

Databases and ontologies

JASPAR RESTful API: accessing JASPAR data from any programming language

Aziz Khan^{1,*} and Anthony Mathelier^{1,2,*}

¹Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, University of Oslo, 0349 Oslo, Norway, ²Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, 0372 Oslo, Norway

*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: JASPAR is a widely used open-access database of curated, non-redundant transcription factor binding profiles. Currently, data from JASPAR can be retrieved as flat files or by using programming language-specific interfaces. Here, we present a programming language-independent application programming interface (API) to access JASPAR data using the Representational State Transfer (REST) architecture. The REST API enables programmatic access to JASPAR by most programming languages and returns data in eight widely used formats. Several endpoints are available to access the data and an endpoint is available to infer the TF binding profile(s) likely bound by a given DNA binding domain protein sequence. Additionally, it provides an interactive browsable interface for bioinformatics tool developers.

Availability and Implementation: This REST API is implemented in Python using the Django REST Framework. It is accessible at <http://jaspar.genereg.net/api/> and the source code is freely available at <https://bitbucket.org/CBGR/jaspar> under GPL v3 license.

Contact: aziz.khan@ncmm.uio.no, anthony.mathelier@ncmm.uio.no

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

JASPAR (<http://jaspar.genereg.net>) is the largest open-access database of curated transcription factor (TF) binding profiles for TFs from six different taxa ((Sandelin *et al.*, 2004; Khan *et al.*, 2018). The underlying JASPAR data, accessible as flat files, SQL tables, or in a R data package, has been used by numerous motif enrichment tools, data analysis pipelines, and several other bioinformatics software (Stormo, 2015). Over the years, programming language-specific packages have been developed in Perl (Lenhard and Wasserman, 2002), Python (Mathelier *et al.*, 2014), R (Tan and Lenhard, 2016), and Ruby (Mathelier *et al.*, 2016) to retrieve and process data from JASPAR. However, all of these tools are language-specific and most require connection to the MySQL database.

We developed a language- and platform-independent Application Programming Interface (API) to programmatically retrieve data stored in the JASPAR database using the Representational State Transfer (REST) architecture. REST is a well-recognized modern web service architecture, which uses HTTP protocol to send and receive data using uniform resource locators (URLs) (Fielding, 2000). This new JASPAR API provides an easy-to-use RESTful web service to query/retrieve data

from several releases of the JASPAR database. The API comes with programmatic and human browsable interfaces. The HTML-based browsable interface provides a useful means for developers to interrogate the RESTful API manually before implementation in their system. From the JASPAR RESTful API, TF binding profiles can be retrieved in eight different formats: JSON, JSONP, JASPAR, MEME, PFM, TRANSFAC, BED, and YAML.

2 Implementation

The JASPAR RESTful API is implemented in Python using the Django REST Framework. The source code is freely available under GPL v3 license at <https://bitbucket.org/CBGR/jaspar>. We used Apache HTTP server as a gateway to the API and MySQL database for persistence storage and retrieval of data requested by API services.

2.1 JASPAR RESTful API overview

The RESTful API is interrogated through simple URLs to provide programmatic and human browsable interface to the JASPAR database.

Khan, A. et al.

For instance, information related to the TF binding matrix associated with the CTCF TF can be accessed at <http://jaspar.genereg.net/api/v1/matrix/MA0139.1> and one can get the position frequency matrix (PFM) in JASPAR format by using <http://jaspar.genereg.net/api/v1/matrix/MA0139.1.jaspar>

To ensure no disruption in the applications and tools using the API, the API is version-based with the version included in the API's URL. New versions will be released with version number-based URLs, and there will be a prior notice before obsoleting any older versions. In order to control the rate of requests that clients can make to the API, we use throttling. It allows 25 requests per second from the same IP address, but no limit on the total number of requests. To enhance performance, we use Memcached (<http://memcached.org>), which is a fast, efficient, and fully memory-based cache server. Further, to provide a faster response to each request, the API provides pagination. The number of records per page can be modified by setting the *page_size* parameter (by default *page_size=10*). The *page* parameter is then used to request a specific page (see examples in Supplementary data). Additionally, the API supports simple query parameters to control ordering of results. For instance, one can set the parameter *order=name* to order TF binding profiles by name. The order can be reversed by prefixing the field name with '-' and multiple orderings may also be specified, separated by ',' (see Supplementary data for examples).

The RESTful API supports several data renderers to return responses with various media types. Currently, eight data formats are available for data associated to TF binding profiles: JSON, JSONP, JASPAR, MEME, PFM, TRANSFAC, BED, and YAML. Users can set the output format in three different ways: (i) by setting a *<format>* parameter in the URL, (ii) by adding a *<.format>* suffix to the URL, or (iii) by using the Accept headers in the get request (see Supplementary data for examples).

2.2 RESTful API endpoints

The RESTful API provides several endpoints to query and retrieve data from the JASPAR database (Table 1 and Figure S1). It also provides endpoints to infer the most likely TF binding profile(s) associated with a given protein sequence (Mathelier et al., 2016). A summary of API endpoints is available in Table 1 and Figure S1, and a list of parameters to filter TF binding profiles from the *'/matrix'* endpoint is provided in Table S1.

Table 1: Summary of REST API endpoints

Method	Endpoint	Description
GET	/matrix	Gets a list of all profiles *
GET	/matrix/:matrix_id	Detail view of the profile
GET	/matrix/:base_id/versions	List of profile versions
GET	/collections	List all JASPAR collections
GET	/collections/:collection	List all profiles in a collection
GET	/infer/:sequence	Infer profiles, given a protein sequence
GET	/taxon	Return all taxonomic groups
GET	/taxon/:tax_group	List all profiles in a tax group
GET	/species	List all species in JASPAR
GET	/species/:tax_id	List all profiles in a specie
GET	/sites/:matrix_id	Return genomic locations of sites used to construct PFMs
GET	/tffm	Return all the TFFMs profiles
GET	/tffm/:tffm_id	Return details of a TFFM

GET	/releases	List all the releases (versions)
GET	/releases/:release_number	Return details of a release

* Profiles can be filtered using several parameters listed in Table S1.

A detailed documentation for each endpoint with sample calls in multiple programming languages (JavaScript, Perl, Python, R, Java, and Ruby, following (Yates et al., 2015)) is available on the API website. Further, an interactive and live API documentation for all the endpoints is available at <http://jaspar.genereg.net/api/v1/docs>.

3 Discussion

The JASPAR RESTful API can be used to query TF binding profiles and associated data from the JASPAR database programmatically, using numerous programming languages. Moreover, the API's web browsable interface brings an easy interface to bioinformatics tool developers.

To interact with the REST API, clients are needed, which handle the underlying details of how network requests are made and how responses are decoded. We are using Core API (<http://www.coreapi.org/>), which is a format-independent Document Object Model for representing web APIs. Core API comes with clients in command line, Python, and JavaScript. It is recommended to use these clients for more robust and meaningful interactions with the JASPAR RESTful API rather than constructing HTTP GET requests and decoding responses. More details about the clients are available on the API website.

The REST architecture has been demonstrated as a sustainable and reliable way to distribute genomic data to many programming languages. In the future, we plan to implement non-disruptive features to the API, such as: adding new endpoints, data fields, and filtering methods. All the disruptive changes will be released as a new version. We will make sure to keep the JASPAR RESTful API up-to-date and functional at any time to avoid any disruption to the tools based on this web service.

Acknowledgements

We thank all JASPAR collaborators and members of the Gene Regulation Ensemble Effort for Knowledge Commons (<http://greekc.org>) for their useful suggestions and testing the API, and Georgios Magklaras for IT support. We thank the four anonymous reviewers for their constructive comments and suggestions.

Funding

This work has been supported by the Norwegian Research Council, Helse Sør-Øst, and the University of Oslo through the Centre for Molecular Medicine Norway (NCMM), which is part of the Nordic European Molecular Biology Laboratory partnership for Molecular Medicine.

Conflict of Interest: none declared.

References

- Fielding, R.T. (2000) Architectural Styles and the Design of Network-based Software Architectures. PhD dissertation: University of California, Irvine.
- Khan, A. et al. (2018) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.*, *gkx1126*, doi: 10.1093/nar/gkx1126
- Lenhard, B. and Wasserman, W.W. (2002) TFBS: Computational framework for transcription factor binding site analysis. *Bioinformatics*, **18**, 1135–1136.
- Mathelier, A. et al. (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **42**, D142–D147.

- Mathelier,A. *et al.* (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **44**, D110-D115.
- Sandelin,A. *et al.* (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.*, **32**, D91–D94.
- Stormo,G.D. (2015) DNA motif databases and their uses. *Curr. Protoc. Bioinform.* **51**, 2.15.1-2.15.6.
- Tan,G. and Lenhard,B. (2016) TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics*, **32**, 1555–1556.
- Yates,A. *et al.* (2015) The Ensembl REST API: Ensembl Data for Any Language. *Bioinformatics*, **31**, 143–145.