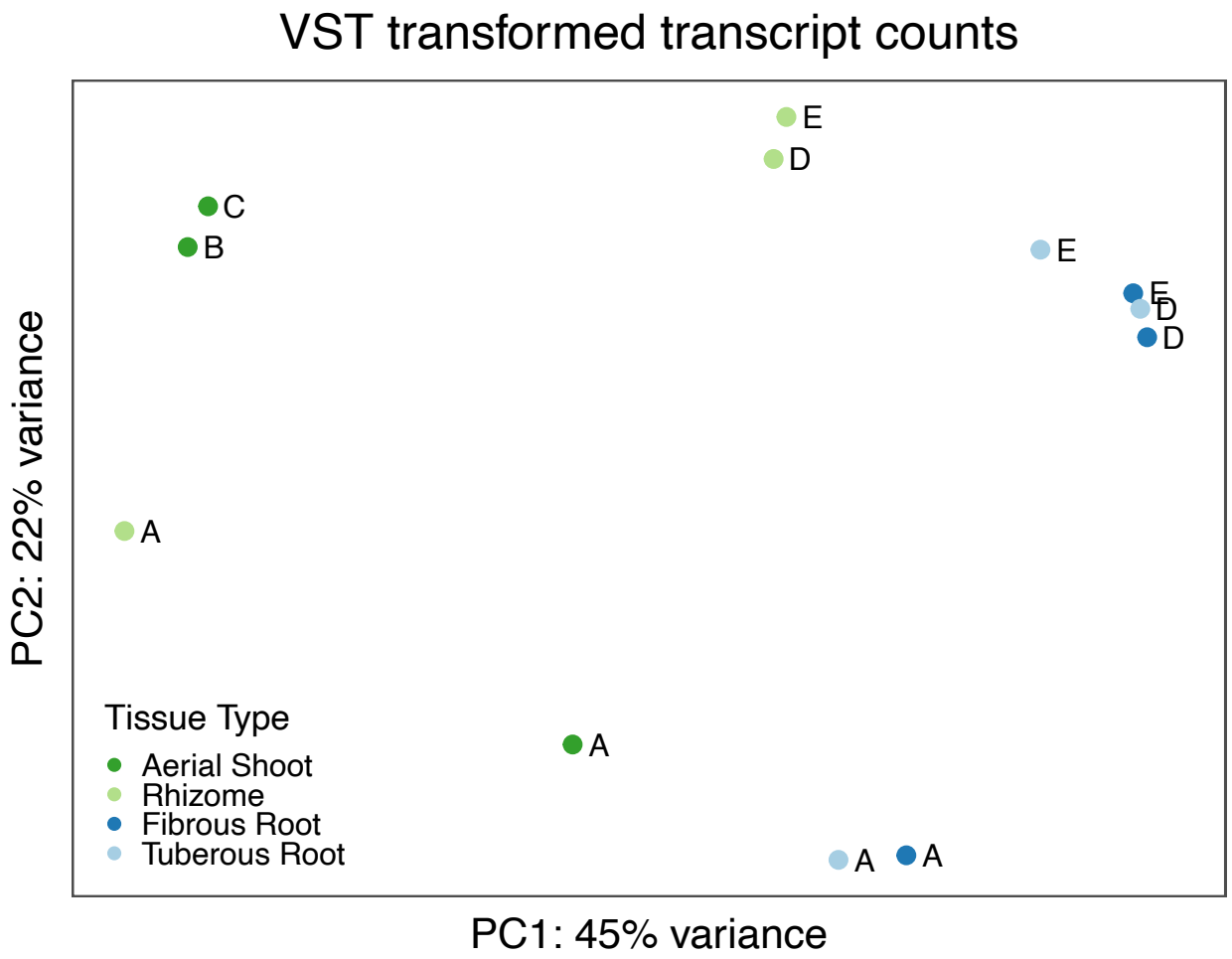
Cuántas réplicas se necesita?

lo más que se puede…

Cuántas réplicas se necesita?

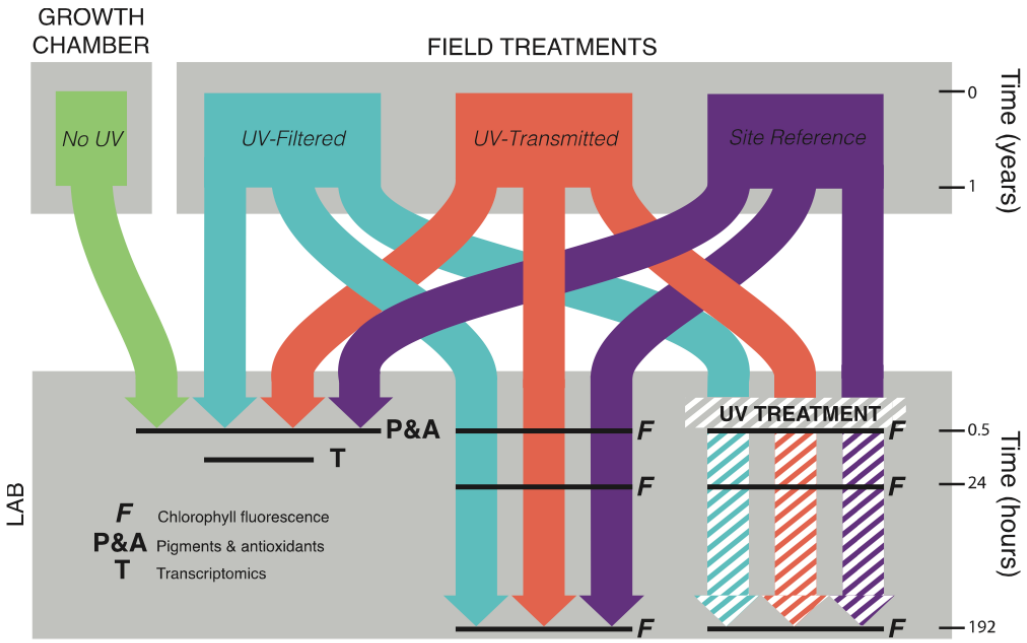


Cuántas réplicas se necesita?

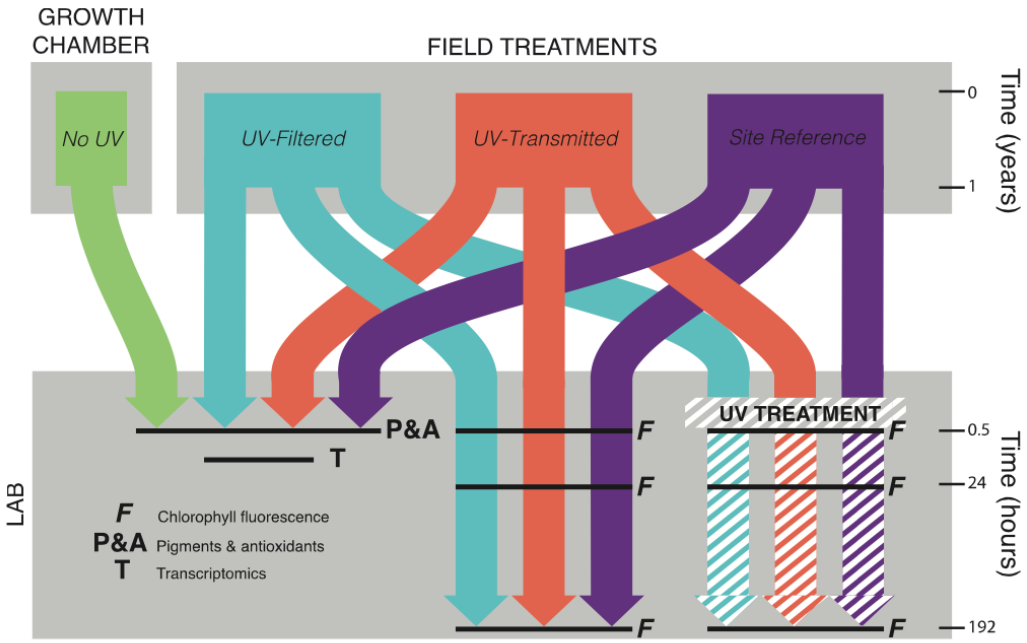


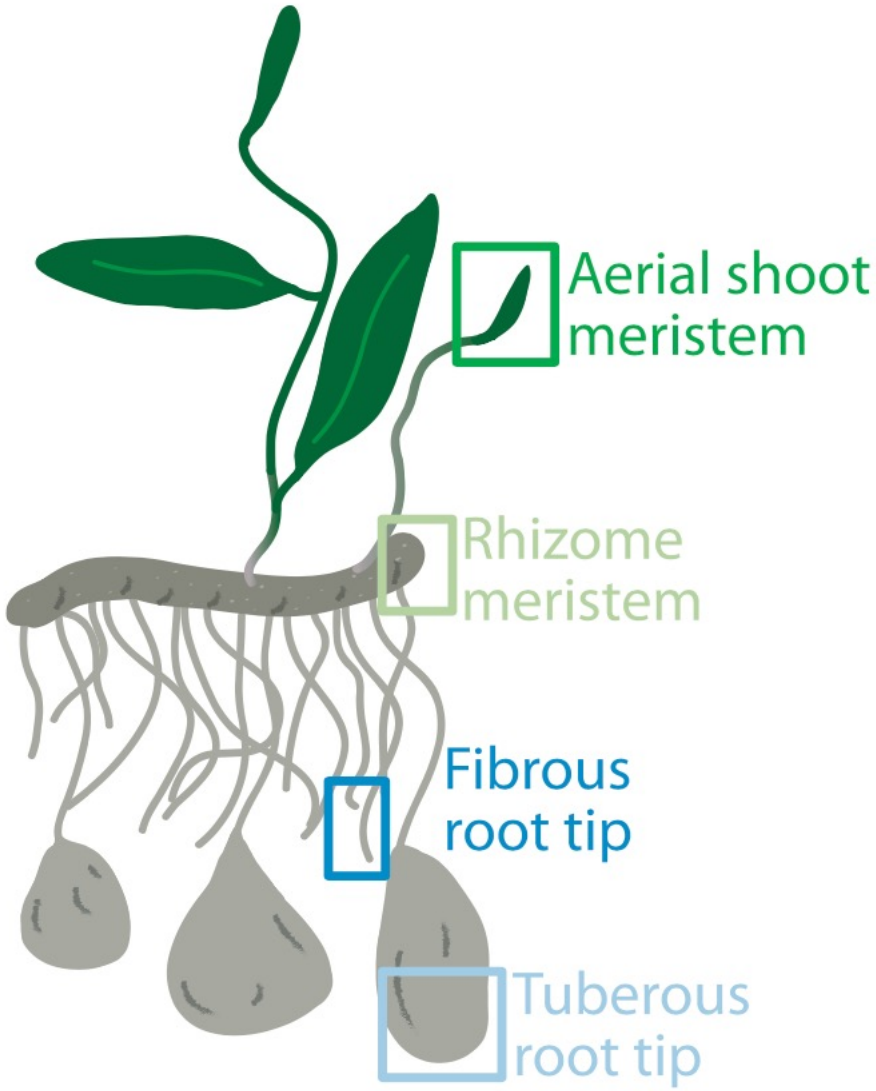
3 mínimo
5-6 mejor

Cuál es el control del experimento?



Cuál es el control del experimento?



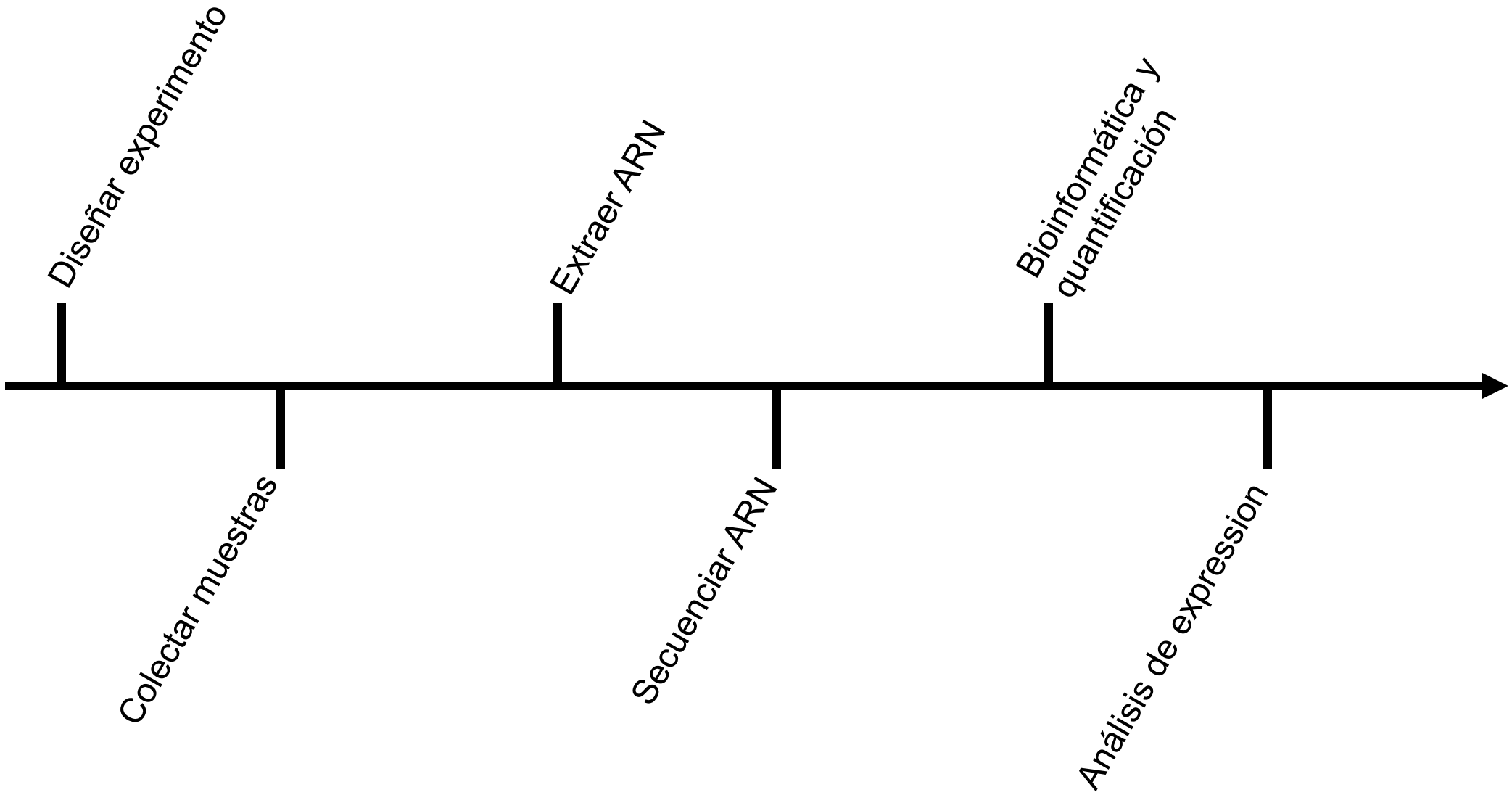


DESCANSO

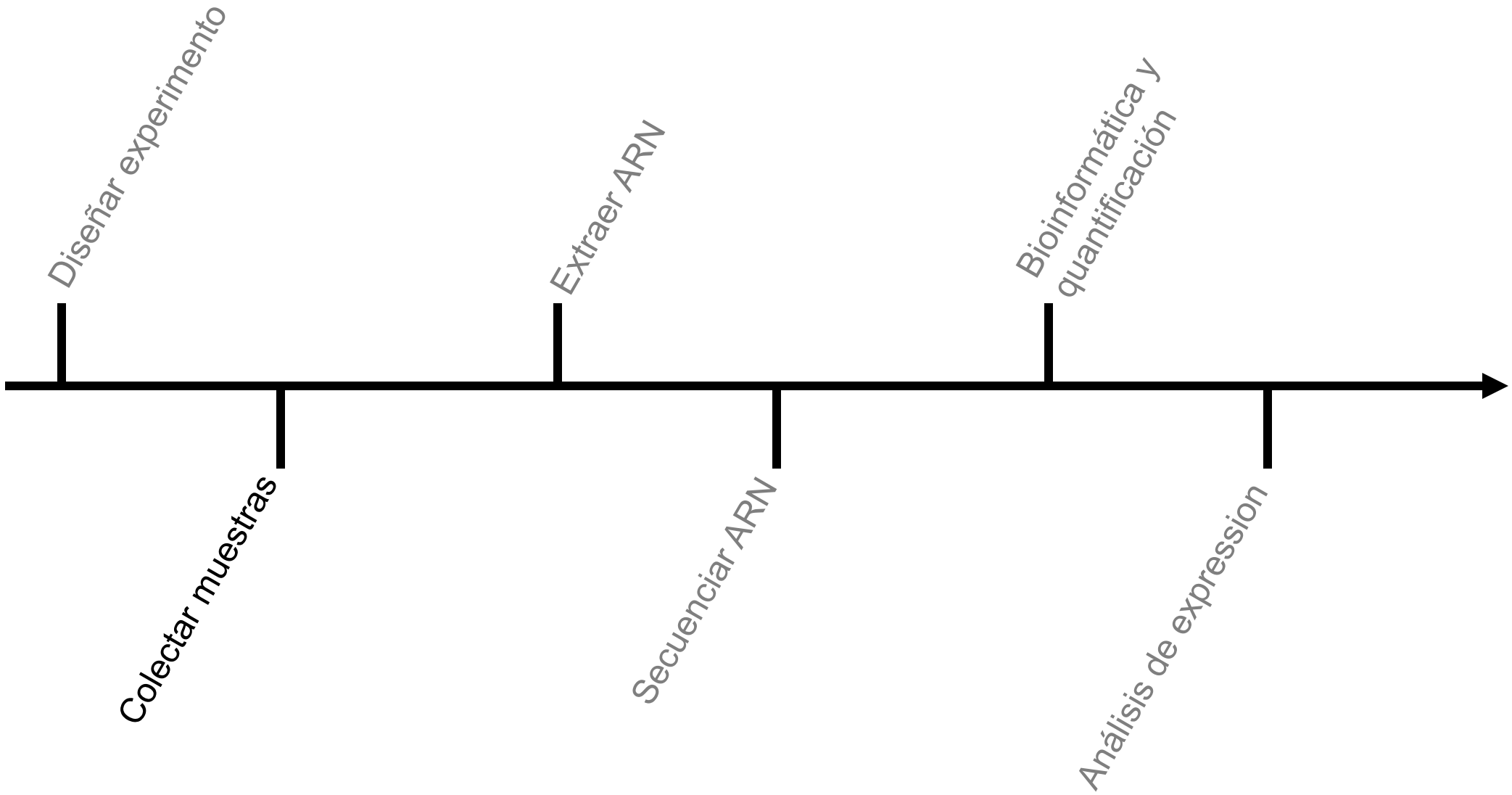
1) Teoría de transcriptomas (1 hora)

2) Proceso de realizar un experimento (0.5 hora)

3) Analizar datos en R (2 horas)


Diseñar experimento
Colectar muestras
Extraer ARN
Secuenciar ARN
Bioinformática y quantificación
Análisis de expression

EL PROCESO


Diseñar experimento
Colectar muestras
Extraer ARN
Secuenciar ARN
Bioinformática y quantificación
Análisis de expression

EL PROCESO

Opción 1: el lab/ invernadero

Opción 2: el campo

EL PROCESO

Opción 2: el campo

An efficient field and laboratory workflow for plant phylotranscriptomic projects[1]

Ya Yang[2,6,8], Michael J. Moore[3], Samuel F. Brockington[4], Alfonso Timoneda[4], Tao Feng[4], Hannah E. Marx[5,7], Joseph F. Walker[2], and Stephen A. Smith[2]

[2]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, 830 North University Avenue, Ann Arbor, Michigan 48109 USA; [3]Department of Biology, Oberlin College, 119 Woodland Street, Oberlin, Ohio 44074-1097 USA; [4]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom; and [5]Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, Idaho 83844 USA

- *Premise of the study:* We describe a field and laboratory workflow developed for plant phylotranscriptomic projects that involves cryogenic tissue collection in the field, RNA extraction and quality control, and library preparation. We also make recommendations for sample curation.
- *Methods and Results:* A total of 216 frozen tissue samples of Caryophyllales and other angiosperm taxa were collected from the field or botanical gardens. RNA was extracted, stranded mRNA libraries were prepared, and libraries were sequenced on Illumina HiSeq platforms. These included difficult mucilaginous tissues such as those of Cactaceae and Droseraceae.
- *Conclusions:* Our workflow is not only cost effective (ca. $270 per sample, as of August 2016, from tissue to reads) and time efficient (less than 50 h for 10–12 samples including all laboratory work and sample curation), but also has proven robust for extraction of difficult samples such as tissues containing high levels of secondary compounds.

Key words: Caryophyllales; cryogenic field sampling; phylogenomics; phylotranscriptomics; RNA.

Family	Taxon name	Collection or accession no. (Herbarium)[a,b]	Collection locality	Field or cultivated?	RNA extraction	SRA accession no. (Publication)
Polygonaceae	*Podopterus mexicanus* Humb. & Bonpl.	Michael J. Moore et al. 2721 (OC)	Jalisco, Mexico: La Huerta, Estacion Biologica Chamela, along Arroyo Colorado near the end of the Eje Central	Field	PureLink	Unpublished
Polygonaceae	*Polygonum* sp.	Michael J. Moore et al. 3219 (OC)	California, USA: San Luis Obispo, Morro Strand State Beach: just W of Hwy. 1 in Cayucos	Field	PureLink	Unpublished
Polygonaceae	*Polygonum* sp.	Michael J. Moore et al. 3263 (OC)	California, USA: alpine, along Round Top Lake Trail	Field	PureLink	Unpublished
Polygonaceae	*Polygonum aviculare* L.	Michael J. Moore 2667 (OC)	Ohio, USA: Lorain, adjacent to Woodland St. parking lot on campus of Oberlin College	Field	Bio-Rad	Unpublished
Polygonaceae	*Polygonum dentoceras* T. M. Schust. & Reveal	Michael J. Moore et al. 3353 (OC)	Florida, USA: Highlands, just S of the Denny's parking lot on the E side of US 27, approx. 0.6 mi. N of jct. w/ Schumacher Rd./Sebring Pkwy.	Field	PureLink	Unpublished
Polygonaceae	*Pterostegia drymarioides* Fisch. & C. A. Mey.	Millennium Seed Bank accession 0496863	Cultivated at Cambridge University Botanic Garden	Cultivated	PureLink	Unpublished

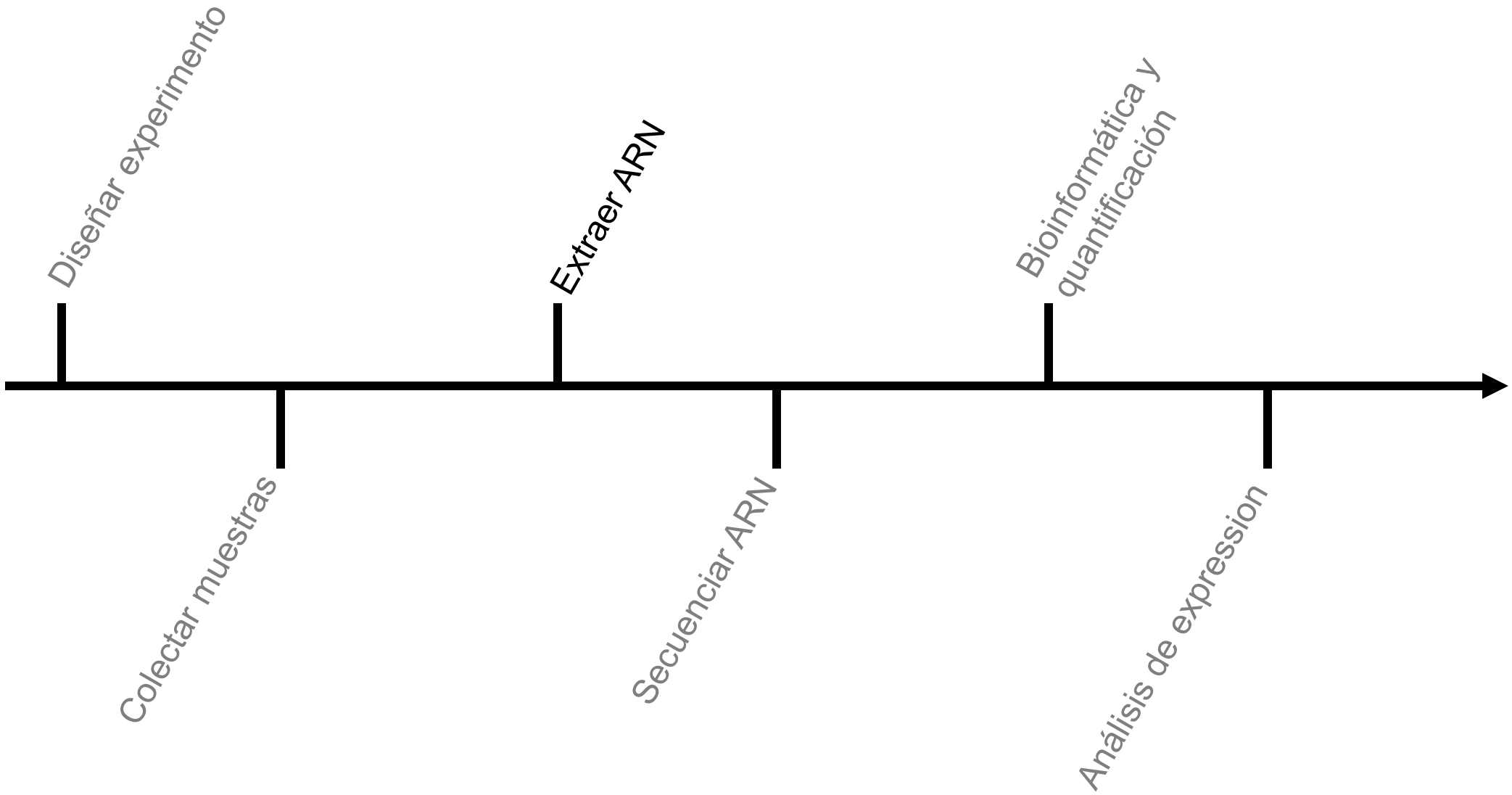
EL PROCESO

Opción 3: el campo

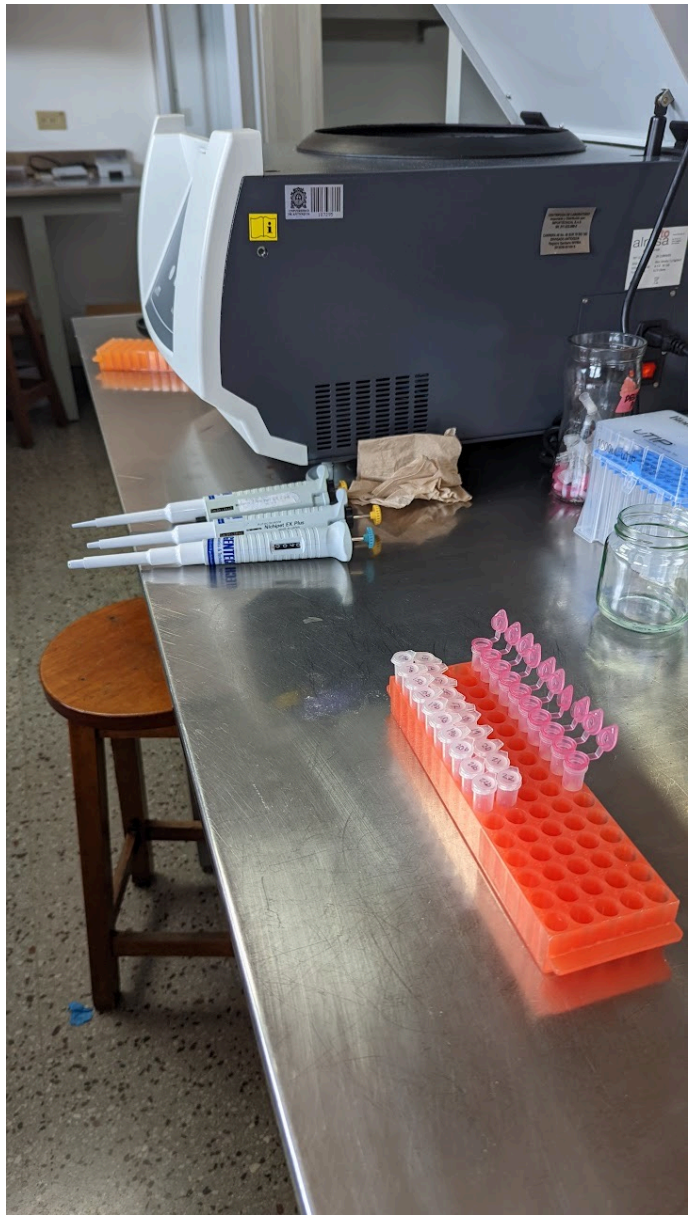
RNAlater, etc.

Este curso:

RNAlater

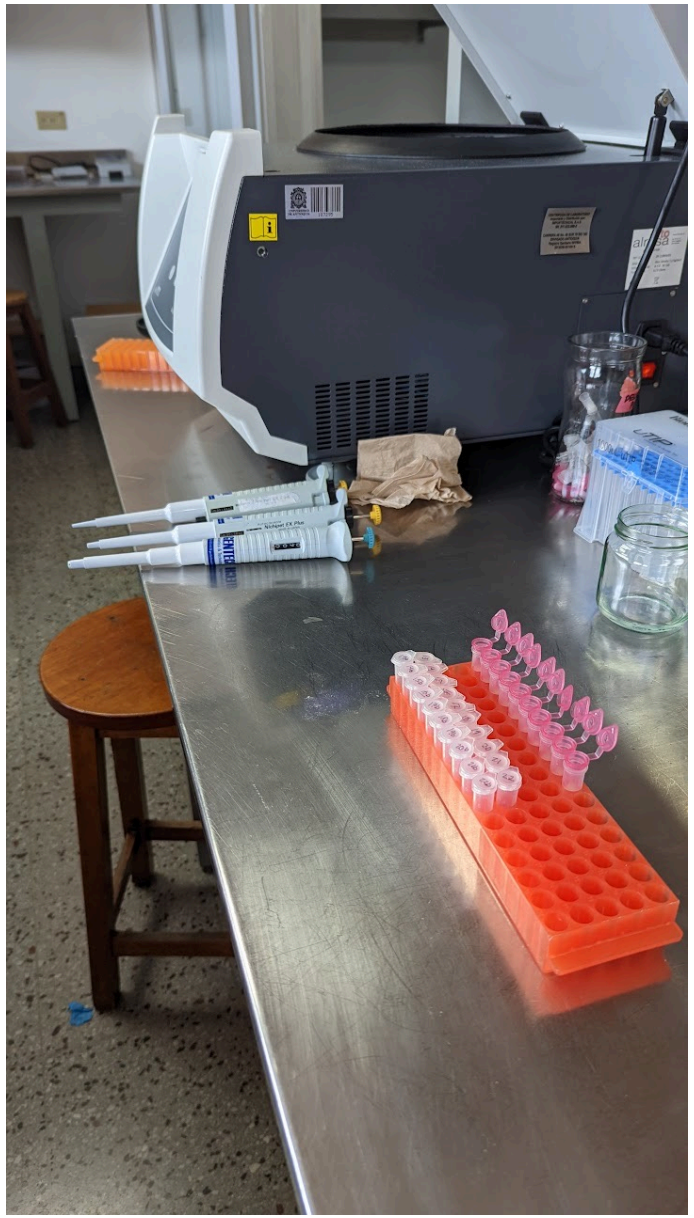

Diseñar experimento
Colectar muestras
Extraer ARN
Secuenciar ARN
Bioinformática y quantificación
Análisis de expression

EL PROCESO



FRIO

EL PROCESO



LIMPIO

EL PROCESO

Qiaigen Rneasy Plant Mini Kit
RNeasy® Plant Mini Kit (50)
Cat. No. 74904
QIAGEN

EL PROCESO

Agilent | Trusted Answers

Plant RNA Isolation Mini Kit

- Sample required per isolation is 25 – 100 mg tissue
- Isolate higher purity RNA even from notoriously challenging sources such as waxy and oil tissues (e.g., seeds & conifer needles) or polysaccharide-containing storage tissues (e.g., tubers)
- Up to a 10,000-fold reduction in genomic DNA contamination compared to kits that don't include DNase
- Resulting RNA is free of known inhibitors of downstream reactions, such as carbohydrates, waxes, and secondary metabolites

The Agilent Plant RNA Isolation Mini Kit delivers intact and high-purity total RNA for use in gene expression and other downstream analyses. This kit is a straightforward, spin-column format that is both DNAse- and phenol-free. It has been specially formulated for plant samples from a wide variety of species and tissue types.

Applications in Plant Sciences
Official publication of the Botanical Society of America

Protocol Note | Open Access

A method for extracting high-quality RNA from diverse plants for next-generation sequencing and gene expression analyses†

Roxana Yockteng, Ana M. R. Almeida, Stephen Yee, Thiago Andre, Colin Hill, Chelsea D. Specht

First published: 09 December 2013 | https://doi.org/10.3732/apps.1300070 | Citations: 34

† The authors thank the Specht Laboratory for sharing ideas and data; T. Renner (carnivores), S. Salzman, I. Terry, and B. Dorsey (cycads), F. Zapata (Burseraceae), A. Jackson and Mi Yeon Lee (*Nicotiana* and *Hordeum*), and K. Morioka (monocots) for troubleshooting and testing; M. Bartlett for introducing the laboratory to Invitrogen's RNA reagent; the Evolutionary Genetics Lab, Museum of Vertebrate Zoology, for supporting next-gen sequencing among the Berkeley Natural History Museums; and T. Gregory for providing cycad tissue and financial support. Additional funding was provided by the National Science Foundation (IOS 0845641, DEB 1110461), the University of California Institute for Mexico (UC MEXUS)–Consejo Nacional de Ciencia y Tecnología (CONACYT), the Hellman Family Fund, and a Prytanean Junior Faculty Award.

Agilent | Trusted Answers

Plant RNA Isolation Mini Kit

- Sample required per isolation is 25 – 100 mg tissue
- Isolate higher purity RNA even from notoriously challenging sources such as waxy and oil tissues (e.g., seeds & conifer needles) or polysaccharide-containing storage tissues (e.g., tubers)
- Up to a 10,000-fold reduction in genomic DNA contamination compared to kits that don't include DNase
- Resulting RNA is free of known inhibitors of downstream reactions, such as carbohydrates, waxes, and secondary metabolites

The Agilent Plant RNA Isolation Mini Kit delivers intact and high-purity total RNA for use in gene expression and other downstream analyses. This kit is a straightforward, spin-column format that is both DNAse- and phenol-free. It has been specially formulated for plant samples from a wide variety of species and tissue types.

EL PROCESO

Modified CTAB and TRIzol protocols improve RNA extraction from chemically complex Embryophyta[1]

Ingrid E. Jordon-Thaden[2,3,6,7], Andre S. Chanderbali[2], Matthew A. Gitzendanner[2], and Douglas E. Soltis[2,4,5]

[2]Department of Biology, University of Florida, Gainesville, Florida 32611 USA; [3]University and Jepson Herbaria, University of California Berkeley, Berkeley, California 94720 USA; [4]Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611 USA; and [5]Genetics Institute, University of Florida, Gainesville, Florida 32611 USA

- *Premise of the study:* Here we present a series of protocols for RNA extraction across a diverse array of plants; we focus on woody, aromatic, aquatic, and other chemically complex taxa.
- *Methods and Results:* Ninety-one taxa were subjected to RNA extraction with three methods presented here: (1) TRIzol/TURBO DNA-*free* kits using the manufacturer's protocol with the addition of sarkosyl; (2) a combination method using cetyltrimethylammonium bromide (CTAB) and TRIzol/sarkosyl/TURBO DNA-*free*; and (3) a combination of CTAB and QIAGEN RNeasy Plant Mini Kit. Bench-ready protocols are given.
- *Conclusions:* After an iterative process of working with chemically complex taxa, we conclude that the use of TRIzol supplemented with sarkosyl and the TURBO DNA-*free* kit is an effective, efficient, and robust method for obtaining RNA from 100 mg of leaf tissue of land plant species (Embryophyta) examined. Our protocols can be used to provide RNA of suitable stability, quantity, and quality for transcriptome sequencing.

Key words: aquatic plants; aromatic plants; extraction methods; RNA; transcriptome; woody plants.

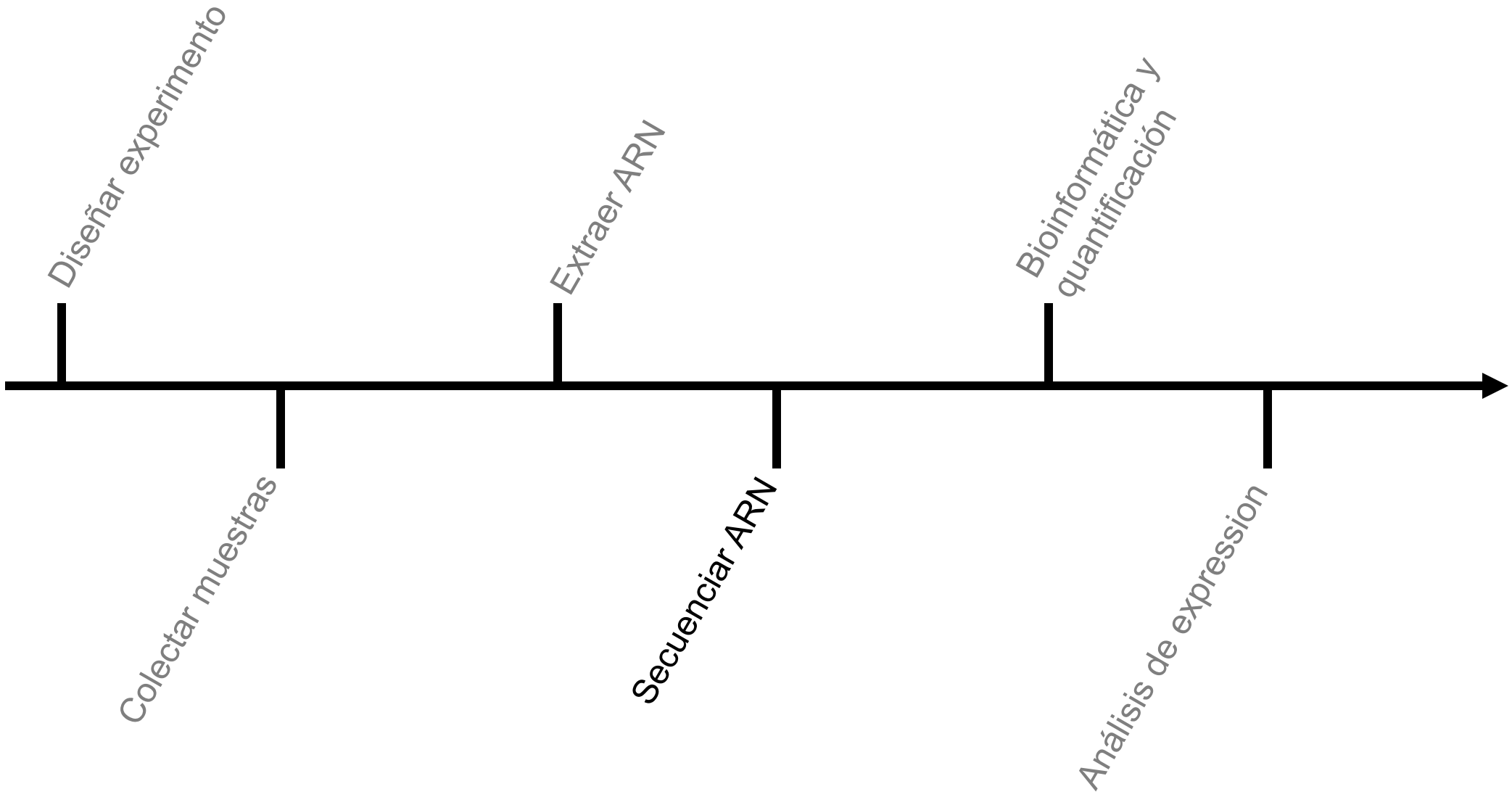
Agilent | Trusted Answers

Plant RNA Isolation Mini Kit

- Sample required per isolation is 25 – 100 mg tissue
- Isolate higher purity RNA even from notoriously challenging sources such as waxy and oil tissues (e.g., seeds & conifer needles) or polysaccharide-containing storage tissues (e.g., tubers)
- Up to a 10,000-fold reduction in genomic DNA contamination compared to kits that don't include DNase
- Resulting RNA is free of known inhibitors of downstream reactions, such as carbohydrates, waxes, and secondary metabolites

The Agilent Plant RNA Isolation Mini Kit delivers intact and high-purity total RNA for use in gene expression and other downstream analyses. This kit is a straightforward, spin-column format that is both DNAse- and phenol-free. It has been specially formulated for plant samples from a wide variety of species and tissue types.

EL PROCESO


Diseñar experimento
Colectar muestras
Extraer ARN
Secuenciar ARN
Bioinformática y quantificación
Análisis de expression

EL PROCESO

Novogene

Advancing Genomics, Improving Life

UC Davis, California

No	Product Name	Description	UNIT PRICE	Currency	QTY	Total price
1	Plant and Animal Eukaryotic mRNA (WOBI)	RNA sample QC	169.00	USD	45	7605.00
2		mRNA library preparation (poly A enrichment)				
3		NovaSeq PE150 (6 G raw data per sample)				
4		Data quality control				
5		FTP	0.00	USD	1	0.00
Total:						7605.00
Grand Total						7605.00

EL PROCESO

Novogene

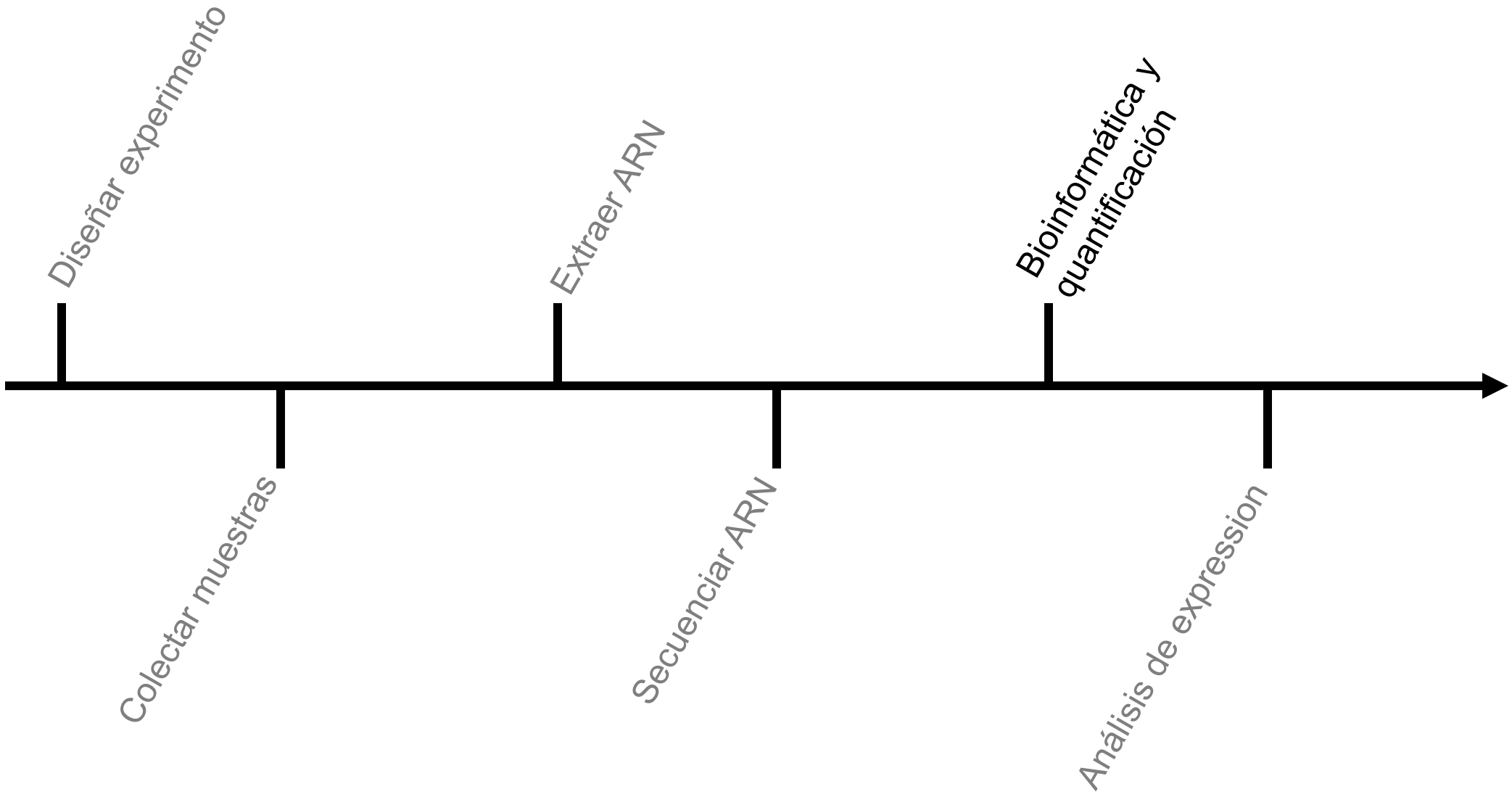
Advancing Genomics, Improving Life

UC Davis, California

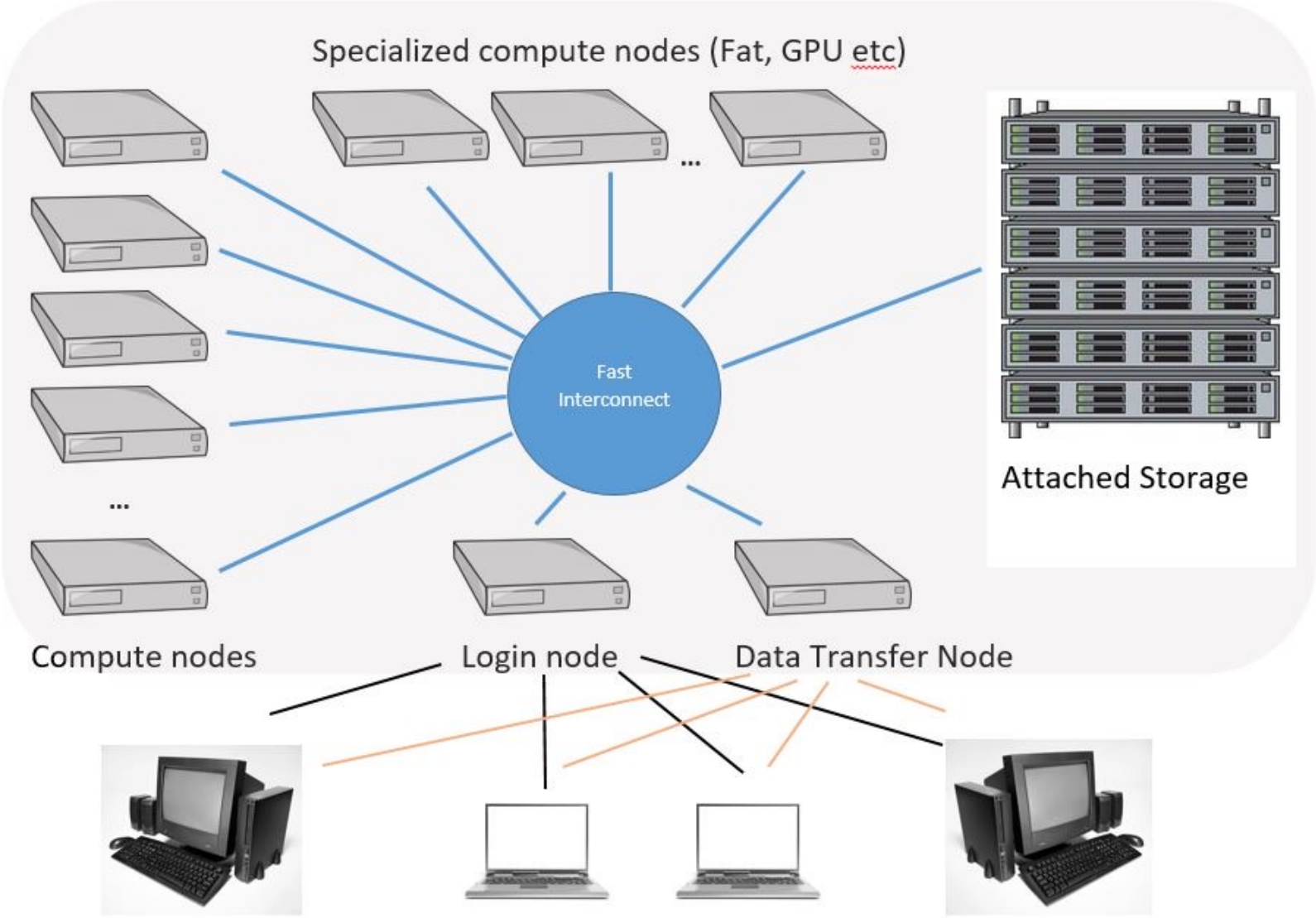
Humanizing Genomics

macrogen

Seoul, South Korea


Diseñar experimento
Colectar muestras
Extraer ARN
Secuenciar ARN
Bioinformática y quantificación
Análisis de expression

EL PROCESO



1 muestras = 1--2 Gb
12 muestras = 24 Gb
varios pasos > 100 Gb

EL PROCESO

Quienes de ustedes han usado un sistema de supercomputadora?

```
@K00188:104:HVHN5BBXX:6:1101:1265:1472 1:N:0:GACACAGT+ACTGTGTC
GTCCTCCGCTTTCCGCTCCTCCCTCGGCGTTCGTCGGAATCGGAGCCGGAGGAGGAGGCGCGATCTGTACGGCTCCGCCCGCTGCGGTGGCTACGGTGGCG
+
AAAFA7JFF<F<FJFF-AFF-FJ-AAFJFJ-FJ<F<-AAA7JJAJF<AF-7<-A<-AAJJAAAJJFFAJJJJJFJJFJJJJJ<AJF7AFJJFAFF-FJ7-A
@K00188:104:HVHN5BBXX:6:1101:1387:1472 1:N:0:GACACAGT+ACTGTGTC
GCTCCAATCTCCTCTTCCTCCTCGTACTCTTCATCATCAGCTGTCGCATCCTGGTATTGCTGGTACTCAGCCACCAAGTCATTCATATTACTCCCGGCCTC
+
AAAFFAFFJFJJ-FFJJJ-FJ<FJF<F7FF-FFFJJFJJJFJJJJJJJJFJ<A<FFJJFJJJJAJJAJ7JFJJJJJJJJ<A-FJJF-FAAJ<F-A7AFJ<J
@K00188:104:HVHN5BBXX:6:1101:2077:1472 1:N:0:GACACAGT+ACTGTGTC
CTTTCAACGGTAAGAACATCGGTCATTTCTTTTATTGGCACAAGCGTGATTTTCGCAGAAGAGTATATATTTCTTAGACCTTTGCAAGCCTCCTTAACATG
+
AAAAFFJJFJFJFF<JFF<J<J7JF-JFFJJAJFJFAFJ<FFFFJFJAJFJFJJJFJJJ7-<FJFJJJJJJJJJJJJFJJJJJJJJJ-FJJJJJFJJ7FJ-
@K00188:104:HVHN5BBXX:6:1101:2361:1472 1:N:0:GACACAGT+ACTGTGTC
CCCGAAACAGAAATACTGATTCATCAGTCGATTTTTGACAAAGCCCGACTTTTGTGGTACATGTTCAGCACCTTGATAAGTTGACGAATACGAGCAGGATG
+
AAFFFJJJJFJJJJJJAAFJJJFJJJJFJJF7JJJJJJJJJFJJJJJJJAJ7AJFFJJJJJJJFJJFFJFFJ<7-F7FAFJJJJJJFFFFJJJJJJJFJJF
@K00188:104:HVHN5BBXX:6:1101:2808:1472 1:N:0:GACACAGT+ACTGTGTC
GTCGACTCGATGATGAAGCCGCCGCCATGGTAGTAGACGAGGACCGGGAGCTTCGATTCGCGGTCCTGTGTGGCGGGGAGGTACAGGCGCGCCGAAACAGG
+
AAFFFJJJJFFJJ-AJJJJJJJJFJJ<AFJFJFAFJFJJJJJJJJJFJFAJ7<FFJJJJJJJJFJJ<J7<FFJJJJAJJAJJFJJJJFJJJJJJJJJJJJ
@K00188:104:HVHN5BBXX:6:1101:3721:1472 1:N:0:GACACAGT+ACTGTGTC
GCCTTCACTATTGTTTCCAACAAACTCATTAGTTTCTAACAAGTGCCTCCTCATTCTCTTCTTAATAGCAGCCTGAACATGTATTTGCAATTATTCAGTAT
+
AAFFFJJJJJJJJJJJJJJJJJJJJ<FJJJJJJFJJFJJJJJJAJJJFJJFJJJJJJJJJJJJJJJJJJJJJJFFFJJJFJJFJJFFJJFJJJJJJJJJF
@K00188:104:HVHN5BBXX:6:1101:3863:1472 1:N:0:GACACAGT+ACTGTGTC
CCCTTGTAAACACGTCCAGGAGTTGTTCCAGCACCAATGGATCCCAAGGCCCTGTGACTCTTAGAACCATGGGTCATTAACCCACGTTTAAAATTGTGTCT
+
AAFAFAFJJAJJJJJJF<FJ<F<FJF<FJJJJ<FFAAAJFFFJJJF7JFJJJ<<7J<F7FAJJFJJJJJJJJJAAA7AFJJJJJJJ<FFJJJJJJJFJ<F-
@K00188:104:HVHN5BBXX:6:1101:4107:1472 1:N:0:GACACAGT+ACTGTGTC
GGTTGCATATGTCACAAAACCAAACCCCTTTGACCTTCCGGACACTCTATCAGTTATAAGTTTTGCATCAACAAGCTGGCCGTGCTTAGCAAATGCCACTT
+
AAFFFJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJ
@K00188:104:HVHN5BBXX:6:1101:4290:1472 1:N:0:GACACAGT+ACTGTGTC
GGTGTCTCCGGCGCGAGAAGCAAATCGCCAAGGCTGCCGCCTCTCGAGCACTTCCTGCTCCTCCCCCCTCTAGATCTCCCATCCTCCGACTCCAAATCCTC
+
AAFFFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ7FJJJJ<JFJJJJJF-FF7<JJJJJ7JFJJJFJFJJJFAJJFJJ<FJ<JJFAAFJJJJ
@K00188:104:HVHN5BBXX:6:1101:4858:1472 1:N:0:GACACAGT+ACTGTGTC
GGGTTTTAATGTGTTAGGATCAACCAGGCCAACGACGGGTAGAAAGATTGTCTGGCGTGTAGAAGTTTTTAAAACAACACATTTTCCTTTCCTCACCGAGT
+
AAFFAFJJJJJFJJFFJFJJJJJJJJJ<JJJJJJJJJJJJFJJJJJJJJJJJJJJJFJJJFJFJJJJJFFFJJFJJJFJJJJJJJJJJJJFJJAJFJJFJF-
@K00188:104:HVHN5BBXX:6:1101:4980:1472 1:N:0:GACACAGT+ACTGTGTC
ATGGTTATGTGCGCCGGGGCAGGAGAACTCGCTGTCTGCCAGAGTGAAGAGGCTTCGTAGGTTGGAGCCGACGGGGGAGAAGTCTAATCCTCCCTAGAAAT
+
<AFFFJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJF
@K00188:104:HVHN5BBXX:6:1101:5731:1472 1:N:0:GACACAGT+ACTGTGTC
CCCGTACCCGTTGTTTTCCACGTTCAAGTGAGCGAAGAGCCCTGCAGGGGAGCTGCTGTGTCTGATGAGATTCGATCCGCCGCCGCTCTTCATTTGATCCG
```

fastq

@ID `@K00188:104:HVHN5BBXX:6:1101:1265:1472 1:N:0:GACACAGT+ACTGTGTC`

Secuencia `GTCCTCCGCTTTCCGCTCCTCCCTCGGCGTTCGTCGGAATCGGAGCCGGAGGAGGAGGCGCGATCTGTACGGCTCCGCCCGCTGCGGTGGCTACGGTGGCG`

`+`

CALIDAD `AAAFA7JFF<F<FJFF-AFF-FJ-AAFJFJ-FJ<F<-AAA7JJAJF<AF-7<-A<-AAJJAAAJJFFAJJJJJFJJFJJJJJ<AJF7AFJJFAFF-FJ7-A`

Limpiar los reads
Combinar datos
Ensamblar el transcriptoma
Cuantificar los reads
Anotar la transcriptoma

https://github.com/trinityrnaseq/trinityrnaseq/wiki

QUANTIFICAR

Limpiar los reads: 1_runTrimGalore.sh
Combinar datos: 2_catForTrinity.sh
Ensamblar el transcriptoma: 3_runTrinityBom.sh
Cuantificar los reads: 4_quantTrans.sh
Anotar la transcriptoma: 5_runTrinotate.sh

```
cd /global/scratch/cmt2/transcriptome_raw_data/transcriptomes_Bomarea
mkdir outputdir36bp

trim_galore -o outputdir36bp --paired CSJM003A_S33_L006_R1_001.fastq.gz CSJM003A_S33_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003B_S34_L006_R1_001.fastq.gz CSJM003B_S34_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003C_S35_L006_R1_001.fastq.gz CSJM003C_S35_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003D_S36_L006_R1_001.fastq.gz CSJM003D_S36_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003E_S37_L006_R1_001.fastq.gz CSJM003E_S37_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003F_S38_L006_R1_001.fastq.gz CSJM003F_S38_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003G_S39_L006_R1_001.fastq.gz CSJM003G_S39_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003H_S40_L006_R1_001.fastq.gz CSJM003H_S40_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003I_S41_L006_R1_001.fastq.gz CSJM003I_S41_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003J_S42_L006_R1_001.fastq.gz CSJM003J_S42_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003L_S44_L006_R1_001.fastq.gz CSJM003L_S44_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
trim_galore -o outputdir36bp --paired CSJM003k_S43_L006_R1_001.fastq.gz CSJM003k_S43_L006_R2_001.fastq.gz --retain_unpaired --
    dont_gzip --length 36
```

1_runTrimGalore.sh

```
cd /global/scratch/cmt2/transcriptome_raw_data/transcriptomes_Bomarea
mkdir bom_mult

cat clean/CSJM003A_1.fq clean/CSJM003B_1.fq clean/CSJM003C_1.fq clean/CSJM003D_1.fq clean/CSJM003E_1.fq clean/CSJM003F_1.fq
    clean/CSJM003G_1.fq clean/CSJM003H_1.fq clean/CSJM003I_1.fq clean/CSJM003J_1.fq clean/CSJM003L_1.fq clean/CSJM003k_1.fq >
    bom_mult/bom_mult_1.fq
cat clean/CSJM003A_2.fq clean/CSJM003B_2.fq clean/CSJM003C_2.fq clean/CSJM003D_2.fq clean/CSJM003E_2.fq clean/CSJM003F_2.fq
    clean/CSJM003G_2.fq clean/CSJM003H_2.fq clean/CSJM003I_2.fq clean/CSJM003J_2.fq clean/CSJM003L_2.fq clean/CSJM003k_2.fq >
    bom_mult/bom_mult_2.fq
```

2_catForTrinity.sh

```
cd /global/scratch/cmt2/transcriptome_raw_data/transcriptomes_Bomarea/bom_mult
Trinity --SS_lib_type RF --seqType fq --max_memory 400G --left bom_mult_1.fq --right bom_mult_2.fq --CPU 20 --output trinity
exit
```

3_runTrinityBom.sh

```
cd /global/scratch/cmt2/trans_diff_analysis/
mkdir output
module load bowtie2/2.3.4.1
module load samtools/1.8

trinityrnaseq-Trinity-v2.8.4/util/align_and_estimate_abundance.pl --transcripts Trinity.fasta --est_method RSEM --aln_method
    bowtie2 --output_dir /output --prep_reference --trinity_mode --samples_file replicate_information.txt --seqType fq
```

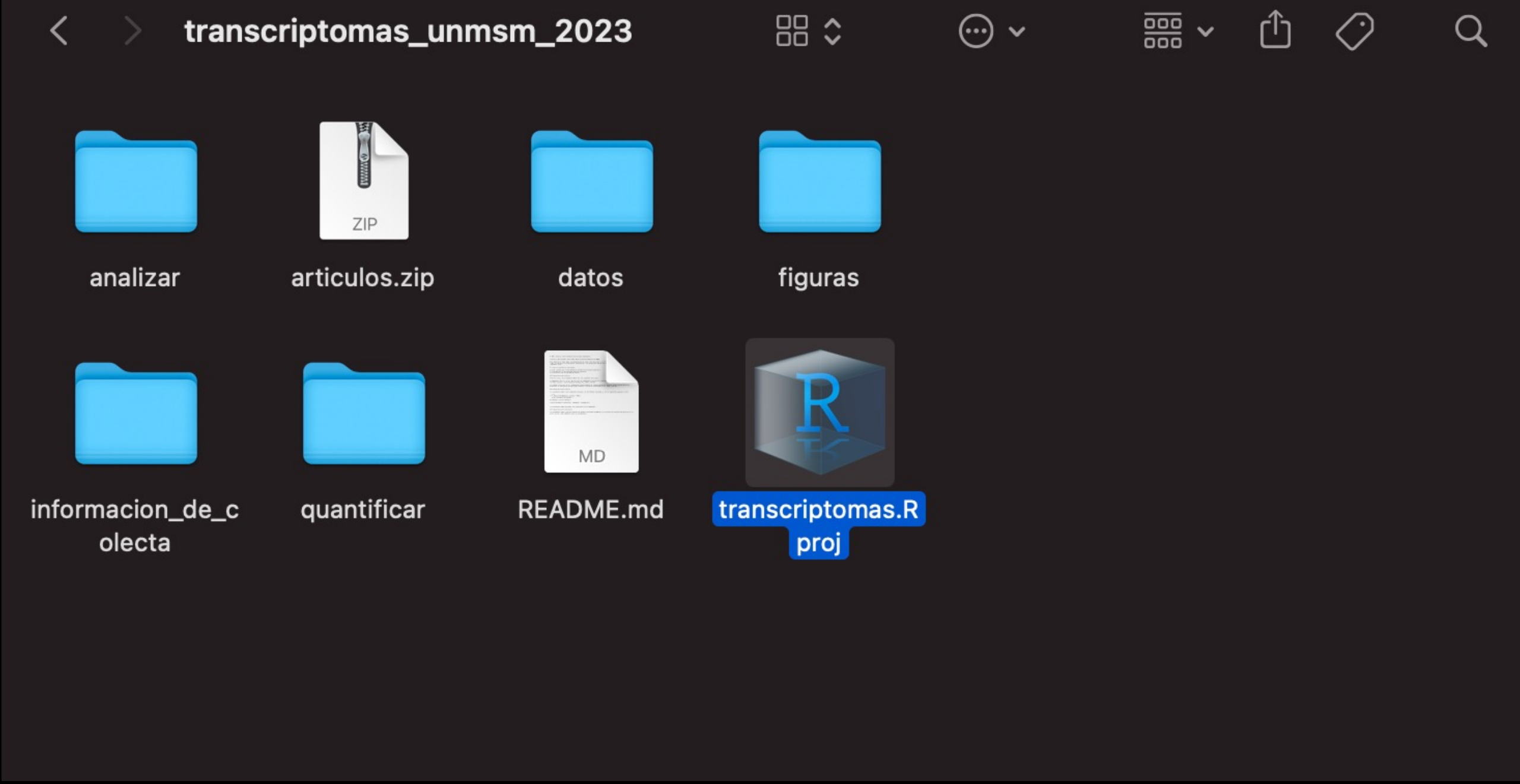
4_quantTrans.sh

```
cd /global/scratch/cmt2/transcriptome_raw_data/transcriptomes_Bomarea/bom_mult/trinity
/global/home/users/cmt2/trinityrnaseq-Trinity-v2.8.4/util/support_scripts/get_Trinity_gene_to_trans_map.pl Trinity.fasta >
    Trinity.fasta.gene_trans_map


module load blast
module load hmmer
module load sqlite

/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/auto/autoTrinotate.pl  --Trinotate_sqlite /global/scratch/cmt2/
    transcriptome_raw_data/transcriptomes_Jesus/allium/trinity/Trinotate.sqlite --transcripts Trinity.fasta --gene_to_trans_map
     Trinity.fasta.gene_trans_map --conf /global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/auto/conf.txt --CPU 10
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/Trinotate Trinotate.sqlite init --gene_trans_map
    Trinity.fasta.gene_trans_map --transcript_fasta Trinity.fasta --transdecoder_pep Trinity.fasta.transdecoder.pep
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/Trinotate Trinotate.sqlite LOAD_swissprot_blastx swissprot.blastx.outfmt6
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/Trinotate Trinotate.sqlite LOAD_swissprot_blastp swissprot.blastp.outfmt6
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/Trinotate Trinotate.sqlite LOAD_pfam TrinotatePFAM.out
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/Trinotate Trinotate.sqlite LOAD_rnammer Trinity.fasta.rnammer.gff
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/Trinotate Trinotate.sqlite LOAD_tmhmm tmhmm.out
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/Trinotate Trinotate.sqlite LOAD_signalp signalp.out
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/Trinotate Trinotate.sqlite report > Trinotate.xls
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/util/extract_GO_assignments_from_Trinotate_xls.pl --Trinotate_xls
    Trinotate.xls -T -I > Trinotate.xls.gene_ontology
/global/home/users/cmt2/Trinotate-Trinotate-v3.1.1/util/annotation_importer/import_transcript_names.pl Trinotate.sqlite
    Trinotate.xls
```

5_runTrinotate.sh

transcript_id	gene_id	length	effective_length	expected_count	TPM	FPKM	IsoPct
TRINITY_DN100002_c0_g1_i1	TRINITY_DN100002_c0_g1	714	517.19	4.00	0.29	0.45	100.00
TRINITY_DN100002_c0_g2_i1	TRINITY_DN100002_c0_g2	270	80.67	0.00	0.00	0.00	0.00
TRINITY_DN100003_c0_g1_i1	TRINITY_DN100003_c0_g1	270	80.67	0.00	0.00	0.00	0.00
TRINITY_DN100004_c0_g1_i1	TRINITY_DN100004_c0_g1	299	106.55	0.00	0.00	0.00	0.00
TRINITY_DN100007_c0_g1_i1	TRINITY_DN100007_c0_g1	640	443.20	0.00	0.00	0.00	0.00
TRINITY_DN100008_c0_g1_i1	TRINITY_DN100008_c0_g1	352	156.78	0.00	0.00	0.00	0.00
TRINITY_DN100011_c0_g1_i1	TRINITY_DN100011_c0_g1	396	199.87	3.00	0.57	0.87	100.00
TRINITY_DN100012_c0_g1_i1	TRINITY_DN100012_c0_g1	291	99.26	0.00	0.00	0.00	0.00
TRINITY_DN100012_c0_g2_i1	TRINITY_DN100012_c0_g2	264	75.53	0.00	0.00	0.00	0.00
TRINITY_DN100013_c0_g1_i1	TRINITY_DN100013_c0_g1	371	175.29	0.00	0.00	0.00	0.00
TRINITY_DN100014_c0_g1_i1	TRINITY_DN100014_c0_g1	228	47.01	0.00	0.00	0.00	0.00
TRINITY_DN100014_c0_g1_i2	TRINITY_DN100014_c0_g1	293	101.07	2.00	0.75	1.15	100.00
TRINITY_DN100015_c0_g1_i1	TRINITY_DN100015_c0_g1	342	147.11	1.00	0.26	0.39	100.00
TRINITY_DN100016_c0_g1_i1	TRINITY_DN100016_c0_g1	226	45.57	0.00	0.00	0.00	0.00
TRINITY_DN100017_c0_g1_i1	TRINITY_DN100017_c0_g1	615	418.21	3.00	0.27	0.42	100.00
TRINITY_DN100017_c0_g1_i2	TRINITY_DN100017_c0_g1	236	52.95	0.00	0.00	0.00	0.00
TRINITY_DN100018_c0_g1_i1	TRINITY_DN100018_c0_g1	710	513.19	2.00	0.15	0.23	100.00
TRINITY_DN100019_c0_g1_i1	TRINITY_DN100019_c0_g1	534	337.25	0.00	0.00	0.00	0.00
TRINITY_DN100019_c0_g2_i1	TRINITY_DN100019_c0_g2	909	712.19	5.00	0.26	0.41	100.00
TRINITY_DN10001_c0_g1_i1	TRINITY_DN10001_c0_g1	480	283.33	0.00	0.00	0.00	0.00
TRINITY_DN100020_c0_g1_i1	TRINITY_DN100020_c0_g1	1062	865.19	0.00	0.00	0.00	0.00
TRINITY_DN100021_c0_g1_i1	TRINITY_DN100021_c0_g1	395	198.88	0.00	0.00	0.00	0.00
TRINITY_DN100025_c0_g1_i1	TRINITY_DN100025_c0_g1	404	207.77	0.00	0.00	0.00	0.00
TRINITY_DN100026_c0_g1_i1	TRINITY_DN100026_c0_g1	357	161.63	0.00	0.00	0.00	0.00
TRINITY_DN100026_c0_g2_i1	TRINITY_DN100026_c0_g2	343	148.08	0.00	0.00	0.00	0.00
TRINITY_DN100029_c0_g1_i1	TRINITY_DN100029_c0_g1	661	464.20	2.00	0.16	0.25	100.00
TRINITY_DN100030_c0_g1_i1	TRINITY_DN100030_c0_g1	469	272.36	0.00	0.00	0.00	0.00
TRINITY_DN100031_c0_g1_i1	TRINITY_DN100031_c0_g1	359	163.58	0.00	0.00	0.00	0.00
TRINITY_DN100031_c0_g2_i1	TRINITY_DN100031_c0_g2	322	128.01	1.00	0.29	0.45	100.00
TRINITY_DN100034_c0_g1_i1	TRINITY_DN100034_c0_g1	269	79.80	0.00	0.00	0.00	0.00

EL PROCESO


transcriptomas_unmsm_2023
analizar
articulos.zip
ZIP
datos
figuras
informacion_de_colecta
quantificar
README.md
MD
transcriptomas.Rproj

EL PROCESO

Cuántos de ustedes han hecho análisis en R?