



# Large Scale Analyses and Visualization of Adaptive Amino Acid Changes Projects

Noé Vázquez<sup>1,2</sup> · Cristina P. Vieira<sup>3,4</sup> · Bárbara S. R. Amorim<sup>3,5</sup> · André Torres<sup>3,4</sup> · Hugo López-Fernández<sup>1,2,3,4</sup> · Florentino Fdez-Riverola<sup>1,2</sup> · José L. R. Sousa<sup>3,4</sup> · Miguel Reboiro-Jato<sup>1,2</sup> · Jorge Vieira<sup>3,4</sup> 

Received: 24 August 2017 / Revised: 9 January 2018 / Accepted: 17 January 2018 / Published online: 30 January 2018  
© The Author(s) 2018. This article is an open access publication

## Abstract

When changes at few amino acid sites are the target of selection, adaptive amino acid changes in protein sequences can be identified using maximum-likelihood methods based on models of codon substitution (such as codeml). Although such methods have been employed numerous times using a variety of different organisms, the time needed to collect the data and prepare the input files means that tens or hundreds of coding regions are usually analyzed. Nevertheless, the recent availability of flexible and easy to use computer applications that collect relevant data (such as BDBM) and infer positively selected amino acid sites (such as ADOPS), means that the entire process is easier and quicker than before. However, the lack of a batch option in ADOPS, here reported, still precludes the analysis of hundreds or thousands of sequence files. Given the interest and possibility of running such large-scale projects, we have also developed a database where ADOPS projects can be stored. Therefore, this study also presents the B+ database, which is both a data repository and a convenient interface that looks at the information contained in ADOPS projects without the need to download and unzip the corresponding ADOPS project file. The ADOPS projects available at B+ can also be downloaded, unzipped, and opened using the ADOPS graphical interface. The availability of such a database ensures results repeatability, promotes data reuse with significant savings on the time needed for preparing datasets, and effortlessly allows further exploration of the data contained in ADOPS projects.

**Keywords** ADOPS · Positive selection · B+ database · Open data

---

Noé Vázquez, Cristina P. Vieira and Bárbara S. R. Amorim have contributed equally to this work

---

✉ Jorge Vieira  
jbvieira@ibmc.up.pt

<sup>1</sup> ESEI – Escuela Superior de Ingeniería Informática, Edificio Politécnico, Universidade de Vigo, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain

<sup>2</sup> CINBIO - Centro de Investigaciones Biomédicas, University of Vigo, Campus Universitario Lagoas-Marcosende, 36310 Vigo, Spain

<sup>3</sup> Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

<sup>4</sup> Instituto de Biologia Molecular e Celular (IBMC), Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

<sup>5</sup> Instituto Nacional de Engenharia Biomédica (INEB), Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

## 1 Introduction

Amino acid changes in protein sequences can be adaptive, and when changes at few amino acid sites are the target of selection they can be detected using maximum-likelihood methods based on the models of codon substitution [1–3]. This approach has been applied numerous times to infer positively selected amino acid sites at numerous proteins including, but not limited to: interleukin-3 (IL3), a protein associated with brain volume variation in general human populations [4]; formyl peptide receptors in mammals [5]; scorpion sodium channel toxins [6]; the *Mimulus* plant CENH3 protein [7]; the oyster *Crassostrea gigas* peptidoglycan recognition proteins [8]; host immune response genes [9, 10]; the envelope glycoprotein of dengue viruses [11]; the attachment glycoprotein of respiratory syncytial virus [12]; measles virus hemagglutinin [13]; influenza B virus hemagglutinin [14]; HIV proteins [15]; hemagglutinin-neuraminidase protein of Newcastle disease virus [16]; *Trypanosoma brucei* proteins [17]; the vertebrate skeletal muscle sodium

channel protein [18]; the p53 protein [19]; the fruitless protein in *Anastrepha* fruit flies [20]; CC chemokine receptor proteins [21]; or the proteins encoded by plant genes that are involved in gametophytic self-incompatibility specificity determination [22–25]. Recently, it has been argued that pharma and biotech industries can successfully use the knowledge generated by such an approach to deal with real-life problems [26].

Although maximum-likelihood methods based on models of codon substitution have been widely used to infer positively selected amino acid sites, the size of the average project is still relatively small mainly due to the time needed to collect the relevant coding sequences and prepare input files for the different software applications. The recent availability of computer applications such as *Blast DataBase Manager* (BDBM; <http://www.sing-group.org/BDBM/>) greatly eases the preparation of large datasets. Moreover, the availability of the *Automatic Detection of Positively Selected Sites* (ADOPS) [27] computer application has allowed the automated execution of all the steps needed to infer positively selected amino acid sites, starting from a FASTA file with non-aligned coding sequences; however, the lack of a batch option in this application still means that it is not practical to run thousands of sequence files.

This study will report the implementation of a batch option in the ADOPS software [27] that allows users to easily run large scale analyses involving thousands of genes, using moderate computer resources. Given this improvement, the next logical step is to make ADOPS projects (especially large-scale projects) available to the research community. To that end, we also present B+ (<http://bpositive.i3s.up.pt/>), a database that has been specifically designed to store and show the information contained in ADOPS project files. Although a database dedicated to positive selection inferences at the codon level has already been published [28], it is dedicated to a specific group of organisms, and the possibility of reusing data is not as easy as with B+ and ADOPS. Both large and small ADOPS datasets can be submitted to B+ (as compressed tar.gz files) along with a description containing the details about how the project was performed. At present, the B+ database hosts the “Closely related *Drosophila* dataset (2016)”, which provides ADOPS projects for 19,652 *Drosophila* transcripts, 14.6% of which show signs of positive selection (1200 genes), although curated analyses must now be performed to validate these results.

## 2 ADOPS Batch Mode

Multiple instances of the ADOPS graphical user interface (GUI) can be launched simultaneously, which correspondingly allows for the possibility of multiple parallel

processing of ADOPS projects. However, this is only possible if a sufficient amount of memory is available, which in turn is determined by the number of sequences used in the project and the total number of individual projects to be run. A single ADOPS batch project with 50 individual projects, each with an average of 10 sequences per individual project, runs in approximately 1–2 days. This suggests that even with limited computational power it is possible to run about 100 individual projects every 2 days.

To launch the new batch option implemented in ADOPS, the user launches the GUI and chooses the ‘*Create Batch Project*’ option under the ‘*Project*’ menu (Fig. 1). The user then gives the name and location of the folder that will contain the individual ADOPS project files. The base configuration can be changed at this time; however, if none is specified the default configuration stored in the ‘*system.conf*’ file will be used. Finally, the user selects the FASTA files that will be used for the experiments and a new window is launched, showing the status (including the detection of positively selected amino acid sites) of each individual ADOPS project, each of which is named according to the names of FASTA files (Fig. 1). The name of the experiment of each individual project will be named “batch”.

## 3 B+ Database

The B+ database is both a data repository and a convenient interface to browse the information contained in each ADOPS project interactively, without the need to download and unzip the corresponding ADOPS project. Thus, B+ allows the effortless exploration of the data contained in ADOPS projects.

The common workflow when performing a study with ADOPS and B+ is summarized in Fig. 2. A researcher with a dataset to be studied, comprising several FASTA files, will use the ADOPS’ batch mode to analyze each FASTA file. On this mode, ADOPS will create a new project for each file, to analyze it. At the same time, the researcher will create a new dataset in B+ by simply providing a name and a description of the related study. Once the dataset is created, the ADOPS projects previously created can be uploaded even if they have not yet been fully analyzed, as they can be reloaded when completed. B+ will store the dataset and projects meta-information in a relational database, while the project files are stored in a local folder. Whenever appropriate, the researcher can make the dataset public, allowing any user to explore it.

B+ is open to any researcher wanting to share the result of large and small-scale analyses done with ADOPS, although contributing credentials are granted under request.

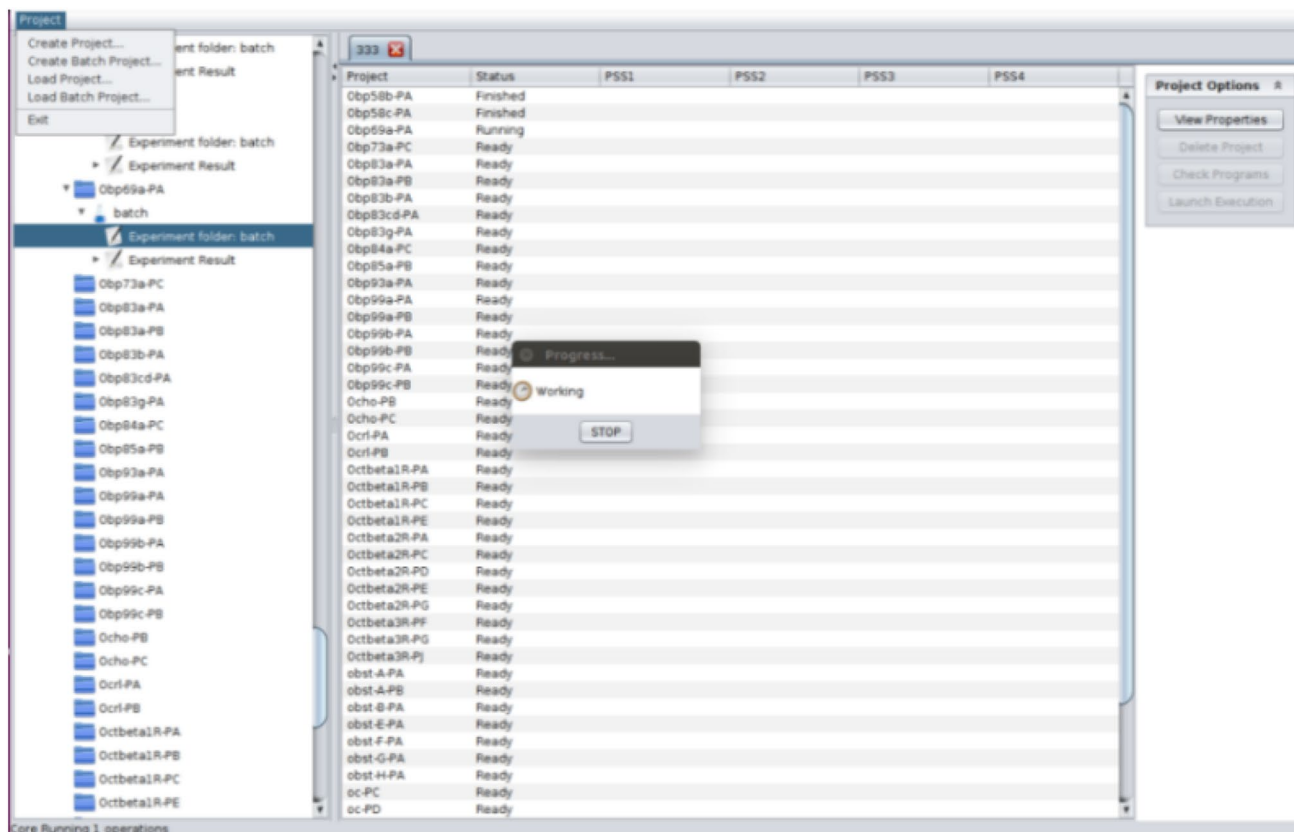


Fig. 1 The ‘Create Batch Project’ option

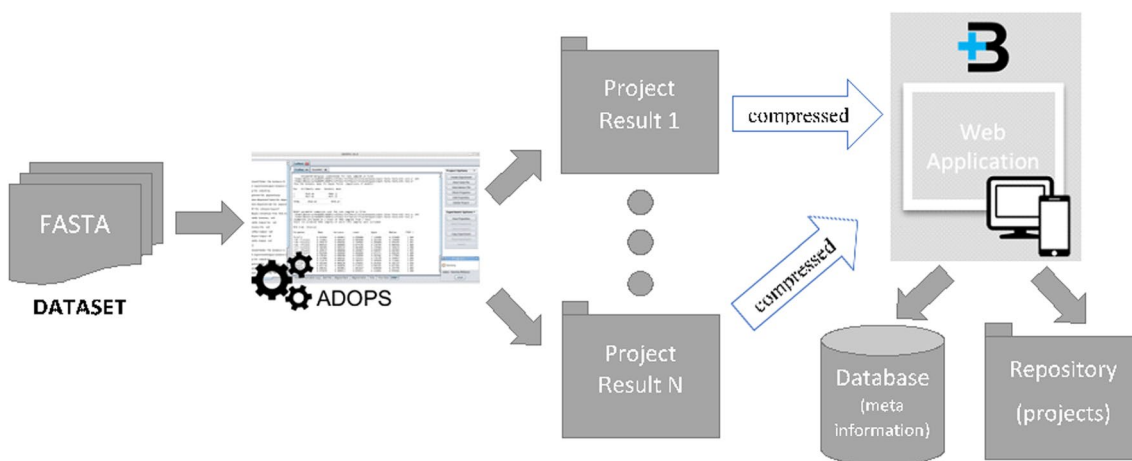


Fig. 2 Graphical representation of the common workflow that researchers follow when performing large studies with ADOPS and B+

### 3.1 Implementation

B+ was developed using the Laravel framework (<https://laravel.com/>) for web development and MySQL for database management. For a richer user interface, the

Bootstrap framework (<http://getbootstrap.com/>) and the jQuery library (<https://jquery.com/>) were also used.

B+ repository is available at <http://bpositive.i3s.up.pt/> and its source code is publicly available at <https://github.com/sing-group/bpositive>, under a GNU GPL 3.0 Open Source License (<http://www.gnu.org/copyleft/gpl.html>).

### 3.2 Database Exploration

The B+ repository exploration interface is divided into three visualization levels: (1) dataset list level, introducing each dataset available under the platform; (2) dataset view, showing the projects which compose a dataset; and (3) the project view, presenting the different result views of a project. Thus, the visualization levels are arranged from the most general to the most detailed view of the stored data.

At the dataset list level (Fig. 3) users have access to all the datasets stored in the B+ repository. For each dataset, a title, a unique identifier and a brief description of the performed analysis are presented. In addition, “Open” or

“Access” buttons are shown, depending on whether the dataset is public or private.

The dataset view (Fig. 4) allows user to explore the projects that compose a dataset. As the full analysis of a dataset can take several days (even months), B+ allows the storage of datasets partially analyzed, which enables users to have early access to the completed ADOPS projects results. Moreover, once a project is analyzed, the presence of positively selected amino acid sites can be confirmed or discarded. On this basis, any project in a dataset can be classified into one of three different states: (1) not analyzed, when the project has not yet been analyzed; (2) analyzed, when the project was analyzed but no positively selected amino acid sites were identified; or (3) positively selected, when



Fig. 3 Dataset list view presenting several datasets

# Closely related Drosophila September 2016 (BP2016000001)

Back

Filters		Rows by page	Search		
<input checked="" type="radio"/> Positively Selected <input type="radio"/> Analyzed <input type="radio"/> Not Analyzed <input type="radio"/> All		10	Type your query here	contains	✖
Name	Experiment	Result	View	Download	Select
ClustalW2	ClustalW2	Positively Selected			<input type="checkbox"/>
ClustalW2	ClustalW2	Positively Selected			<input type="checkbox"/>
AP-1gamma-PE	ClustalW2	Positively Selected			<input type="checkbox"/>
AP-1gamma-PF	ClustalW2	Positively Selected			<input type="checkbox"/>
AP-1gamma-PG	ClustalW2	Positively Selected			<input type="checkbox"/>
AP-1gamma-PH	ClustalW2	Positively Selected			<input type="checkbox"/>
AP-1gamma-PI	ClustalW2	Positively Selected			<input type="checkbox"/>
aPKC-PK	ClustalW2	Positively Selected			<input type="checkbox"/>
Aplip1-PA	ClustalW2	Positively Selected			<input type="checkbox"/>
Aplip1-PB	ClustalW2	Positively Selected			<input type="checkbox"/>

Showing 61 to 70 of 2880 entries.

Fig. 4 Dataset view presenting the ADOPS projects of a dataset

zormin-PF

Model 8: beta>w>1

Download as PDF | Download as PNG

Species	Sequence
D_melanoga	MERRQALDSS MSSLSVAVSS ---SNPNPRQ PHTGRLQQTQ ---QSQQSQQ SQLDAFNYS A TMSSSSPTPS QAMSTAMSTG TGTGTG---TT
D_simulans	MERRQALDSS MSSLSVAVSS ---SNPNPRQ PHTGRLQQTQ ---QSQQSQQ SQLDAFNYS A TMSSSSPTPS QAMSTAMSTG TGTGTG---TT
D_yakuba_2	MERRQALDSS MSSLSVAVSS ---SNPNPRQ PHTGRLQQTQ ---QS--- --LDAFNYS A TMSSSSPTPS QAMSTAMSTG TGTGTGTGTT
D_erecta_2	MERRQALDSS MSSLSVAVSS ---SNPNPRQ PHTGRLQQTQ ---QSQQSQQ SQLDAFNYS A TMSSSSPTPS QAMSTAMSTG TGTG---TT
D_takahashi	MERRQALDSS MSSLSVAVSS ---SNPNPNRQ NPN-RLQLQ ---TQQSQQ SQLDAFNYS A TMSSSSPTPS QAMSTAMSTG TTHIT---P
D_biarmipe	MERRQALDSS MSSLSVAVSS ---SNPNPRQ PPS-RLQLQ Q---SQSQSQ SQLDAFNYS A TMSSSSPTPS QAMSTAMSTG TGTGTGLGSG
D_suzukii	MERRQALDSS MSSLSVAVSS ---SNPNPRQ PQS-RLQLQ Q---SQSQSQ SQLDAFNYS A TMSSSSPTPS QAMSTAMSTG TGAGTIGGSG
D_eugracil	MERRQALDSS MSSLSVAVSS ---SNP--RQ PQT-RLQLQ ---QSQQSQQ SQLDAFNYS A TMSSSSPTPS QAMSTAMSTA MSTG---TT
D_ficusphi	MERRQALDSS MSSLSVAVSS ---SNP--RQ PQT-RLQLQ HSQQSQSQSQ SQLDAFNYS A TMSSSSPTPS QALSTAMSTG ---TG---TT
D_elegans_	MERRQALDSS MSSLSVAVSS SNPNPNPRQ PQM-RLQLQ -----QSQQS SQLDAFNYS A TMSSSSPTPS QAMSTAMSTG MGTGLG---TT

Fig. 5 Project view presenting an ADOPS project with positively selected amino acid sites

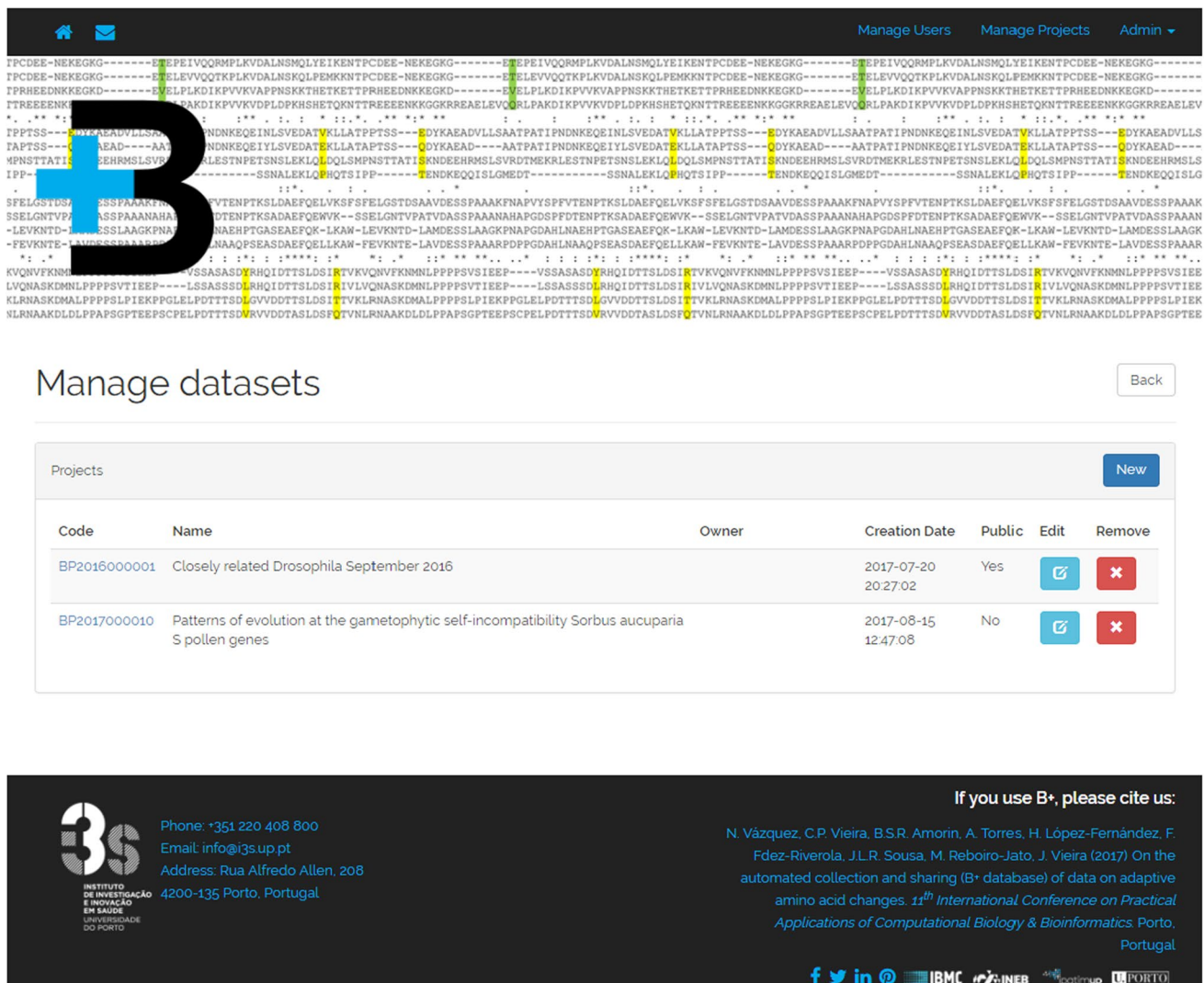


Fig. 6 Manage datasets view

the project was analyzed and positively selected amino acid sites were identified.

As seen in Fig. 4, the dataset view has a table arrangement of ten rows per page that can be changed using the “Number of entries” field. Several search options are provided at the top menu bar of the interface, and they can be executed in the database on the server side to provide maximum performance. The pagination is also handled on the server side to minimize the transfer of unnecessary data to the client. The free text search matches full and partial words using name and description fields of the database, also allowing the use of regular expressions.

The project view presents the results of an ADOPS project in the same way that the desktop application does. Each result is presented in a different tab, allowing users to explore them directly in the B+ web or to download them in the appropriate format.

As seen in Fig. 5, the first tab of the project view is a viewer for positively selected amino acid sites, which can be configured dynamically to match user preferences. It also allows downloading a PDF or PNG file with the result. Another tab that includes a viewer is the “Tree View”. Using the PhyD3 JavaScript library (<https://phyd3.bits.vib.be/>), this view shows an unrooted phylogram/cladogram for each tree available in the record. It can be also configured and the result can be downloaded in PNG or SVG formats. The remaining tabs present plain text documents.

### 3.3 Management Interface

B+ also features a management interface, available only to users with specific privileges, to control users and datasets registered in the repository (Fig. 6). Specifically, B+ allows for two different privileged user profiles: (1) administrators,

## Edit Dataset

Back

**Name**

**Description**

<p>
 We have used data on the <em>D. melanogaster</em> genome annotation (release 6.12) and data on the genome assemblies of 11 closely related <em>Drosophila</em>

Project edition is disabled when dataset is public

[Save](#)

Filters ▾

Rows by page  Search  contains ▾

id	Name	Experiment	Creation Date	View	Download	Analyzed	Positively Selected
106940	128up-PA	ClustalW2	2016-12-20 14:09:30			Yes	No
91348	14-3-3epsilon-PA	ClustalW2	2016-12-20 14:06:11			No	No
91349	14-3-3epsilon-PB	ClustalW2	2016-12-20 14:06:11			No	No
91350	14-3-3epsilon-PC	ClustalW2	2016-12-20 14:06:11			No	No
91351	14-3-3epsilon-PD	ClustalW2	2016-12-20 14:06:11			No	No
106941	14-3-3zeta-PA	ClustalW2	2016-12-20 14:09:30			Yes	No
106942	14-3-3zeta-PB	ClustalW2	2016-12-20 14:09:30			Yes	No
106943	14-3-3zeta-PC	ClustalW2	2016-12-20 14:09:30			Yes	No
106944	14-3-3zeta-PD	ClustalW2	2016-12-20 14:09:30			Yes	No
106945	14-3-3zeta-PE	ClustalW2	2016-12-20 14:09:30			Yes	No

« 1 2 3 4 5 6 7 8 ... 3044 3045 »

Showing 1 to 10 of 30442 entries.

Fig. 7 Edit dataset view

to manage all the users and datasets; and (2) contributors, to manage only their own datasets. Apart from managing the datasets metadata and the projects that compose it, dataset management also includes visibility control. In B+, projects can be public, which are visible by any user, or private, which can be viewed only by administrators, owner contributors, or those using a password set by administrators or contributors.

Every dataset uploaded to the repository is automatically checked to ensure that the ADOPS format is correct and that every result is displayed correctly. Datasets can be uploaded in bundle files including one folder for each ADOPS project, or one file for each project in a multiple upload. Allowable formats are “zip” and “tar.gz”. Once files are uploaded and validated, they are stored in the B+ repository and can be viewed immediately. Administrators can at any time edit the metadata and/or add, update or remove projects belonging

to a dataset, while owner contributors can do the same when the project state is private only (Fig. 7). To ensure the stability of public data, after a project has been made public by the owner contributor, only administrators can make it private again.

## 4 Usefulness

The first large scale dataset available at B+ is the “Closely related Drosophila dataset (2016)”. In brief, ADOPS projects for 19,652 Drosophila transcripts were generated (the details on how the sequence data were obtained and the analyses performed is provided at the B+ database under the project description), 14.6% of which show signs of positive selection (1200 genes), although human curated analyses must now be performed to validate these automatic inferences.

These analyses are the first step toward the identification of the genes and amino acid sites that contribute to adaptation.

While ADOPS is intended to be a flexible and easy to use pipeline aimed at making robust inferences on positively selected amino acid sites, the information contained in the B+ database may serve many additional purposes. For instance, since a Bayesian phylogenetic tree is always generated and the corresponding NEWICK tree file saved, a robust tree for the relationship of the species analyzed could be easily created using applications such as CLANN [29], which would then allow for the construction of supertrees from partially overlapping species datasets. Moreover, ADOPS projects always provide the nucleotide and protein sequences in FASTA format (aligned and non-aligned), which can be used for many other types of analyses, including the identification of invariant (likely functionally important) amino acid sites. Moreover, the codeml tab provides information on the percentage of amino acid sites that are strongly conserved, neutral and adaptive. It should be noted that the “notes.txt” (the information shown in the notes tab) file under the folder with the name of the ADOPS experiment is a convenient way to store plain text results obtained with additional software applications, which may help the user with the interpretation of the data.

The ADOPS projects available at B+ can be downloaded, unzipped, and opened using the ADOPS GUI. When this is done, additional analyses can be performed, such as testing the impact of a different alignment algorithm on the results. Therefore, the availability of such a database ensures results repeatability, promotes data reuse with significant savings in the time needed to prepare datasets, and effortlessly allows further exploration of the data contained in ADOPS projects. In the new ADOPS version, there is also an option for adding new sequences to a given project, a tool that is certainly useful when the sequences that a given researcher needs are not all contained in the original ADOPS project.

## 5 Conclusion

The ADOPS batch option allows running hundreds or even thousands of projects in a short period of time, without human intervention. B+ is both a data repository and a convenient interface to look at the information contained in ADOPS projects. The ADOPS projects can be downloaded, unzipped, and opened using the ADOPS GUI (<https://www.sing-group.org/ADOPS/>). Therefore, researchers can repeat the analyses, reuse the sequence and phylogenetic trees data, and make novel analyses without losing time on input file preparation. B+ currently holds a large dataset and several small datasets, although more will soon be available. Furthermore, the research community is welcome to contribute

with other projects as well, even with small datasets. B+ will increase the repeatability of published analyses on the inference of positively selected amino acid sites, and make the process of reading articles more interactive.

**Acknowledgements** This article is a result of the project Norte-01-0145-FEDER-000008 - Porto Neurosciences and Neurologic Disease Research Initiative at I3S, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). This work has been also funded by the “Platform of integration of intelligent techniques for analysis of biomedical information” project (TIN2013-47153-C3-3- R) from Spanish Ministry of Economy and Competitiveness. SING group thanks CITI (Centro de Investigación, Transferencia e Innovación) from University of Vigo for hosting its IT infrastructure. H. López-Fernández is supported by a post-doctoral fellowship from Xunta de Galicia.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Yang ZH, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol* 46(4):409–418
2. Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13(5):555–556
3. Yang ZH (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591
4. Li M, Huang L, Li KQ, Huo YX, Chen CH, Wang JK, Liu JW, Luo ZW, Chen CS, Dong Q et al (2016) Adaptive evolution of interleukin-3 (IL3), a gene associated with brain volume variation in general human populations. *Hum Genet* 135(4):377–392
5. Muto Y, Guindon S, Umemura T, Kohidai L, Ueda H (2015) Adaptive evolution of formyl peptide receptors in mammals. *J Mol Evol* 80(2):130–141
6. Zhang SF, Gao B, Zhu SY (2015) Target-driven evolution of scorpion toxins. *Sci Rep* 5:14973
7. Finseth FR, Dong YZ, Saunders A, Fishman L (2015) Duplication and adaptive evolution of a key centromeric protein in *mimulus*, a genus with female meiotic drive. *Mol Biol Evol* 32(10):2694–2706
8. Zhang Y, Yu ZN (2013) The first evidence of positive selection in peptidoglycan recognition protein (PGRP) genes of *Crasostrea gigas*. *Fish Shellfish Immun* 34(5):1352–1355
9. Jiggins FM, Kim KW (2007) A screen for immunity genes evolving under positive selection in *Drosophila*. *J Evol Biol* 20(3):965–970
10. Morales-Hojas R, Vieira CP, Reis M, Vieira J (2009) Comparative analysis of five immunity-related genes reveals different levels of adaptive evolution in the virilins and melanogaster groups of *Drosophila*. *Heredity* 102(6):573–578
11. Twiddy SS, Woelk CH, Holmes EC (2002) Phylogenetic evidence for adaptive evolution of dengue viruses in nature. *J Gen Virol* 83:1679–1689
12. Woelk CH, Holmes EC (2001) Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). *J Mol Evol* 52(2):182–192



13. Woelk CH, Jin L, Holmes EC, Brown DWG (2001) Immune and artificial selection in the haemagglutinin (H) glycoprotein of measles virus. *J Gen Virol* 82:2463–2474
14. Shen J, Kirk BD, Ma JP, Wang QH (2009) Diversifying selective pressure on influenza B virus hemagglutinin. *J Med Virol* 81(1):114–124
15. Yang W, Bielawski JP, Yang ZH (2003) Widespread adaptive evolution in the human immunodeficiency virus type 1 genome. *J Mol Evol* 57(2):212–221
16. Gu M, Liu WJ, Xu LJ, Cao YZ, Yao CF, Hu SL, Liu XF: Positive selection in the hemagglutinin-neuraminidase gene of Newcastle disease virus and its effect on vaccine efficacy. *Virol J* 2011, 8
17. Emes RD, Yang ZH (2008) Duplicated paralogous genes subject to positive selection in the genome of *Trypanosoma brucei*. *PLoS One* 3(5):e2295
18. Lu J, Zheng JZ, Xu QG, Chen KP, Zhang CY (2011) Adaptive evolution of the vertebrate skeletal muscle sodium channel. *Genet Mol Biol* 34(2):323–U304
19. Khan MMG, Ryden AM, Chowdhury MS, Hasan MA, Kazi JU (2011) Maximum likelihood analysis of mammalian p53 indicates the presence of positively selected sites and higher tumorigenic mutations in purifying sites. *Gene* 483(1–2):29–35
20. Sobrinho IS, de Brito RA (2010) Evidence for positive selection in the gene fruitless in *Anastrepha* fruit flies. *BMC Evol Biol* 10:293
21. Metzger KJ, Thomas MA (2010) Evidence of positive selection at codon sites localized in extracellular domains of mammalian CC motif chemokine receptor proteins. *Bmc Evol Biol* 10:139
22. Vieira CP, Charlesworth D, Vieira J (2003) Evidence for rare recombination at the gametophytic self-incompatibility locus. *Heredity* 91(3):262–267
23. Nunes MDS, Santos RAM, Ferreira SM, Vieira J, Vieira CP (2006) Variability patterns and positively selected sites at the gametophytic self-incompatibility pollen SFB gene in a wild self-incompatible *Prunus spinosa* (Rosaceae) population. *New Phytol* 172(3):577–587
24. Vieira J, Morales-Hojas R, Santos RAM, Vieira CP (2007) Different positively selected sites at the gametophytic self-incompatibility pistil S-RNase gene in the solanaceae and rosaceae (*Prunus*, *Pyrus*, and *Malus*). *J Mol Evol* 65(2):175–185
25. Vieira J, Santos RAM, Ferreira SM, Vieira CP (2008) Inferences on the number and frequency of S-pollen gene (SFB) specificities in the polyploid *Prunus spinosa*. *Heredity* 101(4):351–358
26. Anisimova M (2015) Darwin and Fisher meet at biotech: on the potential of computational molecular evolution in industry. *Bmc Evol Biol* 15:76
27. Reboiro-Jato D, Reboiro-Jato M, Fdez-Riverola F, Fonseca NA, Vieira J (2012) On the development of a pipeline for the automatic detection of positively selected sites. *Adv Intel Soft Comput* 154:225–229
28. Nickel GC, Tefft D, Adams MD (2008) Human PAML browser: a database of positive selection on human genes using phylogenetic methods. *Nucl Acids Res* 36:D800–D808
29. Creevey CJ, McInerney JO (2005) Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics* 21(3):390–392