

## ПРИЛОЖЕНИЕ

к статье Р.А. Иванова, А.М. Мухина, Ф.В. Казанцева, З.С. Мустафина, Д.А. Афонникова, Ю.Г. Матушкина, С.А. Лашина «Orthoweb: программный комплекс для эволюционного анализа геновых сетей»

### Описание API запросов к базе данных для хранения результатов Orthoweb:

- <https://orthoweb.sysbio.cytogen.ru/api-db> – стартовая страница с описанием доступа к данным через API. Этот путь в дальнейшем используется для конструирования запросов как `<server_name>`.
- `<server_name>/organisms` – запрос списка имеющихся организмов в базе.
- `<server_name>/organisms/<name>` – запрос имеющихся данных по конкретному организму с идентификатором `<name>`. Это стартовая точка для дополнительных запросов для этого организма.
- `<server_name>/organisms/<name>/pai` – получить список имеющихся значений филостратиграфического индекса генов (PAI) из базы для организма с идентификатором `<name>`.
- `<server_name>/pai/organism/<name>` – получить список имеющихся значений филостратиграфического индекса генов (PAI) в базе для организма с идентификатором `<name>`.
- `<server_name>/ko_pai/organism/<name>` – получить список имеющихся значений филостратиграфического индекса генов (PAI), посчитанного с помощью КО групп, из базы для организма с идентификатором `<name>`.
- `<server_name>/go_id/organism/<name>` – получить список имеющихся идентификаторов терминов геновой онтологии в базе для организма с идентификатором `<name>`.
- `<server_name>/go_name/organism/<name>` – получить список имеющихся наименований терминов геновой онтологии в базе для организма с идентификатором `<name>`.
- `<server_name>/pai/organism/<name>/identity/<value>` – получить список имеющихся значений филостратиграфического индекса генов (PAI) в базе для организма с идентификатором `<name>` с параметром идентичности, равным `<value>`.
- `<server_name>/di/organism/<name>` – получить список имеющихся значений индекса эволюционной изменчивости генов (DI) в базе для организма с идентификатором `<name>`.
- `<server_name>/di/organism/<name>/identity/<value>` – получить список имеющихся значений индекса эволюционной изменчивости генов (DI) в базе для организма с идентификатором `<name>` с параметром идентичности, равным `<value>`.
- `<server_name>/snp/organism/hsa` – получить список имеющихся значений количества найденных SNP для генов человека.

**Пример кода на языке Python для выполнения запроса к базе через API с целью выгрузить все PAI, посчитанные на основе КО групп для организма с идентификатором 'hsa'.**

```
import requests
pathToAPI_start = "https://orthoweb.sysbio.cytogen.ru/api-db"
PathToOrg = 'ko_pai/organism'
pathToAPI = "/" .join((pathToAPI_start, PathToOrg))

organisms = ['hsa',]

# GenesOfOrg - переменная содержащая 'словарь' со списком генов для выбранного организма.
GenesOfOrg = dict()
for org in organisms:
    myRequest = "/" .join((pathToAPI, org))
    print(myRequest)
    r = requests.get(myRequest)
    theResponse = r.json()
    GenesOfOrg[org] = theResponse
GenesOfOrg
```

**Рис. П1.** Пример кода для выполнения запроса к базе данных для получения списка всех генов *Homo sapiens*.

В результате успешного выполнения приведенного кода (рис. П1) переменная GenesOfOrg будет содержать словарь, в котором ключом является организм, а значением – список объектов PAI для генов. Каждый объект PAI представляет собой словарь, из которого можно извлечь необходимую информацию через соответствующие ключи (рис. П2):

```
{ '_url_gene': '/api-db/genes/hsa:100616498',
  '_url_organism': '/api-db/organisms/hsa',
  '_url_self': '/api-db/ko_pai/79334',
  'gene': 'hsa:100616498',
  'genename': '100616498',
  'ko': '08_Eutheria',
  'koid': '79334',
  'organism': 'hsa' }
```

**Рис. П2.** Пример структуры объекта PAI с ключами для API запроса.

После получения данных можно искать интересующие гены без дополнительных онлайн-запросов к базе данных. Например, можно выбрать гены, связанные с определенным метаболическим путем, таким как *Wnt signaling pathway* (рис. П3).

```
import pandas as pd

# listOfGenes - список генов, для которых необходимо выбрать данные
listOfGenes = ['hsa: 5535', 'hsa: 1488', 'hsa:8326', 'hsa:9350', ... 'hsa:4776']

# myData - объект типа pandas.DataFrame, содержащий таблицу с выбранными генами и их характеристиками
myData = pd.DataFrame( [ [gn['gene'], gn['ko'].split('_')[0], gn['ko'].split('_')[1] ] for gn
                        in GenesOfOrg['hsa'] if gn['gene'] in listOfGenes ],
                      columns = ['KEGG gene ID', 'PAI', 'Phylostratum'])

# сохранить полученные данные в *.CSV формате в файле "data.csv"
import csv
myData.to_csv("data.csv", sep=';', quoting=csv.QUOTE_NONNUMERIC)
```

**Рис. П3.** Пример кода на языке Python для получения информации по значениям PAI, рассчитанным с использованием анализа групп ортологии KEGG, и сохранения ее в файле формата .csv.

В результате выполнения этого фрагмента кода переменная myData будет содержать объект типа DataFrame, который широко используется в вычислительных конвейерах для анализа данных. Приведенные фрагменты кода были использованы для визуализации сети Wnt- $\beta$  катенинового каскада, представленного на рис. 3 (см. таблицу).

Фрагмент данных с выгрузкой PAI индексов, посчитанных на основе КО групп,  
 для генов из метаболического пути *Wnt signaling pathway*

KEGG gene ID	PAI	Phylostratum	KEGG gene ID	PAI	Phylostratum
hsa:5535	0	Cellular Organisms	hsa:79718	1	Eukaryota
hsa:147111	0	Cellular Organisms	hsa:6885	1	Eukaryota
hsa:6500	0	Cellular Organisms	hsa:343637	1	Eukaryota
hsa:25776	0	Cellular Organisms	hsa:8607	1	Eukaryota
hsa:4609	0	Cellular Organisms	hsa:9475	1	Eukaryota
hsa:5534	0	Cellular Organisms	hsa:8945	1	Eukaryota
hsa:1487	0	Cellular Organisms	hsa:8323	1	Eukaryota
hsa:1488	0	Cellular Organisms	hsa:7855	1	Eukaryota
hsa:8326	1	Eukaryota	hsa:1457	1	Eukaryota
hsa:90665	1	Eukaryota	hsa:1459	1	Eukaryota
hsa:283106	1	Eukaryota	hsa:8454	1	Eukaryota
hsa:29964	1	Eukaryota	hsa:5880	1	Eukaryota
hsa:102800317	1	Eukaryota	hsa:5879	1	Eukaryota
hsa:5881	1	Eukaryota	hsa:894	1	Eukaryota
hsa:817	1	Eukaryota	hsa:5578	1	Eukaryota
hsa:8321	1	Eukaryota	hsa:6477	1	Eukaryota
hsa:6907	1	Eukaryota	hsa:387	1	Eukaryota
hsa:5602	1	Eukaryota	hsa:2033	1	Eukaryota
hsa:7976	1	Eukaryota	hsa:818	1	Eukaryota
hsa:5330	1	Eukaryota	hsa:84870	1	Eukaryota
hsa:23236	1	Eukaryota	hsa:1452	1	Eukaryota
hsa:816	1	Eukaryota	hsa:23002	1	Eukaryota
hsa:166336	1	Eukaryota	hsa:5663	1	Eukaryota
hsa:23500	1	Eukaryota	hsa:23411	1	Eukaryota
hsa:4007	1	Eukaryota	hsa:5566	1	Eukaryota
hsa:122011	1	Eukaryota	hsa:8324	1	Eukaryota
hsa:5601	1	Eukaryota	hsa:8322	1	Eukaryota
hsa:5532	1	Eukaryota	hsa:59352	2	Metazoa
hsa:9978	1	Eukaryota	hsa:7479	2	Metazoa
hsa:5332	1	Eukaryota	hsa:7482	2	Metazoa
hsa:7157	1	Eukaryota	hsa:7473	2	Metazoa
hsa:5568	1	Eukaryota	hsa:2239	2	Metazoa
hsa:8325	1	Eukaryota	hsa:80319	2	Metazoa
hsa:5533	1	Eukaryota	hsa:55366	2	Metazoa
hsa:5331	1	Eukaryota	hsa:7484	2	Metazoa
hsa:11211	1	Eukaryota	hsa:81029	2	Metazoa
hsa:5567	1	Eukaryota	hsa:85407	2	Metazoa
hsa:815	1	Eukaryota	hsa:57216	2	Metazoa
hsa:27101	1	Eukaryota	hsa:59343	2	Metazoa
hsa:1454	1	Eukaryota	hsa:27121	2	Metazoa
hsa:27130	1	Eukaryota	hsa:7089	2	Metazoa
hsa:1460	1	Eukaryota	hsa:7483	2	Metazoa
hsa:144165	1	Eukaryota	hsa:79816	2	Metazoa
hsa:6424	1	Eukaryota	hsa:23401	2	Metazoa
hsa:51701	1	Eukaryota	hsa:284654	2	Metazoa
hsa:5599	1	Eukaryota	hsa:2535	2	Metazoa
hsa:23291	1	Eukaryota	hsa:85409	2	Metazoa
hsa:595	1	Eukaryota	hsa:10023	2	Metazoa
hsa:5579	1	Eukaryota	hsa:7090	2	Metazoa
hsa:3725	1	Eukaryota	hsa:1501	2	Metazoa
hsa:57680	1	Eukaryota	hsa:102723796	2	Metazoa
hsa:2932	1	Eukaryota	hsa:6259	2	Metazoa
hsa:5530	1	Eukaryota	hsa:896	2	Metazoa
hsa:1387	1	Eukaryota	hsa:7091	2	Metazoa

Окончание таблицы

KEGG gene ID	PAI	Phylostratum	KEGG gene ID	PAI	Phylostratum
hsa:64840	2	Metazoa	hsa:4919	2	Metazoa
hsa:7475	2	Metazoa	hsa:7481	2	Metazoa
hsa:27123	2	Metazoa	hsa:22943	2	Metazoa
hsa:51384	2	Metazoa	hsa:8312	2	Metazoa
hsa:8840	2	Metazoa	hsa:340419	2	Metazoa
hsa:7088	2	Metazoa	hsa:6934	2	Metazoa
hsa:80326	2	Metazoa	hsa:1499	2	Metazoa
hsa:8061	2	Metazoa	hsa:7474	2	Metazoa
hsa:6422	2	Metazoa	hsa:4089	2	Metazoa
hsa:4316	2	Metazoa	hsa:324	2	Metazoa
hsa:11197	2	Metazoa	hsa:7472	2	Metazoa
hsa:8313	2	Metazoa	hsa:4088	2	Metazoa
hsa:6932	2	Metazoa	hsa:7476	2	Metazoa
hsa:81839	2	Metazoa	hsa:6425	2	Metazoa
hsa:51176	2	Metazoa	hsa:4041	2	Metazoa
hsa:2487	2	Metazoa	hsa:6423	2	Metazoa
hsa:1857	2	Metazoa	hsa:7478	2	Metazoa
hsa:7477	2	Metazoa	hsa:7471	2	Metazoa
hsa:4040	2	Metazoa	hsa:25805	2	Metazoa
hsa:1855	2	Metazoa	hsa:5582	3	Chordata
hsa:64321	2	Metazoa	hsa:9350	5	Vertebrata
hsa:89780	2	Metazoa	hsa:56998	5	Vertebrata
hsa:54361	2	Metazoa	hsa:4775	5	Vertebrata
hsa:84133	2	Metazoa	hsa:4773	5	Vertebrata
hsa:4920	2	Metazoa	hsa:54894	5	Vertebrata
hsa:8549	2	Metazoa	hsa:5176	5	Vertebrata
hsa:10297	2	Metazoa	hsa:50964	5	Vertebrata
hsa:5467	2	Metazoa	hsa:4772	5	Vertebrata
hsa:7480	2	Metazoa	hsa:4776	6	Euteleostomi
hsa:1856	2	Metazoa	hsa:83439	6	Euteleostomi