Using R in Tiki for Bioinformatics

Xavier de Pedro[1,1], Alex Sánchez[1],

[1] Departament d'Estadística. Facultat de Biologia. University of Barcelona

Avda Diagonal 645, 08028 Barcelona. Catalonia. Spain

Abstract. The need to work with colleagues from other institutions is quite common in Science. Teams often look for tools which allow coordinating with others in web platforms which enhance collaboration across space and time. Although data analysis and visualization is getting very popular using the free statistical software R, it still seems to lack some easy complementary program to allow creating quickly a web interface to support the usual workflows in Bioinformatic tasks such as Microarray analysis. This communication focuses on analyzing the state of the art of web interfaces to R scripts in Life Sciences, with a new add-on developed for a mature general purpose and free software Wiki CMS/Groupware platform: PluginR for Tiki. A use case for Microarray analysis will be described. More information: htp://estbioinfo.stat.ub.es

Keywords: R, Tiki, Web Interface, Free Software, Microarray Analysis.

1 Introduction

1.1 "R" for Bioinformatics

R is a free/open source software environment for statistical computing and graphics, and it is commonly used by many life scientists in (omics) data analysis, among other disciplines [1]. R can be seen as an strategic tool to be adopted to perform data analysis and visualization in science, because it is being used worldwide and its growth is exponential [2]). The CRAN (Comprehensive R Archive Network) system is the key contributing factor in this growth: R's package system together with the CRAN infrastructure provides a standardized process for authoring, documenting, validating, building and distributing packages to millions of users worldwide. As an example in the Life Sciences ground, researchers using R benefit also from Biocondcutor, one the most commonly used repositories in R for Bioinformatics.

1.2 Our needs

The research groups in which we participate often require from some web interface to interact with R scripts. Or if they don't require it, they would greatly benefit from it, if it existed. Of course we can manually create web pages based on reports generated from R packages or by other means, but some web application which was

---

[1] Corresponding author: xavier.depedro@ub.edu

able to interact with R allowing to write reports on the fly was desired. Like what can be achieved with with SWeave package but with the flexibility of Wiki (quick) syntax, and the versatility of content management systems and groupware applications on-line.

In this area, we were seeking some general purpose tool to extend R usage through a web interface, which could be useful for Bioinformatics (and in fact, for any field where data analysis and visualization is required). We wanted it to comply with the following requirements to be sustainable enough for us in the mid and long-runs:

- It had to be free/open source software, multi-platform and multi-browser
- It had to be based on some mature and maintained software package
- It had to offer extensive-enough documentation on its usage
- It had to require standard technology and programming languages so that it was easily extend-able by us or on demand by many companies or consultants, and not just a single company.
- It had to be versatile enough, so that if we could use the same tool for other purposes in our research groups and departments, even better.
- It had to allow quick and easy web output or reports from those scripts.

1.3 Web interfaces for R

There are a few interfaces listed in the R documentation and "frequently asked questions" pages [3], which could apparently fulfill some of our needs. However, our experience is that most of them are either unmaintained (and thus, risky to rely on them in the mid term), or doesn't work anymore, or they are too difficult to make them work for an averaged researcher or technician. Other researchers came to similar conclusions to ours (Table 1). Besides the cited applications in the previous references, there a few others which would deserve some attention.

Galaxy and Rgalaxy [5]: a Web application written in perl that is highly customized for genomic analysis for pipelines, which allows users to do analyses that you cannot do anywhere else without the need to install or download anything. R is used as an external tool, that can be queried from Galaxy using some additional R toolkits. However, it is highly specific for just some Bioinformatics tasks.

Taverna and Plugin Rshell [6] imply an open source and domain independent Workflow Management System written in Java (a suite of tools used to design and execute scientific workflows and aide in silico experimentation), and an additional plugin to manage R scripts from within the Taverna Platform. Again, it is highly specific for just some Bioinformatics tasks.

Mediawiki and R Extension [7], represent a well known free software wiki engine written in PHP and Mysql, with an optional extension which allows to integrate R output (raw, html and graphics) into mediawiki pages. This is general enough for our needs, and easy to install and manage. However, Mediawiki is (just) a very powerful wiki engine to produce encyclopedia-like sites, and not versatile enough as a CMS/Groupware solution.

Zope and R [8] are another combination of a free application framework used by many collaborative projects (such as Plone, among others), and some tools for integrating R with it for developing WWW-based statistical environments. These include RSOAP (simple API for managing and communicating with multiple concurrent R processes), RSessionDA (Zope objects for interacting with R), and others. However, Zope is not installed as a standard set of packages in most commercial and university servers, and it seems only an interesting path if you are already using Zope application framework for other purposes, which is not our case.

RevoDeployR [9] was announced in public in late summer 2010 by Revolution Analytics as one add-on for R in their 'Revolution R Enterprise' product. RevoDeployR is a Web Services framework for R in Java and Javascript, designed to make it easy to scalably and securely integrate computations done in R into other applications like spreadsheets and web pages. We have not been able to review this new tool (too recently announced, by the time of this communication), and we wonder how "free" (as in freedom) this tool will be. To be analyzed in the future.

Since there is no clear option which fulfilled all our needs, we finally came out to the conclusion that extending some general-purpose free software web application to interact with R was also a potentially suitable option for us. We chose for such task a general-purpose Wiki CMS/Groupware application called “Tiki” [10].

2 Tiki and PluginR

2.1 Why Tiki

Tiki can be defined as a "Tightly Integrated Knowledge Infrastructure". It was born in 2002 as a versatile and general purpose free software platform written in PHP and Mysql. It has many plugins which extend the basic wiki syntax to perform many advanced operations, taking profit of the long list of features available in any Tiki installation. One of the main differences as a software package is that almost all those features and plugins are bundled within each Tiki release, highly integrated to take profit of the other features from Tiki. Thus, if this software had some interface to query R in the server, it would mean somewhat a similar solution than the one found for MediaWiki engine, but within some full featured web application. This way we could define our own workflows for Bioinformatics or any other area, and the investment in time to learn a new tool could be profited for other purposes of our research groups and projects.

But why Tiki, among all those general purpose free software Wiki [11] and CMS [12] web applications? Because of:

- Wiki way of making software, with 3 basic rules to successfully embrace Erik Raymonds "bazaar model" [13], allowing a few awards [14].
- Smarty template engine to split presentation and layout from php code
- Plugin based system of extensions, bundled within the software
- All-in-one design, except when license restrictions apply, even within free software applications like in this case with PluginR

Table 1. Review from Neil Saunders (personal communication in his blog [4]).

Software	Brief notes
Rweb	Page last updated 1999. Of the 3 example links on the page one ran very slowly, the second not at all and the third is broken.
R-Online	Or rather, not online. Unless this CGI form is the same thing. I tried Example 1, it returned a server error.
Rcgi	Links to several CGI forms, none of which worked for me.
CGI-based R access	Link did not load.
CGIwithR	Package now maintained at Omegahat. Did not attempt installation. Last updated 2005.
Rpad	I could not connect to this URL.
RApache	The pick of the bunch. Provides server-side access to R through an Apache module. I was able to install RApache on 32-bit (but not 64-bit) Ubuntu 9.10 and get it running. Could use more documentation.
Rserve	Serves R via TCP/IP. Last updated 2006.
OpenStatServer	Broken link. No longer exists, so far as I can tell.
R PHP Online	Link out of date (but you can follow it to the newer page). Last updated 2003, so unlikely to be much use.
R-php	Last updated 2006; the example that I tried gave a server error.
Webbioc	A Bioconductor package. Did not investigate further.
Rwui	An application to create R web interfaces. My browser hung at 'waiting for cache'. I gave up.

- Subversion based repositories, like in R
- Fine-grained permission management on groups of users at a global level, single object level and/or content-category level.
- Configuration profiles: Community-created configuration profiles applicable in one click to your Tiki installation. Hosted at profiles.tiki.org, a public site which allows to share your highly customized configuration also.
- Using GNU/LGPL copyright license: Free to copy, study, extend, share...

2.2 Tiki Trackers

Trackers in Tiki are another key feature for our needs, since they act as a form and database generator. Combined with Wiki pages and plugin calls to query and report on trackers, you can define simple workflows to collect data and show it later filtered on multiple criteria, allowing to produce highly customized reports based on "Pretty Trackers" [15]. Also, the category transition system in Tiki can be used to produce customized workflows.

2.3 PluginR

Our own R scripts are run through Tiki by means of a newly developed plugin called “PluginR” [16], which queries R in the server and shows the results in wiki pages within the same platform. PluginR is still in its early stages of development at its version 0.5, but it already provides basic functionality to safely run R scripts from within Tiki Wiki CMS/Groupware, showing the output as text or images on demand.

The execution of potentially risky plugins like PluginR is managed by a permission system, and only those trusted users (usually administrators of the Tiki site or which ever group of users has that permission granted) can approve the execution of custom R scripts, once ensured that they are safe for the server.

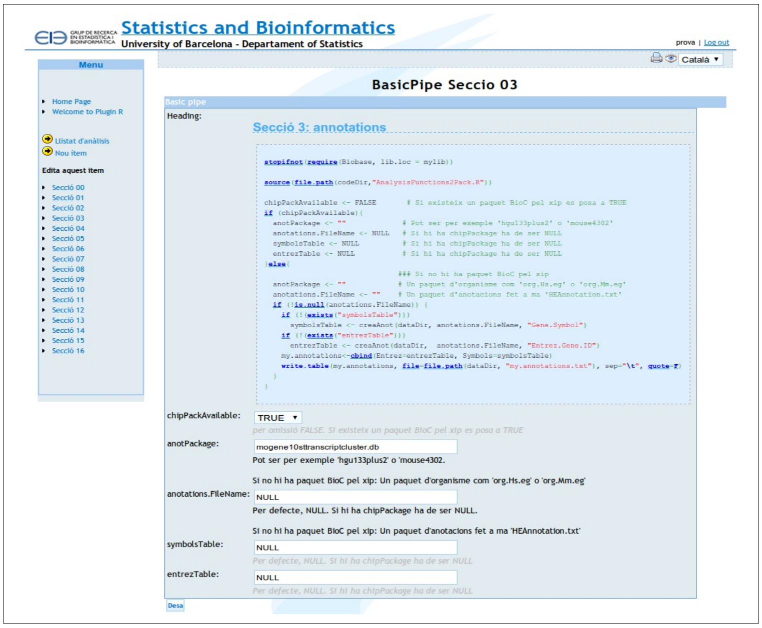
In its current state, PluginR is released under the GNU/GPL license, because some parts of the code come from the R-PHP package and the Mediawiki R Extension, so that it has to be downloaded and installed separately of the main Tiki installation.

A working example of Tiki and PluginR combined together in a production environment can be found in the research field of educational technologies for experimental sciences [17].

2.4 Creating custom workflows: Basic Pipelines in the EiB group

There are many customized software packages to create and manage workflows in Bioinformatics [18]. However, in our Statistics and Bioinformatics research group (EiB from its Catalan name) with the "use one versatile tool for all" approach, we can do our own workflows for Bioinformatics (or for any discipline) combining Tiki trackers with R scripts in wiki pages through PluginR calls (Figure 1). In addition, we have full control to grant permissions to set "who can see or do what" as we want.

Thus, we are designing a set of forms to collect data for microarray pipeline analysis, and we provide these values to custom R scripts online to produce the required reports. Further details will be given at the communication.



Heading: Secció 3: annotations

```
stopifnot(require(Biobase, lib.loc = mylib))

source(file.path(codeDir,"AnalysisFunctions2Pack.R"))

chipPackAvailable <- FALSE       # Si existeix un paquet BioC pel xip es posa a TRUE
if (chipPackAvailable){
  anotPackage <- ""             # Pot ser per exemple 'hgu133plus2' o 'mouse4302'
  anotations.FileName <- NULL   # Si hi ha chipPackage ha de ser NULL
  symbolsTable <- NULL          # Si hi ha chipPackage ha de ser NULL
  entrezTable <- NULL           # Si hi ha chipPackage ha de ser NULL
}else{
                               ### Si no hi ha paquet BioC pel xip
  anotPackage <- ""            # Un paquet d'organisme com 'org.Hs.eg' o 'org.Mm.eg'
  anotations.FileName <- ""    # Un paquet d'anotacions fet a ma 'HEAnnotation.txt'
  if (!is.null(anotations.FileName)) {
    if (!(exists("symbolsTable")))
      symbolsTable <- creaAnot(dataDir, anotations.FileName, "Gene.Symbol")
    if (!(exists("entrezTable")))
      entrezTable <- creaAnot(dataDir,  anotations.FileName, "Entrez.Gene.ID")
    my.annotations<-cbind(Entrez=entrezTable, Symbols=symbolsTable)
    write.table(my.annotations, file=file.path(dataDir, "my.annotations.txt"), sep="\t", quote=F)
  }
}
```

chipPackAvailable: TRUE ▾

per omissió FALSE. Si existeix un paquet BioC pel xip es posa a TRUE

anotPackage: mogene10sttranscriptcluster.db

Pot ser per exemple 'hgu133plus2' o 'mouse4302.

Si no hi ha paquet BioC pel xip: Un paquet d'organisme com 'org.Hs.eg' o 'org.Mm.eg'

anotations.FileName: NULL

Per defecte, NULL. Si hi ha chipPackage ha de ser NULL.

Si no hi ha paquet BioC pel xip: Un paquet d'anotacions fet a ma 'HEAnnotation.txt'

symbolsTable: NULL

Per defecte, NULL. Si hi ha chipPackage ha de ser NULL

entrezTable: NULL

Per defecte, NULL. Si hi ha chipPackage ha de ser NULL

Figure 1: Sample form in Tiki to collect data for PluginR.

2.5 Future work

We have identified a few areas which can be improved to facilitate even more the usage of R scripts for Bioinformatics within Tiki. These comprise a higher degree of compatibility between the Wiki & Smarty parsers in Tiki and R native syntax, allowing the usage of curly brackets in R scripts when used through Smarty Templates, as it can be used with Wiki pages. Moreover, some support for persistent sessions to limit data transfer would probably be implemented in the future.

Acknowledgments

We would like to thank the funding from Generalitat de Catalunya-CUR (ref. 2009SGR932), "Traces" project from REDICE at University of Barcelona, the researchers from the Cochise project where we have been using PluginR in its early stages of development. And Rodrigo Sampaio, Luciash, Louis-Philippe Huberdeau and Sylvie Greverend, from Tiki for their supportive help with PluginR development.

References

1. R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
2. John Fox. Aspects of the Social Organization and Trajectory of the R Project. The R Journal, 1(2):5-13, December 2009.
3. R FAQ: Frequently Asked Questions on R. http://cran.r-project.org/doc/FAQ/R-FAQ.html#R-Web-Interfaces
4. What You're Doing Is Rather Desperate: Notes from the life of a bioinformatics researcher (Neil Saunders). http://nsaunders.wordpress.com/2009/11/30/a-brief-survey-of-r-web-interfaces/
5. Rgenetics: statistical and population genetics for Bioconductor. http://rgenetics.org
6. Wassink et al. 2009. "Using R in Taverna: Rshell v1.2"
7. Extension:R for MediaWiki - http://www.mediawiki.org/wiki/Extension:R
8. Zope and R. http://rsoap.sf.net
9. RevoDeployR—Application Integration, Deployment & Administration For R. http://www.revolutionanalytics.com/products/RevoDeployR.php
10. Tiki Wiki CMS Groupware: Software made the wiki way. http://tiki.org
11. WikiMatrix - Compare them all. http://www.wikimatrix.org
12. CMS Matrix - The Compare Stuff Network. http://www.cmsmatrix.org
13. Hakan Erdogmus, "A Process That Is Not," IEEE Software, pp. 4-7, November/December, 2009.
14. Tiki Community: Awards. http://tiki.org/Awards
15. Tiki Documentation: Pretty Trackers. http://doc.tiki.org/Pretty+Tracker
16. Tiki Documentation: PluginR. http://doc.tiki.org/PluginR
17. X. de Pedro, M. Calvo, A. Carnicer, J. Cuadros, A. Miñarro. 2010. Assessing Student Activity Through Log Analysis From Computer Supported Learning Assignments. In International Congress of University Teaching and Innovation - CIDUI 2010, Barcelona July 2010. http://cochise.bib.ub.es
18. Wikipedia: Bioinformatics workflow management systems. http://en.wikipedia.org/wiki/Bioinformatics_workflow_management_systems