

Vidjil, a platform for the interactive analysis of immune repertoires

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



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

Hôpital Rennes
GBMHM/Inca

November 21, 2017
contact@vidjil.org

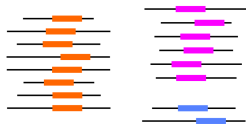
Vidjil Platform

High-throughput Repertoire Sequencing (RepSeq) analysis

Web Application

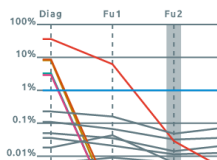
Patient database
Server

Algorithm

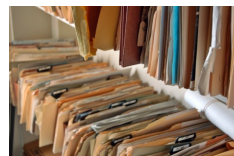


C++

Client



Javascript, d3.js



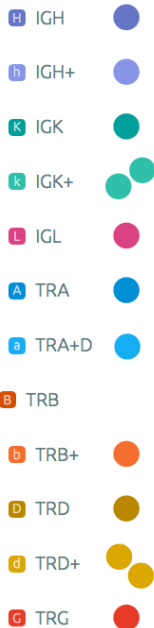
Python, web2py,
AJAX

- ▶ 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

Vidjil analyzes recombinations on all human TR/Ig locus

complete recombinations		incomplete/special recombinations	
TRA	Va-Ja		
TRB	Vb-(Db)-Jb	TRB+	Db-Jb
TRD	Vd-(Dd)-Jd	TRD+	Vd-Dd3, Dd2-(Dd)-Jd, Dd2-Dd3
		TRA+D	Vd-(Dd)-Ja, Dd-Ja
TRG	Vg-Jg		
IGH	Vh-(Dh)-Jh	IGH+	Dh-Jh
IGL	Vi-Ji		
IGK	Vk-Jk	IGK+	Vk-KDE, INTRON-KDE



V(D)J Designation

Comparison against IMGT V(D)J germline genes

. . . TGGGATGGGCCTATTATCAG . . .

V(D)J Designation

Comparison against IMGT V(D)J germline genes

... TGGGATGGGCCTATTATCAG ...

V4 TGGGATGGG

V(D)J Designation

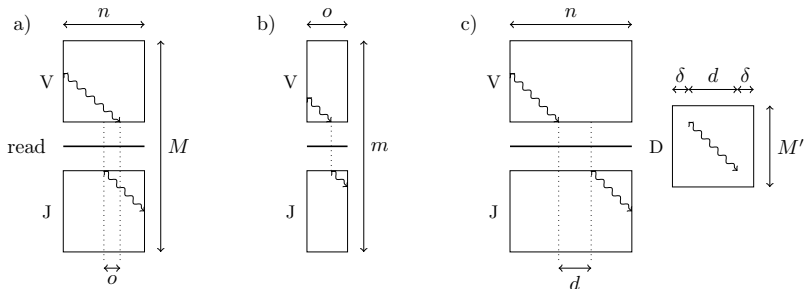
Comparison against IMGT V(D)J germline genes

```
... TGGGATGGG CCTATTATCAG ...  
V4 TGGGATGGG TATTATAAG J2
```

V(D)J Designation

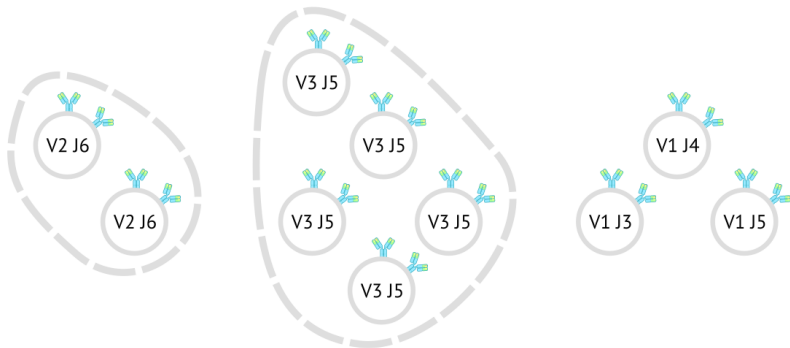
Comparison against IMGT V(D)J germline genes, dynamic programming

... TGGGATGGG CCTATTATCAG ...
V4 TGGGATGGG TATTATAAG J2



Immune Repertoire Sequencing (RepSeq)

Identification of all VDJ recombinations



20%

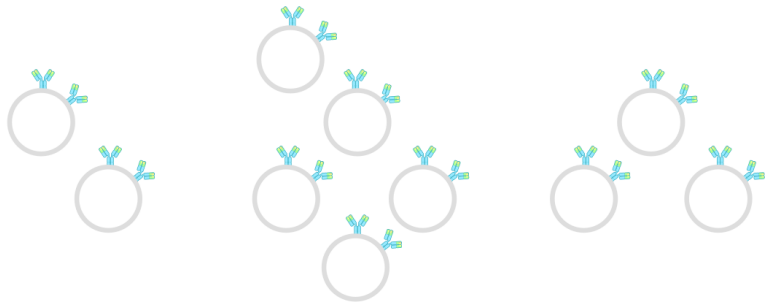
50%

30%

1 000 000 VDJ = 100 000 s

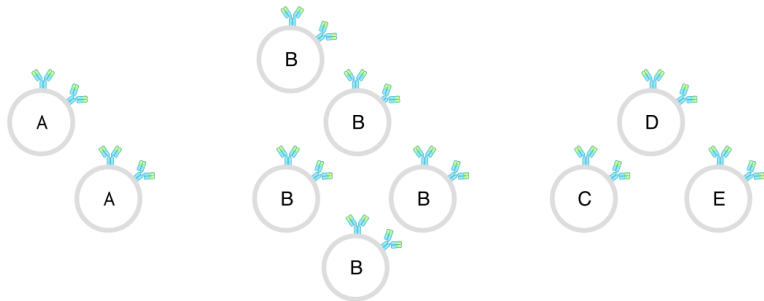
Immune Repertoire Sequencing (RepSeq)

Clone clustering



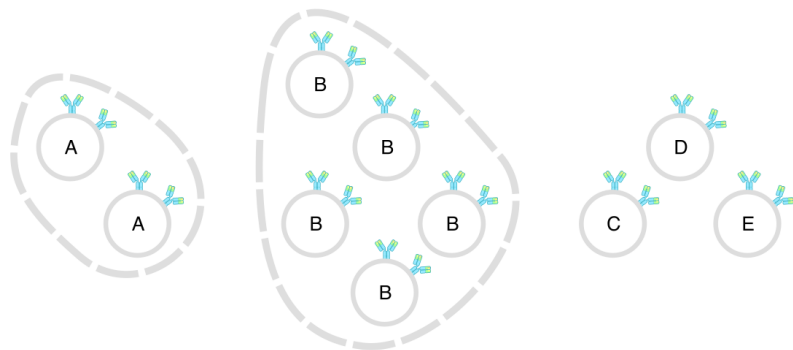
Immune Repertoire Sequencing (RepSeq)

Clone clustering



Immune Repertoire Sequencing (RepSeq)

Clone clustering



20%

50%

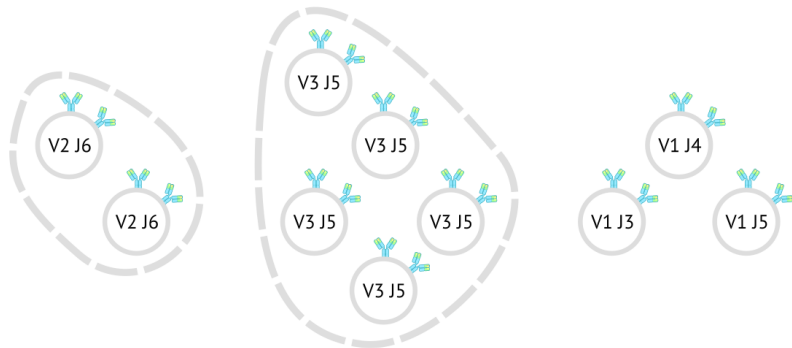
30%

1 000 000 VDJ = 100 s

Giraud, Salson et al., BMC Genomics, 2014

Immune Repertoire Sequencing (RepSeq)

Clone clustering



20%

50%

30%

1 000 000 VDJ = 100 s

Giraud, Salson et al., BMC Genomics, 2014

Fast identification of a window centered on the CDR3

Clone clustering

ACACGGCCGTGTATTACTGTGCGAGAGAGCTGAATACTTCCAGCACTGGGGCC

$O(n)$ linear time

Giraud, Salson et al., BMC Genomics, 2014

Fast identification of a window centered on the CDR3

Clone clustering

parts of V genes

ACAC CACG ACGG CGGC GGCC
GCCG TCTT CTTC TTCC TCCA
CCAA CAAC AACC ACCT CCTT
CTTG TTGG TGGA ACTT ...

parts of J genes

ATAC TACT ACTT CCAG CAGC
AGCA GCAC TGGG GGGC GGCA
GCAA CAAG AAGA AGAG GAGT
AGTT GTTG TTGG ...

ACACGGCCGTGTATTACTGTGCGAGAGAGCTGAATACTTCCAGCACTGGGGCC

$O(n)$ linear time

Fast identification of a window centered on the CDR3

Clone clustering

parts of V genes

ACAC CACG ACGG CGGC GGCC
GCCG TCTT CTTC TTCC TCCA
CCAA CAAC AACC ACCT CCTT
CTTG TTGG TGGA ACTT ...

parts of J genes

ATAC TACT ACTT CCAG CAGC
AGCA GCAC TGGG GGGC GGCA
GCAA CAAG AAGA AGAG GAGT
AGTT GTTG TTGG ...

ACACGGCCGTGTATTACTGTGCGAGAGAGCTGAATACTTCCAGCACTGGGGCC



$O(n)$ linear time

Fast identification of a window centered on the CDR3

Clone clustering

parts of V genes

ACAC CACG ACGG CGGC GGCC
GCCG TCTT CTTC TTCC TCCA
CCAA CAAC AACC ACCT CCTT
CTTG TTGG TGGG ACTT ...

parts of J genes

ATAC TACT ACTT CCAG CAGC
AGCA GCAC TGGG GGGC GGCA
GCAA CAAG AAGA AGAG GAGT
AGTT GTTG TTGG ...

Fenêtre de 50 nucléotides

ACACGGCCGTGTATTACTGTGCGAGAGAGCTGAATACTTCCAGCACTGGGGCC

$O(n)$ linear time

Fast identification of a window centered on the CDR3

Clone clustering

parts of V genes

ACAC CACG ACGG CGGC GGCC
GCCG TCTT CTTC TTCC TCCA
CCAA CAAC AACC ACCT CCTT
CTTG TTGG TGGG ACTT ...

parts of J genes

ATAC TACT ACTT CCAG CAGC
AGCA GCAC TGGG GGGC GGCA
GCAA CAAG AAGA AGAG GAGT
AGTT GTTG TTGG ...

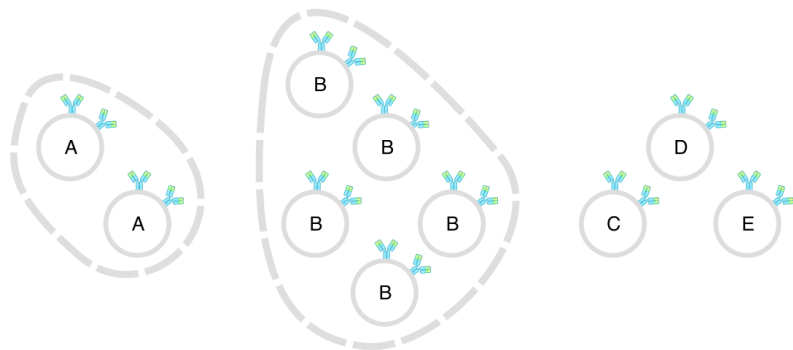
Fenêtre de 50 nucléotides

ACACGGCCGTGTATTACTGTGCGAGAGAGCTGAATACTTCCAGCACTGGGGCC

$O(n)$ linear time

Immune Repertoire Sequencing (RepSeq)

Clone clustering



20%

50%

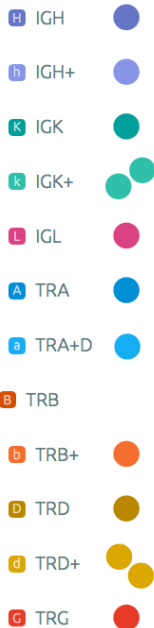
30%

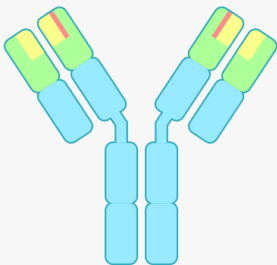
1 000 000 VDJ = 100 s

Giraud, Salson et al., BMC Genomics, 2014

Vidjil analyzes recombinations on all human TR/Ig locus

complete recombinations		incomplete/special recombinations	
TRA	Va-Ja		
TRB	Vb-(Db)-Jb	TRB+	Db-Jb
TRD	Vd-(Dd)-Jd	TRD+	Vd-Dd3, Dd2-(Dd)-Jd, Dd2-Dd3
		TRA+D	Vd-(Dd)-Ja, Dd-Ja
TRG	Vg-Jg		
IGH	Vh-(Dh)-Jh	IGH+	Dh-Jh
IGL	Vi-Ji		
IGK	Vk-Jk	IGK+	Vk-KDE, INTRON-KDE





CTGTGCTACGTG TRGJ2*01



The search engine that tracks... your clones!

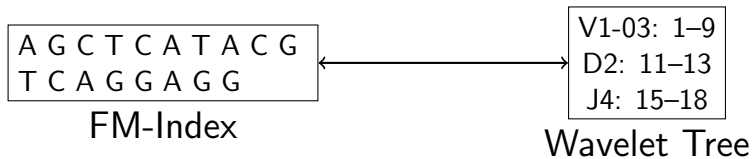


Indexing labelled sequences (current PhD)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
AGCTCATACGTCAGGAGG
V1-03 D2 J4

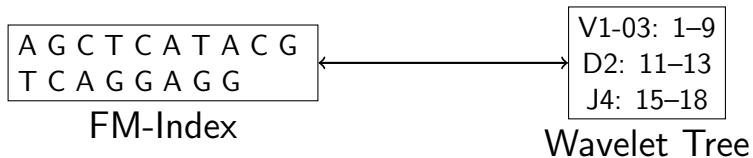
Indexing labelled sequences (current PhD)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
AGCTCATACGTCAGGAGG
V1-03 D2 J4



Indexing labelled sequences (current PhD)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
AGCTCATACGTCAGGAGG
V1-03 D2 J4



Combined sequence/label queries

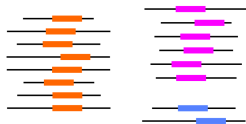
Vidjil Platform

High-throughput Repertoire Sequencing (RepSeq) analysis

Web Application

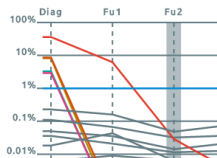
Patient database
Server

Algorithm

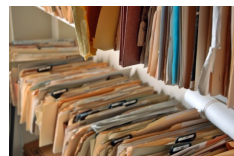


C++

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