

Vidjil, a platform for the interactive analysis of immune repertoires

Mathieu Giraud, Mikaël Salson, Florian Thonier



Bonsai bioinformatics, CRISAL (Université Lille, CNRS),
VidjilNet consortium, Inria

December 7, 2022
contact@vidjil.org

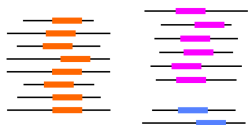
Vidjil

High-throughput Repertoire Sequencing (AIRR/RepSeq) analysis

Web Application

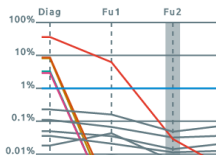
Patient database
Server

Vidjil-algo



C++

Client



Javascript, d3.js

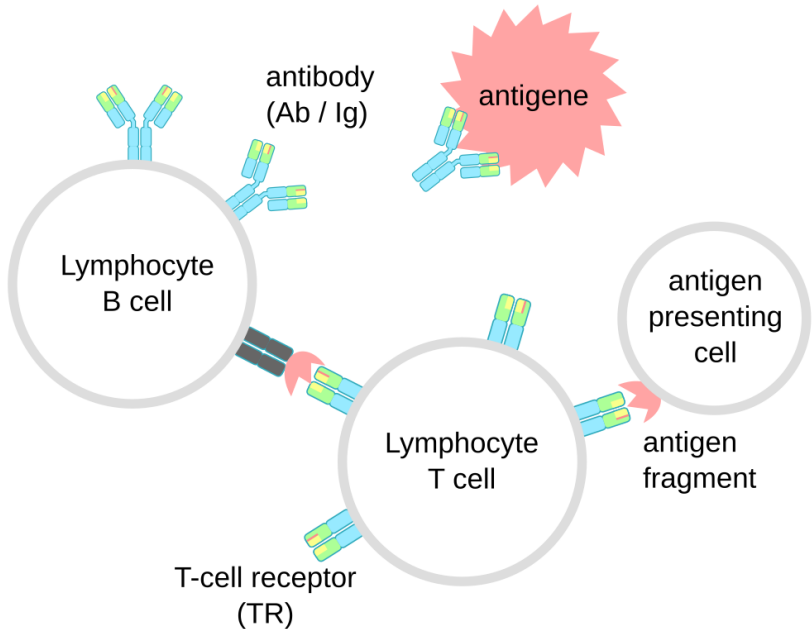


Python, web2py,
AJAX

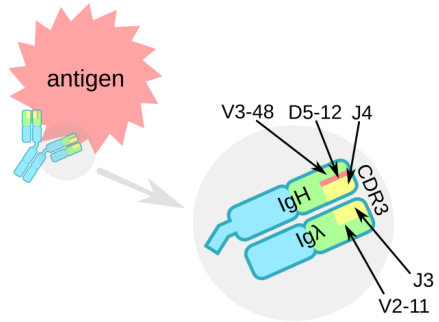
Duez et al., PLoS One, 2016

- ▶ <http://gitlab.vidjil.org/>: open-source (GPL v3), public issue tracker
- ▶ Software quality, continuous integration, > 3,000 unit and functional tests
- ▶ VidjilNet: certification/accreditation, health data hosting

TCR and Antibody Specificity – V(D)J Recombination



The Adaptive Immune System and V(D)J Recombinations



...TTCCTGTGTGAGA TC TATAGTGGCTA CTG GATGCTTTT...
...TTACTGTGCGAGAGA TGATGCTTTT...
V3-48 ...GATATAGTGGCTACG... J4
D5-12

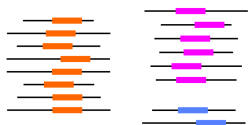
Vidjil

High-throughput Repertoire Sequencing (AIRR/RepSeq) analysis

Web Application

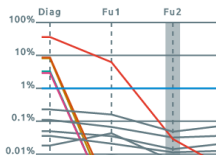
Patient database
Server

Vidjil-algo



C++

Client



Javascript, d3.js



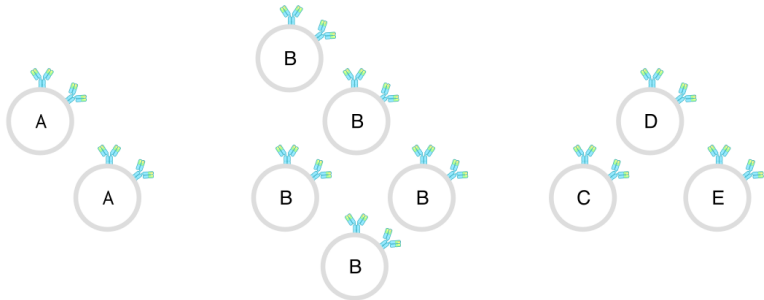
Python, web2py,
AJAX

Duez et al., PLoS One, 2016

- ▶ <http://gitlab.vidjil.org/>: open-source (GPL v3), public issue tracker
- ▶ Software quality, continuous integration, > 3,000 unit and functional tests
- ▶ VidjilNet: certification/accreditation, health data hosting

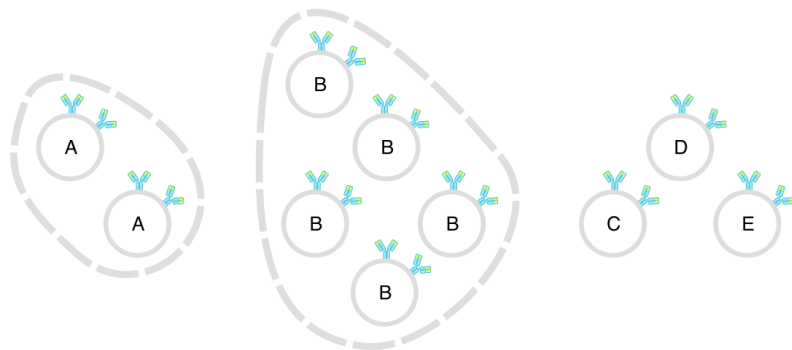
Vidjil-algo

Clonotype clustering



Vidjil-algo

Clonotype clustering



20%

50%

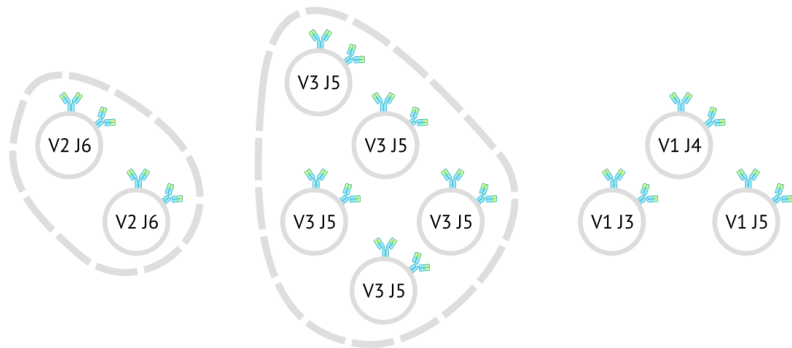
30%

1 000 000 VDJ = 100 s

Giraud, Salson et al., BMC Genomics, 2014

Vidjil-algo

Clonotype clustering



20%

50%

30%

1 000 000 VDJ = 100 s

Giraud, Salson et al., BMC Genomics, 2014

Vidjil client

Demo patient (multi-inc+xxx) samples import/export cluster color by Locus filter settings System Administrator (logout) palette help Vidjil ^(beta)

Demo data of a patient

TRG IGH

LIL-L3-0 analyzed reads 1 965 949 (82.25%)
 selected locus 1 964 587 (82.19%)

2022-01-01

lil-l3-0 #type=demo

search sort by

TRG smaller clonotypes	20.92%	★
IGH smaller clonotypes	10.17%	★
IGHV3-9 7/6/17 J6*02	9.670%	★ i
TRGV10 13//5 JP1	8.858%	★ i
TRGV10 4//8 JP2	6.694%	★ i
TRGV10 13//5 JP1	5.108%	★ ⚠
TRGV10 4//8 JP2	4.994%	★ ⚠
TRGV10 4//8 JP2	4.933%	★ ⚠
TRGV2*02 0/14/2 J1	4.623%	★ ⚠
TRGV2*02 0/14/2 J1	4.107%	★ ⚠
TRGV3 3/11/3 JP2	3.305%	★ i
IGHV3-13*05 1/55/16 J6*02	2.405%	★ i

LIL-L3-0 LIL-L3-1 LIL-L3-2 LIL-L3-3 LIL-L3-4 5/5

Reads length

160 180 200 220 240 260 280 300 320 ?

Locus

IGH

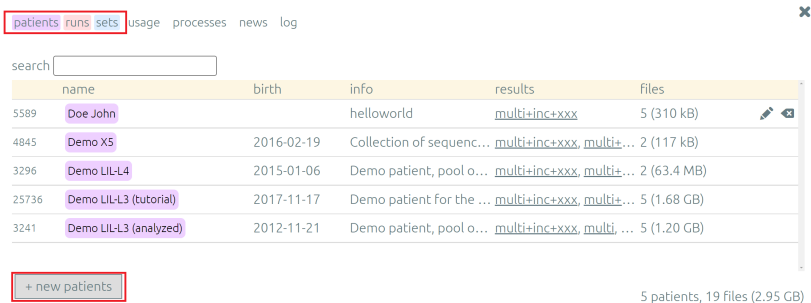
TRG

Cluster Align II 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100



IGH smaller clonotypes 3 clonotypes, 369 544 reads (18.81%)

IGHV3-9 7/6/17 J6*02	9.670%	★ i	A A G G C C T G G A G T G G G T C T C A G G T A T T A G T T G G A A A T A G T G G T A G C C A T A G G C T A T G C G G
TRGV10 13//5 JP1	8.858%	★ i	C T A C T G T G C G G A C T G G T T G G T T C A A G A T A T T T G C T G A A G G G A C T A A G C T C A T A G T A A C
TRGV2 0/7/0 JP1	0.280%	★ i	A C T G C A A A A T C T A A T T G A A A A T G A C T C T G G G G T C T A T A C T G T G C C A C C C T G G G A C G G

Server, database



The screenshot shows a web interface with a navigation bar at the top containing links for 'patients', 'runs', 'sets', 'usage', 'processes', 'news', and 'log'. The 'patients' link is highlighted with a red box. Below the navigation bar is a search input field. The main content area displays a table with columns for 'name', 'birth', 'info', 'results', and 'files'. The table contains five rows of patient data. The 'name' column for each row is highlighted with a purple box. Below the table is a button labeled '+ new patients', which is also highlighted with a red box. In the bottom right corner, there is a summary text: '5 patients, 19 files (2.95 GB)'.

	name	birth	info	results	files	
5589	Doe John		helloworld	multi+inc+xxx	5 (310 kB)	 
4845	Demo X5	2016-02-19	Collection of sequenc...	multi+inc+xxx , multi+...	2 (117 kB)	
3296	Demo LIL-L4	2015-01-06	Demo patient, pool o...	multi+inc+xxx , multi+...	2 (63.4 MB)	
25736	Demo LIL-L3 (tutorial)	2017-11-17	Demo patient for the ...	multi+inc+xxx , multi+...	5 (1.68 GB)	
3241	Demo LIL-L3 (analyzed)	2012-11-21	Demo patient, pool o...	multi+inc+xxx , multi, ...	5 (1.20 GB)	

+ new patients

5 patients, 19 files (2.95 GB)

- ▶ Organize your samples
- ▶ Associate Tags to facilitate future research

Upload samples / Assignment to patients/runs/sets

Patient, run or set association

Samples have to be associated with at least one patient, run or set. You can also associate them with any combination of the three. All the samples added here will be associated to the "common sets". Moreover, in the sample list, you can associate individually some samples to some specific patient, run or set.

Sample list

Common sets: ✘ Lib-9871 (26932)

Click on add other sample to add at once more than one sample.

Sample 1	Parcourir...	2018-04-01	#ALL #diagnosis	✘ John Doe (1234)	✘ Diag-experiment
Sample 2	Parcourir...	2018-04-02	#pre-B-ALL #MRD	✘ Alice Doe (4321)	

Genetic data

- ▶ Amplicon; specific, 10Mo/sample
- ▶ capture; non-specific; 100Mo (less than 1mo to keep)
- ▶ WES/WGS; non-specific; Go (some Mo to keep)

Main Analysis Process

patients runs sets usage processes news log

< > 6586 run 190305

(None)

file name	sampling date	info	size	pre-process
Sample 1			3.64 MB	M+R2: Merge ...
Sample 2			2.46 MB	M+R2: Merge ...
Sample 3			3.16 MB	M+R2: Merge ...
Sample 4			2.85 MB	M+R2: Merge ...
Sample 5			2.90 MB	M+R2: Merge ...
5 sample(s)				

+ add samples

process config | IGH

last processing	status
2019-03-07 09:07:53	COMPL
	RUNNIN
	QUEUE
	QUEUE
	PREPR

- Human V(D)J recombinations
 - Clonality
 - IGH
 - multi
 - multi+inc+xxx
 - TRG
- Other recombinations
 - IKZF1/ERG
- Analysis with/for other software
 - Export all clones (AIRR)
- Other species
 - Mus musculus

Use customizable config to suit your needs

- ▶ Software (Vidjil, MiXCR, ...)
- ▶ Settings
- ▶ Germlines (Human, mice, ...)

Pre/post-Process (optional)

Additional steps to pre/post-process all samples or results data

Pre-process

- ▶ Filter data of bad quality
- ▶ Merge paired-end reads
- ▶ Fastq trimming (removing adaptaters)

Post-process

- ▶ Computing sample stats
- ▶ Detecting MRD for a patient
- ▶ Detecting contamination in a run

(2022) New server API

For bioinformaticians

```
1 vidjil = Vidjil(SERVER, SSL_CERTIFICATE)
2 vidjil.login(user, password)
3
4 patient = vidjil.createPatient(first_name='Jane',
5                               last_name='Austeen',
6                               birth_date='1775-12-16')
7
8 sample = vidjil.createSample(file_filename='Diag-53363.fa',
9                              file_info="#LAL-B, EuroClonality-NGS primers",
10                             pre_process=None, ...)
11
12 analysis = vidjil.launchAnalysisOnSample (patient, sample[0], analysis = "multi-inc-xxx")
13
14 if analysis.SUCCESS:
15     print(analysis)
```

Patient/run/set creation, file upload, analysis launch... all server tasks can be launched through the API.

Software quality: A golden set of manually curated sequences

~ 250 sequences with hard-copy analysis available at vidjil.org/curated-vdj

```
>IGHV5-10-1*01 2/CTTC/3 IGHD1-14*01 1/GTTA/5 IGHJ5*01  
ggacaccgccatgtattactgtgCGaCTTCataaccggaaccaGTTA  
tggttcGactcctggggccaaggaaccctgggtcaccgtctcctcag
```

Contribute by sending any manually curated sequences

- ▶ Lille (A. Caillault, Y. Ferret)
- ▶ GOSH, London (G. Wright)
- ▶ Paris-Necker (F. Thonier)
- ▶ CLIP, Prague (M. Kotrova)
- ▶ Bristol (S. Wakeman, M. Duez)
- ▶ Rennes (A. Fievet)

Presented during RepSeq 2016

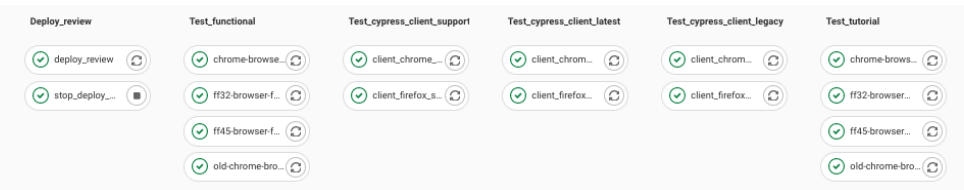
Ability to encode alternative solutions:

```
>TRGV4*02 (4/4/4 TRGJ1*01, 4/4/1 TRGJ1*02)
```

Software quality: Automated tests, as much as possible

Vidjil-algo \simeq 2,000 tests (including “curated sequences”)

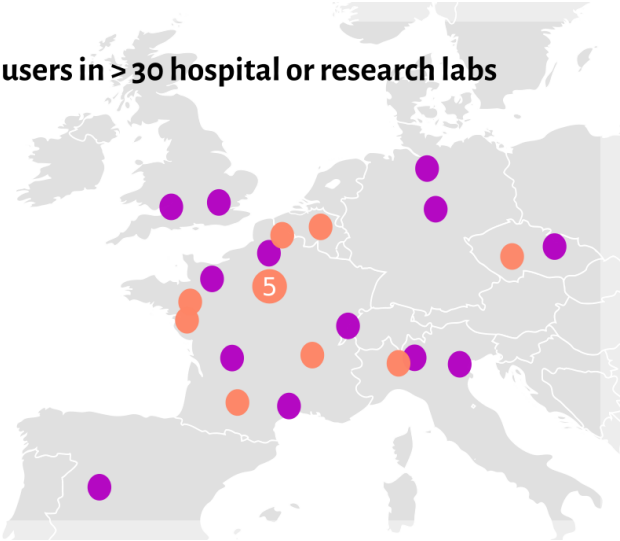
Web app \simeq 1,000 tests on internal behaviour,
 \simeq 100 tests on practical behaviour (“when I click here, I should see the clone sequence”)



Usage map

> 50 regular users in > 30 hospital or research labs

- Brasil ●
- Lithuania ●
- Japan ● ●
- South Korea ●



Securing bioinformatics for hospital routine

Vidjil is open source, we are public researchers.

To secure software part of clinical practise, a contracted service could include:

- ▶ upgrade and one-day support to hosted instances of the server
- ▶ some custom development
- ▶ possibly certification
- ▶ possibly patient-related data agreement

This is **not** the job of a university research team in bioinformatics.

VidjilNet, a non-profit consortium at Inria

www.vidjil.net

The VidjilNet consortium gathers onco-hematological laboratories, research centers, and industrial and institutional partners

VidjilNet is dedicated to

- ▶ **sustain and support open-source development of the Vidjil platform** and possibly other software to study immune repertoires
- ▶ **provide bioinformatics support and services** to the members for these software, whether for clinical purpose or for hematological or immunological research.

The consortium works into a **quality approach** concerning both software development and testing, the relationship with its members, and ethics on health data processing.



Patient database and server – Hosting

app.vidjil.org

Public test server: research use only

Data center supported by Mésocentre U. Lille

health.vidjil.org

Production/clinical server, Health regulation compliance

Secure data center in Gravelines (F)

Data hosting: ISO 27001, HDS Health-compliant regulations
for VidjilNet members

In your lab/hospital

Production/clinical server

Installed by local IT staff

with remote support for VidjilNet members

Problèmes et difficultés

- ▶ Dette technique à resorber (projet de plus de 10 ans).
- ▶ Utilisateurs biologistes/cliniciens, peu de bioinformaticiens
- ▶ Meilleure gestion du support
- ▶ Bugtracker pas tout à fait ouvert
- ▶ Relation avec les administrations hospitalières difficile.

