



## Mid Term Project – Team 1

Aakash Asli | Shao Yu Chen | Tushar Asnani

# Scope of Project

Use R program to extract required information from specified website

Main page:

<https://genomebiology.biomedcentral.com>

Total articles: 6261

Required information:

10 fields: DOI, Title, Authors, Author Affiliations, Corresponding Author, Corresponding Author's Email, Publication Date, Abstract, Keywords, Full Text (Textual format)

The screenshot shows the homepage of BMC Genome Biology, part of Springer Nature. The header includes navigation links: Home, About, Articles, and Submission Guidelines. A search bar is present with the text "Search articles within this journal". A prominent "Genome Biology" title is displayed in a yellow box. Below the header, there are three main sections: a "Call for Papers: Microbiome Biology" featuring a colorful 3D visualization of a microbial community; an "Article collection: Insights from Genome Editing" featuring a chromosome map; and a "2017 Journal Metrics" section listing citation impact statistics.

BMC Part of Springer Nature Explore Journals Get Published About BMC Login Search

## Genome Biology

Home About Articles Submission Guidelines

Search articles within this journal Submit a manuscript

**Call for Papers: Microbiome Biology**

**Article collection: Insights from Genome Editing**

**2017 Journal Metrics**

**Citation Impact**  
13.2 - 2-year Impact Factor  
16.5 - 5-year Impact Factor  
3.1 - Source Normalized Impact per Paper (SNIP)  
12.7 - SCImago Journal Rank (SJR)

Contribution of Group Members

Aakash Asli	main program: (loadAllArticles.R) Functions(util.R): analysisArticle extractAuthors extractAffiliation	Testing and debugging Project report(PDF file)
Tushar Asnani	Function(util.R): • loadArticleList	Testing and debugging Readme.txt
Shao Yu Chen	Function(util.R): extract extracAttribute	Testing and debugging

Challenges

▪ Carefully check the HTML of the article page and find the xpath	• Unfamiliar with R packages, like XML
▪ Use a 2 <sup>nd</sup> time extraction to get more details or related data.	• Check the online documents and samples
▪ Used httr package to use URL() as getURL() needed tlv error	• Extract author, corresponding author, corresponding author's email



Thank you