

Gestión de datos de investigación

**El papel de las bibliotecas biomédicas en el contexto del
acceso abierto**

FECYT-OpenAire

Escuela Nacional de Sanidad. Instituto de Salud Carlos III

14 junio 2019

Entendiendo el contexto



- Gestión de **datos personales y sensibles**
- Legislaciones, regulaciones, protocolos
- Técnicas de **anonimización y pseudoanonimización**
- **Licencias** restrictivas, **control** de accesos



- Crecimiento exponencial, proliferación de colecciones de **big data**
- **Dispersión** en distintos tipos de repositorios y bases de datos, que usan **estrategias diferentes** de agregación, esquemas semánticos, taxonomías, formatos, PIDs...

¿De qué datos hablamos?

Protein

PDB

UniProt:Swiss-Prot

SBGrid

Phenotype

ClinVar

dbGaP

MPD

MorphoBank

Gene Expression

GEO

ArrayExpress

GEMMA

NURSA

LSDB

GeneNetwork

GTex

Nucleotide Sequence

SRA

GigaDB

Morphology

NeuroMorpho.Org

Clinical Trials

ClinicalTrials.gov

CTN

Proteomics Data

PeptideAtlas

ProteomeXchange

YPED

Physiological Signals

PhysioBank

Epigenetic Data

Epigenomics

Data from Papers

PeerJ

Ndar Papers

Omics Data

OmicsDI

Survey Data

FDZ-DZA

Cell Signaling

NSGM

Imaging Data

CVRG

NeuroMorpho.Org

TCIA

Unspecified

LINCS

BioProject

Dryad

La dispersión de los datos de investigación en Biomedicina

Como material suplementario en revistas (ej, [Journal of Immunology](#))

En los “data papers” (ej, [BMC Research Notes](#))

En repositorios temáticos (ej, [Protein Data Bank](#)), institucionales ([Apollo](#)) y generalistas ([Figshare](#))

En agregadores y bases de datos (ej, [Database of Genomic Variants](#))

¿Cuántos datasets de Biomedicina?

bioCADDIE
biomedical and healthCARE Data Discovery Index Ecosystem

Home About Feedback Web API Submit Login

Search for data set Search for repository

Data Types

- Unspecified (1,267,256)
- Protein (569,714)
- Phenotype (210,743)
- Gene Expression (83,448)
- Omics Data (78,201)
- Imaging Data (67,950)
- Morphology (50,356)
- Nucleotide Sequences (43,408)
- Data from Papers (6,279)
- Clinical Trials (5,718)
- Proteomics Data (3,499)
- Epigenetic Data (128)
- Physiological Signals (82)
- Survey Data (21)
- Cell Signaling (10)

Repositories

- CCDC (656,223)
- UniProt:Swiss-Prot (438,182)
- BioProject (226,434)
- Figshare (192,552)
- PDB (131,204)
- Dryad (100,545)
- omicsDI (78,201)
- ArrayExpress (69,996)
- Dataverse (60,303)
- NeuroMorpho.Org (50,356)

Displaying 20 of 2,122,179 results for "*" 20 Per Page

Results by year

Word Cloud

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Search NCBI

NCBI Databases

Literature

Genes

Genetics

Proteins

Genomes

Chemicals

EMBL-EBI

The home for big data in biology

8,7 petabytes

Total volume of data downloaded in 2016.

More about EMBL-EBI's impact in our annual report

Find a tool for your data analysis.

Share your scientific data with the world.

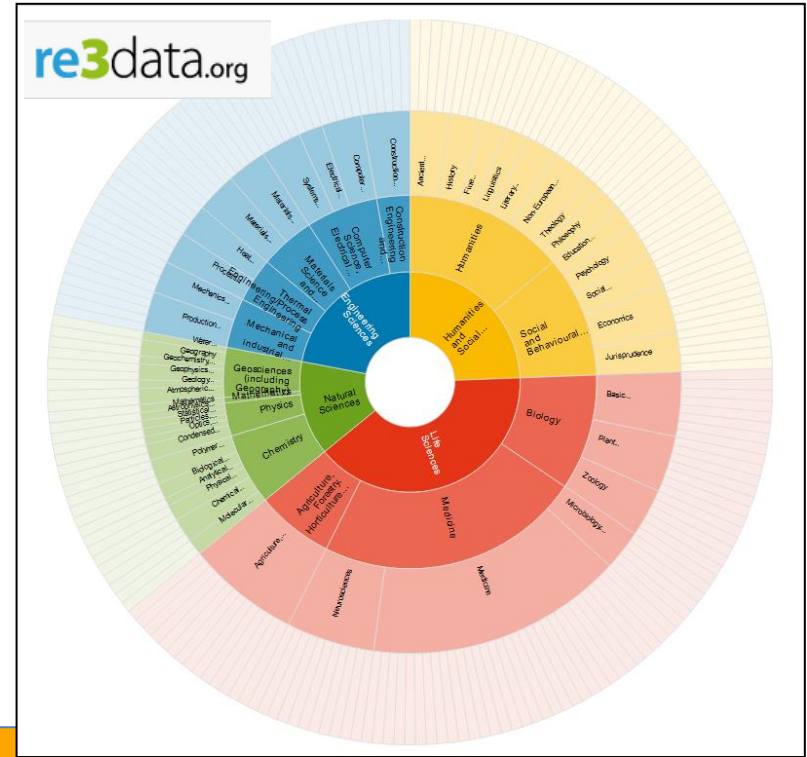
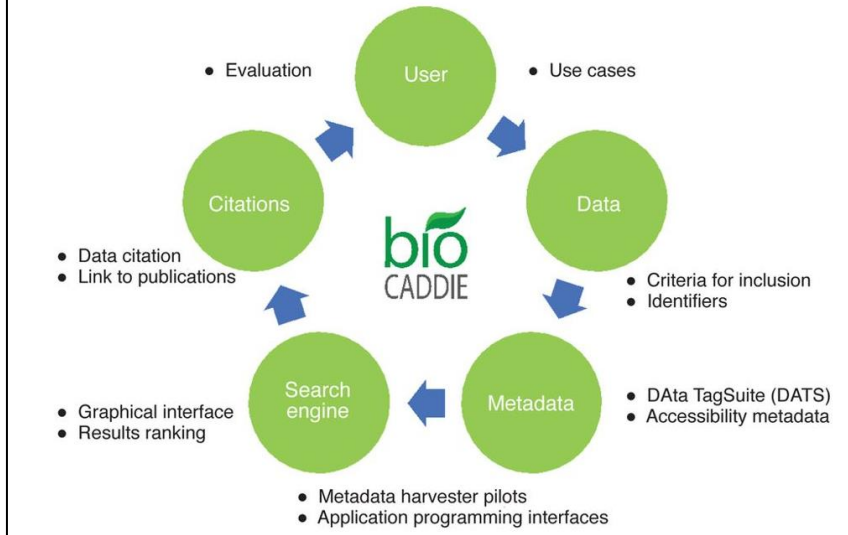
Find a gene, protein or chemical

Example searches: [blast](#) [keratin](#) [bfl1](#)

Si te pierdes buscando...

Figure 2 : Community input to the Data Discovery Index Consortium.

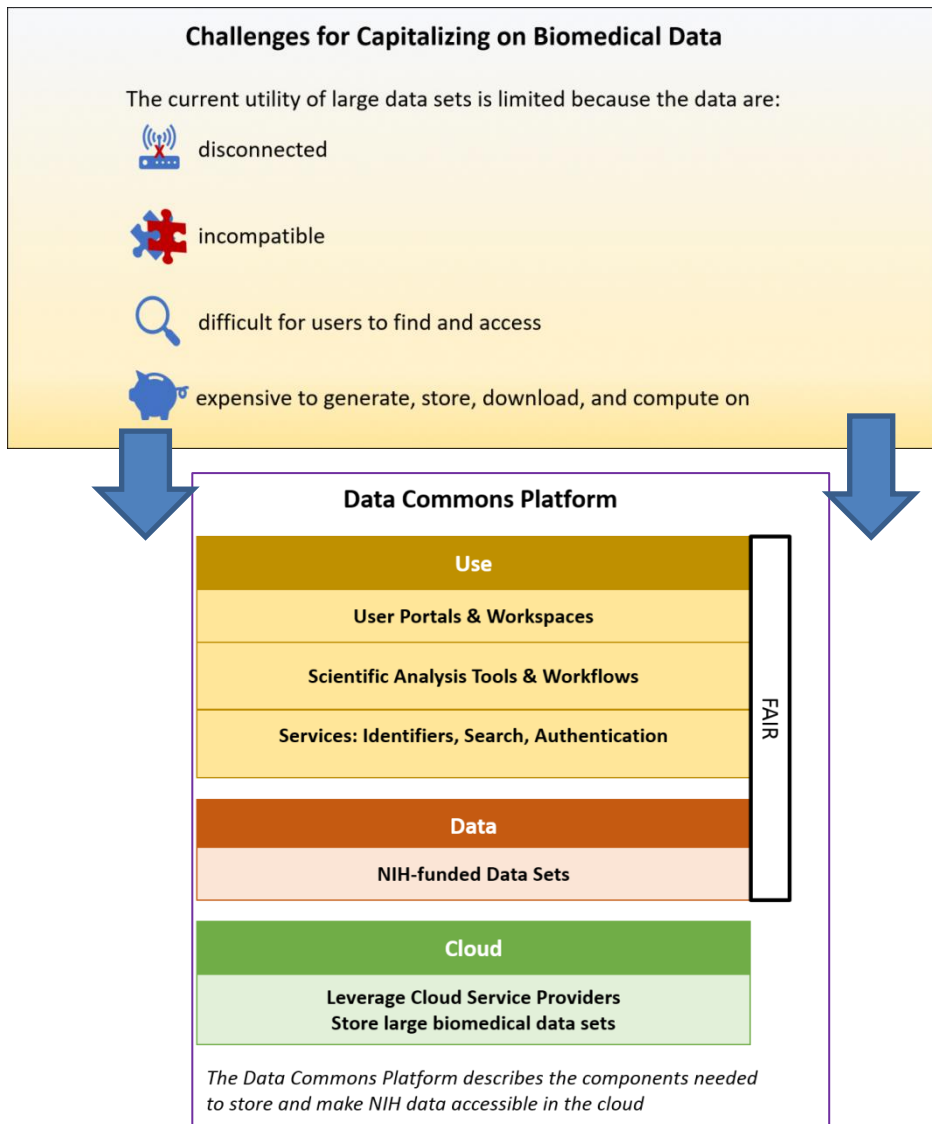
From: Finding useful data across multiple biomedical data repositories using DataMed



<https://biocaddie.org/>

The screenshot shows the **bio.tools** website. At the top left is the **elivr bio.tools** logo. A cookie consent banner reads: "This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here." Below the banner, the site title **bio.tools** is displayed, followed by the tagline: "Essential scientific and technical information about software tools, databases and services for bioinformatics and the life sciences." A search bar with the placeholder text "Search bio.tools" and a magnifying glass icon is present. At the bottom, there are navigation buttons for "Standards", "Databases", "Policies", and "Collections".

Grandes proyectos en marcha



- **NIH Data Commons: Its main objective is to develop the ability to make data FAIR through use of a shared virtual space to store and work with biomedical research data and analytical tools.**
- The NIH Data Commons will leverage currently available cloud-computing environments in a flexible and scalable way, aiming to increase the value of NIH-supported data by democratizing access and use of data and analytical tools and allowing multiple datasets to be queried together.
- <https://commonfund.nih.gov/commons>

POLÍTICA DE DATOS DE DIGITAL.CSIC: INICIOS, DESARROLLO, MANDATO



BUENAS PRÁCTICAS, AYUDA Y RECURSOS



POLÍTICAS Y
SERVICIOS DE
DIGITAL.CSIC



ACCESO
ABIERTO



DATOS DE
INVESTIGACIÓN



COPYRIGHT



MANDATOS
DE ACCESO
ABIERTO



SUBIR
TRABAJOS A
DIGITAL.CSIC

Gestión de datos en los servicios de DIGITAL.CSIC

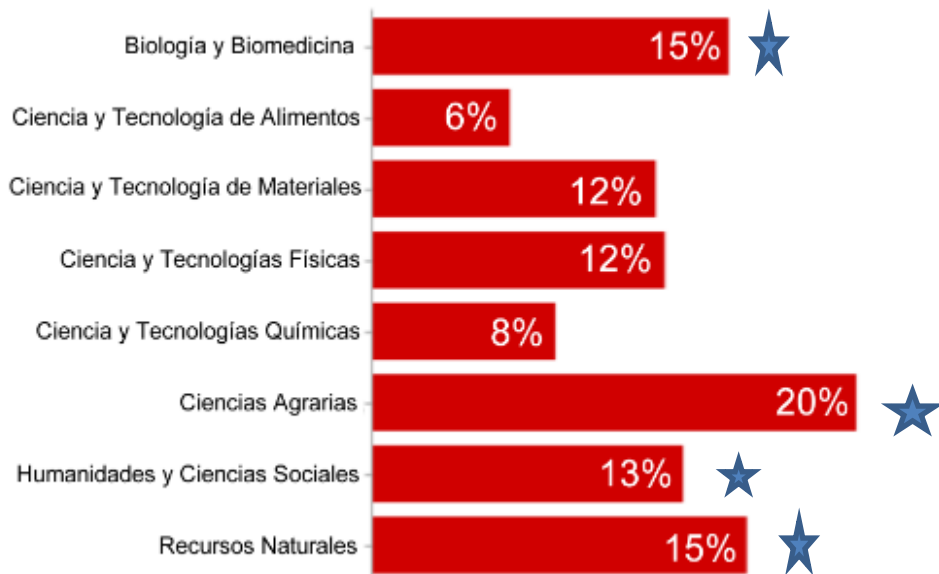


- **Desarrollo de planes de gestión de datos**
- **Ayuda en el cumplimiento de los datos FAIR**
- **Asignación de DOI a datasets, software de investigación, preprints y cuadernos de laboratorio**
- **Promoción de uso de esquemas de metadatos y vocabularios controlados**
- **Gestión, descripción y publicación de datos**
- **Promoción de licencias abiertas (Creative Commons, Open Data Commons..)**
- **Asesoramiento en la gestión de datos personales/ sensibles**
- **Formación a bibliotecas e investigadores**
- **Agregación de recursos en infraestructuras abiertas de investigación (p.e, EOSC)**

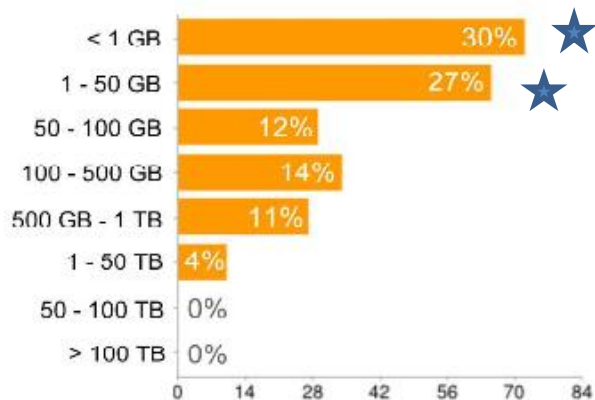
Encuesta DIGITAL.CSIC a los investigadores CSIC (2013)

(2013)

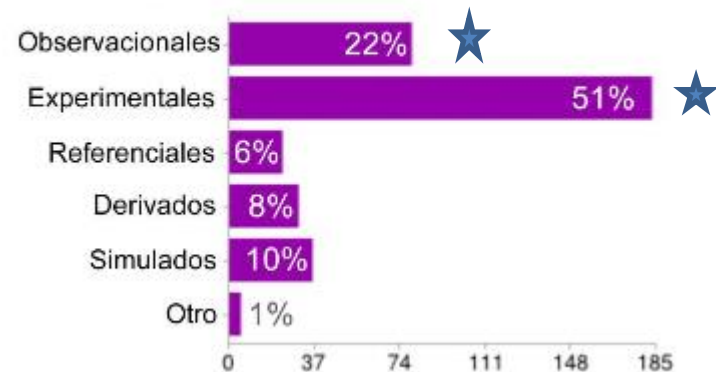
Área de investigación



¿Qué tamaño de datos digitales está generando/manteniendo en la actualidad?



¿Cómo son los datos puros que produce en sus proyectos de investigación?



¿Qué formato tienen los datos?



Arranque y evolución

Política de gestión de datos publicada en 2014

Criterios de aceptación de datos

Aceptación, preservación y conversión de formatos

Actividades formativas presenciales y online para bibliotecas e investigadores CSIC desde 2015

Directrices generales y multidisciplinarias

Plantilla para describir datos y recomendaciones

Gestión de copyright y licencias de uso

Data Seal of Approval en 2015

Asignación de DOIs desde 2016

Apoyo en la preparación de planes de gestión de datos desde 2017

Revisión importante en marcha

Mandato de acceso abierto y datos en abril 2019

Revisión de la política en curso

[Recursos de interés](#)

Primeros datasets en 2010 sobre niveles de sequía en el mundo

11.738 items en DIGITAL.CSIC son datasets (en dc.type), de los que algo más de 8.300 están en acceso abierto

En colecciones de institutos de Ciencias Agrarias, Recursos Naturales, Humanidades, Biología/Biomedicina, Ciencias Marinas principalmente

Datasets que están asociados a artículos de revistas con una política de datos y/o a proyectos H2020

Datasets que no están necesariamente vinculados como “underlying data” a publicaciones

Datasets, en su mayoría imágenes/dibujos/videos, que son resultado de proyectos de investigación de muy largo recorrido

¿Qué tipologías en DIGITAL.CSIC pueden tener tratamiento de datos de investigación?

Tipologías	Nº registros
Artículo	108.295
Comunicación de congreso	16.730
Dataset	11.737
Capítulo de libro	8.696
Póster	6.524
Tesis	3.677
Trabajo de divulgación	2.410
Informe Técnico	2.303
Documento de trabajo	1.986
Patente	1.962
Solicitud de patente	1.937
Libro	1.559
Presentación	1.513
Partitura	887
Pre-print	775
Material didáctico	705
Proyecto fin de carrera	523
Reseña bibliográfica	465
Tesina	344
Video	261
Revista	261
Imagen	231
Trabajo fin de Máster	221
Índice	146
Entrada de blog	135
Sitio web	111
Blog	78
Traducción de patente	66
Planimetría	62
Software	47
Mapa	46
Circuito Integrado	43
Modelo de utilidad	14
Revisión	8
Comentario	5
Journal Article	3
Imagen 3-D	3
Manuscrito	2

Además de los datos observacionales, experimentales, puros los siguientes:

- Video científico
- Imágenes de carácter científico
- Mapas
- Resultados de encuestas
- Resultados de trabajo de campo
- Entrevistas
- Datos de excavaciones arqueológicas
- ¿Qué más?

¿Dónde están los datasets de Biomedicina del CSIC?

The screenshot shows the bioCADDIE search interface. The search bar contains 'csic'. The results page displays 25 results for 'csic'. The 'Repositories' section is circled in blue and lists: ArrayExpress (15), Figshare (5), OmicsDI (3), and BioProject (2). The 'Data Types' section lists: Gene Expression (15), Unspecified (7), and Omics Data (3). The 'Accessability' section lists: Download (25). The 'Authorization' section lists: None (25). The main results area shows three entries from 'Instituto de Biología Evolutiva (CSIC-UPF)'. The first entry has ID: EGA000000000328 and Description: EGA Organisation EGA000000000328. The second entry has ID: EGA000000000331 and Description: EGA Organisation EGA000000000331. The third entry has ID: PRJNA54033 and Description: Estacion Experimental del Zaidin (CSIC) Granada, Spain. The 'Word Cloud' section contains terms like 'generar', 'compartir', 'científicos', 'bibliométricos', 'cchs', 'aneca', 'agencias', 'científica', 'cneai', and 'actualidad'.

A grid of logos for various scientific data repositories and databases, including: Cambridge Structural Database (CSD), NIST Data Gateway, Protein Data Bank (PDB), Electron Microscopy Data Bank (EMDB), UniProt, British Oceanographic Data Center, Biolisis, National Geophysical Data Center (NOAA), Academic Seismic Portal (ASP) at UTIG, Marine Geoscience Data System, Ocean Data Portal, Gene Expression Omnibus (GEO), Spanish Virtual Observatory (SVO), Virtual Observatory, Centre de Données Astronomiques de Strasbourg, Simbad Astronomical Database, Inorganic Crystal Structure Database, GenBank, Calar Alto Archive, data.NASA, esa Earth Online, National Center for Biotechnology Information (NCBI), and Global Diversity Information Facility (GBIF).

DIGITAL.CSIC
CIENCIA ABIERTA



Principales motivaciones para depositar datos de investigación en DIGITAL.CSIC

Publicación de
datos de
investigación

Asignación de DOIs

Cumplimiento de
políticas de datos
de revistas

Cumplimiento de
políticas de
agencias
financiadoras

Almacenamiento y
descripción de
datasets según
estándares

Vías de carga de datos en DIGITAL.CSIC



CARGAS AUTOMATIZADAS

- Gestionadas por la Oficina Técnica del repositorio
- Importación masiva de metadatos y ficheros
- Labor previa de organización de los objetos digitales, red denominación de ficheros y de mapeos de metadatos, habitualmente larga
- Esencial la colaboración de los productores de datos

CARGAS MANUALES

- Por la Oficina Técnica del repositorio, bibliotecas CSIC e investigadores
- En general, son datasets (1 o más ficheros) en 1 solo item
- No hay límite en la dimensión de los ficheros, suelen ocupar pocos kB/GB

Descripción de datasets según estándares

Datasets: plantilla normalizada para la descripción de registros en DIGITAL.CSIC
Oficina Técnica de DIGITAL.CSIC
13/02/2019

La descripción recomendada para datos de investigación debe ser en inglés si es posible e incluye los siguientes aspectos:

- Contexto, descripción del proyecto y propósito de la investigación, metodología utilizada
- Naturaleza de los datos, historia de los datos, contenido y estructura, terminología, software, fecha de creación y fechas de modificación, versiones, responsables y participantes
- Formatos de ficheros, estructura y nomenclatura de los ficheros
- Aspectos legales, políticas de acceso y seguridad

La descripción de los datos debe ser lo suficientemente completa como para poder responder las siguientes preguntas, si no, su utilidad para otros investigadores es relativa:

- ¿Quiénes han producido los datos?
- ¿Es el título lo suficientemente específico?
- ¿Por qué han sido creados los datos?
- ¿Qué limitaciones tienen los datos (por ejemplo, datos confidenciales han sido eliminados)?
- ¿Cómo deben interpretarse los datos?
- ¿Hay lagunas en los datos o dan una visión completa del tema estudiado?
- ¿Qué procesos han generado los datos?
- ¿Qué miden los datos en las columnas de los ficheros?
- ¿Qué software es necesario para poder leer los datos?
- ¿Cómo deben citarse los datos?
- ¿Pueden reutilizarse los datos? ¿Qué licencia de uso tienen asignada?
- ¿Existen más versiones de los datos? ¿Dónde?
- ¿Se han definido los términos técnicos y acrónimos a los que hacen referencia los datos?
- ¿Se han cualificado los parámetros geográficos y cronológicos de los datos?
- ¿Las palabras clave son suficientemente específicas a los datos? ¿Se basan en algún tesoro?
- ¿Cómo se llama el proyecto de investigación en que se encuadraron los datos?
- ¿Quién ha financiado la producción y la gestión de los datos?

Además, la política de datos de DIGITAL.CSIC recomienda a sus autores la generación de un fichero txt `read.me` con más información.

Se recomienda ver [Buenas prácticas y política de datos de investigación de DIGITAL.CSIC](#) antes de describir y depositar un conjunto de datos en el repositorio.

AUTORES

DESCRIPTOR	METADATO DUBLIN CORE	CUALIFICADOR	CARÁCTER
AUTOR	dc.contributor	Author	Obligatorio

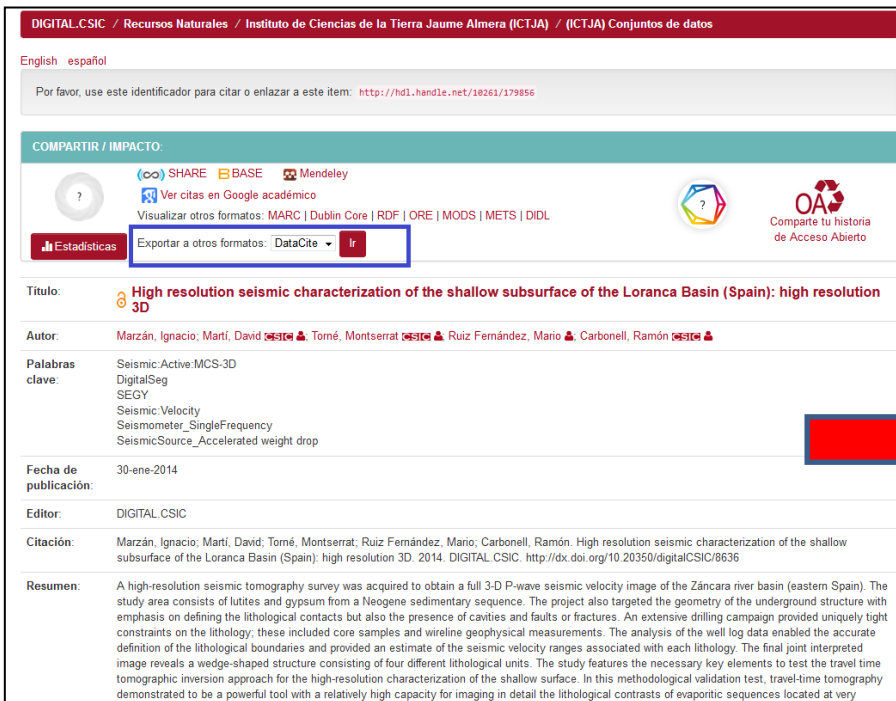
Nombres de los autores de los datos.

Deben citarse en campos independientes todos los autores en el orden en que aparecen. En la generación y gestión de datos científicos pueden estar involucrados distintos tipos de autores y todos deben ser reconocidos en el registro.

- La plantilla se basa en el esquema de Dublin Core Cualificado con extensiones de DataCite (por ejemplo, Vocabulario de Tipos de Contribución) y formato estándar de citación de FORCE11
- Marca un mínimo común denominador para la descripción de datasets de cualquier disciplina

<http://digital.csic.es/handle/10261/81323>

Exportación en formato DataCite



DIGITAL.CSIC / Recursos Naturales / Instituto de Ciencias de la Tierra Jaume Almera (ICTJA) / (ICTJA) Conjuntos de datos

English español

Por favor, use este identificador para citar o enlazar a este ítem: <http://hdl.handle.net/10261/179856>

COMPARTIR / IMPACTO:

Visualizar otros formatos: MARC | Dublin Core | RDF | ORE | MODS | METS | DIDL

Exportar a otros formatos: DataCite

Título: High resolution seismic characterization of the shallow subsurface of the Loranca Basin (Spain): high resolution 3D

Autor: Marzán, Ignacio; Martí, David; Torné, Montserrat; Ruiz Fernández, Mario; Carbonell, Ramón

Palabras clave: Seismic.Active:MCS-3D; DigitalSeg; SEGY; Seismic.Velocity; Seismometer_SingleFrequency; SeismicSource_Accelerated weight drop

Fecha de publicación: 30-ene-2014

Editor: DIGITAL.CSIC

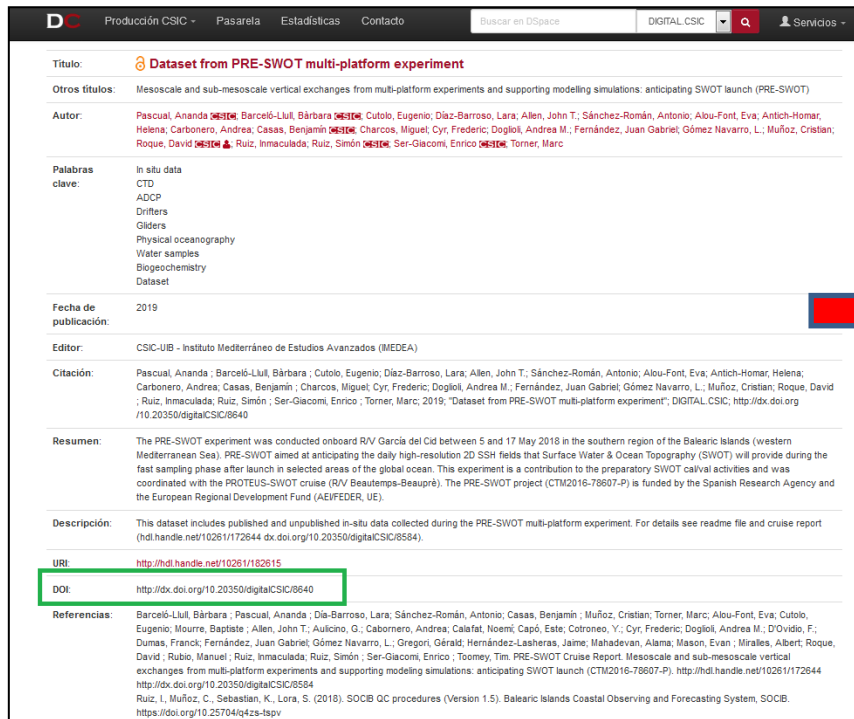
Citación: Marzán, Ignacio; Martí, David; Torné, Montserrat; Ruiz Fernández, Mario; Carbonell, Ramón. High resolution seismic characterization of the shallow subsurface of the Loranca Basin (Spain): high resolution 3D. 2014. DIGITAL.CSIC. <http://dx.doi.org/10.20350/digitalCSIC/8636>

Resumen: A high-resolution seismic tomography survey was acquired to obtain a full 3-D P-wave seismic velocity image of the Záncara river basin (eastern Spain). The study area consists of lutites and gypsum from a Neogene sedimentary sequence. The project also targeted the geometry of the underground structure with emphasis on defining the lithological contacts but also the presence of cavities and faults or fractures. An extensive drilling campaign provided uniquely tight constraints on the lithology; these included core samples and wireline geophysical measurements. The analysis of the well log data enabled the accurate definition of the lithological boundaries and provided an estimate of the seismic velocity ranges associated with each lithology. The final joint interpreted image reveals a wedge-shaped structure consisting of four different lithological units. The study features the necessary key elements to test the travel time tomographic inversion approach for the high-resolution characterization of the shallow surface. In this methodological validation test, travel-time tomography demonstrated to be a powerful tool with a relatively high capacity for imaging in detail the lithological contrasts of evaporitic sequences located at very

Por defecto, los registros bibliográficos de DIGITAL.CSIC se generan sobre el esquema Dublin Core Cualificado

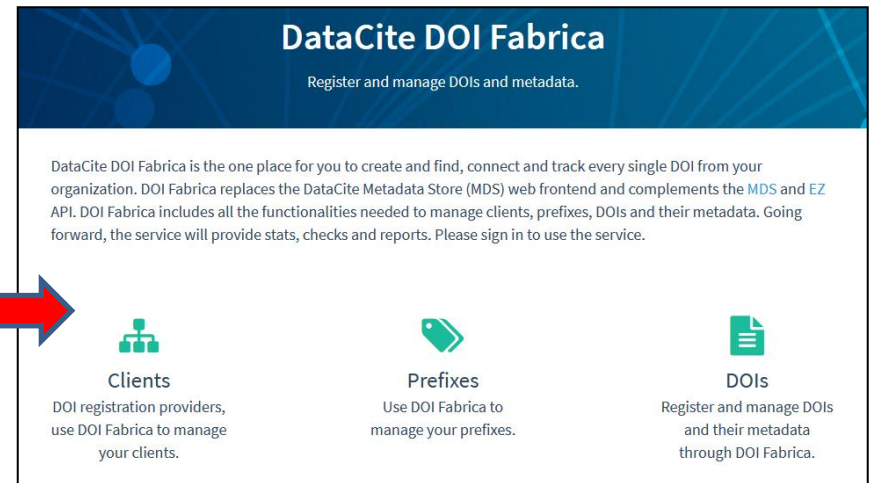
```
<?xml version='1.0' encoding='utf-8'?>
<resource xmlns="http://datacite.org/schema/kernel-4"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://datacite.org/schema/kernel-4
http://schema.datacite.org/meta/kernel-4.1/metadata.xsd">
<creators>
<creator>
<creatorName>Marzán, Ignacio</creatorName>
<creator>
<creatorName>Martí, David</creatorName>
</creator>
<creator>
<creatorName>Torné, Montserrat</creatorName>
</creator>
<creator>
<creatorName>Ruiz Fernández, Mario</creatorName>
</creator>
<creator>
<creatorName>Carbonell, Ramón</creatorName>
</creator>
</creators>
<publicationYear>2014</publicationYear>
<dates>
<date dateType="Issued">2014-01-30</date>
</dates>
<alternateIdentifiers>
<alternateIdentifier
alternateIdentifierType="Handle">http://hdl.handle.net/10261/179856</alternateIdentifier
>
</alternateIdentifiers>
<descriptions>
<description descriptionType="Other">The data acquisition contract was awarded to
ENRESA and took place in January 2014. Seismic data was successfully collected in the
Záncara river basin (Cuenca, Spain). This is a high-resolution seismic tomography survey to
obtain a full 3-D P-wave seismic velocity image of the studied area. A regular and dense grid
of 676 shots and 1200 receivers was used to image a 500 m x 500 m area of the shallow
surface. A 240-channel system and a seismic source, consisting of an accelerated weight
```

Asignación de DOIs a través de DataCite



The screenshot shows a dataset page on the Digital.CSIC website. The dataset title is "Dataset from PRE-SWOT multi-platform experiment". The DOI field is highlighted with a red box and a red arrow pointing to the right.

Título:	Dataset from PRE-SWOT multi-platform experiment
Otros títulos:	Mesoscale and sub-mesoscale vertical exchanges from multi-platform experiments and supporting modeling simulations: anticipating SWOT launch (PRE-SWOT)
Autor:	Pascual, Ananda ORCID ; Barceló-Luigi, Bárbara ORCID ; Cutolo, Eugenio; Díaz-Barroso, Lara; Allen, John T.; Sánchez-Román, Antonio; Alou-Font, Eva; Antich-Homar, Helena; Carbonero, Andrea; Casas, Benjamín ORCID ; Charcos, Miguel; Cyr, Frederic; Doglioli, Andrea M.; Fernández, Juan Gabriel; Gómez Navarro, L.; Muñoz, Cristian; Roque, David ORCID ; Ruiz, Inmaculada; Ruiz, Simón ORCID ; Ser-Giacomi, Enrico ORCID ; Torner, Marc
Palabras clave:	In situ data CTD ADCP Drifters Glanders Physical oceanography Water samples Biogeochemistry Dataset
Fecha de publicación:	2019
Editor:	CSIC-UIB - Instituto Mediterráneo de Estudios Avanzados (IMEDEA)
Citación:	Pascual, Ananda ; Barceló-Luigi, Bárbara ; Cutolo, Eugenio; Díaz-Barroso, Lara; Allen, John T.; Sánchez-Román, Antonio; Alou-Font, Eva; Antich-Homar, Helena; Carbonero, Andrea; Casas, Benjamín ; Charcos, Miguel; Cyr, Frederic; Doglioli, Andrea M.; Fernández, Juan Gabriel; Gómez Navarro, L.; Muñoz, Cristian; Roque, David ; Ruiz, Inmaculada; Ruiz, Simón ; Ser-Giacomi, Enrico ; Torner, Marc; 2019, "Dataset from PRE-SWOT multi-platform experiment"; DIGITAL.CSIC; http://dx.doi.org/10.20350/digitalCSIC/8640
Resumen:	The PRE-SWOT experiment was conducted onboard R/V García del Cid between 5 and 17 May 2018 in the southern region of the Balearic Islands (western Mediterranean Sea). PRE-SWOT aimed at anticipating the daily high-resolution 2D SSH fields that Surface Water & Ocean Topography (SWOT) will provide during the fast sampling phase after launch in selected areas of the global ocean. This experiment is a contribution to the preparatory SWOT cal/val activities and was coordinated with the PROTEUS-SWOT cruise (R/V Beaufort-Beaupré). The PRE-SWOT project (CTM2016-78607-P) is funded by the Spanish Research Agency and the European Regional Development Fund (AEI/FEDER, UE).
Descripción:	This dataset includes published and unpublished in-situ data collected during the PRE-SWOT multi-platform experiment. For details see readme file and cruise report (hdl.handle.net/10261/172644 dx.doi.org/10.20350/digitalCSIC/8584).
URI:	http://hdl.handle.net/10261/182615
DOI:	http://dx.doi.org/10.20350/digitalCSIC/8640
Referencias:	Barceló-Luigi, Bárbara ; Pascual, Ananda ; Díaz-Barroso, Lara; Sánchez-Román, Antonio; Casas, Benjamín ; Muñoz, Cristian; Torner, Marc; Alou-Font, Eva; Cutolo, Eugenio; Mourou, Baptiste; Allen, John T.; Auslino, G.; Carbonero, Andrea; Calafat, Noemi; Capó, Ester; Cotronero, Y.; Cyr, Frederic; Doglioli, Andrea M.; D'Ovidio, F.; Dumas, Franck; Fernández, Juan Gabriel; Gómez Navarro, L.; Gregori, Gérald; Hernández-Lasheras, Jaime; Mahadevan, Alama; Mason, Evan; Miralles, Albert; Roque, David ; Rubio, Manuel ; Ruiz, Inmaculada; Ruiz, Simón ; Ser-Giacomi, Enrico ; Toomey, Tim. PRE-SWOT Cruise Report. Mesoscale and sub-mesoscale vertical exchanges from multi-platform experiments and supporting modeling simulations: anticipating SWOT launch (CTM2016-78607-P). http://hdl.handle.net/10261/172644 http://dx.doi.org/10.20350/digitalCSIC/8584 Ruiz, I., Muñoz, C., Sebastian, K., Lora, S. (2018). SOCB QC procedures (Version 1.5). Balearic Islands Coastal Observing and Forecasting System, SOCB. https://doi.org/10.25704/ig4z8-1spv



The screenshot shows the DataCite DOI Fabrica website. The main heading is "DataCite DOI Fabrica" with the subtext "Register and manage DOIs and metadata." Below this, there are three icons representing "Clients", "Prefixes", and "DOIs".

- Clients:** DOI registration providers, use DOI Fabrica to manage your clients.
- Prefixes:** Use DOI Fabrica to manage your prefixes.
- DOIs:** Register and manage DOIs and their metadata through DOI Fabrica.

Con la asignación del DOI se pueden explotar diversos servicios de valor añadido, como:

- Agregación en el buscador DataCite
- DataCite Event Data service
- Alimentación de ORCID
- Exportación en diversos estilos

La creación del DOI es posterior a la subida del dataset en DIGITAL.CSIC

Se asigna DOI a los items que llevan determinados valores en dc.type

Licencias en datasets de DIGITAL.CSIC

Licencias Creative Commons

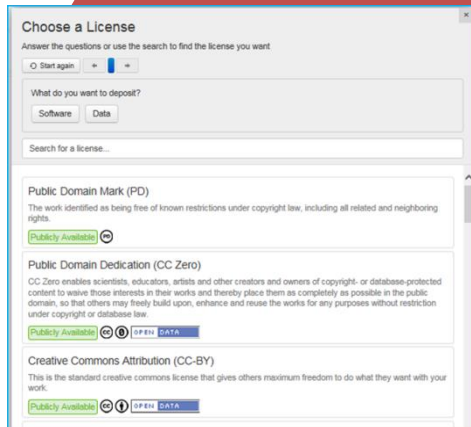
- **CC BY 4.0**
- CC BY NC SA
- CC BY SA

Licencias Open Data Commons

- Para datasets de estructura compleja
- [Ejemplo](#)

Licencia estándar de DIGITAL.CSIC

- Acceso gratuito a los datasets para uso personal/investigación



Centro de Biología Molecular Severo

Ochoa

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted.

DC Producción CSIC Pasarela Estadísticas Contacto Buscar en DSpace DIGITAL CSIC Servicios

Título: 2018 CIC Samples KCNQ2 exon 7 partial duplication fastq files

Autor: Lazo, Pedro A.; Sisti, A.; García-Hernández, Juan L.; Gómez-Puertas, Paulino; Marcos-Alcalde, Iñigo; Arjona, César; Villarreal, Álvaro; González-Sarmiento, Rogelio; Fons, Carmen

Palabras clave: Dystonia
Epilepsy
Neurodevelopmental delay
KCNQ2
CNV

Fecha de publicación: sep-2018

Editor: DIGITAL CSIC

Citación: Lazo, Pedro A.; García-Hernández, Juan L.; Gómez Puertas, Paulino; Marcos-Alcalde, Iñigo; Arjona, César; Villarreal, Álvaro; González-Sarmiento, Rogelio; Fons, Carmen; 2018, "2018 CIC Samples KCNQ2 exon 7 partial duplication fastq files [Dataset]", DIGITAL CSIC; <http://dx.doi.org/10.20350/digitalCSIC/8569>

Descripción: Complex neurodevelopmental syndromes in newborns frequently have an unknown etiology, in which genetic factors play a pathogenic role. We have studied a family with an affected child presenting since birth epileptic-like crisis combined with severe neuromotor and developmental delay, dystonia, and hyper-excitability of unknown origin. Genomic studies by Whole-exome sequencing (WES) were performed in the four members of the family. In the patient, WES detected a de novo in-frame duplication of thirty-three nucleotides within exon seven of KCNQ2 (potassium II-channel). This insertion results in the duplication of the first twelve aminoacids of the calmodulin binding site I, which impairs its regulation by calcium. The dataset includes the WES raw sequence reads (fastq format) of the four family members: mother, father, daughter and son (patient).

URI: <http://hdl.handle.net/10261/170405>

DOI: <http://dx.doi.org/10.20350/digitalCSIC/8569>

Referencias: Lazo, Pedro A.; García-Hernández, Juan L.; Gómez Puertas, Paulino; Marcos-Alcalde, Iñigo; Arjona, César; Villarreal, Álvaro; González-Sarmiento, Rogelio; Fons, Carmen. Phenotypic cooperation of a KCNQ2 exon 7 partial duplication and compound copy number variations in genes associated to a severe epileptic and neurodevelopmental delay.

Aparece en las colecciones: (CBM) Conjuntos de datos

Ficheros en este ítem:

Fichero	Descripción	Tamaño	Formato	
2018_CIC_Sample_MOTHER_1.fastq.gz		4,37 GB	Unknown	Visualizar/Abrir
2018_CIC_Sample_MOTHER_2.fastq.gz		4,34 GB	Unknown	Visualizar/Abrir
2018_CIC_Sample_FATHER_1.fastq.gz		4,16 GB	Unknown	Visualizar/Abrir



NCBI GEO Gene Expression Omnibus

HOME SEARCH SITE MAP GEO Publications FAQ MIAME Email GEO

NCBI > GEO > Accession Display Not logged in | Login

Scope: Self Format: HTML Amount: Quick GEO accession: GSE122584 GO

Series GSE122584 Query DataSets for GSE122584

Status Public on May 22, 2019

Title Phenotypic cooperation of a KCNQ2 exon 7 partial duplication and compound copy number variations in genes associated to a severe epileptic and neurodevelopmental delay

Organism [Homo sapiens](#)

Experiment type Genome variation profiling by SNP array

Summary A 6-year-old boy, second son of healthy parents affected with epileptic encephalopathy of neonatal onset. Pregnancy with gestational diabetes controlled with diet. Delivery was uneventful. Since 48 hours of life, he presented episodes of cyanosis, generalized hypertonia, and tonic asymmetric postures followed by apnea. Video-EEG at 5 days of life showed bilateral and asynchronous spike-and-wave. Seizures were refractory to phenobarbital but were controlled with phenytoin. Since 3 months of age, he presented with startle episodes without EEG correlate related to the wake and sleep transition together with dystonic postures. One month later, epileptic spasms without hypsarrhythmia were observed. No response to levetiracetam and valproic acid but stopped after lacosamide (LCS) treatment was initiated (11 months of age). Metabolic tests (including CSF studies), karyotype, and brain MRI were normal. Epileptic encephalopathy gene panel was negative including hyperekplexia related genes. Progressive increase of TSH was treated with oral L-t4. At 20 months he presented recurrence of spasms that were controlled with LCS dose adjustment. V-EEG: abnormal background activity and paroxysmal multifocal discharges that showed activation and generalization during sleep. After several years of good seizure control, LCS was discontinued. He suffered from decompensation of seizures with febrile viral infection and at 5 years of age he presented startle episodes associated to auditory stimulus without EEG correlate, but during sleep, he presented epileptic spasms with EEG correlate. Treatment with vigabatrine was initiated with a clinical and EEG improvement. At physical exam he presents severe axial hypotonia without head control. Hypertonia of 4 limbs with dystonic movements of upper limbs. No hand use and is not able to sit or crawl. No development of expressive language. GMFCS: V. No development of expressive language. Whole exome sequencing was performed and partial exon 7 duplication, heterozygous variant, in KCNQ2 was found. aCGH study detected several CNV in genes associated to neural functions (losses: LRRCS5, PCDH9, NALCN, RYR3, ELAVL2, CDH13, ATP1A2, SLC17A5, ANO3. Gains : PCDH19, TENM1, EFNA5).


Overall design A high-density CytoScan® HD Array (Thermo Fisher Scientific, In), which includes 2.67 million markers for copy number (CN) analysis, 750,000 SNP probes and 1.9 million non-polymorphic probes for comprehensive whole-genome coverage was used. Following hybridization, the laser scanner (GeneChip® 3000 Scanner) was used for scanning the arrays, and the images were

Instituto de Bioquímica Vegetal y Fotosíntesis

Producción CSIC - Pasarela Estadísticas Contacto

Buscar en DSpace DIGITAL.CSIC Servicios

Título: **Elements of the heterocyst-specific transcriptome unraveled by co-expression analysis in Nostoc sp. PCC 7120 [Dataset]**

Autor: Brenes-Álvarez, Manuel; Mitschke, Jan; Olmedo-Verd, Elvira; Georg, Jens; Hess, Wolfgang R.; Vioque, Agustín ; Muro-Pastor, Alicia M.

Fecha de publicación: 2019

Citación: Manuel Brenes-Álvarez, Jan Mitschke, Elvira Olmedo-Verd, Jens Georg, Wolfgang R. Hess, Agustín Vioque and Alicia M. Muro-Pastor. 2019. Elements of the heterocyst-specific transcriptome unraveled by co-expression analysis in Nostoc sp. PCC 7120 [Dataset]. DIGITAL.CSIC, <http://dx.doi.org/10.20350/digitalCSIC/8631>

Descripción: Supplementary data file 1 "genomeplot_WT_hetR_N_chromosome.pdf". Overview (chromosome) combining log2-normalized expression values (left y axis) from the microarray analysis of cultures of wild type and hetR strain subjected to nitrogen deprivation for 0-6-12-24 hours as indicated by coloured lines, with the reads numbers from the RNA-Seq dataset in (Mitschke, J., Vioque, A., Haas, F., Hess, W.R., and Muro-Pastor, A.M. (2011). Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in Anabaena sp. PCC7120. Proc. Natl. Acad. Sci. USA 108, 20130-20135) (right y axis).

Supplementary data file 2 "genomeplot_WT_hetR_N_plasmids.pdf". Overview (six plasmids combined) combining log2-normalized expression values (left y axis) from the microarray analysis of cultures of wild type and hetR strain subjected to nitrogen deprivation for 0-6-12-24 hours as indicated by coloured lines, with the reads numbers from the RNA-Seq dataset in (Mitschke, J., Vioque, A., Haas, F., Hess, W.R., and Muro-Pastor, A.M. (2011). Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in Anabaena sp. PCC7120. Proc. Natl. Acad. Sci. USA 108, 20130-20135) (right y axis).

Supplementary data file 3 "genomeplot_WT_N_N_chromosome.pdf". Overview (chromosome) combining log2-normalized expression values (left y axis) from the microarray analysis of cultures of wild type at 0 and 8 h after nitrogen depletion and at 0 and 8 h after NH4 addition as indicated by coloured lines, with the reads numbers from the RNA-Seq dataset in (Mitschke, J., Vioque, A., Haas, F., Hess, W.R., and Muro-Pastor, A.M. (2011). Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in Anabaena sp. PCC7120. Proc. Natl. Acad. Sci. USA 108, 20130-20135) (right y axis).

Supplementary data file 4 "genomeplot_WT_N_N_plasmids.pdf". Overview (six plasmids combined) combining log2-normalized expression values (left y axis) from the microarray analysis of cultures of wild type at 0 and 8 h after nitrogen depletion and at 0 and 8 h after NH4 addition as indicated by coloured lines, with the reads numbers from the RNA-Seq dataset in (Mitschke, J., Vioque, A., Haas, F., Hess, W.R., and Muro-Pastor, A.M. (2011). Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in Anabaena sp. PCC7120. Proc. Natl. Acad. Sci. USA 108, 20130-20135) (right y axis).

Supplementary data file 5 "Co-expression network". Cytoscape (www.cytoscape.org) file corresponding to the co-expression network constructed.

URI: <http://hdl.handle.net/10261/179430>

DOI: <http://dx.doi.org/10.20350/digitalCSIC/8631>

Aparece en las colecciones: (IBVF) Conjuntos de datos


Ficheros en este ítem:

Fichero	Descripción	Tamaño	Formato
Data_file_1_genomeplot_WT_hetR_N_chromosome.pdf		128,76 MB	Adobe PDF





environmental microbiology

Environmental Microbiology (2019) 00(00), 00-00

doi:10.1111/1462-2920.14647



Elements of the heterocyst-specific transcriptome unraveled by co-expression analysis in Nostoc sp. PCC 7120

Manuel Brenes-Álvarez ¹, Jan Mitschke,^{2†} Elvira Olmedo-Verd,¹ Jens Georg,² Wolfgang R. Hess ^{2,3}, Agustín Vioque ¹ and Alicia M. Muro-Pastor ^{1*}

¹Instituto de Bioquímica Vegetal y Fotosíntesis, Consejo Superior de Investigaciones Científicas and Universidad de Sevilla, E-41092, Sevilla, Spain.

²Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, D-79104, Freiburg, Germany.

³Freiburg Institute for Advanced Studies, University of Freiburg, D-79104, Freiburg, Germany.

Summary

Nitrogen is frequently limiting microbial growth in the environment. As a response, many filamentous cyanobacteria differentiate heterocysts, cells devoted to N₂ fixation. Heterocyst differentiation is under the control of the master regulator HetR. Through the characterization of the HetR-dependent transcriptome in Nostoc sp. PCC 7120, we identified the new candidate genes likely involved in heterocyst differentiation. According to their maximum induction, we defined E-DF (early in differentiation) and L-DF (late in differentiation) genes. Most of the genes known to be involved in the critical aspects of heterocyst differentiation or function were also classified into these groups, showing the validity of the approach. Using fusions to gfp, we verified the heterocyst-specific transcription of several of the found genes, antisense transcripts and potentially transcribing sRNAs. Through comparative sequence analysis of promoter regions, we noticed the prevalence of the

The elements identified here are of interest for understanding cell differentiation, engineering of biological nitrogen fixation or production of O₂-sensitive molecules in cyanobacteria.

Introduction

In the absence of combined nitrogen sources, many filamentous cyanobacteria, including Nostoc sp. PCC 7120 (a.k.a. Anabaena sp. PCC 7120) the model organism used in this work, differentiate a one-dimensional pattern of specialized cells called heterocysts, that are devoted to the fixation of atmospheric nitrogen (N₂) (Muro-Pastor and Hess, 2012). Heterocyst differentiation relies on a transcriptional program that is exclusive of the cells becoming heterocysts (Flansburg et al., 2011; Mitschke et al., 2011; Flores et al., 2019). Major landmarks of the process include the sequential synthesis and deposition of specific heterocyst envelopes as well as the synthesis and maturation of nitrogenase, the enzymatic complex for N₂ fixation. In parallel, precise coordination between the nitrogen-fixing heterocysts and vegetative cells becomes established because these two cell types depend on each other in order to sustain growth of the filament as a whole. Growth at the expense of N₂ involves not only the morphological and metabolic transformation of selected vegetative cells to become heterocysts, but also the precise regulation of the pattern and timing of differentiation (Flores and Herrero, 2019). Thus, a high number of different regulatory factors can be expected to be involved in these processes. Classical approaches involving transposon or chemical mutagenesis followed by screening for Fox⁻ mutants,

Peer Access to Data and Materials

We recommend that data for which public repositories are widely used, and are accessible to all, should be deposited in such a repository prior to publication. The appropriate linking details and identifier(s) should then be included in the publication and where possible the repository, to facilitate linking between the journal article and the data

<https://onlinelibrary.wiley.com/page/journal/14622920/homepage/forauthors.html>

Estación Experimental de Zonas Áridas

Título: Experimentally broken faecal sacs affect nest bacterial environment, development and survival of spotless starling nestlings [Dataset]

Autor: Azcárate, Manuel; Ruiz-Rodríguez, Magdalena; Díaz-Lora, Silvia; Ruiz-Castellano, Cristina; Soler, Juan José

Palabras clave: Bacteria, Ectoparasites, Faecal sacs, Nestling growth, Nest predation, Nest sanitation, Stumus unicolor

Fecha de publicación: 1-feb-2019

Citación: Azcárate-García, Manuel; Ruiz-Rodríguez, Magdalena; Díaz-Lora, Silvia ; Ruiz-Castellano, Cristina; Soler, Juan José. "Experimentally broken faecal sacs affect nest bacterial environment, development and survival of spotless starling nestlings [Dataset]". DIGITAL.CSIC. <http://hdl.handle.net/10261/175087>

Resumen: Nestlings of most avian species produce faecal sacs, which facilitate the removal of nestlings' excrements by parents, thereby reducing proliferation of potentially pathogenic microorganisms and/or detectability by predators and parasites. The nest microbial environment that birds experience during early life might also affect their development and thus, faecal sacs facilitating parental removal may be a strategy to decrease bacterial contamination of nests that could harm developing nestlings. Here, we tested this hypothesis by experimentally broken faecal sacs and spreading them in nests of spotless (Stumus unicolor), thereby avoiding their removal by adults. In accordance with the hypothesis, experimental nests harboured higher bacterial control nests. Nestlings in experimental nests were of smaller size (tarsus length) and experienced lower probability of survival (predation) than control nests. Moreover, nestlings in experimental nests tended to suffer more from ectoparasites than those in control nests. We discuss the pivotal role of bacteria producing chemical volatiles that ectoparasites and predators might use to find avian nests, and that could explain our experimental results in starlings

URI: <http://hdl.handle.net/10261/175087>

DOI: <http://dx.doi.org/10.20350/digitalCSIC/8610>

Aparece en las colecciones: (EEZA) Conjuntos de datos

Fichero	Descripción	Tamaño	Formato
Data_Experimentally broken faecal sacs affect nest bacterial environment, development and survival of spotless starling nestlings.xlsx		80.99 KB	Microsoft Excel XML

Visualizar/Abrir

Petición de una copia

Petición de una copia del trabajo Experimentally broken faecal sacs affect nest bacterial environment, development and survival of spotless starling nestlings [Dataset]

Dado que el trabajo está actualmente embargado, puede pedir una copia personal con fines de investigación/estudio al autor clicando en "Enviar". Es importante tener en cuenta que esta funcionalidad depende de que el email del depositador del ítem esté aún operativo. Si no lo estuviera, por favor contacte con la Oficina Técnica de DIGITAL.CSIC, digital.csic@bib.csic.es

Nombre del peticionario:

Correo electrónico del peticionario:

Ficheros: todos los ficheros de este documento en acceso restringido Los ficheros solicitados

Mensaje:

Cancelar Enviar

Dataset embargado

A través de la petición de copia privada:
Acceso gratuito a grupos restringidos, a individuos o revisores

Los datos de investigación en el mandato institucional de acceso abierto

- Solicita que las referencias bibliográficas (...) de los *datasets* asociados a artículos de revistas sean hechas públicas de manera permanente en DIGITAL.CSIC desde el momento de la aceptación para su publicación de los artículos asociados.
- Solicita que se ofrezcan en modalidad de acceso abierto en DIGITAL.CSIC aquellos *datasets* asociados a publicaciones siempre y cuando no se den legítimas razones de confidencialidad, propiedad intelectual y/o seguridad. Estos *datasets* en acceso abierto deben ser FAIR (“findable, accessible, interoperable, reproducible”) e ir acompañados de una licencia estándar que explícitamente indique las condiciones de uso y favorezca la reproducibilidad científica (por ejemplo, Creative Commons y Open Data Commons)
- Solicita el depósito público de los metadatos de los *datasets* asociados a artículos de revistas desde el momento de su aceptación
- Acceso abierto a los *datasets* asociados a artículos de revistas en DIGITAL.CSIC tan pronto como sea posible (a menos que haya excepciones)
- DIGITAL.CSIC no es una infraestructura de big data
- Los metadatos de los *datasets* deben estar bien descritos según estándares internacionales y los *datasets* debe llevar **licencias de uso estándar para que sean FAIR**
- **Alineamiento con la política EOSC** (European Open Science Cloud) de la Comisión Europea

Oportunidades para las bibliotecas: servicios que demandan los investigadores



Larga experiencia en la descripción de recursos científicos según **estándares, vocabularios** y en recomendar **herramientas de apoyo a la investigación de calidad**



Un nuevo perfil profesional en bibliotecas es el **especialista en gestionar datos de investigación**: limpiar y analizar datos, crear obras derivadas, geolocalizar datasets, lanzar proyectos innovadores, abrir los datasets a la sociedad



Oportunidad estratégica de jugar un **papel destacado en el apoyo, monitorización y cumplimiento del mandato institucional y en incluir los servicios bibliotecarios en los planes de gestión de datos**



Apoyo y monitorización del mandato



Promoción y enriquecimiento de datos en proyecto INEOS



Exposición de colecciones de datasets de DIGITAL.CSIC en EOSC (EPOS, EOSC-Synergy)



Explotación de servicios de valor añadido



Formación en gestión y publicación de datos para bibliotecas y otros técnicos CSIC



Isabel.bernal@bib.csic.es

GRACIAS