cirad
LA RECHERCHE AGRONOMIQUE
POUR LE DÉVELOPPEMENT

data2019opensci : En route vers la science ouverte – 12 nov 2019

data2019opensci

Publier ses données, un point sur les Data papers

Laurence Dedieu

CIRAD - Délégation à l'Information Scientifique et Technique

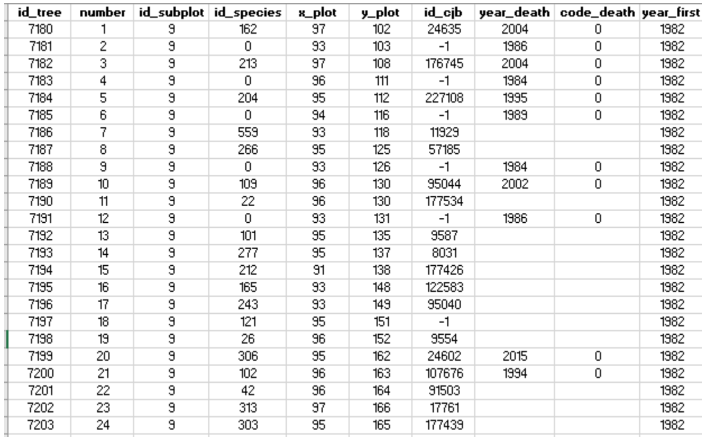
Journal of open health data

A peer-reviewed open-access journal
Biodiversity Data Journal
Making your data count!
ISSN 1314-2828 (online)

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Principe du Data paper

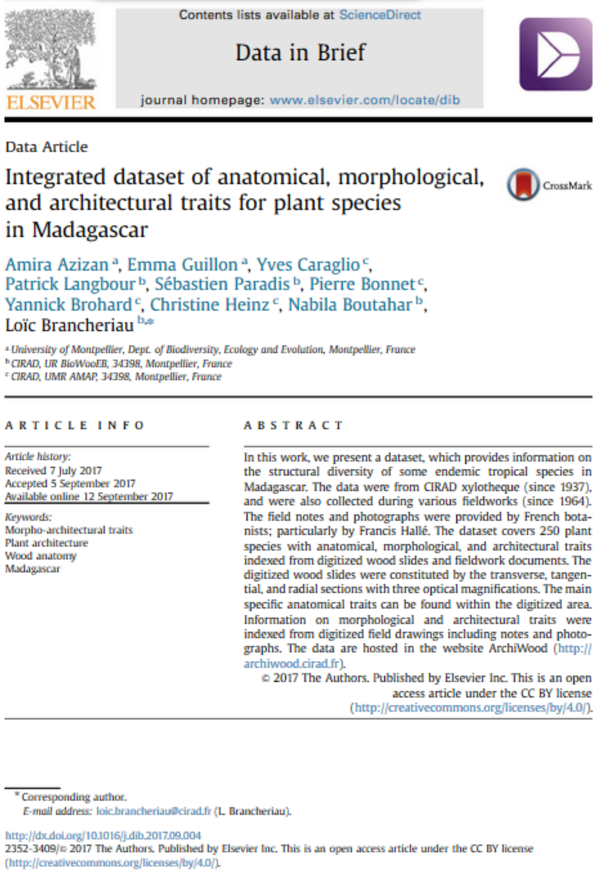
Jeu de données



id_tree	number	id_subplot	id_species	x_plot	y_plot	id_cjb	year_death	code_death	year_first
7180	1	9	162	97	102	24635	2004	0	1982
7181	2	9	0	93	103	-1	1986	0	1982
7182	3	9	213	97	108	176745	2004	0	1982
7183	4	9	0	96	111	-1	1984	0	1982
7184	5	9	204	95	112	227108	1995	0	1982
7185	6	9	0	94	116	-1	1989	0	1982
7186	7	9	559	93	118	11929			1982
7187	8	9	266	95	125	57185			1982
7188	9	9	0	93	126	-1	1984	0	1982
7189	10	9	109	96	130	95044	2002	0	1982
7190	11	9	22	96	130	177534			1982
7191	12	9	0	93	131	-1	1986	0	1982
7192	13	9	101	95	135	9587			1982
7193	14	9	277	95	137	8031			1982
7194	15	9	212	91	138	177426			1982
7195	16	9	165	93	148	122583			1982
7196	17	9	243	93	149	95040			1982
7197	18	9	121	95	151	-1			1982
7198	19	9	26	96	152	9554			1982
7199	20	9	306	95	162	24602	2015	0	1982
7200	21	9	102	96	163	107676	1994	0	1982
7201	22	9	42	96	164	91503			1982
7202	23	9	313	97	166	17761			1982
7203	24	9	303	95	165	177439			1982

Décrire un jeu de données en vue de sa réutilisation

Data paper



Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

ELSEVIER

Data Article

Integrated dataset of anatomical, morphological, and architectural traits for plant species in Madagascar

CrossMark

Amira Azizan[a], Emma Guillon[a], Yves Caraglio[c], Patrick Langbour[b], Sébastien Paradis[b], Pierre Bonnet[c], Yannick Brohard[c], Christine Heinz[c], Nabila Boutahar[b], Loïc Brancheriau[b,*]

[a] University of Montpellier, Dept. of Biodiversity, Ecology and Evolution, Montpellier, France
[b] CIRAD, UR BioWooEB, 34398, Montpellier, France
[c] CIRAD, UMR AMAP, 34398, Montpellier, France

ARTICLE INFO

Article history:
Received 7 July 2017
Accepted 5 September 2017
Available online 12 September 2017

Keywords:
Morpho-architectural traits
Plant architecture
Wood anatomy
Madagascar

ABSTRACT

In this work, we present a dataset, which provides information on the structural diversity of some endemic tropical species in Madagascar. The data were from CIRAD xylotheque (since 1937), and were also collected during various fieldworks (since 1964). The field notes and photographs were provided by French botanists; particularly by Francis Hallé. The dataset covers 250 plant species with anatomical, morphological, and architectural traits indexed from digitized wood slides and fieldwork documents. The digitized wood slides were constituted by the transverse, tangential, and radial sections with three optical magnifications. The main specific anatomical traits can be found within the digitized area. Information on morphological and architectural traits were indexed from digitized field drawings including notes and photographs. The data are hosted in the website ArchiWood (http://archiwood.cirad.fr).

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

---

[*] Corresponding author.
E-mail address: loic.brancheriau@cirad.fr (L. Brancheriau).

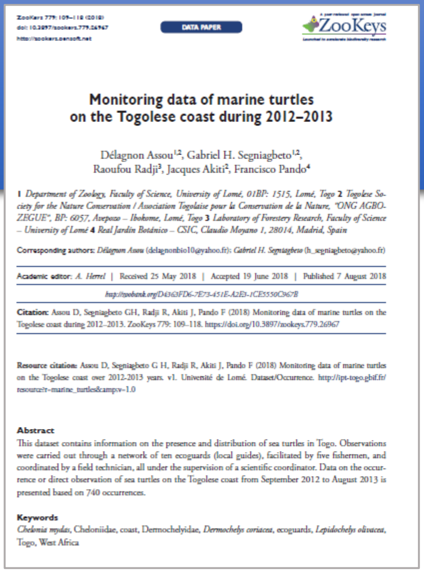
http://dx.doi.org/10.1016/j.dib.2017.09.004
2352-3409/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Informer la communauté scientifique

Mettre à disposition d'une communauté scientifique

Entrepôt de données

Principe du Data paper



ZooKeys 779: 109–118 (2018)
doi: 10.3897/zookeys.779.26967
http://zookeys.pensoft.net

DATA PAPER

Monitoring data of marine turtles on the Togolese coast during 2012–2013

Délagnon Assou[1,2], Gabriel H. Segniagbeto[1,2],
Raoufou Radji[3], Jacques Akiti[2], Francisco Pando[4]

1 *Department of Zoology, Faculty of Science, University of Lomé, 01BP: 1515, Lomé, Togo* **2** *Togolese Society for the Nature Conservation / Association Togolaise pour la Conservation de la Nature, "ONG AGBO-ZEGUE", BP: 6057, Avepozo – Ibokome, Lomé, Togo* **3** *Laboratory of Forestery Research, Faculty of Science – University of Lomé* **4** *Real Jardín Botánico – CSIC, Claudio Moyano 1, 28014, Madrid, Spain*

Corresponding authors: *Délagnon Assou* (delagnonbio10@yahoo.fr); *Gabriel H. Segniagbeto* (h_segniagbeto@yahoo.fr)

Academic editor: *A. Herrel* | Received 25 May 2018 | Accepted 19 June 2018 | Published 7 August 2018

http://zoobank.org/D4363FD6-7E73-451E-A2E3-1CE5550C967B

Citation: Assou D, Segniagbeto GH, Radji R, Akiti J, Pando F (2018) Monitoring data of marine turtles on the Togolese coast during 2012–2013. ZooKeys 779: 109–118. https://doi.org/10.3897/zookeys.779.26967

Resource citation: Assou D, Segniagbeto G H, Radji R, Akiti J, Pando F (2018) Monitoring data of marine turtles on the Togolese coast over 2012-2013 years. v1. Université de Lomé. Dataset/Occurrence. http://ipt-togo.gbif.fr/resource?r=marine_turtles&v=1.0

Abstract

This dataset contains information on the presence and distribution of sea turtles in Togo. Observations were carried out through a network of ten ecoguards (local guides), facilitated by five fishermen, and coordinated by a field technician, all under the supervision of a scientific coordinator. Data on the occurrence or direct observation of sea turtles on the Togolese coast from September 2012 to August 2013 is presented based on 740 occurrences.

Keywords

Chelonia mydas, Cheloniidae, coast, Dermochelyidae, *Dermochelys coriacea*, ecoguards, *Lepidochelys olivacea*, Togo, West Africa

- Décrire un jeu de données
- Décrire les méthodes d’obtention
- Montrer le potentiel de réutilisation des données
- Donner accès aux données
- Ni résultats, ni analyses, ni interprétation, ni discussion

Objectif: la réutilisation des données

- Décrire un jeu de données
  - → suffisamment pour permettre la réutilisation des données
- Décrire les méthodes d'obtention
  - → suffisamment pour permettre de reproduire l'étude
  - → montrer la rigueur scientifique et la qualité des données
- Montrer le potentiel de réutilisation des données
  - → originalité et valeur scientifique des données
  - → préciser pour qui et pour quoi elles peuvent être utiles

Objectif: la réutilisation des données

- Décrire un jeu de données
  - → **suffisamment** pour permettre la réutilisation des données
- Décrire les méthodes d'obtention
  - → **suffisamment** pour permettre de reproduire l'étude
  - → montrer la **rigueur** scientifique et la **qualité** des données
- Montrer le **potentiel de réutilisation** des données
  - → **originalité** et **valeur scientifique** des données
  - → préciser pour qui et pour quoi elles peuvent être utiles

➡ **Critères d'évaluation**

Data paper: 1 texte + 1 jeu de données

Data paper

Ecol Res (2015) 30: 415
DOI 10.1007/s11284-015-1252-8

CrossMark

DATA PAPER

Itsarapong Voraphab · Yupa Hanboonsong
Youichi Kobori · Hiroaki Ikeda · Takeshi Osawa

Insect species recorded in sugarcane fields of Khon Kaen Province, Thailand, over three seasons in 2012

Résumé

Données

Méthodes

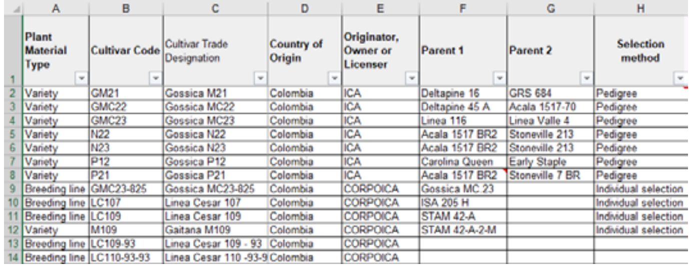
Potentiel

Références

T. Osawa
Japan Node of the Global Biodiversity Information Facility,
Tsukuba, Japan

+

Jeu de données



	A	B	C	D	E	F	G	H
1	Plant Material Type	Cultivar Code	Cultivar Trade Designation	Country of Origin	Originator, Owner or Licenser	Parent 1	Parent 2	Selection method
2	Variety	GM21	Gossica M21	Colombia	ICA	Deltapine 16	GRS 684	Pedigree
3	Variety	GMC22	Gossica MC22	Colombia	ICA	Deltapine 45 A	Acala 1517-70	Pedigree
4	Variety	GMC23	Gossica MC23	Colombia	ICA	Linea 116	Linea Valle 4	Pedigree
5	Variety	N22	Gossica N22	Colombia	ICA	Acala 1517 BR2	Stoneville 213	Pedigree
6	Variety	N23	Gossica N23	Colombia	ICA	Acala 1517 BR2	Stoneville 213	Pedigree
7	Variety	P12	Gossica P12	Colombia	ICA	Carolina Queen	Early Staple	Pedigree
8	Variety	P21	Gossica P21	Colombia	ICA	Acala 1517 BR2	Stoneville 7 BR	Pedigree
9	Breeding line	GMC23-825	Gossica MC23-825	Colombia	CORPOICA	Gossica MC 23		Individual selection
10	Breeding line	LC107	Linea Cesar 107	Colombia	CORPOICA	ISA 205 H		Individual selection
11	Breeding line	LC109	Linea Cesar 109	Colombia	CORPOICA	STAM 42-A		Individual selection
12	Variety	M109	Gaitana M109	Colombia	CORPOICA	STAM 42-A-2-M		Individual selection
13	Breeding line	LC109-93	Linea Cesar 109 - 93	Colombia	CORPOICA			
14	Breeding line	LC110-93-93	Linea Cesar 110 -93-9	Colombia	CORPOICA			

Organisation et documentation des données

+

Entrepôts

GEO Gene Expression Omnibus
Dataverse
PRIDE Archive
ICPSR
ENA European Nucleotide Archive
NCBI GenBank
gbif France
DRYAD
TreeBASE
LTER Network Data Portal
PANGAEA. Data Publisher for Earth & Environmental Science
MOVEBANK For Animal Tracking Data
knb
WormBase

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Identifiez au + tôt la revue et l’entrepôt

(Plan de gestion des données → gain de temps)

Les revues scientifiques

Data in Brief


Journal
Data in Brief

Title, Authors

Abstract ; keywords

Data Description

Describe the data and each file

Experimental Design, Materials and Methods

Complete description of the Experimental design and methods used to acquire the data.

Link to the deposited data

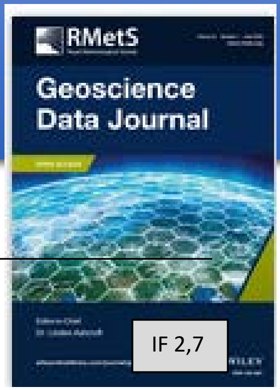
Value of the data : *3-6 points to describe :*

Why these data are of value to the scientific community ?
Who can benefit ?
How can these data be used for further …..

Acknowledgements

References

Geoscience Data Journal



IF 2,7

DATA PAPER

Title

Author names, affiliations

Abstract, Key words

Dataset : name, data centre, and unique identifier.

Identifier: xxxxx
Creator: xxxxx
Publisher: xxxxxx
Publication year: xxxxxx

Introduction

Rationale for collecting the dataset.

1. Data production methods

Description of the experimental set-up for the acquisition of the data.

2. Dataset location and format

Description of location, format and accessibility of the dataset.

3. Dataset use and reuse

Description of actual and potential uses for the dataset.

Link to the deposited data

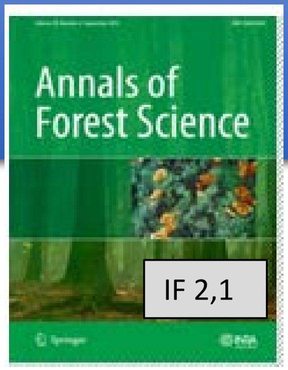
Acknowledgements

The research presented in this paper was funded by xxxxxx. The authors are grateful to xxxx for xxxx.

References

Authors. [year of publication]. Article title. *Journal Name* **[vol]**: pp–pp, doi:xxxxxxxx.

Annals of Forest Science



IF 2,1

Title, Authors

Abstract ; keywords

Background : *objectives and context of data collection*

Methods

Detailed description of methods used to produce data

Metadata description : *fichier Excel obligatoire – norme INSPIRE*

couvertures spatiale et temporelle
espèce, type de données,
protocoles, équipement/logiciel,
variables, unités, …

Link to the deposited data

Reuse potential and limits

Potential of the datasets

Instructions to facilitate the reuse of data-sets

Acknowledgements

References

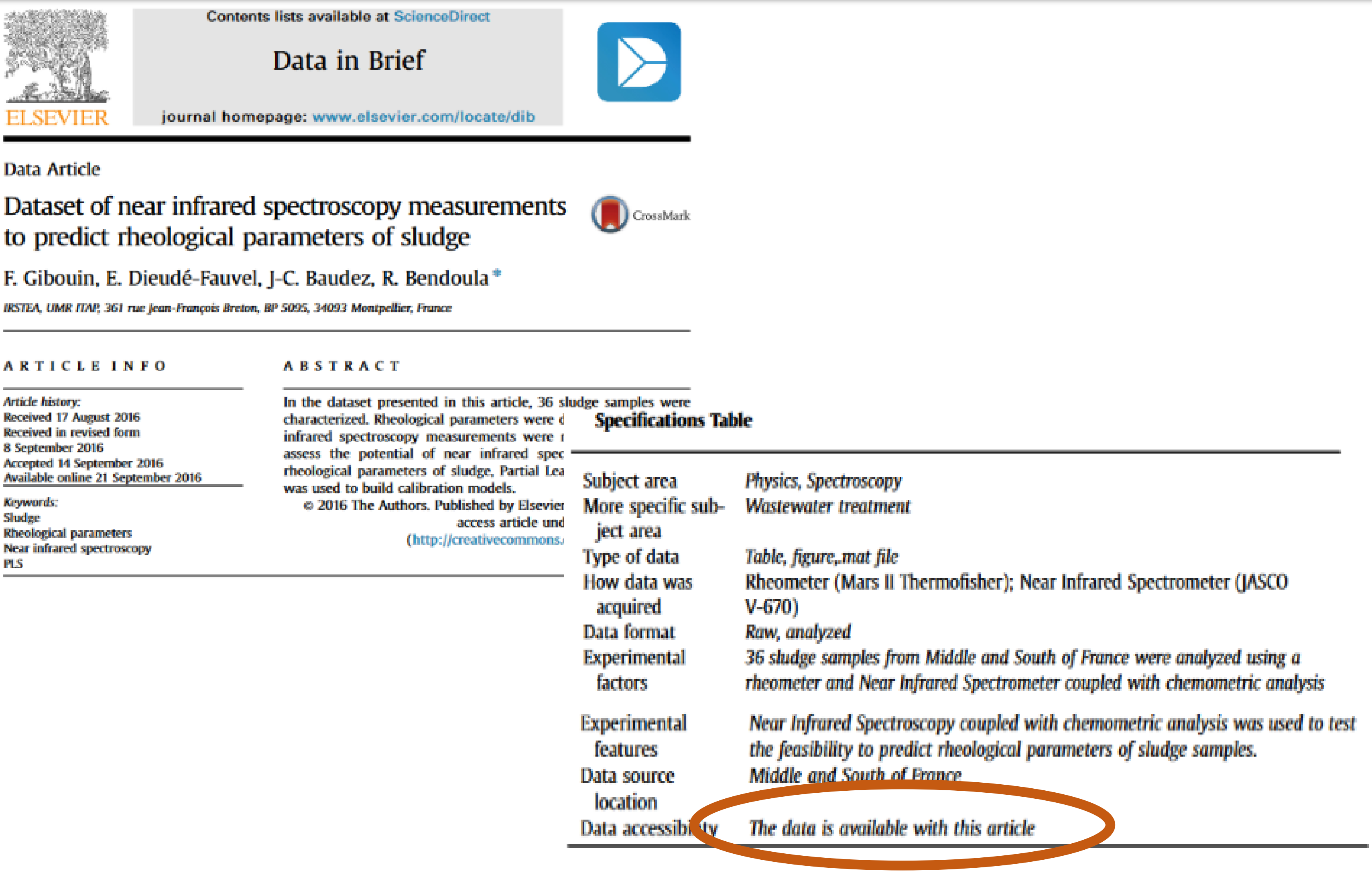
Critères de choix d'une revue : 1

Influence → **sur**

Localisation des données

- Intégrées dans le *Data paper*
- Entrepôt de données

Data in Brief – données dans l'article



Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of near infrared spectroscopy measurements to predict rheological parameters of sludge

F. Gibouin, E. Dieudé-Fauvel, J-C. Baudez, R. Bendoula *

IRSTEA, UMR ITAP, 361 rue Jean-François Breton, BP 5095, 34093 Montpellier, France

ARTICLE INFO

Article history:
Received 17 August 2016
Received in revised form
8 September 2016
Accepted 14 September 2016
Available online 21 September 2016

Keywords:
Sludge
Rheological parameters
Near infrared spectroscopy
PLS

ABSTRACT

In the dataset presented in this article, 36 sludge samples were characterized. Rheological parameters were d infrared spectroscopy measurements were r assess the potential of near infrared spec rheological parameters of sludge, Partial Lea was used to build calibration models.

© 2016 The Authors. Published by Elsevier access article und (http://creativecommons.

Specifications Table

Subject area	*Physics, Spectroscopy*
More specific subject area	*Wastewater treatment*
Type of data	*Table, figure,.mat file*
How data was acquired	Rheometer (Mars II Thermofisher); Near Infrared Spectrometer (JASCO V-670)
Data format	*Raw, analyzed*
Experimental factors	*36 sludge samples from Middle and South of France were analyzed using a rheometer and Near Infrared Spectrometer coupled with chemometric analysis*
Experimental features	*Near Infrared Spectroscopy coupled with chemometric analysis was used to test the feasibility to predict rheological parameters of sludge samples.*
Data source location	*Middle and South of France*
Data accessibility	*The data is available with this article*

Ecological Research et entrepôt LTER

DATA PAPER

ECOLOGICAL RESEARCH WILEY

Long-term fauna and flora records of the experimental forests of the Forest Research Station of Hokkaido University, Japan

Chisato Terada[1] | TaeOh Kwon[2] | Nobuko Kazahari[2] | Osamu Kishida[3] | Shunsuke Utsumi[2]

[1]Wakayama Experimental Forest, Field Science Center for Northern Biosphere, Hokkaido University, Kozagawa, Wakayama, Japan

[2]Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Hokkaido, Japan

[3]Tomakomai Experimental Forest, Field Science Center for Northern Biosphere, Hokkaido University, Tomakomai, Hokkaido, Japan

Correspondence

Shunsuke Utsumi, Field Science Center for Northern Biosphere, Hokkaido University. North 9, West 9, Sapporo, Hokkaido 060-0809, Japan.

Email: utsumi@fsc.hokudai.ac.jp

Funding information

Japan Society for the Promotion of Science, Grant/Award Numbers: 16H06179, 17HP8035

Abstract

The Forest Research Station of Hokkaido University owns a vast area of experimental forests, which are composed of a large variety of ecosystems ranging from aquatic ecosystems, such as wetlands and river basins, to various types of forests, such as primary, secondary and artificial forests. Additionally, these forests are representative of numerous climatic zones, such as warm temperate, cool temperate and subarctic. Since the initial establishment of the Forest Research Station in 1901, huge efforts have been devoted to recording the vegetation structure and vertebrate assemblages of these experimental forests. Thus, a large body of literature and long-term data on fauna and flora of these forests has been accumulated in our archives. However, most of these records have been written in Japanese and are not opened to the public or well structured. Therefore, we comprehensively reviewed these records and related scientific articles from the 1910s to the 2010s to build up the database for vascular plant and vertebrate animals that inhabited (in the past) and/or are currently inhabiting in the experimental forests. Additional site-specific information was also listed, including geological and topographical characteristics where species were found as well as the localities, survey area and year in which species were recorded. These databases, which span a large temporal and spatial scale, are expected to provide useful data for research or educational purposes and for understanding the flora and fauna of Japan. These can also contribute to a greater understanding of the historical transition of biodiversity in Japan.

KEYWORDS

forest, historical record, long-term, species occurrence, vascular plants, vertebrate animals

---

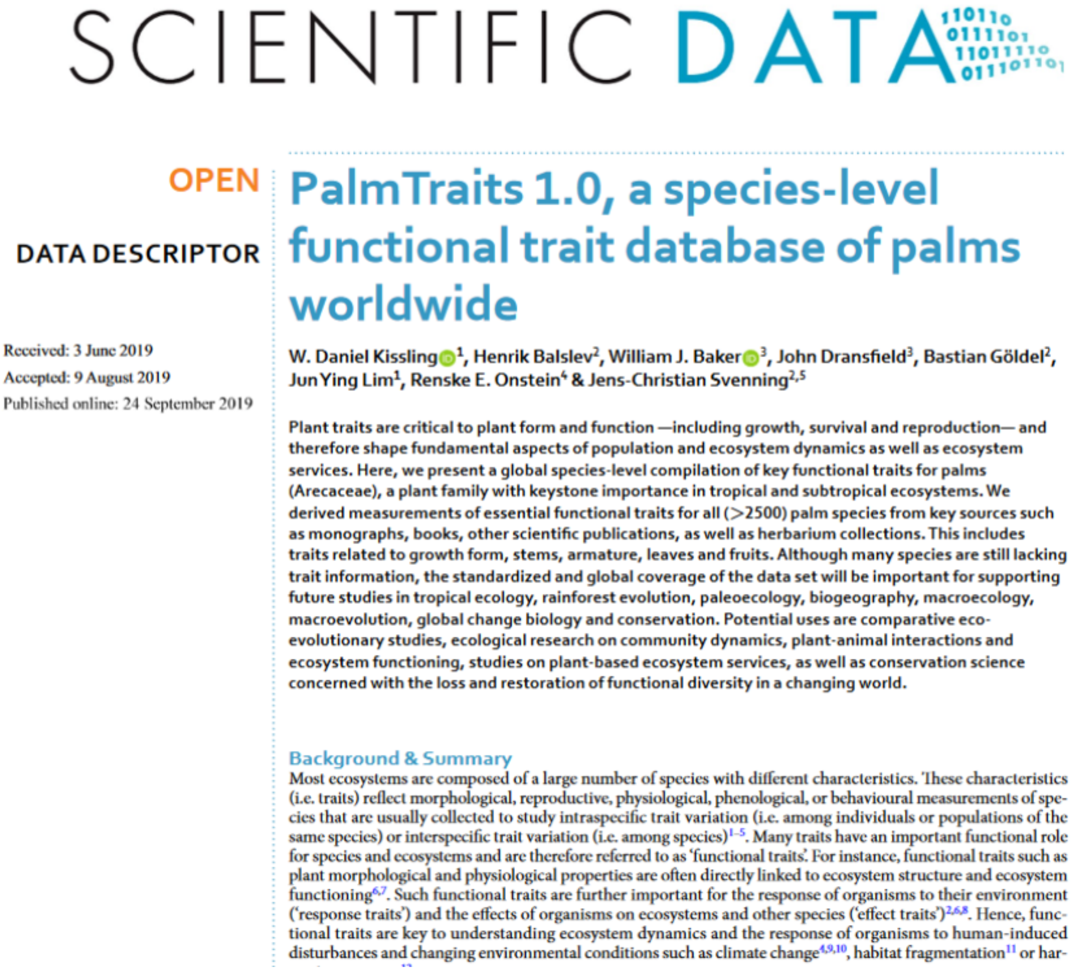
The complete data set for this abstract published in the Data Paper section of the journal is available in electronic format in Ecological Research Data Paper Archives at http://db.cger.nies.go.jp/JaLTER/ER_DataPapers/ archives/2019/ERDP-2019-01.

Paper Archives at http://db.cger.nies.go.jp/JaLTER/ER_DataPapers/ archives/2019/ERDP-2019-01.

LTER

International Long-Term Ecological Research Network

Scientific data et entrepôts recommandés


SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

PalmTraits 1.0, a species-level functional trait database of palms worldwide

Received: 3 June 2019
Accepted: 9 August 2019
Published online: 24 September 2019

W. Daniel Kissling[1], Henrik Balslev[2], William J. Baker[3], John Dransfield[3], Bastian Göldel[2], Jun Ying Lim[1], Renske E. Onstein[4] & Jens-Christian Svenning[2,5]

Plant traits are critical to plant form and function —including growth, survival and reproduction— and therefore shape fundamental aspects of population and ecosystem dynamics as well as ecosystem services. Here, we present a global species-level compilation of key functional traits for palms (Arecaceae), a plant family with keystone importance in tropical and subtropical ecosystems. We derived measurements of essential functional traits for all (>2500) palm species from key sources such as monographs, books, other scientific publications, as well as herbarium collections. This includes traits related to growth form, stems, armature, leaves and fruits. Although many species are still lacking trait information, the standardized and global coverage of the data set will be important for supporting future studies in tropical ecology, rainforest evolution, paleoecology, biogeography, macroecology, macroevolution, global change biology and conservation. Potential uses are comparative eco-evolutionary studies, ecological research on community dynamics, plant-animal interactions and ecosystem functioning, studies on plant-based ecosystem services, as well as conservation science concerned with the loss and restoration of functional diversity in a changing world.

Background & Summary

Most ecosystems are composed of a large number of species with different characteristics. These characteristics (i.e. traits) reflect morphological, reproductive, physiological, phenological, or behavioural measurements of species that are usually collected to study intraspecific trait variation (i.e. among individuals or populations of the same species) or interspecific trait variation (i.e. among species)[1–5]. Many traits have an important functional role for species and ecosystems and are therefore referred to as 'functional traits'. For instance, functional traits such as plant morphological and physiological properties are often directly linked to ecosystem structure and ecosystem functioning[6,7]. Such functional traits are further important for the response of organisms to their environment ('response traits') and the effects of organisms on ecosystems and other species ('effect traits')[2,6,8]. Hence, functional traits are key to understanding ecosystem dynamics and the response of organisms to human-induced disturbances and changing environmental conditions such as climate change[4,9,10], habitat fragmentation[11] or har-

Taxonomy & species diversity

- [Integrated Taxonomic Information System (ITIS)]
- [KNB: The Knowledge Network for Biocomplexity]
- [NCBI Taxonomy]
- [Global Biodiversity Information Facility (GBIF)]
- [Morphobank.org]
- [Movebank Data Repository]

Ecology

- [EDI - LTER]
- [Global Biodiversity Information Facility (GBIF)]
- [KNB: The Knowledge Network for Biocomplexity]

Earth & environmental sciences

- [NERC Data Centres]
- [PANGAEA]

Generalist repositories

- [Zenodo]
- [Dryad Digital Repository]
- [Harvard Dataverse]
- [Figshare]

Critères de choix d’une revue : 2

Influence → **sur**

Localisation des données

- Intégrées dans le *Data paper*
- Entrepôt de données

Modalités de diffusion des données

- Délai d’accès aux données
- Licences de diffusion

Revue et modalités de diffusion des données

Data in Brief ELSEVIER


Global Ecology and Biogeography

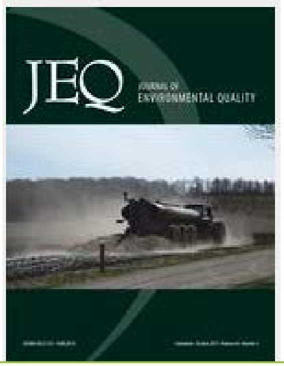
Données accessibles dès publication


STUDIES IN FAMILY PLANNING

Accès aux données sur demande

SCIENTIFIC DATA

Licence la + ouverte possible
(CC-0 ou CC-BY)
aucune restriction possible
(« pas d’utilisation commerciale »: non accepté)


JEQ JOURNAL OF ENVIRONMENTAL QUALITY

Licence CC-BY-NC
uniquement utilisation en recherche

Critères de choix d'une revue : 3

Influence → **sur**

Localisation des données

- Intégrées dans le *Data paper*
- Entrepôt de données

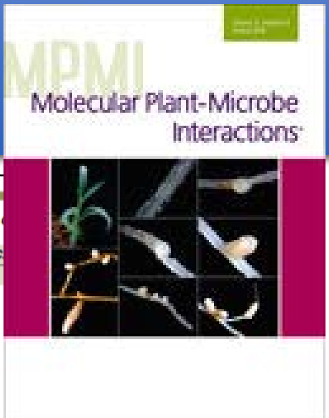
Modalités de diffusion des données

- Délai d'accès aux données
- Licences de diffusion

Rédaction du Data paper

- Texte libre
- Métadonnées structurées

Molecular Plant-Microbe Interactions




MPMI Vol. 32, No. 2, 2019, pp. 139–141. https://doi.org/10.1094/MPMI-05-18-0144-A

RESOURCE ANNOUNCEMENT

A High-Quality Draft Genome Sequence of *Colletotrichum gloeosporioides* sensu stricto SMCG1#C, a Causal Agent of Anthracnose on *Cunninghamia lanceolata* in China

Lin Huang,[1] Ki-Tae Kim,[2] Ji-Yun Yang,[1] Hyeunjeong Song,[3] Gobong Choi,[3] Jongbum Jeon,[3] Kyeongchae Cheong,[3] Jaeho Ko,[2] Haibin Xu,[4,†] and Yong-Hwan Lee[2,3,5,†]

[1] Co-Innovation Center for Sustainable Forestry in Southern China, College of Forestry, Nanjing Forestry University, Nanjing, Jiangsu 210037, China;
[2] Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Korea;
[3] Interdisciplinary Program in Agricultural Genomics, Seoul National University;
[4] College of Biology and the Environment, Nanjing Forestry University;
[5] Center for Fungal Genetic Resources, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, Seoul National University

Abstract

Colletotrichum has a broad host range and causes major yield losses of crops. The fungus *Colletotrichum gloeosporioides* is associated with anthracnose on Chinese fir. In this study, we present a high-quality draft genome sequence of *C. gloeosporioides* sensu stricto SMCG1#C, providing a reference genomic data for further research on anthracnose of Chinese fir and other hosts.

Colletotrichum is one of the most important groups of phytopathogenic fungi in the world because of its scientific and economic importance (Dean et al. 2012). *Colletotrichum gloeosporioides* is a ubiquitous plant pathogen that infects a wide range of plant species (Weir et al. 2012). According to recent advances in taxonomy, *C. gloeosporioides* is considered a species complex and is segregated into 22 species and one subspecies (Weir et al. 2012). Chinese fir (*Cunninghamia lanceolata*) has been cultivated for over 3,000 years and contributes about 40% of timber in southern China (Huang et al. 2018; Shi et al. 2010). Anthracnose caused by *C. gloeosporioides* is one of the most serious fungal diseases on Chinese fir, which is widely distributed in the cultivated areas of Chinese fir, and causes enormous economic losses (Lan et al. 2015). *C. gloeosporioides* SMCG1#C was isolated from the infected leaves of Chinese fir in Nanjing, China. Based on the phylogenetic tree calculated from the alignment of concatenated sequences of ITS, ACT, CAL, CHS-1, and GAPDH, strain SMCG1#C was identified as *C. gloeosporioides* sensu stricto (*unpublished data*).

The genome of *C. gloeosporioides* SMCG1#C was sequenced, using both PacBio
Sequel System (Tianjin Biochip Corporation, Tianjin, China) and Illumina Hiseq X Ten
(Novogene Corporation, Beijing, China). A total of 519,294 reads and 171,464,766
end 150-bp Illumina reads were generated, with respective coverages of 71× and 4
novo assemblies were performed using Velvet version 1.2.10 (Zerbino and Birney 20
obtained 28 contigs with an average length of 2,210,112 bp, an N50 of 4,696,547 bp,
of 5. Finally, a draft genome of 18 scaffolds was produced by using BLASR and
algorithms (Camacho et al. 2009; Chaisson and Tesler 2012), a total of 61.9 Mb
content of 50.3%, N50 of 5,209,244 bp, and L50 of 5 (Table 1). The validation of a

Funding:
This study was financially supported by the National Key R & D Program of China (2017YFD0600102), the Major Project of Jiangsu Province University Natural Science Research (16KJA220002), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). This work was supported by grants from National Research Foundation of Korea (NRF-2017R1A2A1A17069504, NRF-2015M3A9B8028679). Ki-Tae is grateful for a graduate fellowship through the Brain Korea 21 Plus Program.

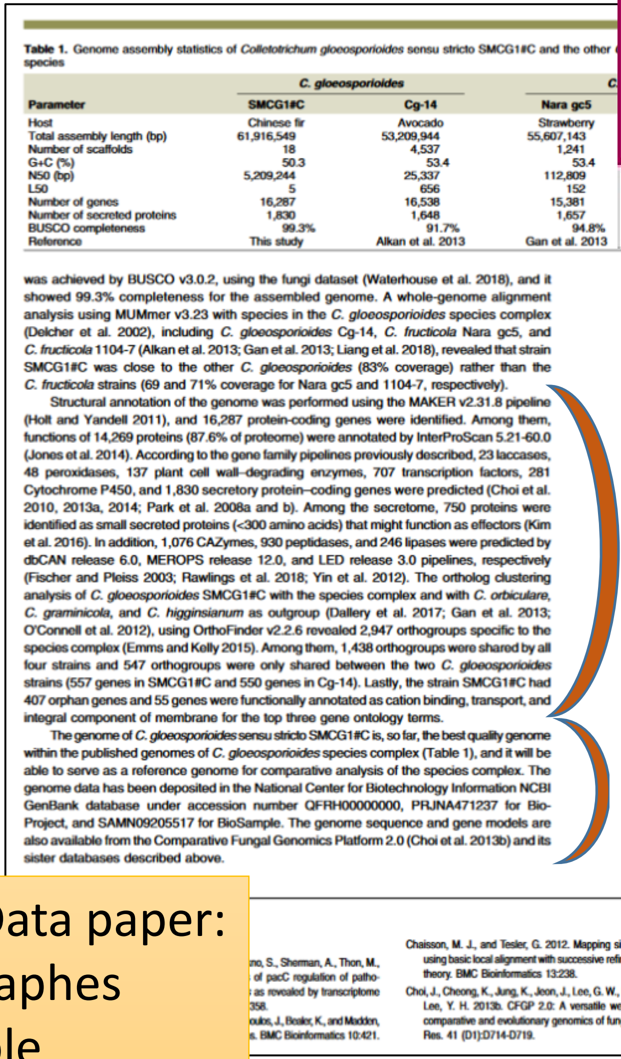


Table 1. Genome assembly statistics of *Colletotrichum gloeosporioides* sensu stricto SMCG1#C and the other
species

	C. gloeosporioides		*C*
Parameter	**SMCG1#C**	**Cg-14**	**Nara gc5**
Host	Chinese fir	Avocado	Strawberry
Total assembly length (bp)	61,916,549	53,209,944	55,607,143
Number of scaffolds	18	4,537	1,241
G+C (%)	50.3	53.4	53.4
N50 (bp)	5,209,244	25,337	112,809
L50	5	656	152
Number of genes	16,287	16,538	15,381
Number of secreted proteins	1,830	1,648	1,657
BUSCO completeness	99.3%	91.7%	94.8%
Reference	This study	Alkan et al. 2013	Gan et al. 2013

was achieved by BUSCO v3.0.2, using the fungi dataset (Waterhouse et al. 2018), and it showed 99.3% completeness for the assembled genome. A whole-genome alignment analysis using MUMmer v3.23 with species in the *C. gloeosporioides* species complex (Delcher et al. 2002), including *C. gloeosporioides* Cg-14, *C. fructicola* Nara gc5, and *C. fructicola* 1104-7 (Alkan et al. 2013; Gan et al. 2013; Liang et al. 2018), revealed that strain SMCG1#C was close to the other *C. gloeosporioides* (83% coverage) rather than the *C. fructicola* strains (69 and 71% coverage for Nara gc5 and 1104-7, respectively).

Structural annotation of the genome was performed using the MAKER v2.31.8 pipeline (Holt and Yandell 2011), and 16,287 protein-coding genes were identified. Among them, functions of 14,269 proteins (87.6% of proteome) were annotated by InterProScan 5.21-60.0 (Jones et al. 2014). According to the gene family pipelines previously described, 23 laccases, 48 peroxidases, 137 plant cell wall–degrading enzymes, 707 transcription factors, 281 Cytochrome P450, and 1,830 secretory protein–coding genes were predicted (Choi et al. 2010, 2013a, 2014; Park et al. 2008a and b). Among the secretome, 750 proteins were identified as small secreted proteins (<300 amino acids) that might function as effectors (Kim et al. 2016). In addition, 1,076 CAZymes, 930 peptidases, and 246 lipases were predicted by dbCAN release 6.0, MEROPS release 12.0, and LED release 3.0 pipelines, respectively (Fischer and Pleiss 2003; Rawlings et al. 2018; Yin et al. 2012). The ortholog clustering analysis of *C. gloeosporioides* SMCG1#C with the species complex and with *C. orbiculare*, *C. graminicola*, and *C. higginsianum* as outgroup (Dallery et al. 2017; Gan et al. 2013; O'Connell et al. 2012), using OrthoFinder v2.2.6 revealed 2,947 orthogroups specific to the species complex (Emms and Kelly 2015). Among them, 1,438 orthogroups were shared by all four strains and 547 orthogroups were only shared between the two *C. gloeosporioides* strains (557 genes in SMCG1#C and 550 genes in Cg-14). Lastly, the strain SMCG1#C had 407 orphan genes and 55 genes were functionally annotated as cation binding, transport, and integral component of membrane for the top three gene ontology terms.

The genome of *C. gloeosporioides* sensu stricto SMCG1#C is, so far, the best quality genome within the published genomes of *C. gloeosporioides* species complex (Table 1), and it will be able to serve as a reference genome for comparative analysis of the species complex. The genome data has been deposited in the National Center for Biotechnology Information NCBI GenBank database under accession number QFRH00000000, PRJNA471237 for BioProject, and SAMN09205517 for BioSample. The genome sequence and gene models are also available from the Comparative Fungal Genomics Platform 2.0 (Choi et al. 2013b) and its sister databases described above.

no, S., Sherman, A., Thon, M.,
of pacC regulation of patho-
as revealed by transcriptome
358.
oulos, J., Bealer, K., and Madden,
s. BMC Bioinformatics 10:421.

Chaisson, M. J., and Tesler, G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. BMC Bioinformatics 13:238.

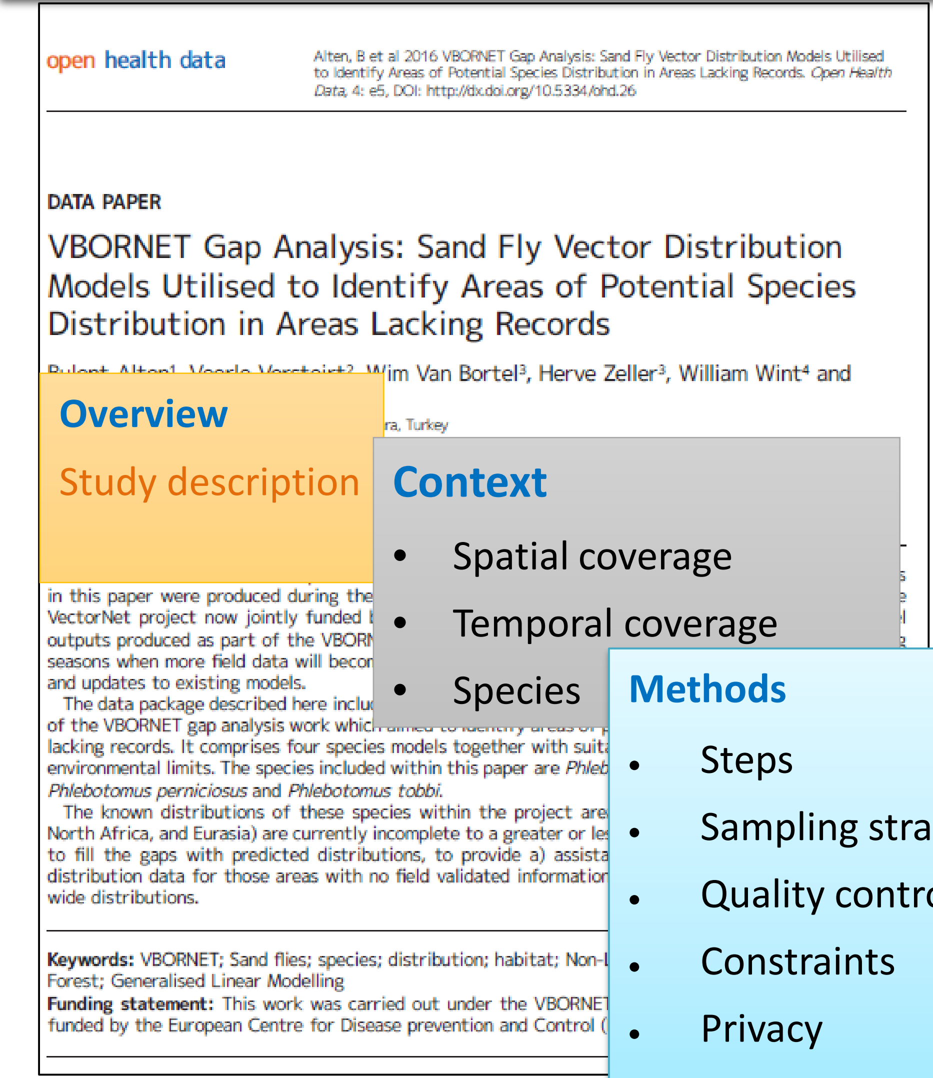
Choi, J., Cheong, K., Jung, K., Jeon, J., Lee, G. W., Kang, S., Kim, S., Lee, Y. W., and Lee, Y. H. 2013b. CFGP 2.0: A versatile web-based platform for supporting comparative and evolutionary genomics of fungi and Oomycetes. Nucleic Acids Res. 41 (D1):D714-D719.

GenBank

Contenu du Data paper:
4 paragraphes
1 table
lien vers données

Journal of Open Health Data

Journal of open health data



open health data

Alten, B et al 2016 VBORNET Gap Analysis: Sand Fly Vector Distribution Models Utilised to Identify Areas of Potential Species Distribution in Areas Lacking Records. *Open Health Data*, 4: e5, DOI: http://dx.doi.org/10.5334/ohd.26

DATA PAPER

VBORNET Gap Analysis: Sand Fly Vector Distribution Models Utilised to Identify Areas of Potential Species Distribution in Areas Lacking Records

Wim Van Bortel[3], Herve Zeller[3], William Wint[4] and

ra, Turkey

in this paper were produced during the
VectorNet project now jointly funded
outputs produced as part of the VBORN
seasons when more field data will becon
and updates to existing models.
The data package described here inclu
of the VBORNET gap analysis work which
lacking records. It comprises four species models together with suita
environmental limits. The species included within this paper are *Phleb*
Phlebotomus perniciosus and *Phlebotomus tobbi*.
The known distributions of these species within the project are
North Africa, and Eurasia) are currently incomplete to a greater or les
to fill the gaps with predicted distributions, to provide a) assista
distribution data for those areas with no field validated information
wide distributions.

Keywords: VBORNET; Sand flies; species; distribution; habitat; Non-
Forest; Generalised Linear Modelling
Funding statement: This work was carried out under the VBORNET
funded by the European Centre for Disease prevention and Control (

Overview

Study description

Context

- Spatial coverage
- Temporal coverage
- Species

Methods

- Steps
- Sampling strategy
- Quality control
- Constraints
- Privacy
- Ethics

Dataset description

- Object name
- Data type
- Ontologies
- Format names and versions
- Creation dates
- Dataset creators
- Licence
- Repository location
- Publication date

Reuse potential

- within your field
- outside of your field

Critères de choix d'une revue

- Champ scientifique et lectorat (= futurs utilisateurs)
- Exigences de la revue (Instructions aux auteurs)
  - portée des données: global/local; générique/spécifique d'espèce
  - contraintes: choix entrepôts; licences de diffusion
- Article en libre accès / coût
- Facteur d'impact
- Vérifiez que:
  - la publication du Data paper ne nuit pas à la publication de votre futur article de recherche
  - *vous avez le droit de publier les données*
  - *vous avez l'accord de tous les contributeurs*

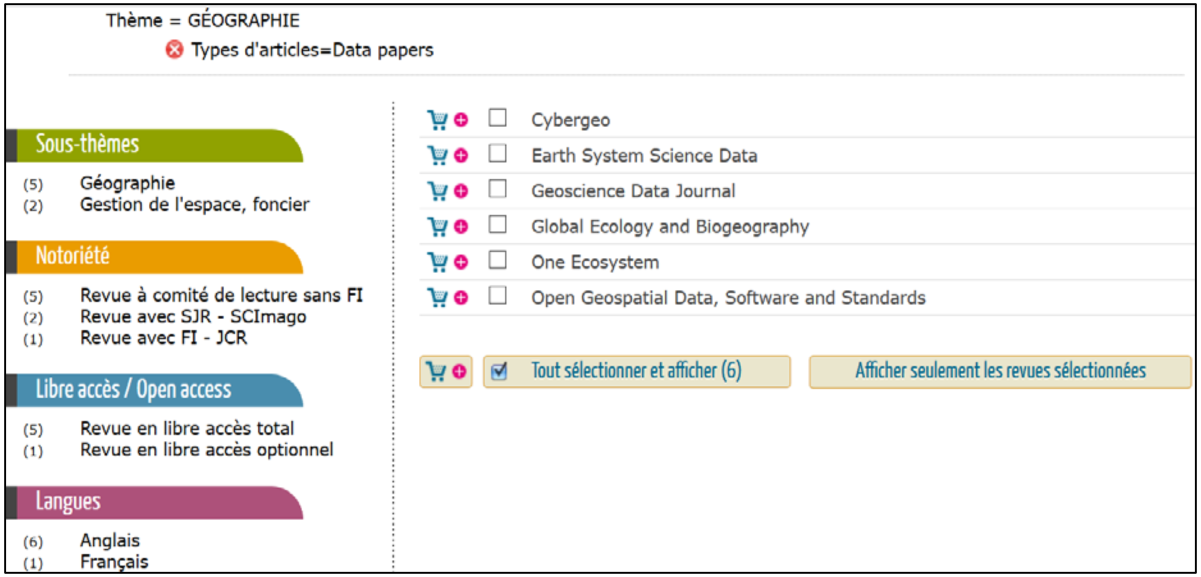
Ressources pour choisir une revue

- **Où publier ?**
  Base d'informations sur les revues en sciences du vivant, sciences sociales et sciences de l'ingénieur appliquées à l'agriculture

http://ou-publier.cirad.fr/

> 2000 revues

- thèmes / sous-thèmes
- FI, libre accès
- types d'articles



Thème = GÉOGRAPHIE

⊗ Types d'articles=Data papers

Sous-thèmes

(5)	Géographie
(2)	Gestion de l'espace, foncier

Notoriété

(5)	Revue à comité de lecture sans FI
(2)	Revue avec SJR - SCImago
(1)	Revue avec FI - JCR

Libre accès / Open access

(5)	Revue en libre accès total
(1)	Revue en libre accès optionnel

Langues

(6)	Anglais
(1)	Français

- ☐ Cybergeo
- ☐ Earth System Science Data
- ☐ Geoscience Data Journal
- ☐ Global Ecology and Biogeography
- ☐ One Ecosystem
- ☐ Open Geospatial Data, Software and Standards

- Liste de revues publiant des *Data papers* (themes du Cirad)

  Accessible sur le site CoopIST du Cirad: https://coop-ist.cirad.fr/

  https://coop-ist.cirad.fr/gerer-des-donnees/rediger-un-data-paper/1-qu-est-ce-qu-un-data-paper

Bénéfices du Data paper

paternité

visibilité

Valoriser
votre travail et
vos données

reconnaissance

notoriété

citations

Article publié

- évalué par les pairs
- citable (DOI)
- Revue à FI
- complémentaire
  d'1 article de recherche

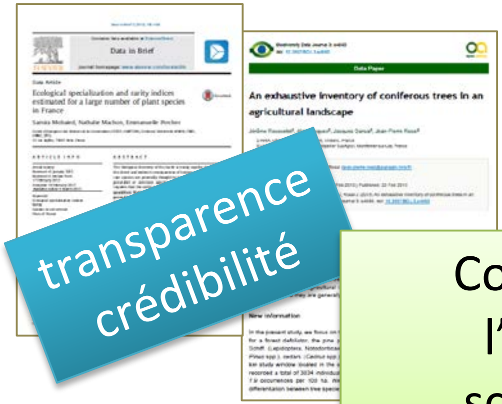
↗ visibilité sur vos données

↗ chances de réutilisation de vos données

Data paper

Entrepôt de données

collaborations



Data in Brief

Ecological specialization and rarity indices estimated for a large number of plant species in France

Data Paper

An exhaustive inventory of coniferous trees in an agricultural landscape

transparence
crédibilité

Contribuer à
l'intégrité
scientifique

Préservation pérenne
des données

financements

Merci pour votre attention

Laurence.dedieu@cirad.fr

Où publier ?

Base d'informations sur les revues en sciences du vivant, sciences sociales et sciences de l'ingénieur appliquées à l'agriculture

Rechercher une revue par ...

● tous les mots
○ au moins un mot

AGRICULTURE, AGRO-ALIMENTAIRE, FORÊTS
- Agriculture
- Production végétale et filières
- Protection des plantes
- Foresterie, agroforesterie, bois
- Production et santé animales, filières
- Technologie, sciences des aliments

SCIENCES SOCIALES
- Économie, sociologie, développement
- Géographie

SCIENCES DE L'ENVIRONNEMENT
- Environnement, ressources naturelles
- Écologie, biodiversité, conservation

SCIENCES DU VIVANT
- Biologie, biochimie
- Génétique, biotechnologie, biologie moléculaire
- Microbiologie, parasitologie
- Entomologie, nématologie

SCIENCES ET TECHNIQUES
- Sciences et techniques
- Géomatique, télédétection
- Modélisation, mathématiques, informatique
- Physique, chimie
- Sciences de l'information

http://ou-publier.cirad.fr/

Liste de revues publiant des datapapers :

https://coop-ist.cirad.fr/gerer-des-donnees/rediger-un-data-paper/1-qu-est-ce-qu-un-data-paper