Centre Hospitalier Régional
Universitaire de Lille

Vidjil, a platform for the interactive analysis of immune repertoires

Aurélien Béliard, Marc Duez, Mathieu Giraud, Ryan Herbert,
Tatiana Rocher, Mikaël Salson, Florian Thonier

Bonsai bioinformatics
CRIStAL (Université Lille, CNRS), Inria, GBMHM/Inca

Hôpital Rennes
GBMHM/Inca

November 21, 2017

contact@vidjil.org

Vidjil Platform

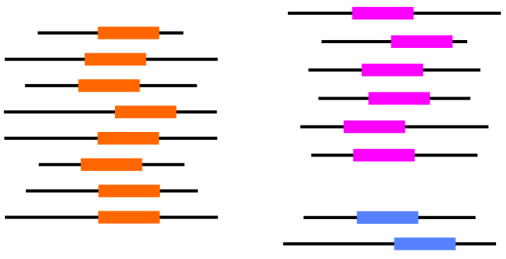
High-throughput Repertoire Sequencing (RepSeq) analysis

Web Application

Patient database

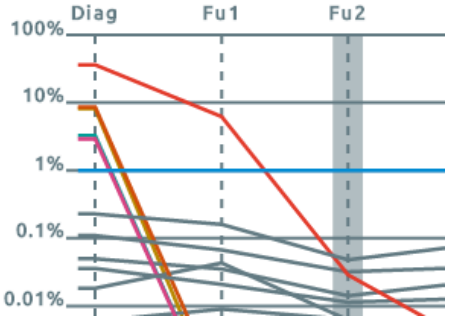
Server

Algorithm



C++

Client

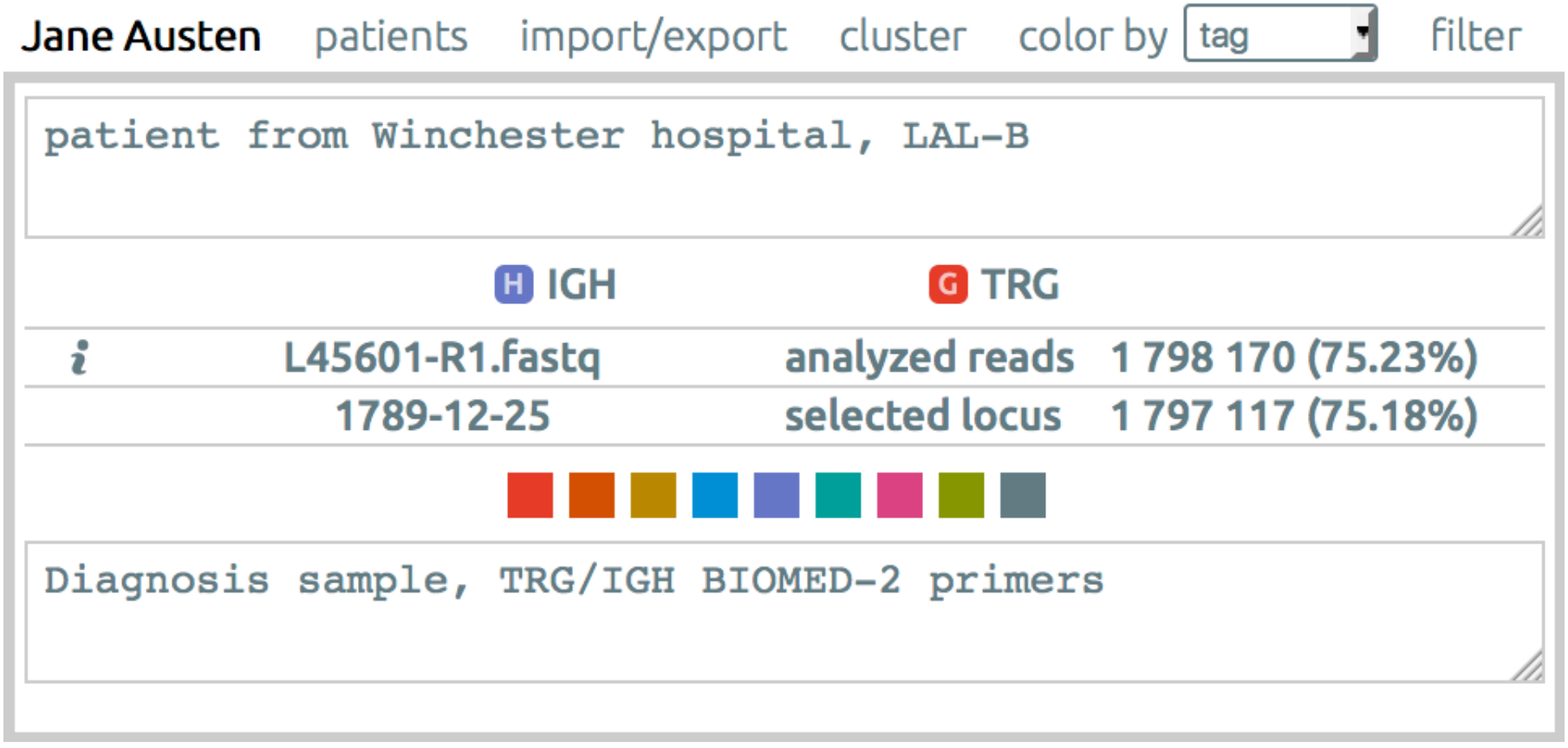


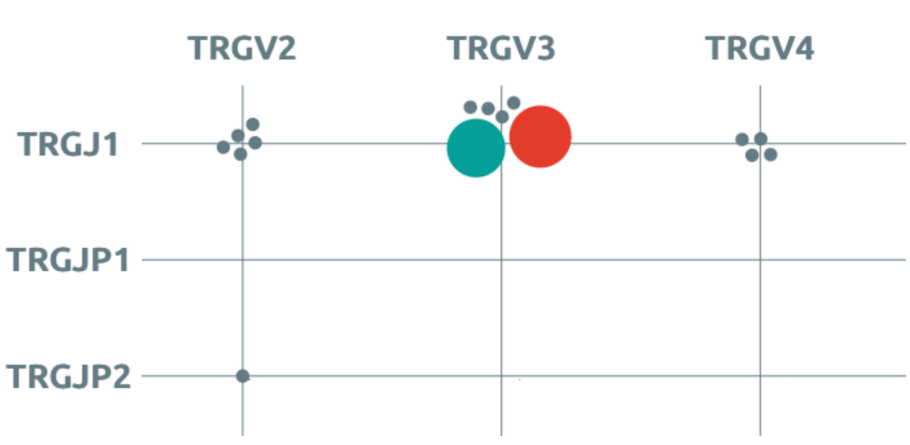
Javascript, d3.js

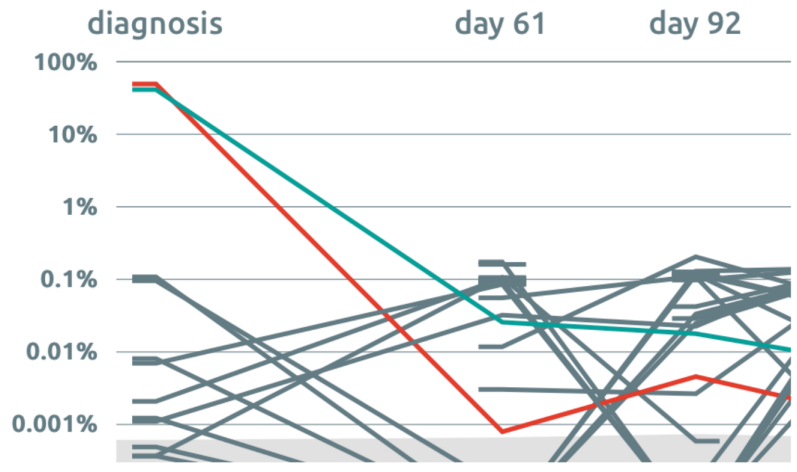
Python, web2py, AJAX

- all the Vidjil components are open-source (GPL v3)
- code on `http://www.vidjil.org/` and on GitHub
- continuous integration, $> 2,000$ unit and functional tests

Duez et al., PLOS One, 2016


Jane Austen
patients
import/export
cluster
color by
tag
filter
patient from Winchester hospital, LAL-B
H IGH
G TRG
L45601-R1.fastq
analyzed reads
1 798 170 (75.23%)
1789-12-25
selected locus
1 797 117 (75.18%)
Diagnosis sample, TRG/IGH BIOMED-2 primers



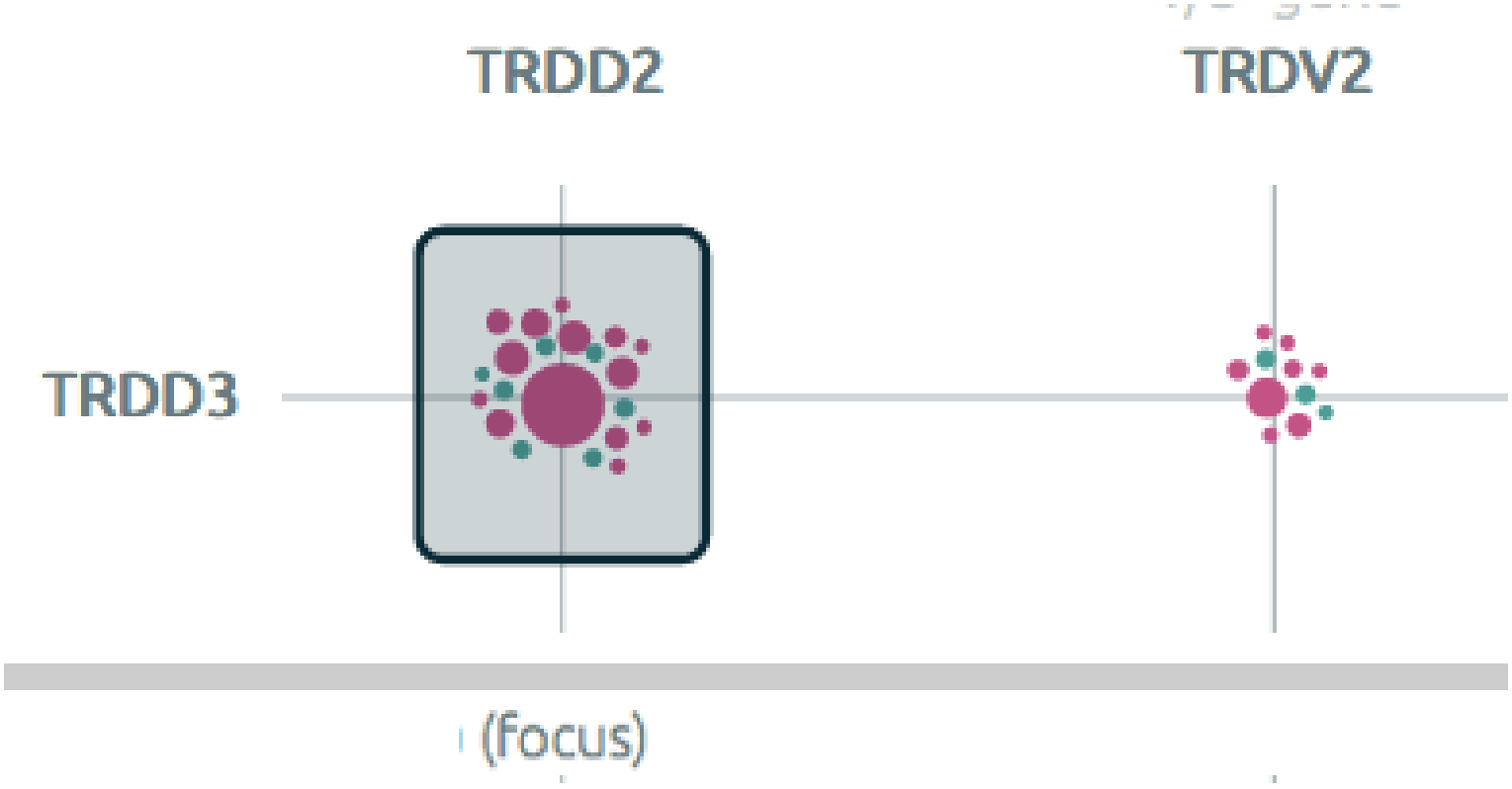


merge align › to IMGT/V-QUEST › to IgBlast › to Blast

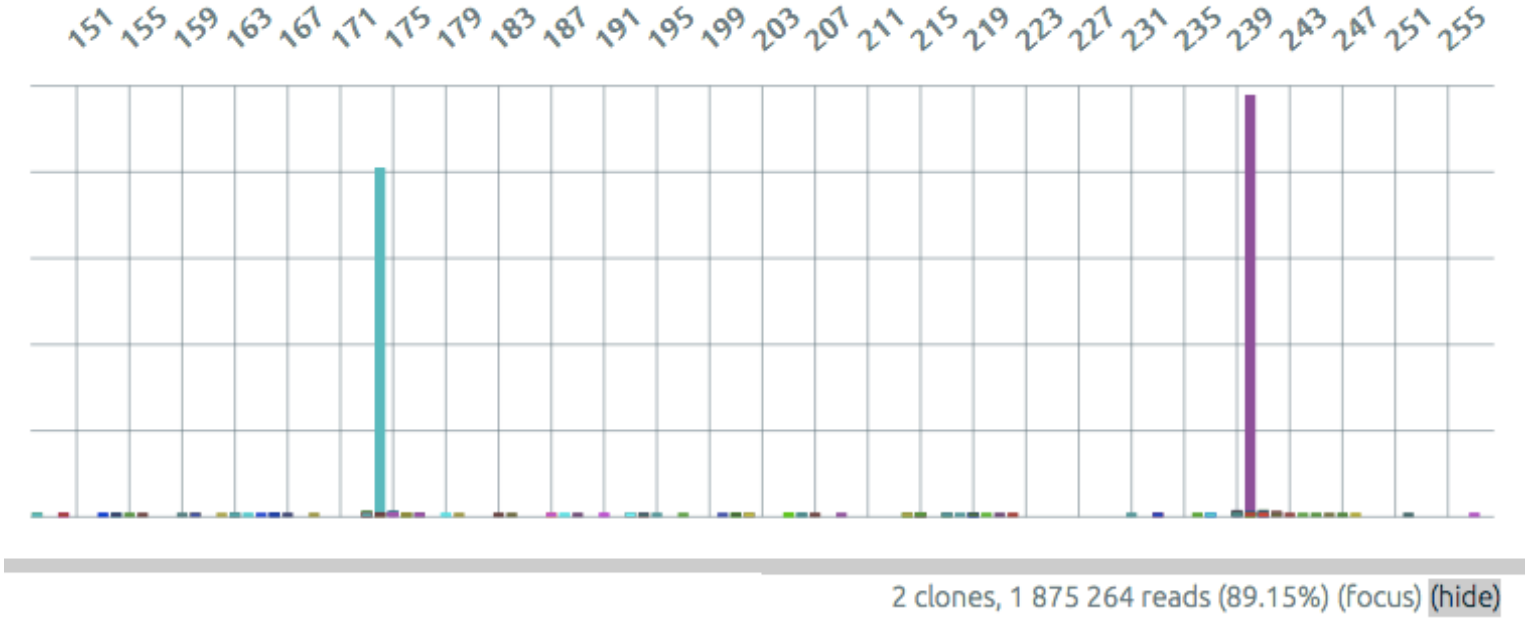
4 clones, 738 494 reads (90.53%)

✕	TRGV3 1/11/2 J1	49.30%	i	GCCACCTGGGACAGCTCCC-TT-GTTC--ATTATAAGAAACTCTTTGGCAGTG
✕	TRGV3 4/1/2 J1	41.23%	i	GCCACCTGGG--A--T--A--T--T--T--ATTATAAGAAACTCTTTGGCAGTG
✕	TRGV3 3/16/3 J1	0.0021%	i	GCCG-CTTGGA-ACCCCAATTTGGTACGGGTTATAAGAAACTCTTTGGCAGTG
✕	TRGV3 5/4/2 J1	+	i	GCCACCTGGG---GC--CA-AT--T--A-TA--AGAAACTCTTTGGCAGTG

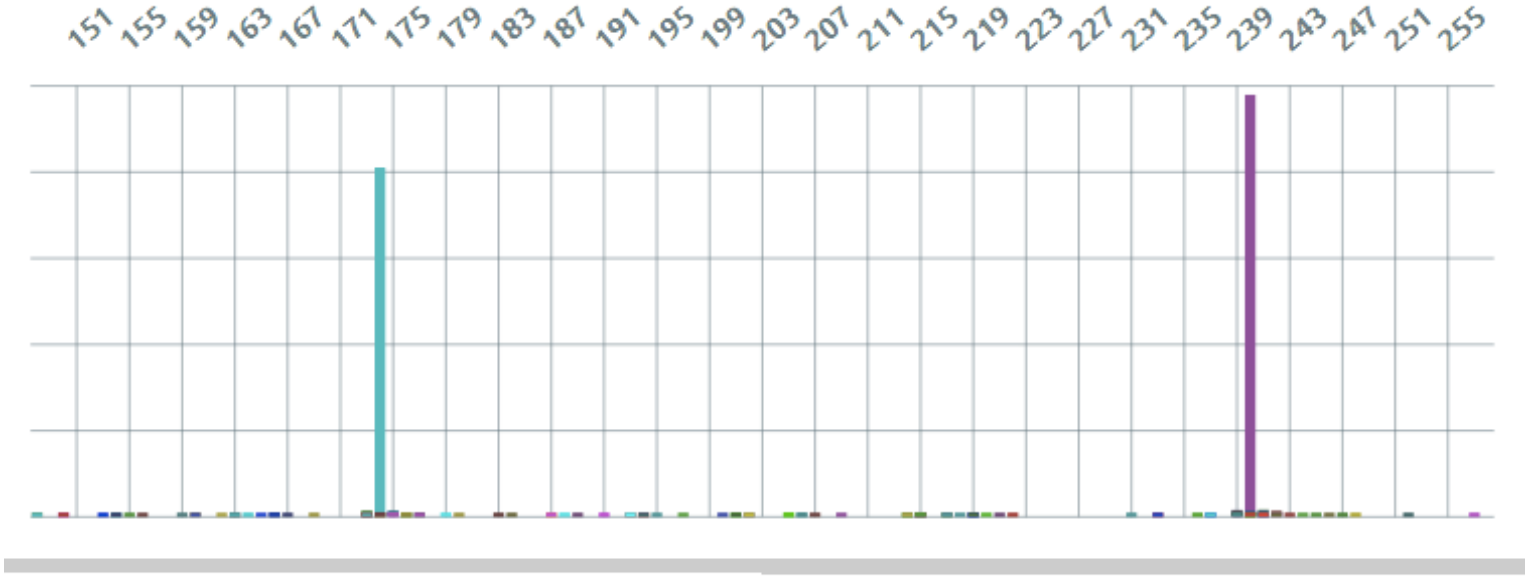
Web application – Browsing and filtering clones



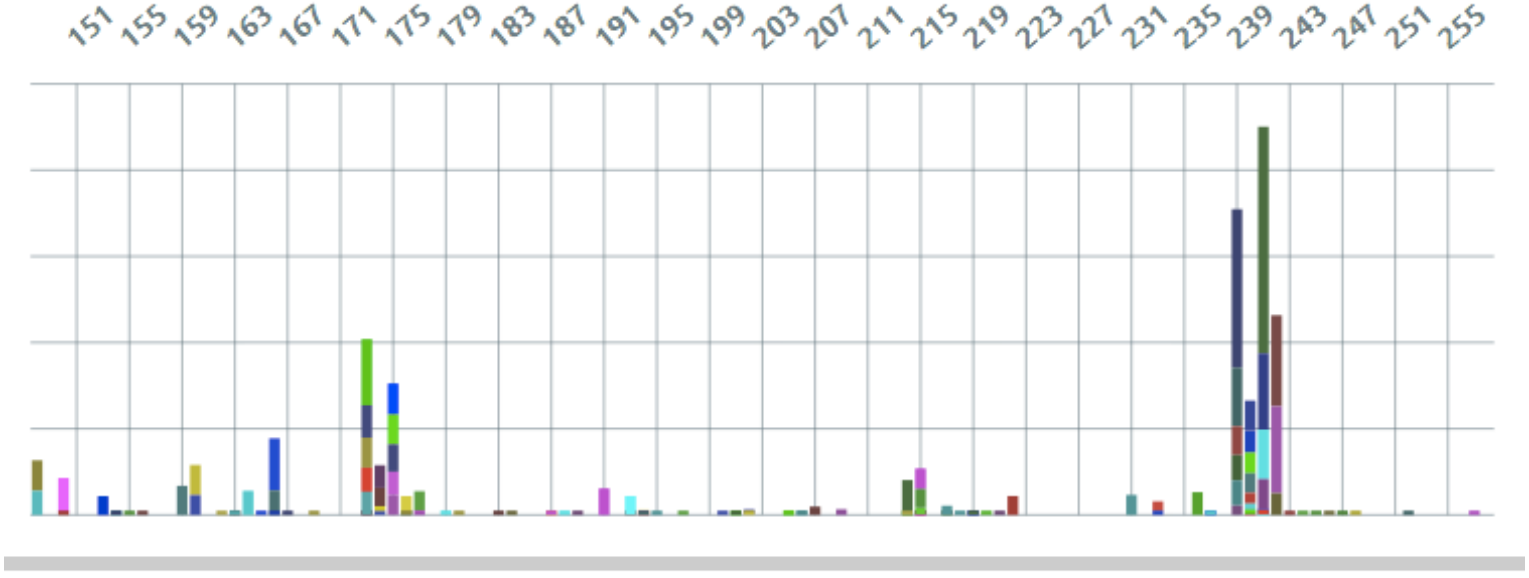
Web application – Browsing and filtering clones



Web application – Browsing and filtering clones



2 clones, 1 875 264 reads (89.15%) (focus) (hide)



Web application: traceability– Reports for medical records

Sample info (IonXpress_086)

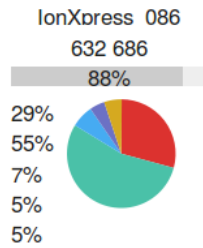
Filename:	IonXpress_086.L1413629.fastq
Sample date:	2014-12-19
Software used:	vidjil dev 0cf35de (2015-02-17)
Parameters:	-i -g germline -c clones -z 100 -r 1
Analysis date:	2015-02-17 21:50:09

User note

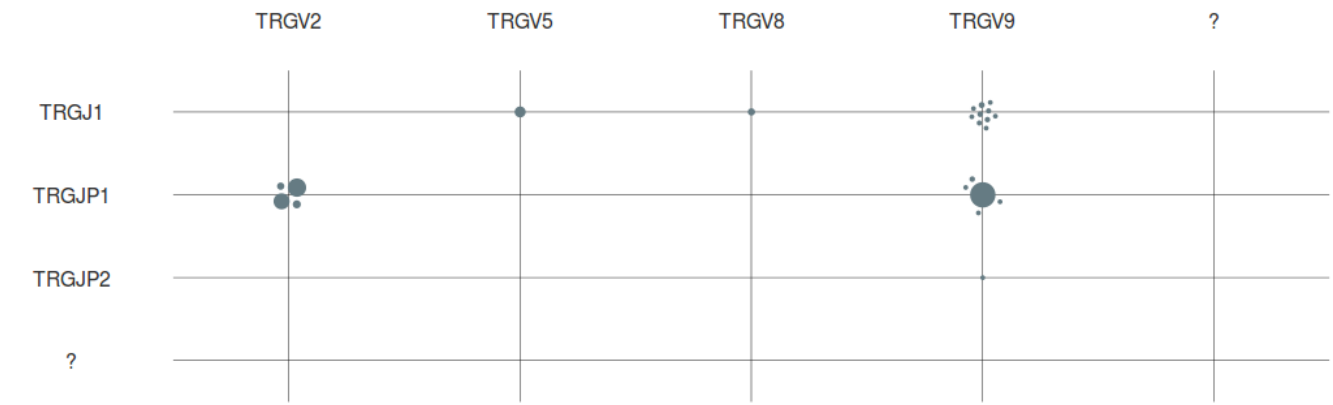
```
PCR avec 250ng d'ADN, Ion Fgt+ 50µL
```

Reads distribution

total
segmented
TRG
IGK+
VdJa
IGH
TRD+



IonXpress_086 – TRG



>TRGV9*01 -8/10/-3 TRGJP1*01 13.44% (45.99% of TRG)

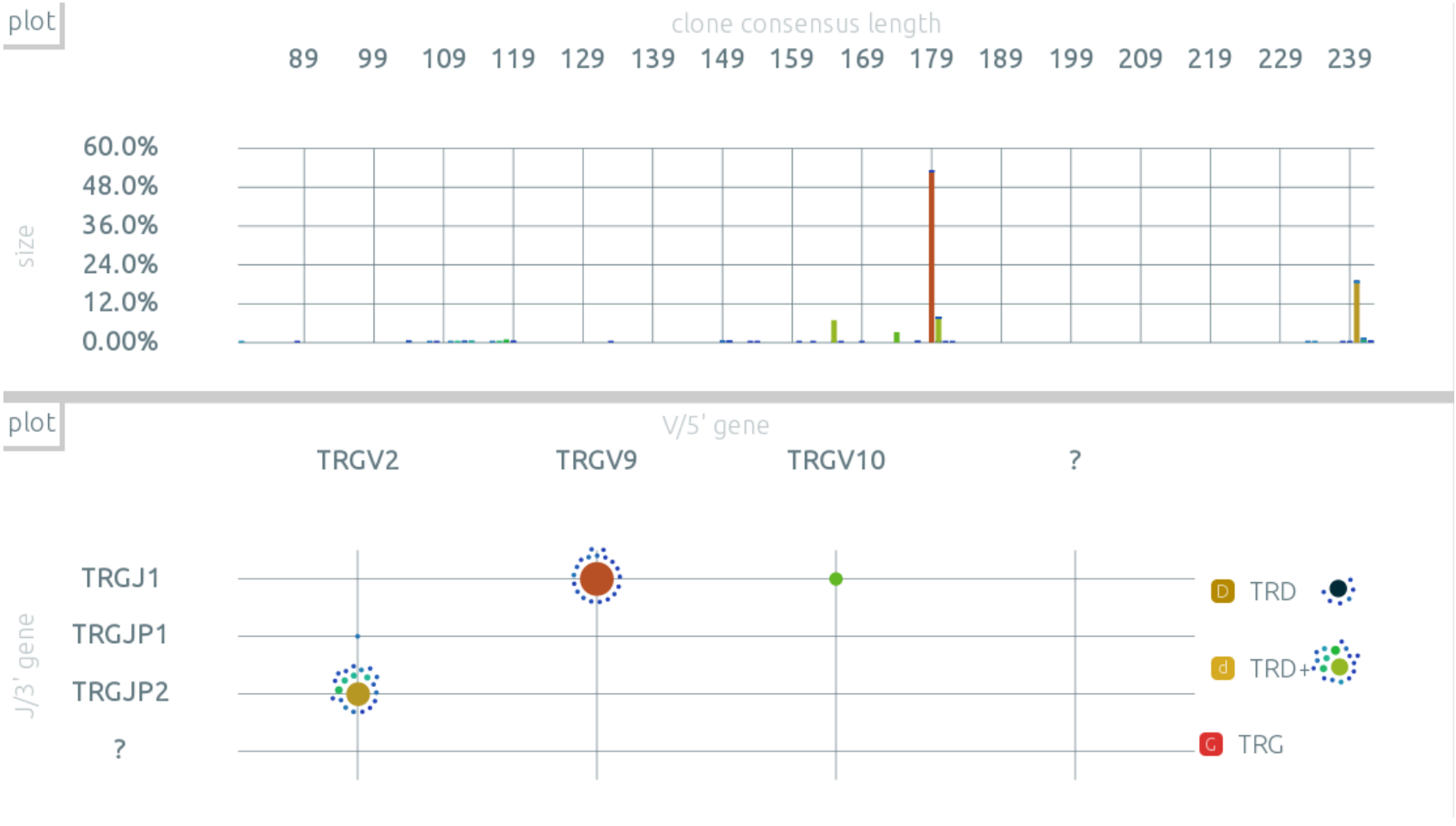
CGGCACTGTCAGAAAGGAATCCGGCATTCCGTCAGGCAAATTTGAGGTGGATAGGATACCTGAAACGTCTACATCCACTCTCACCATTCACAATGTAGAGAAACAGGACATAGCT
ACCTACTACTGTGCCTGT CTCACGAGGG CCACTGGTTGGTTCAAGATATTTGCTGAAGGGACTAAGCTCATAGTAACTTCGCCTGGTAA

>TRGV2*01 -9/6/-1 TRGJP1*01 5.062% (17.32% of TRG)

GGAAGGCCCCACAGCGTCTTCAGTACTATGACTCCTACAACTCCAAGGTTGTGTTGGAATCAGGAGTCAGTCCAGGGAAGTATTATACTTACGCAAGCACAAGGAACAACTTGAG
ATTGATACTGCGAAATCTAATTGAAAATGACTTTGGGGTCTATTACTGTGCCACC CCCTGG TACCACTGGTTGGTTCAAGATATTTGCTGAAGGGACTAAGCTCATAGTAA
CTTCGCCTGGTAA

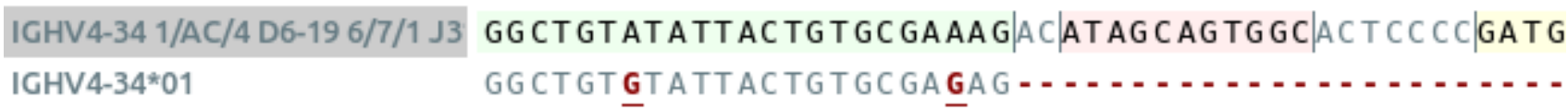
2017 News – Single samples displayed with two plots

Both Genescan-like view and V/J genes grid, or any other plot



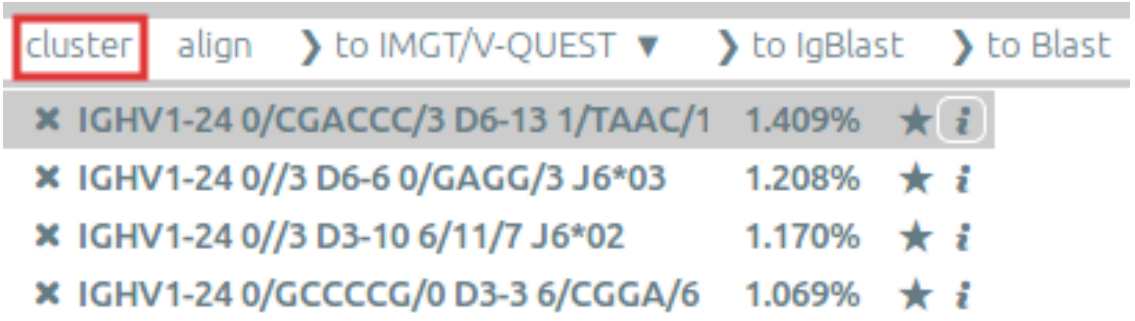
2017 News – Germline Sequences (experimental)

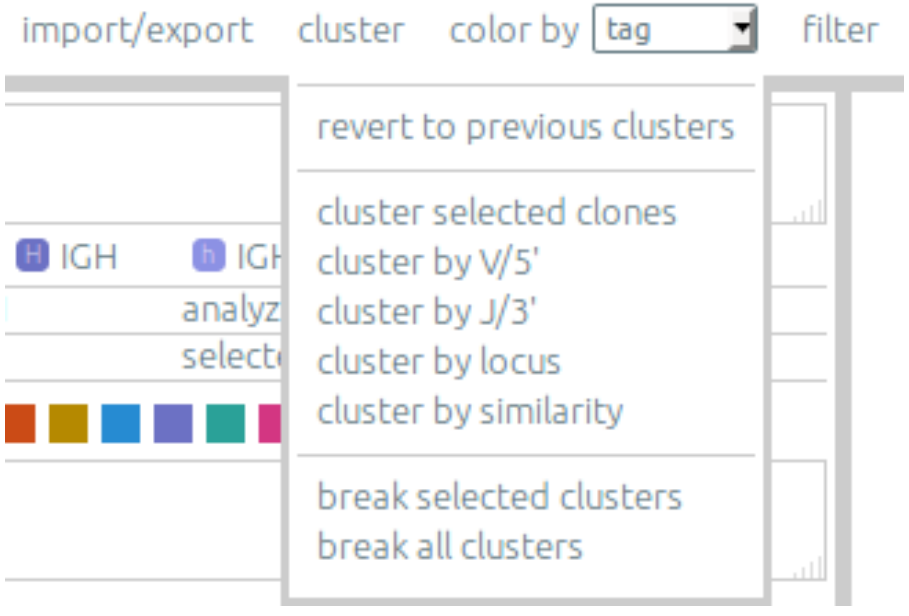
Compare clone sequences to the original germlines



2017 News – More clustering options, including undo

(was previously “merge”)





2017 News – Improved permanent URLs

Saving elments of the client state: set, plot, clones...

Discuss data with colleagues or with us

`http://app.vidjil.org/?set=3241&config=39`

`http://app.vidjil.org/?set=3241&config=39&plot=v,size,bar&clone=11,31`

More elements to come

Exchanging data between RepSeq software

Encoding a set of clones with V(D)J recombinations, possibly with user annotations

```
{
  "id": "clone-001",
  "sequence": "CTCATACACCCAGGAGGTGGAGCTGGATATTGATACTACGAAATCTAATTG(...)",
  "reads" : [ 243241 ],
  "germline": "TRG",
  "top": 1,
  "seg":
  {
    "5": {"name": "TRGV5*01",  "start": 1,   "stop": 86},
    "3": {"name": "TRGJ1*02",  "start": 89,  "stop": 118},
    "cdr3": { "start": 77, "stop": 104, "seq": "gccacctgggccttatt(...)" }
  }
}
```

This .json format is described in `doc/format-analysis.org`. We are open to make evolve this format to better match other software needs.

Upcoming: encoding distributions.

Exchanging data between RepSeq software

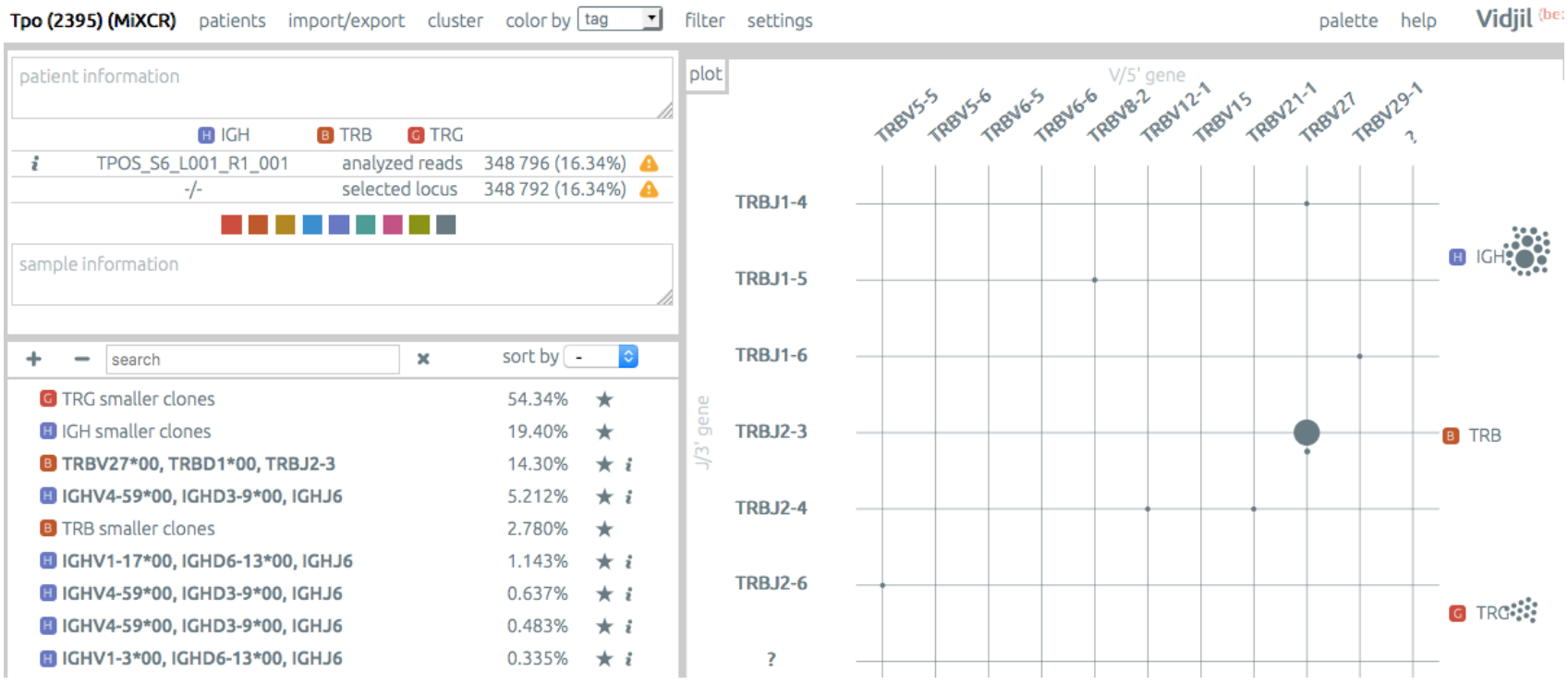
Encoding a set of clones with V(D)J recombinations, possibly with user annotations

What is a clone?

There are several definitions of what may be a clonotype, depending on different RepSeq software or studies. The .json format (and the Vidjil web application) takes a string identifier that may correspond to any kind of clone definition.

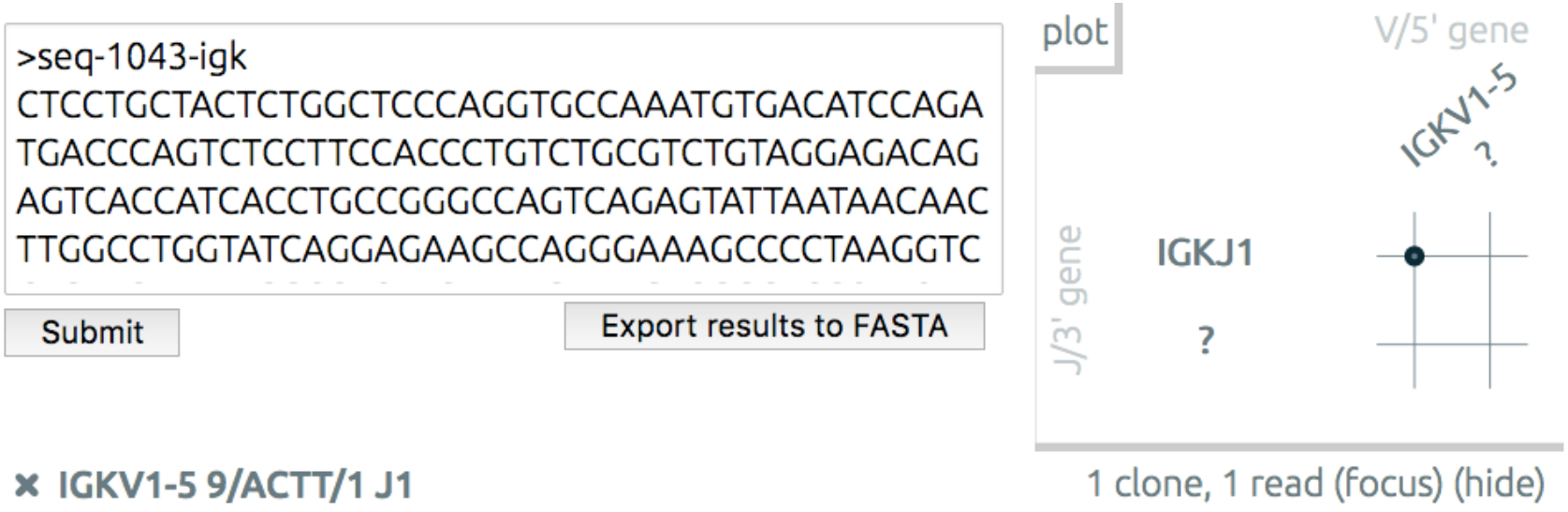
- `CGAGAGGTTACTAT...GGTAGGGCAGTACTAC` (Vidjil, 50 nt window centered on the CDR3)
- `CARPRDWNTYYYYGMDVW` (CDR3 AA sequence)
- `CARPRDWNTYYYYGMDVW IGHV3-11*00 IGHJ6*00` (CDR3 AA sequence with additional V/J gene information, MiXCR)
- `clone sequence` as computed by the ARReST
- `IMGT clonotype (AA) or (nt)`

MiXCR output on Vidjil



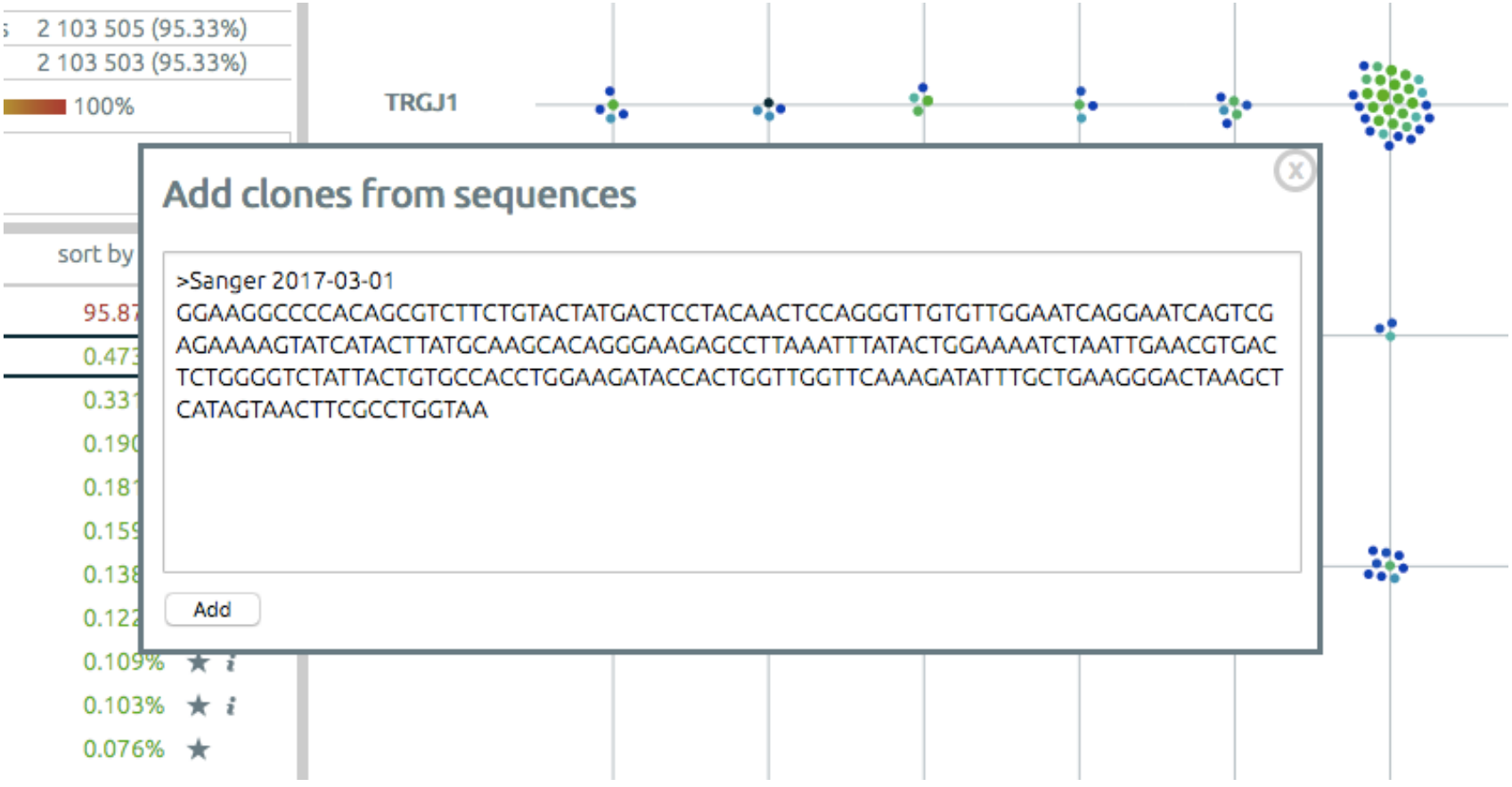
2018: We plan to display results from other software.

app.vidjil.org/analyze – analyzing few sequences



No login, no connection to patient/sample database

Next: Add sequences/clones to one sample



Vidjil Platform

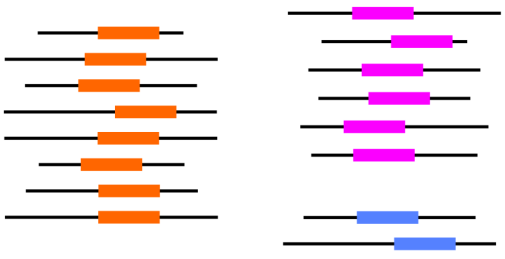
High-throughput Repertoire Sequencing (RepSeq) analysis

Web Application

Patient database

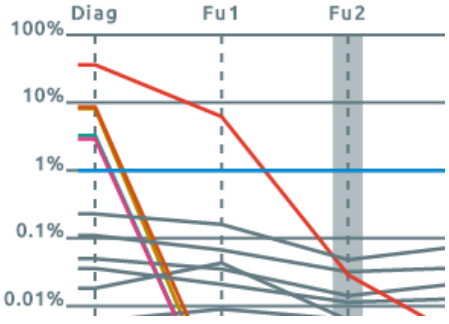
Server

Algorithm



C++

Client



Javascript, d3.js

Python, web2py, AJAX

Patient database and server

Autonomous RepSeq analysis pipeline in a clinical/research lab

Patient database and server

Autonomous RepSeq analysis pipeline in a clinical/research lab

Upload

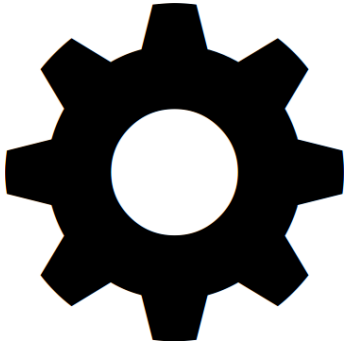
upload list

L1413893.fastq

Patient database and server

Autonomous RepSeq analysis pipeline in a clinical/research lab

Upload



Process

upload list

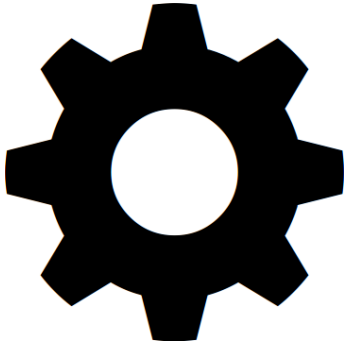
L1413893.fastq

last processing	status
2015-02-09	RUNNING

Patient database and server

Autonomous RepSeq analysis pipeline in a clinical/research lab

Upload



Process

Analyze

upload list

L1413893.fastq

last processing	status
2015-02-09	RUNNING

see the result: multi

Patient database and server – Hosting

app.vidjil.org

Test server: research use only

Secure data center in Lille

Supported by U. Bristol, SIRIC OncoLille and U. Lille

Patient database and server – Hosting

app.vidjil.org

Test server: research use only

Secure data center in Lille

Supported by U. Bristol, SIRIC OncoLille and U. Lille

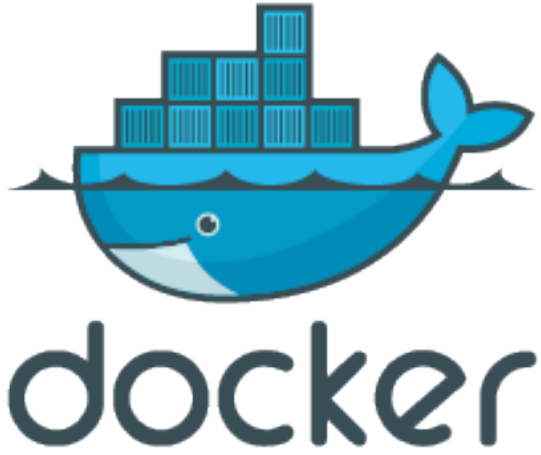
In your lab/hospital

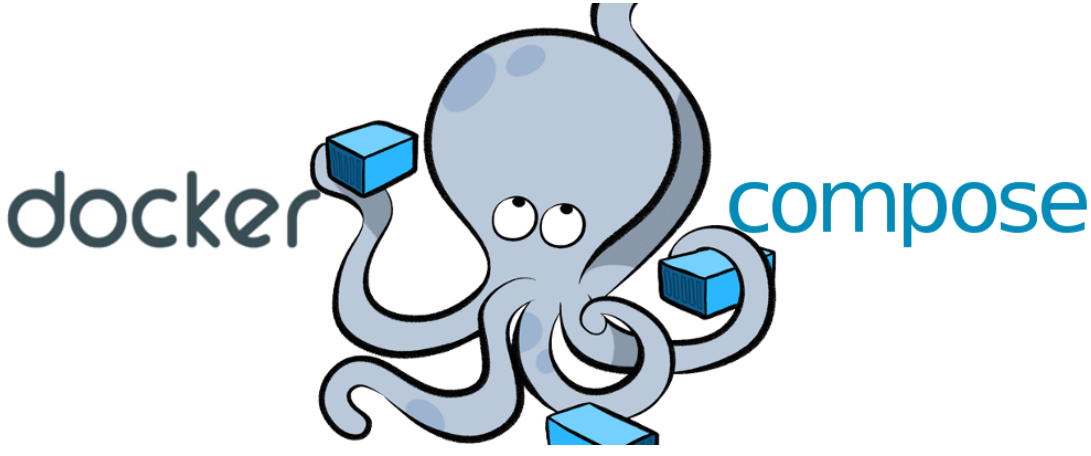
A standard server is enough

Supported by local IT staff

2017 News – Improvements to **Docker** images

Stability and usability





2017 News – Classify and filter your data with **tags**

search

name	birth	info
Austeen Jane (43)	1775-12-16	#T-ALL #diagnosis #cr Winchester hospital
de Cervantes Miguel (42)	1547-09-29	#relapse #pre-SCT #post-SCT
von Schiller Friedrich (41)	1759-01-01	#B-ALL #FR1 blood sample, 10µM primer
Alighieri Dante (40)	1265-01-01	#B-ALL #diagnosis #relapse #KDE
Sand Geroge (39)	1804-02-01	#WM
Goethe Johann Wolfgang (38)	1749-08-28	#T-ALL
Bilderdijk Willem (37)	1756-09-07	
36) ابن خلدون)	1332-05-27	#T-ALL #diagnosis 60% blasts
Němcová Božena (35)	1820-02-04	#T-ALL
Aarestrup Emil (34)	1800-12-04	
Držić Marin (33)	1508-01-01	#CLL 98.4%
Brenner Sophia Elisabet (32)	1659-04-29	#diagnosis #B-ALL #BCR-ABL
Alecsandri Vasile (31)	1801-07-21	#B-ALL #IKAROS
Σολωμός Διονύσιος (30)	1798-04-08	#T-ALL #diagnosis #pre-SCT
Collett Camilla (29)	1813-01-23	
Wilkońska Paulina (28)	1815-01-01	#CLL
Espanca Florbela (27)	1894-12-08	#T-ALL 70% blasts
老子 Laozi (26)		

2017 News – Classify and filter your data with **tags**

search #ALL

ALL

B-ALL

T-ALL

name	birth	info
Austee	1775-12-16	#T-ALL #diagnosis #cr Winchester hospital
de Cerv	1547-09-29	#relapse #pre-SCT #post-SCT
von Sch	1759-01-01	#B-ALL #FR1 blood sample, 10µM primer
Alighieri Dante (40)	1265-01-01	#B-ALL #diagnosis #relapse #KDE
Sand Geroge (39)	1804-02-01	#WM
Goethe Johann Wolfgang (38)	1749-08-28	#T-ALL
Bilderdijk Willem (37)	1756-09-07	
36) ابن خلدون)	1332-05-27	#T-ALL #diagnosis 60% blasts
Němcová Božena (35)	1820-02-04	#T-ALL
Aarestrup Emil (34)	1800-12-04	
Držić Marin (33)	1508-01-01	#CLL 98.4%
Brenner Sophia Elisabet (32)	1659-04-29	#diagnosis #B-ALL #BCR-ABL
Alecsandri Vasile (31)	1801-07-21	#B-ALL #IKAROS
Σολωμός Διονύσιος (30)	1798-04-08	#T-ALL #diagnosis #pre-SCT
Collett Camilla (29)	1813-01-23	
Wilkońska Paulina (28)	1815-01-01	#CLL
Espanca Florbela (27)	1894-12-08	#T-ALL 70% blasts
老子 Laozi (26)		

2017 News – Classify and filter your data with **tags**

search #B-ALL

name	birth	info
von Schiller Friedrich (41)	1759-01-01	#B-ALL #FR1 blood sample, 10µM primer
Alighieri Dante (40)	1265-01-01	#B-ALL #diagnosis #relapse #KDE
Brenner Sophia Elisabet (32)	1659-04-29	#diagnosis #B-ALL #BCR-ABL
Alecsandri Vasile (31)	1801-07-21	#B-ALL #IKAROS

2017 News – Download auxiliary **output files**

May include bigger lists of clones, statistic analysis...

last processing	status			
2017-10-05 15:43:39	COMPLETED	dl	out	run ...
2017-10-05 16:07:02	COMPLETED	dl	out	run ...

file	size
000078.edges	0
000078.vidjil.log	58.8 kB
000078.vidjil	178 kB
000078.windows.fa	6.33 kB
000078.vdj.fa	36.9 kB
seq	4.10 kB
000078-34.fuse.log	1.09 kB
000078-26.fuse.log	1.72 kB

Working collaboratively on shared data

Lille hospital

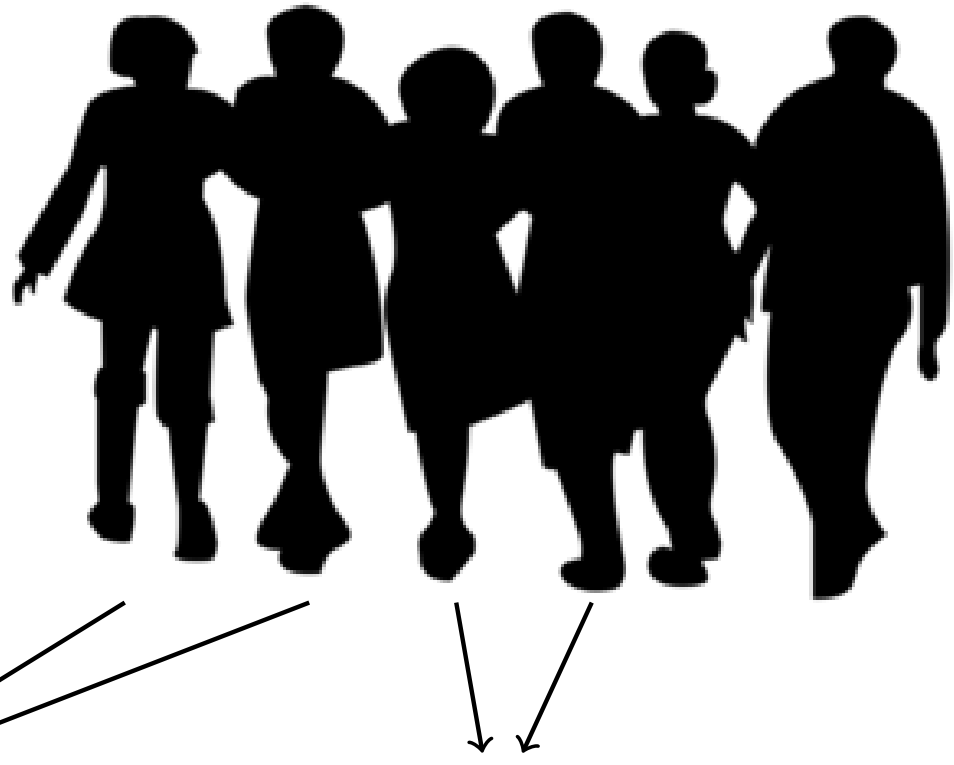
Working collaboratively on shared data

Lille hospital

Upload data
Create patient

Working collaboratively on shared data

Lille hospital

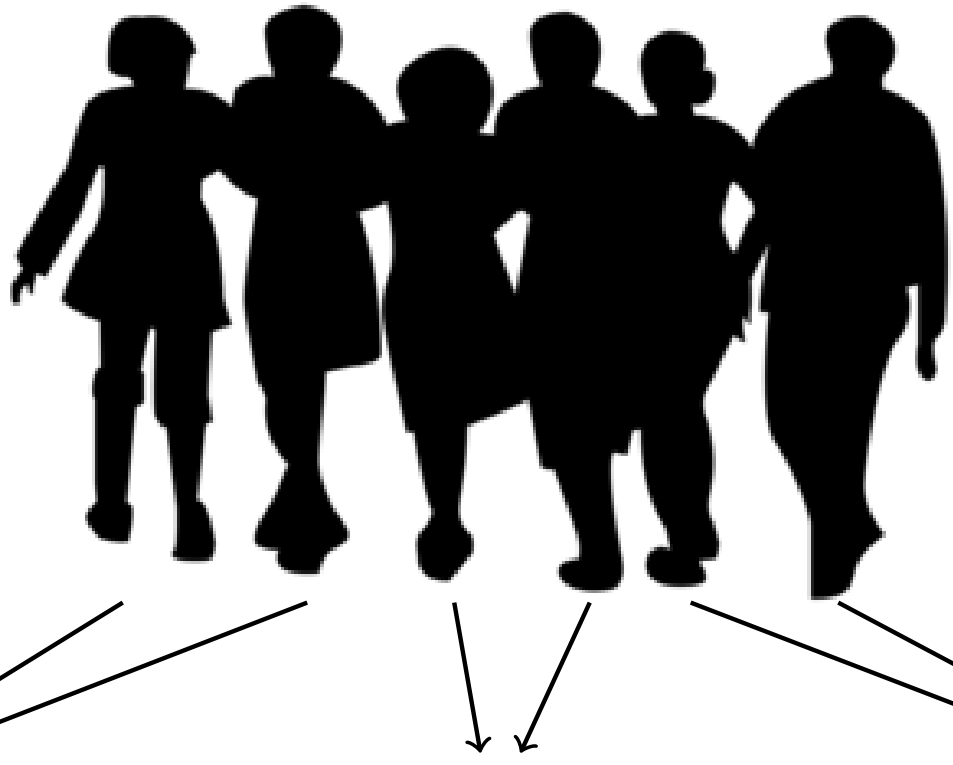


Upload data
Create patient

Annotate results
Freeze/sign (upcoming)

Working collaboratively on shared data

Lille hospital



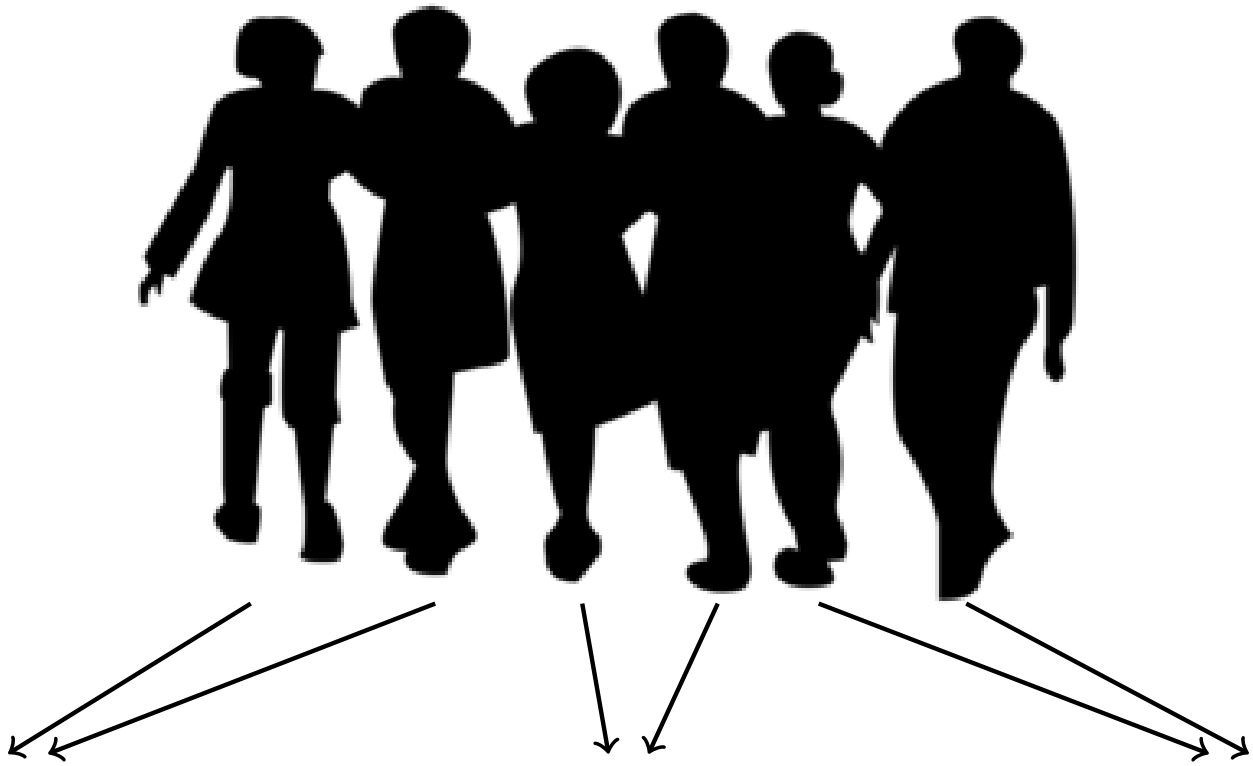
Upload data
Create patient

Annotate results
Freeze/sign (upcoming)

Read-only access

Working collaboratively on shared data

Lille hospital



Upload data
Create patient

Annotate results
Freeze/sign (upcoming)

Read-only access

They can also belong to other organizations

Track updates to patient records

Lille hospital

Track updates to patient records

Lille hospital

2016-09-29 10:15 – **Léa** created patient Jane Doe

Track updates to patient records

Lille hospital

2016-09-29 10:15 – **Léa** created patient Jane Doe

2016-09-29 11:45 – **Guy** uploaded seqs.fastq (15 MB)

Track updates to patient records

Lille hospital

2016-09-29 10:15 – **Léa** created patient Jane Doe

2016-09-29 11:45 – **Guy** uploaded seqs.fastq (15 MB)

2016-09-29 12:12 – **Élodie** annotated the results

Track updates to patient records

Lille hospital

2016-09-29 10:15 – **Léa** created patient Jane Doe

2016-09-29 11:45 – **Guy** uploaded seqs.fastq (15 MB)

2016-09-29 12:12 – **Élodie** annotated the results

2016-09-30 10:47 – **Cindy** froze and signed the report