



CS636 Data Analytics with R Programming

Mid Term Project

By

Gabriel Gonalez (geg7)

Kelvinkumar Gandhi (ksg27)

Vikramsingh Sunil Nimbalkar (vsn22)

Source Journal:

<https://academic.oup.com/dnaresearch>

OXFORD
ACADEMIC


Sign In Register

DNA Research

Kazusa DNA Res. Inst.

Issues More Content Submit Purchase Alerts About

All DNA Research Advanced Search



Latest Issue
Volume 26, Issue 1
February 2019

Impact Factor
5.415

5 year Impact Factor
5.525

Editor-in-Chief
Michio Oishi


Editorial Board

About the journal

DNA Research is an internationally peer-reviewed journal which aims at publishing papers of highest quality in broad aspects of DNA and genome-related research ...

[Find out more](#)


Altmetric data now available on all articles.



Find out more about how Altmetric is used to demonstrate impact.

Browse issues

Year 2019 Issue Volume 26, Issue 1, February 2019, Pages 1–103 [Browse by volume](#)



Volume 26, Issue 1
February 2019
[Cover image](#)


ISSN 1340-2838
EISSN 1756-1663

Full papers

[< Previous](#)

Volume 26, Issue 1, February 2019


FULL PAPERS

Overcoming challenges in variant calling: exploring sequence diversity in candidate genes for plant development in perennial ryegrass (*Lolium perenne*) 

Elisabeth Veeckman; Sabine Van Glabeke; Annelies Haegeman; Hilde Muylle; Frederik R D van Parijs ...

DNA Research, Volume 26, Issue 1, February 2019, Pages 1–12,
<https://doi.org/10.1093/dnares/dsy033>


[Abstract](#) [View article](#) [Supplementary data](#)

Multi-tissue transcriptomes of caecilian amphibians highlight incomplete knowledge of vertebrate gene families 

María Torres-Sánchez; Christopher J Creevey; Etienne Kornobis; David J Gower; Mark Wilkinson ...

Review: Biological roles of glycans

By Ajit Varki

 [Read free online](#)

GLYCOBIOLOGY

Email alerts

[New issue alert](#)

[Advance article alerts](#)

[Receive exclusive offers and updates from Oxford Academic](#)

RSS Feeds

DNA Research - Latest Issue Only

Given a journal, your R code should be capable of fetching html pages of all articles automatically. For each article, you are required to extract the following 10 fields: **DOI, Title, Authors, Author Affiliations, Corresponding Author, Corresponding Author's Email, Publication Date, Abstract, Keywords, Full Text.**

□ Contribution of Group Members:

Vikramsingh Nimbalkar	<ul style="list-style-type: none"> • Developed R script to store a results into memory with the help of File Handling • Executed final script with help of Exception Handling in R. • Development and Testing and Debugging 	<ul style="list-style-type: none"> • dna_research_midterm_proj.R
Kelvinumar Gandhi	<ul style="list-style-type: none"> • Developed primary R scraping script • Development and Performance Optimization 	<ul style="list-style-type: none"> • dna_research_midterm_proj.R
Gabriel Gonzalez	<ul style="list-style-type: none"> • Generated Regular Expressions to match various patterns in text from web • Developed R script to extract predefined 10 fields from scraped html articles. • Development, Testing and Debugging 	<ul style="list-style-type: none"> • dna_research_midterm_proj.R

□ Challenges:

- First challenge in this project was huge number of articles. There are more than 20 years of articles in journal. And articles are not properly organized on journal webpage. So scraping each of those articles was very challenging for us.
- Second challenge in this project was unfamiliarity with some of R packages. Because we have first time use some of R packages like rvest, XML, etc to scrap the webpage and extracting required information.