

# Project Milestone #1

## Project Details:

### DatSci Werx

Team #39

Tristan De Alwis (tdealwis) 50%

Weiran Lin (wlin25) 50%

## Problem Statement:

We would like to create a database that stores and links published researched papers within the field of data science. The user will be able to search/query a term(s) for relevant documents. These documents can then be viewed along with important information such as its relevant topic, the author, date of publishing, publishing Journal, organization and funding agency. More importantly, this database is citation based, meaning that we also have links to one document's citations/references and its own citation information in other works forming a spider web of data science research.

## Target User:

- The main users will be librarians, and researchers who would likely be searching for relevant scientific research of a particular study for their own work.
- Computer Science or Data Science students who want to know what are hot research topics in data science.
- An administrator would likely be an information technologist attached to a scientific library or other place that collectively stores published works.

## List of Relations:

1. Research Paper w/ its corresponding information  
Attributes: Title, DOI(primary key), Author, Date, Source Titles, Abstract, Keywords  
Data Types: VARCHAR, Float, VARCHAR, DATE, VARCHAR, TEXT, VARCHAR
2. Citations: The citations used in each paper CHAR, TEXT  
Attributes: DOI, Citation Title, Citation DOI  
Data Types: FLOAT, VARCHAR, FLOAT
3. Author  
Attributes: Name, ORCID(primary key), Affiliated Organization, Publications, Citations  
Data Types: VARCHAR, FLOAT, VARCHAR, INT, INT
4. Source Titles/Publisher  
Attributes: Name, ISSN(primary key), Impact Factor, Research Domain  
Data Types: VARCHAR, CHAR, FLOAT, VARCHAR

# Web-Interface:

1. Search papers by topic, title, author, source titles, year published and organization

## Web of Science

Clarivate Analytics

Tools ▾ Searches and alerts ▾ Search History Marked List

Select a database Web of Science Core Collection ▾

Learn about alerting enhancements!

Basic Search

Author Search BETA

Cited Reference Search

Advanced Search

Structure Search

Larracuente A\*

✕

Author ▾

Select from Index

And ▾

University of Rochester

✕

Organization-Enhanced ▾

Search

Search tips

Select available organizations from the Index  
Finds papers from organizations with identified name variants.

+ Add row | Reset

Timespan

All years (1900 - 2020) ▾

More settings ▾

University of Rochester

Who are the most influential researchers of the last decade? See the list of Highly Cited Researchers 2019

The result is shown like below:

Web of Science InCites Journal Citation Reports Essential Science Indicators EndNote Publons Kopernio

Sign In ▾ Help ▾ English ▾

Web of Science

Clarivate Analytics

Search

Tools ▾ Searches and alerts ▾ Search History Marked List

Results: 18  
(from Web of Science Core Collection)

View author record(s) for:  
larracuente a\*

You searched for: AUTHOR: (Larracuente A\*) AND ORGANIZATION-ENHANCED: (University of Rochester)  
...More

Create an alert

Refine Results

Search within results for... 🔍

Filter results by:

☐ Highly Cited in Field (1)

☐ Open Access (16)

☐ Associated Data (2)

Refine

Publication Years ▲

Sort by: Date Times Cited **17** Usage Count Relevance More ▾

1 of 2 ▹

☐ Select Page

Export to Other File Formats

More ▾

Add to Marked List

☐ 1. Evolution of genes and genomes on the Drosophila phylogeny

[Associated Data](#)

By: Clark, Andrew G.; Eisen, Michael B.; Smith, Douglas R.; et al.  
Group Author(s): Drosophila 12 Genomes Consor; Broad Inst Genome Sequencing; Broad Inst Whole Genome Ass  
NATURE Volume: 450 Issue: 7167 Pages: 203-218 Published: NOV 8 2007

Find@UoR

Free Full Text from Publisher

View Abstract ▾

☐ 2. The Ecology and Evolutionary Dynamics of Meiotic Drive

By: Lindholm, Anna K.; Dyer, Kelly A.; Firman, Renee C.; et al.  
TRENDS IN ECOLOGY & EVOLUTION Volume: 31 Issue: 4 Pages: 315-326 Published: APR 2016

Find@UoR

Full Text from Publisher

Free Published Article From Repository

View Abstract ▾

☐ 3. The Selfish Segregation Distorter Gene Complex of Drosophila melanogaster

By: Larracuente, Amanda M.; Presgraves, Daven C.  
GENETICS Volume: 192 Issue: 1 Pages: 33-53 Published: SEP 2012

Find@UoR

Free Full Text from Publisher

View Abstract ▾

☐ 4. Genetics on the Fly: A Primer on the Drosophila Model System

By: Hales, Karen G.; Korey, Christopher A.; Larracuente, Amanda M.; et al.  
GENETICS Volume: 201 Issue: 3 Pages: 815-842 Published: NOV 2015

Analyze Results

Create Citation Report

Times Cited: 1,303  
(from Web of Science Core Collection)

Usage Count ▾

Times Cited: 100  
(from Web of Science Core Collection)

Highly Cited Paper

Usage Count ▾

Times Cited: 90  
(from Web of Science Core Collection)

Usage Count ▾

Times Cited: 44  
(from Web of Science Core Collection)

2. In each paper, we can click to see its basic information and links to the papers it citing from and the papers cited from it.

[Web of Science](#) [InCites](#) [Journal Citation Reports](#) [Essential Science Indicators](#) [EndNote](#) [Publons](#) [Kopernio](#) [Sign In](#) [Help](#) [English](#)

# Web of Science

[Search](#) [Search Results](#) [Tools](#) [Searches and alerts](#) [Search History](#) [Marked List](#)

[FindFull](#) [Free Full Text from Publisher](#) [Look Up Full Text](#) [Full Text Options](#) [Export to Other File Formats](#) [More](#) [Add to Marked List](#)

1 of 18

## Evolution of genes and genomes on the Drosophila phylogeny

[Associated Data](#)

By: Clark, AG (Clark, Andrew G.); Eisen, MB (Eisen, Michael B.); Smith, DR (Smith, Douglas R.); Bergman, CM (Bergman, Casey M.); Oliver, B (Oliver, Brian); Markow, TA (Markow, Theresa A.); Kaufman, TC (Kaufman, Thomas C.); Kellis, M (Kellis, Manolis); Gelbart, W (Gelbart, William); Iyer, VN (Iyer, Venky N.)...More

Group Author(s): Drosophila 12 Genomes Consor ; Broad Inst Genome Sequencing ; Broad Inst Whole Genome Ass

[View Web of Science ResearcherID and ORCID](#)

**NATURE**  
Volume: 450 Issue: 7167 Pages: 203-218  
DOI: 10.1038/nature06341  
Published: NOV 8 2007  
Document Type: Article  
[View Journal Impact](#)

**Abstract**  
Comparative analysis of multiple genomes in a phylogenetic framework dramatically improves the precision and sensitivity of evolutionary inference, producing more robust results than single-genome analyses can provide. The genomes of 12 Drosophila species, ten of which are presented here for the first time (sechellia, simulans, yakuba, erecta, ananassae, persimilis, willistoni, mojavensis, virilis and grimshawi), illustrate how rates and patterns of sequence divergence across taxa can illuminate evolutionary processes on a genomic scale. These genome sequences augment the formidable genetic tools that have made Drosophila melanogaster a pre-eminent model for animal genetics, and will further catalyse fundamental research on mechanisms of development, cell biology, genetics, disease, neurobiology, behaviour, physiology and evolution. Despite remarkable similarities among these Drosophila species, we identified many putatively non-neutral changes in protein-coding genes, non-coding RNA genes, and cis-regulatory regions. These may prove to underlie differences in the ecology and behaviour of these diverse species.

**Keywords**  
**KeyWords Plus:** AMINO-ACID SUBSTITUTION; ADAPTIVE PROTEIN EVOLUTION; SYNONYMOUS CODON USAGE; FRUIT-FLY; MOLECULAR EVOLUTION; MITOCHONDRIAL-DNA; NATURAL-SELECTION; FAMILY EVOLUTION; NONCODING DNA; HOST-DEFENSE

### Citation Network

In Web of Science Core Collection

**1,303**  
Times Cited

[Create Citation Alert](#)

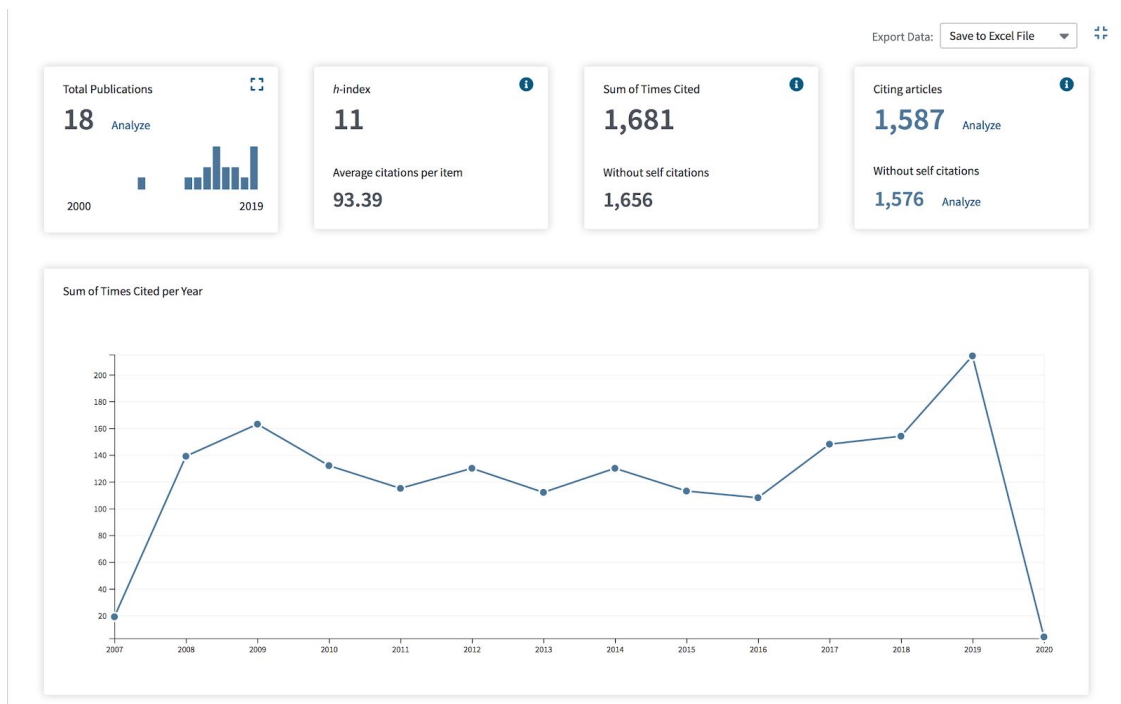
**All Times Cited Counts**  
3,597 in All Databases  
[See more counts](#)

**151**  
Cited References  
[View Related Records](#)

**Associated Data: 12**  
[View the data associated with this record \(from Data Citation Index\)](#)

**Most recently cited by:**

### 3. Author Information Page



#### 4. Publisher Information Page

**NATURE**

**Impact Factor**  
**43.07** **45.819**  
2018 5 year

JCR @ Category	Rank in Category	Quartile in Category
MULTIDISCIPLINARY SCIENCES	<b>1 of 69</b>	<b>Q1</b>

Data from the 2018 edition of [Journal Citation Reports](#)

**Publisher**  
NATURE PUBLISHING GROUP, MACMILLAN BUILDING, 4 CRINAN ST, LONDON N1 9XW, ENGLAND  
**ISSN:** 0028-0836  
**eISSN:** 1476-4687  
**Research Domain**  
Science & Technology - Other Topics

Close Window

### Data:

We will likely populate the database with some searches from Web of Science and other collective databases. This will give us data science papers together with its corresponding information such as Authors, DOI, Keywords, etc; citation information; author information; publisher information. Therefore we could import those data to our 4 relation tables.