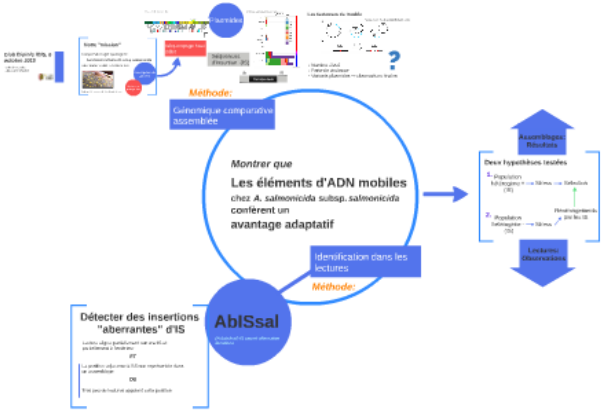
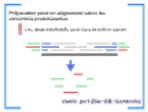




- AbISsal **est** un script
- AbISsal **identifie** les régions flanquantes d'IS données par l'utilisateur
- AbISsal **devrait** détecter des transpositions rares par des régions flanquantes peu représentées

Est-ce que les IS chez *A. salmonicida* sont sollicitées en période de stress pour produire des innovations adaptatives?

Et après...

Détection d'insertions rares
Comparaisons

Analyses sur lecture
Comparaisons
"Transformable"

Transpositions
Connaissances
????

Orienter AbISsal vers la comparaison de jeux de données

- à la sortie
- à l'entrée

Incorporer des méthodes de clustering plus efficaces

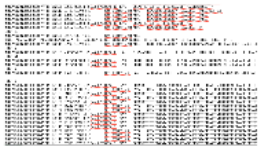
Autres suggestions?

CRSNG NSERC

Ressources Aquatiques Québec RAQ

C$ub Bioinfo IBIS

Les étapes d'AbISsal



Club Bioinfo IBIS, 6 octobre 2015

Katherine Tanaka
Laboratoire Charette

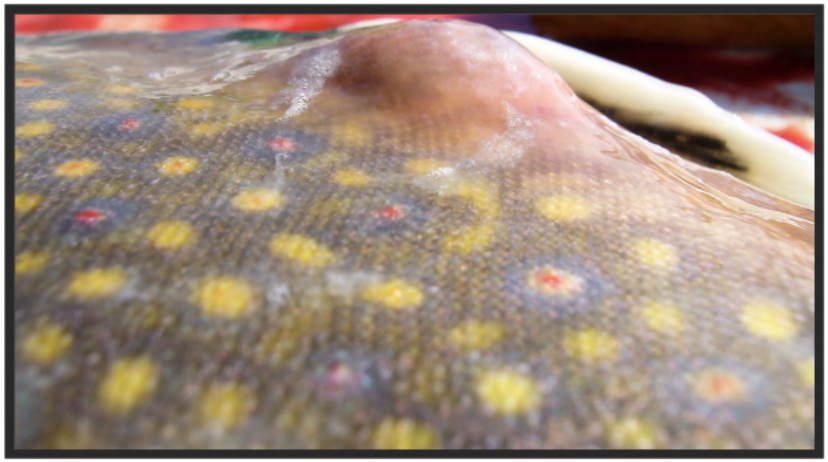
INSTITUT DE BIOLOGIE INTÉGRATIVE ET DES SYSTÈMES
IBIS

Notre "mission"

Comprendre l'agent pathogène

Aeromonas salmonicida subsp. ***salmonicida***

pour mieux traiter la furonculose

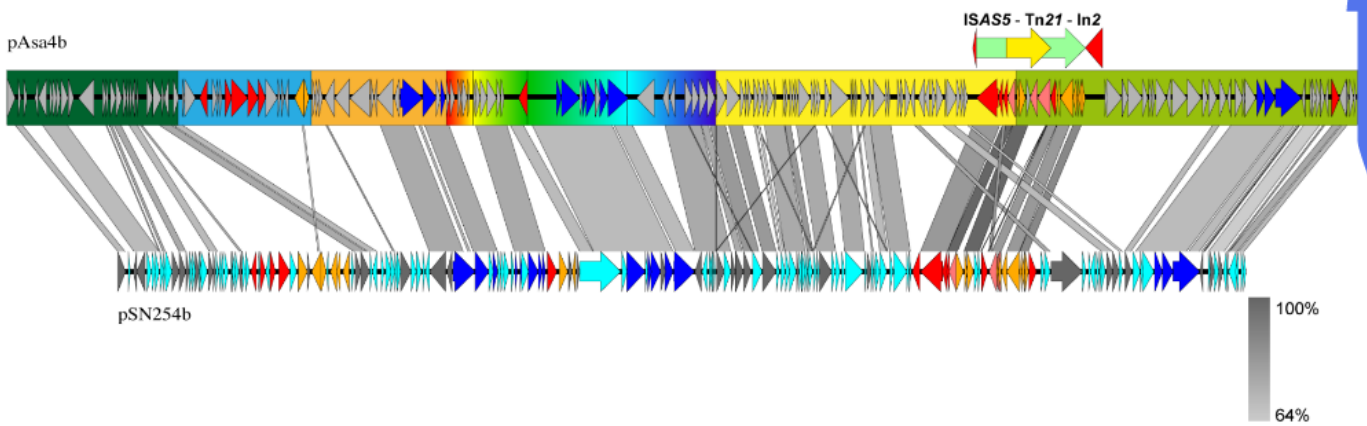


Dallaire-Dufresne et coll., Vet Microbiol, 2013

Investigation du génome

Virulence et pathogenèse

Tanaka et coll., en préparation



Plasmides

Séquençage haut débit

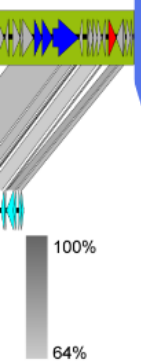
nicida

gation du
nome

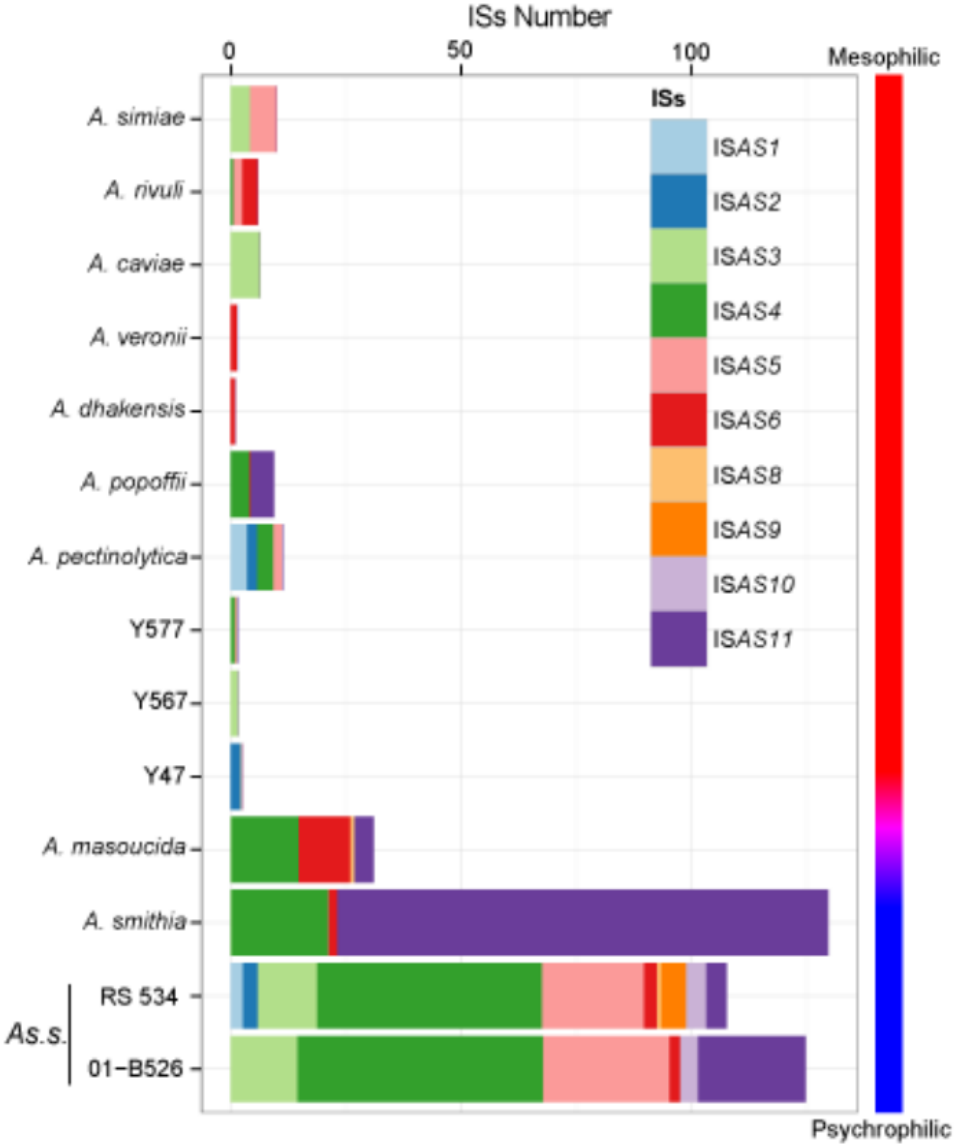
Séquences
d'insertion (I

I

Plasmides



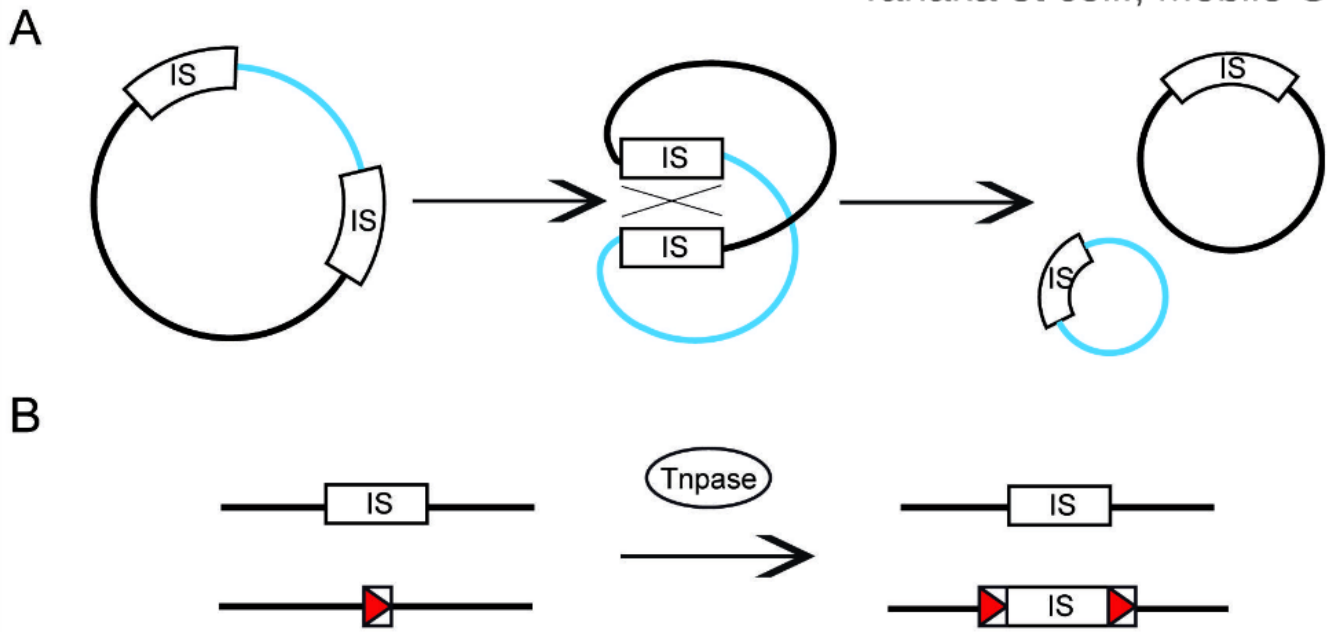
A. Vincent, communication personnelle



Séquences d'insertion (IS)

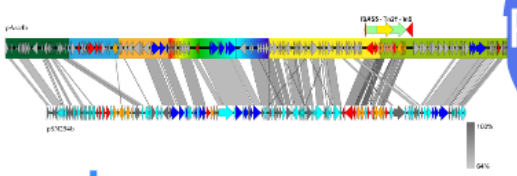
Les fauteuses de trouble

Tanaka et coll., Mobile Genetic Elements, 2013



- Nombre élevé
- Perte de virulence
- Variants plasmides → observations finales

Tanaka et coll., en préparation

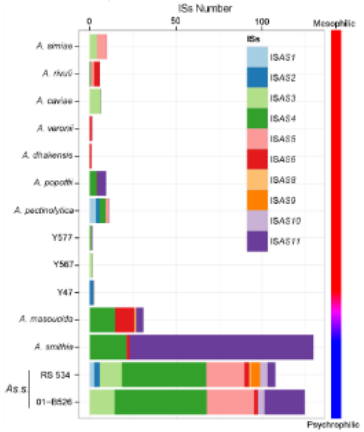


Plasmides

Séquençage haut débit

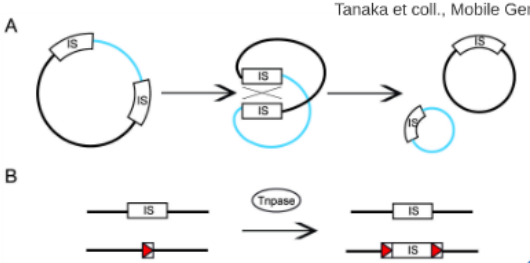
Séquences d'insertion (IS)

A. Vincent, communication personnelle



Les fauteuses de trouble

Tanaka et coll., Mobile Genetic Elements, 2013



?

- Nombre élevé
- Perte de virulence
- Variants plasmides → observations finales

onicida

tigation du génome

Méthode:

Génomique comparative assemblée

Montrer que

Méthode:

Génomique comparative assemblée

Montrer que

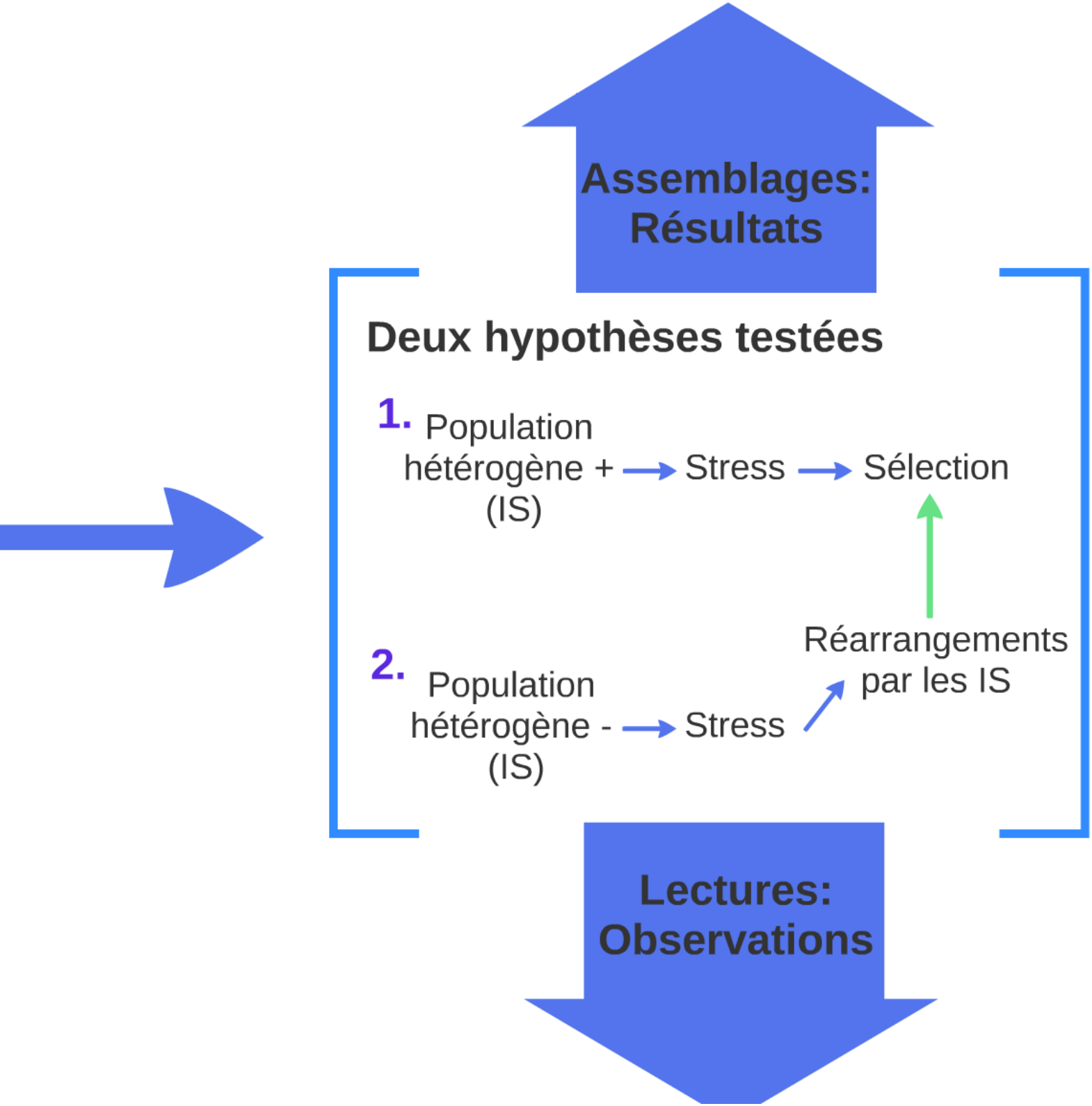
Les éléments d'ADN mobiles

chez *A. salmonicida* subsp. *salmonicida* confèrent un

avantage adaptatif

Identification dans les lectures

Méthode:


Assemblages:
Résultats
Deux hypothèses testées
1.
Population hétérogène + (IS)
Stress
Sélection
2.
Population hétérogène - (IS)
Stress
Réarrangements par les IS
Lectures:
Observations

Les éléments d'ADN mobiles

chez *A. salmonicida* subsp. *salmonicida*
confèrent un
avantage adaptatif

Identification dans les lectures

Méthode:

AbISsal

(A batch of) IS secret alternative locations

des insertions
rrantes" d'IS

ellement sur une IS et
érieur

ET

e à l'IS non représentée dans

OU

s appuient cette position

Détecter des insertions "aberrantes" d'IS

Lecture aligne partiellement sur une IS et partiellement à l'extérieur

ET

La position adjacente à l'IS non représentée dans un assemblage

OU

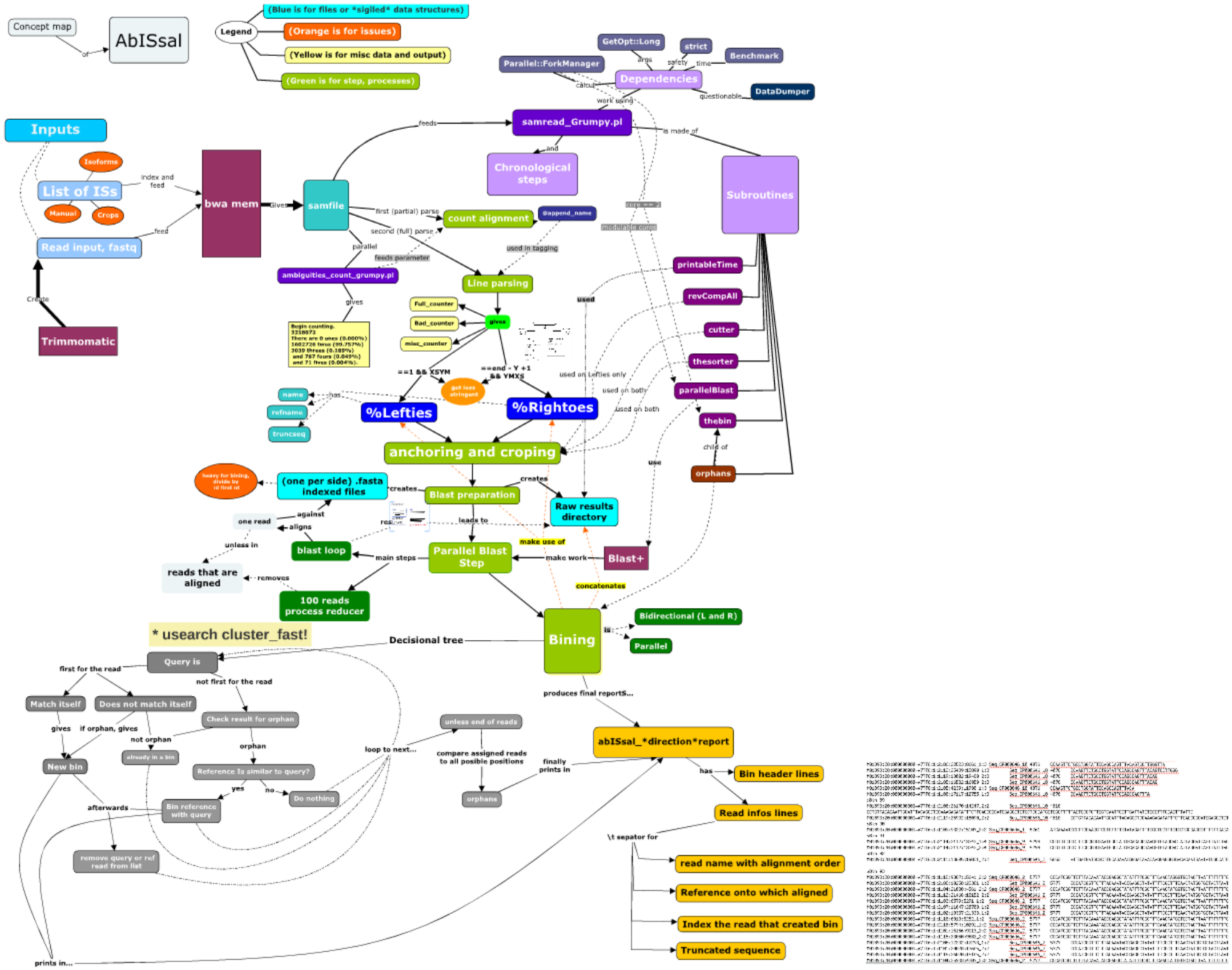
Très peu de lectures appuient cette position

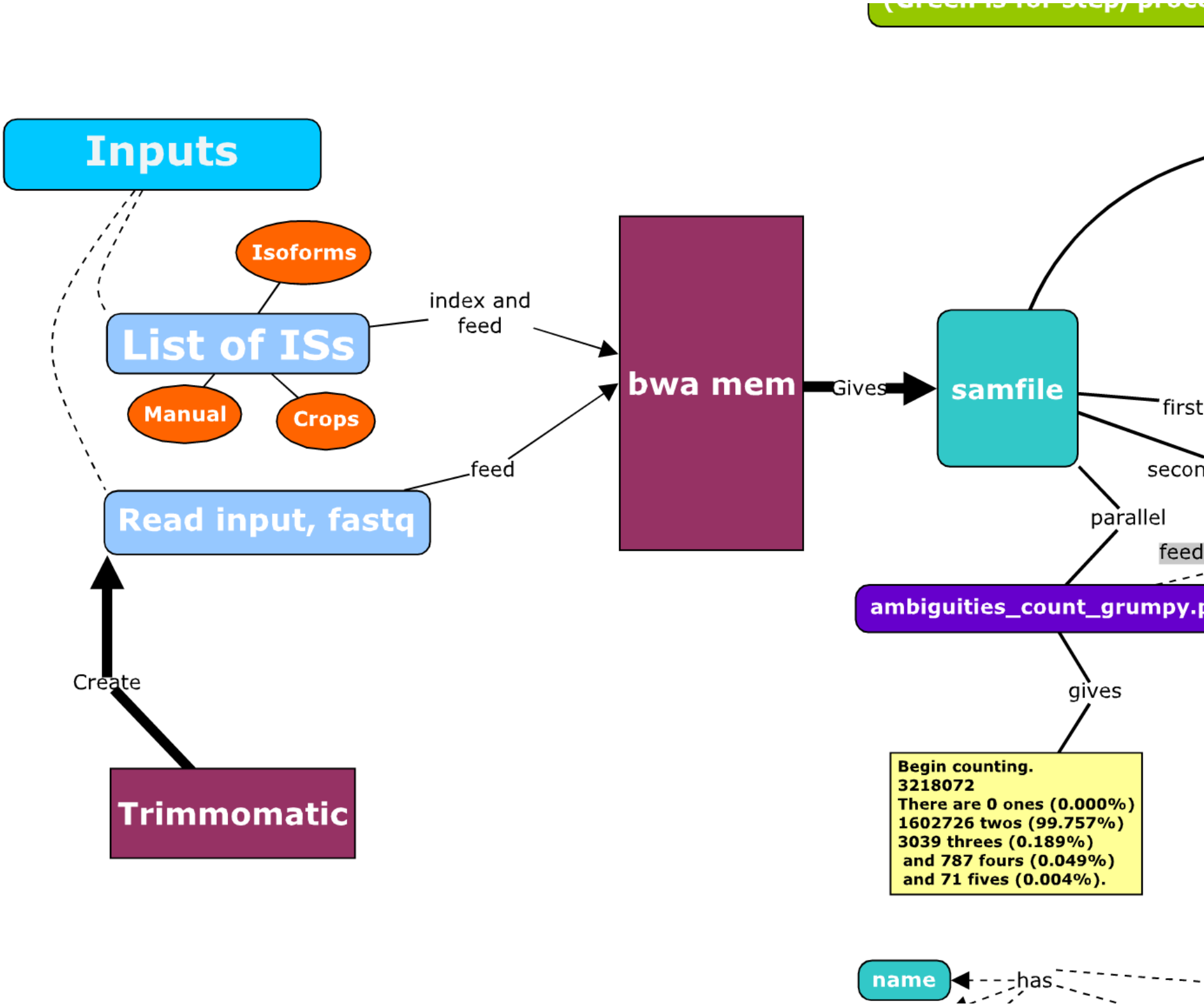
Très peu de lectures appuient cette position

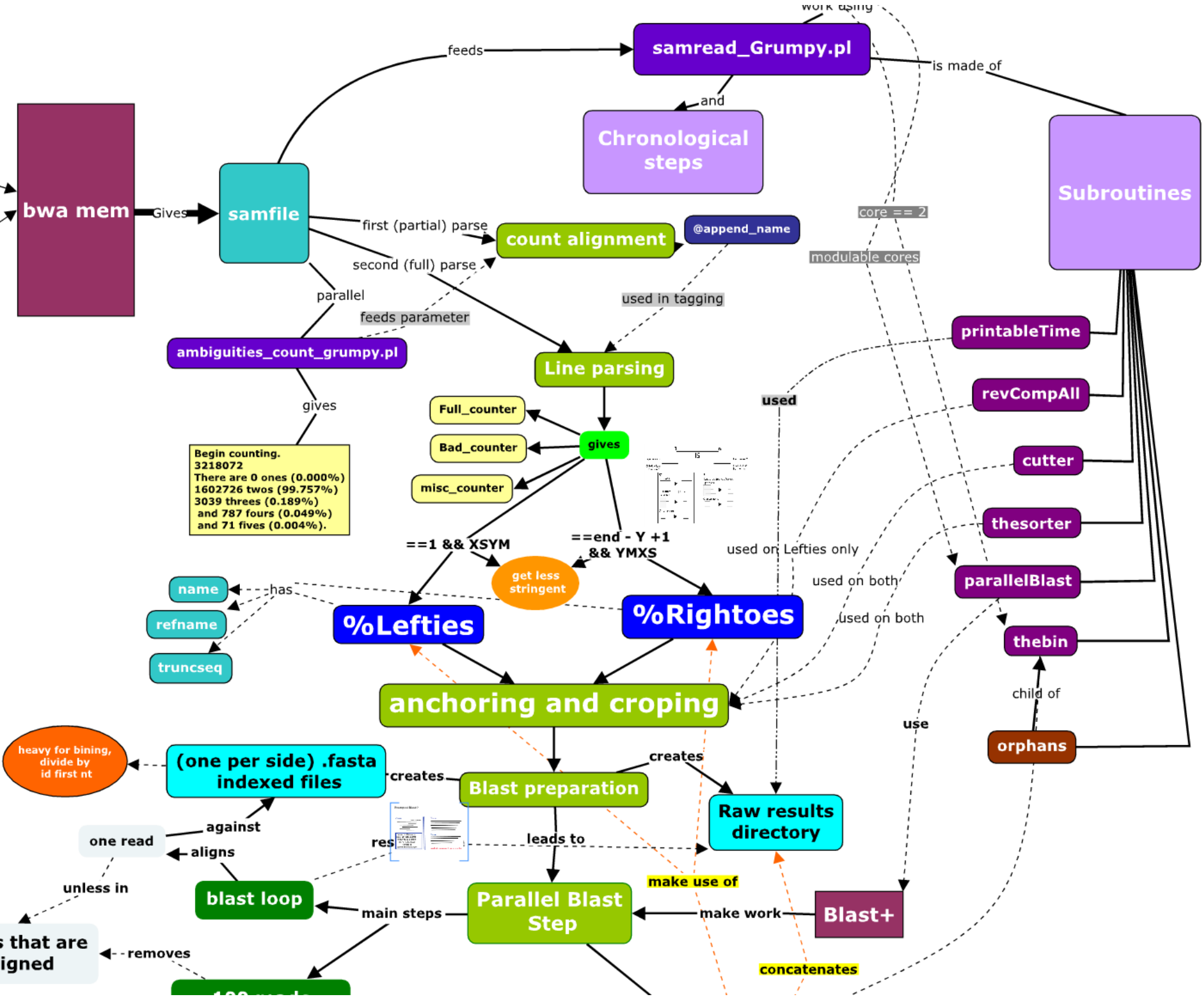
- AbISsal **est** un script
- AbISsal **identifie** les régions flanquantes d'IS données par l'utilisateur
- AbISsal **devrait** détecter des transpositions rares par des régions flanquantes peu représentées

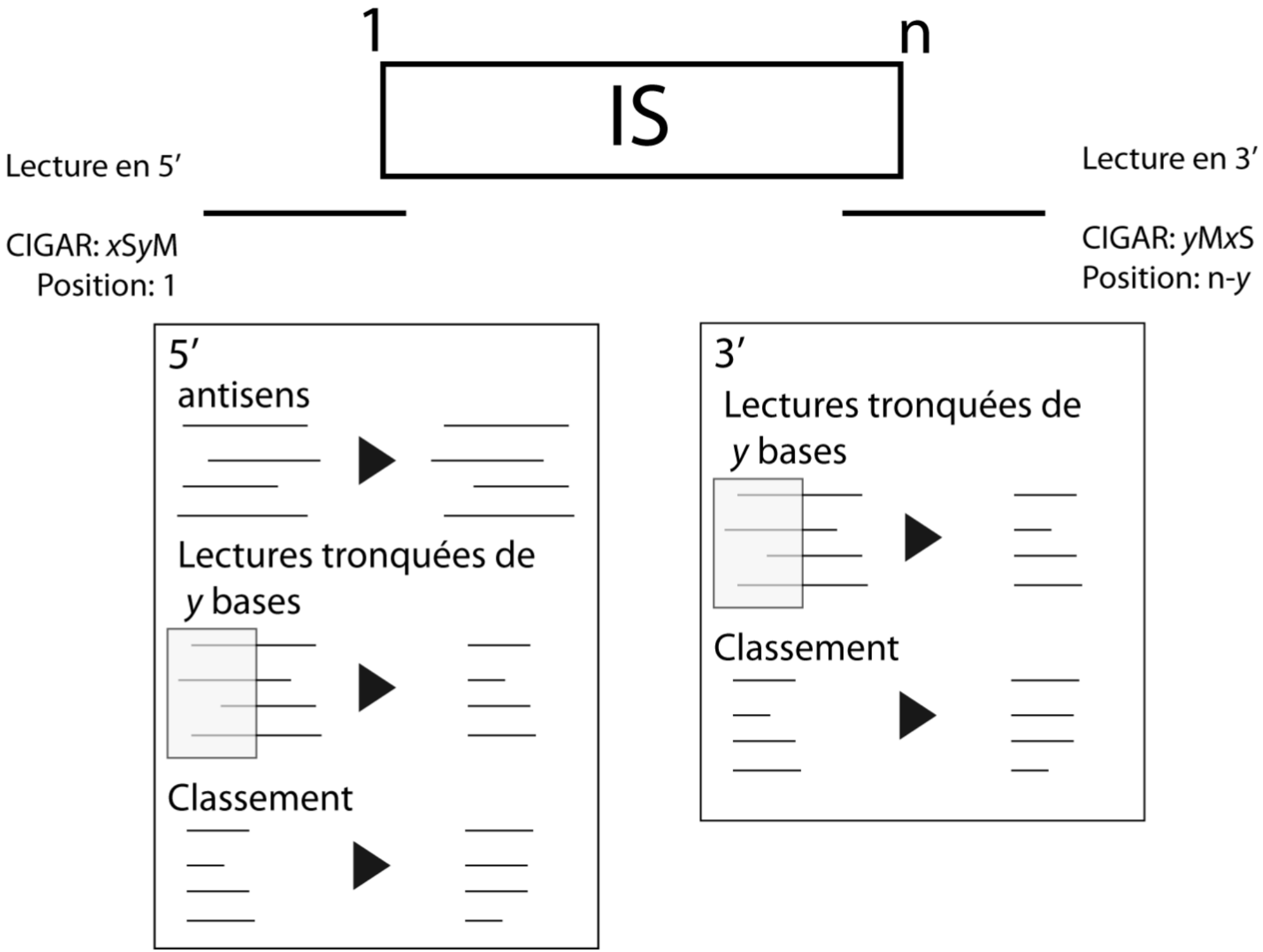
Est-ce que les IS chez *A. salmonicida* sont sollicitées en période de stress pour produire des innovations adaptatives?

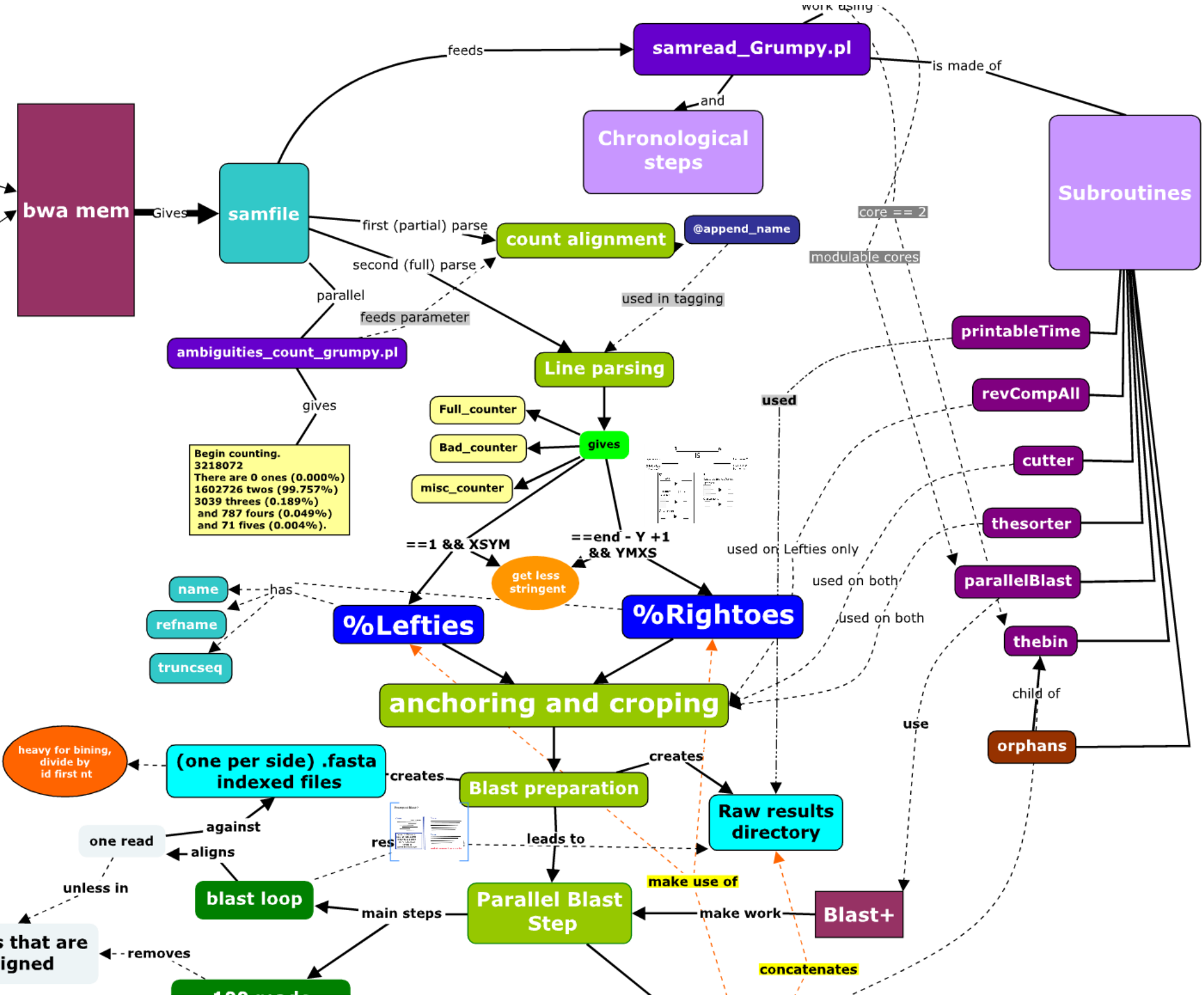
Les étapes d'AbISsal




Inputs
Isoforms
List of ISs
Manual
Crops
index and feed
feed
Read input, fastq
Create
Trimmomatic
bwa mem
Gives
samfile
first
secon
parallel
feed
ambiguities_count_grumpy.p
gives
Begin counting.
3218072
There are 0 ones (0.000%)
1602726 twos (99.757%)
3039 threes (0.189%)
and 787 fours (0.049%)
and 71 fives (0.004%).
name
has


samread_Grumpy.pl
feeds
is made of
and
Chronological steps
Subroutines
bwa mem
Gives
samfile
first (partial) parse
second (full) parse
parallel
feeds parameter
count alignment
@append_name
used in tagging
core == 2
modulable cores
ambiguities_count_grumpy.pl
gives
Begin counting.
3218072
There are 0 ones (0.000%)
1602726 twos (99.757%)
3039 threes (0.189%)
and 787 fours (0.049%)
and 71 fives (0.004%).
Line parsing
Full_counter
Bad_counter
misc_counter
gives
==1 && XSYM
==end - Y +1
&& YMXS
get less stringent
used
used on Lefties only
used on both
used on both
printableTime
revCompAll
cutter
thesorter
parallelBlast
thebin
child of
orphans
use
name
refname
truncseq
has
%Lefties
%Rightoes
anchoring and croping
heavy for bining, divide by id first nt
(one per side) .fasta indexed files
creates
Blast preparation
creates
Raw results directory
one read
against
aligns
unless in
res
leads to
blast loop
main steps
Parallel Blast Step
make work
Blast+
make use of
concatenates
removes

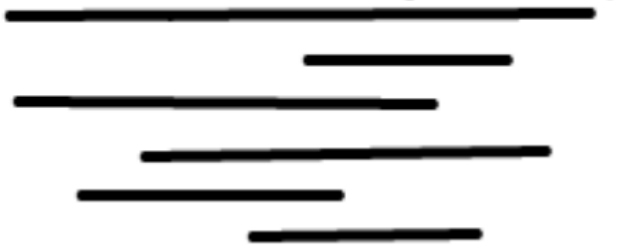

1
n
IS
Lecture en 5′
Lecture en 3′
CIGAR: xSyM
Position: 1
CIGAR: yMxS
Position: n-y
5′
antisens
Lectures tronquées de
y bases
Classement
3′
Lectures tronquées de
y bases
Classement


samread_Grumpy.pl
feeds
is made of
and
Chronological steps
Subroutines
bwa mem
Gives
samfile
first (partial) parse
second (full) parse
parallel
feeds parameter
count alignment
@append_name
used in tagging
core == 2
modulable cores
ambiguities_count_grumpy.pl
gives
Begin counting.
3218072
There are 0 ones (0.000%)
1602726 twos (99.757%)
3039 threes (0.189%)
and 787 fours (0.049%)
and 71 fives (0.004%).
Line parsing
Full_counter
Bad_counter
misc_counter
gives
==1 && XSYM
==end - Y +1
&& YMXS
get less stringent
used
used on Lefties only
used on both
used on both
printableTime
revCompAll
cutter
thesorter
parallelBlast
thebin
child of
orphans
use
name
refname
truncseq
has
%Lefties
%Rightoes
anchoring and croping
heavy for bining, divide by id first nt
(one per side) .fasta indexed files
creates
Blast preparation
creates
Raw results directory
one read
against
aligns
unless in
res
leads to
blast loop
main steps
Parallel Blast Step
make work
Blast+
make use of
concatenates
removes

Pourquoi Blast?

Cluster

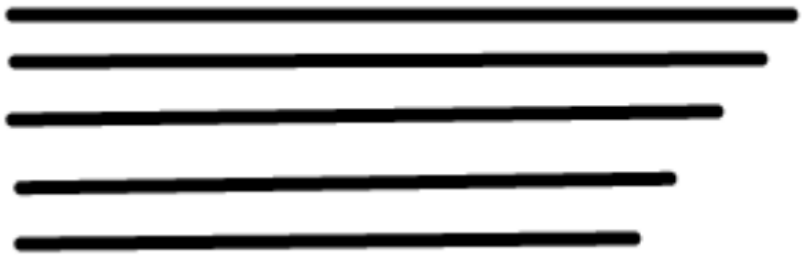
Longest read = rep



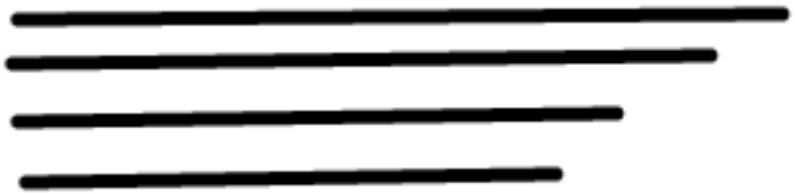
Ce qui m'intéresse le plus: les reads ancrés à gauche ou à droite

Cd-hit ne tient pas compte de cette réalité (l'ancrage)

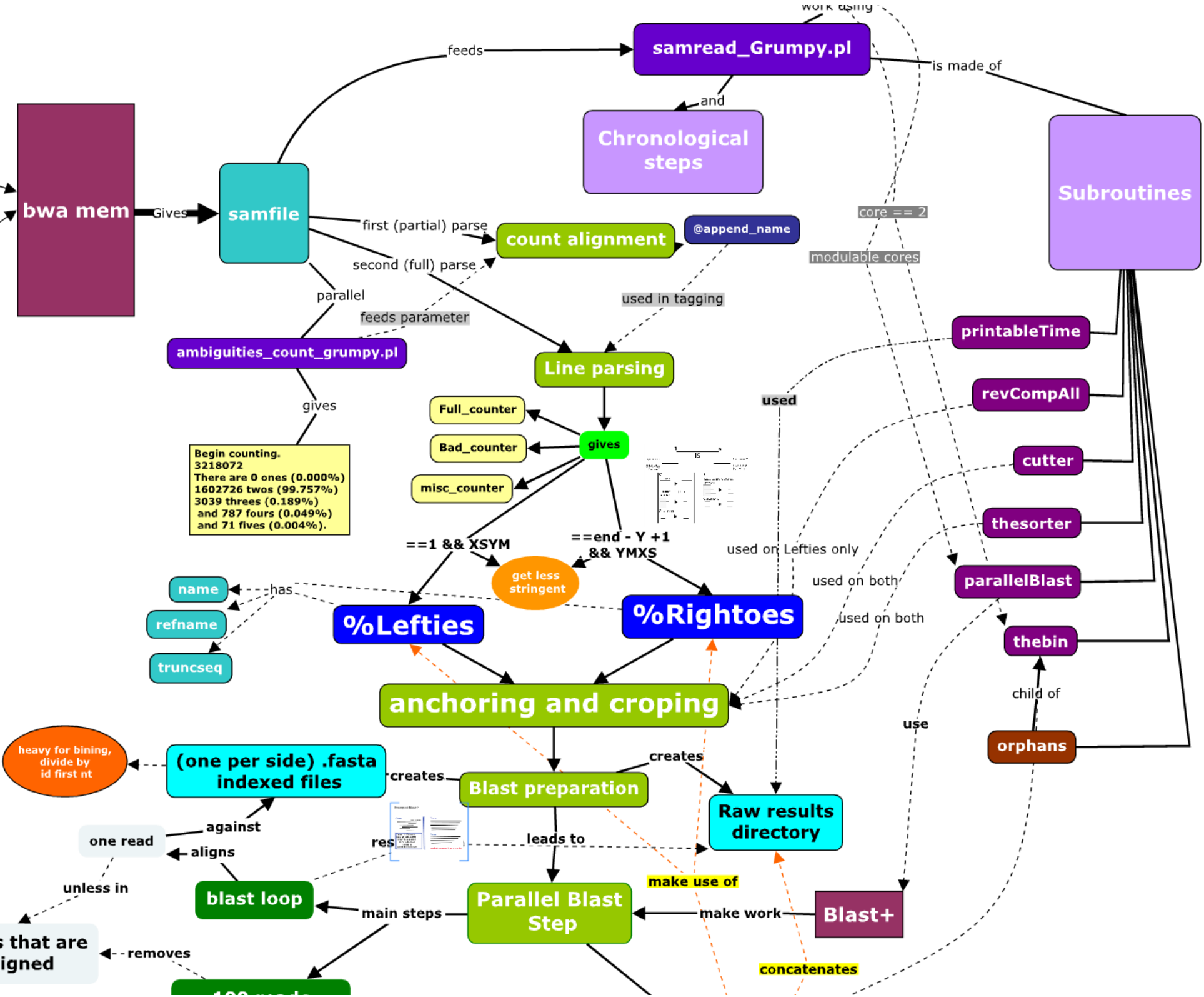
Cluster

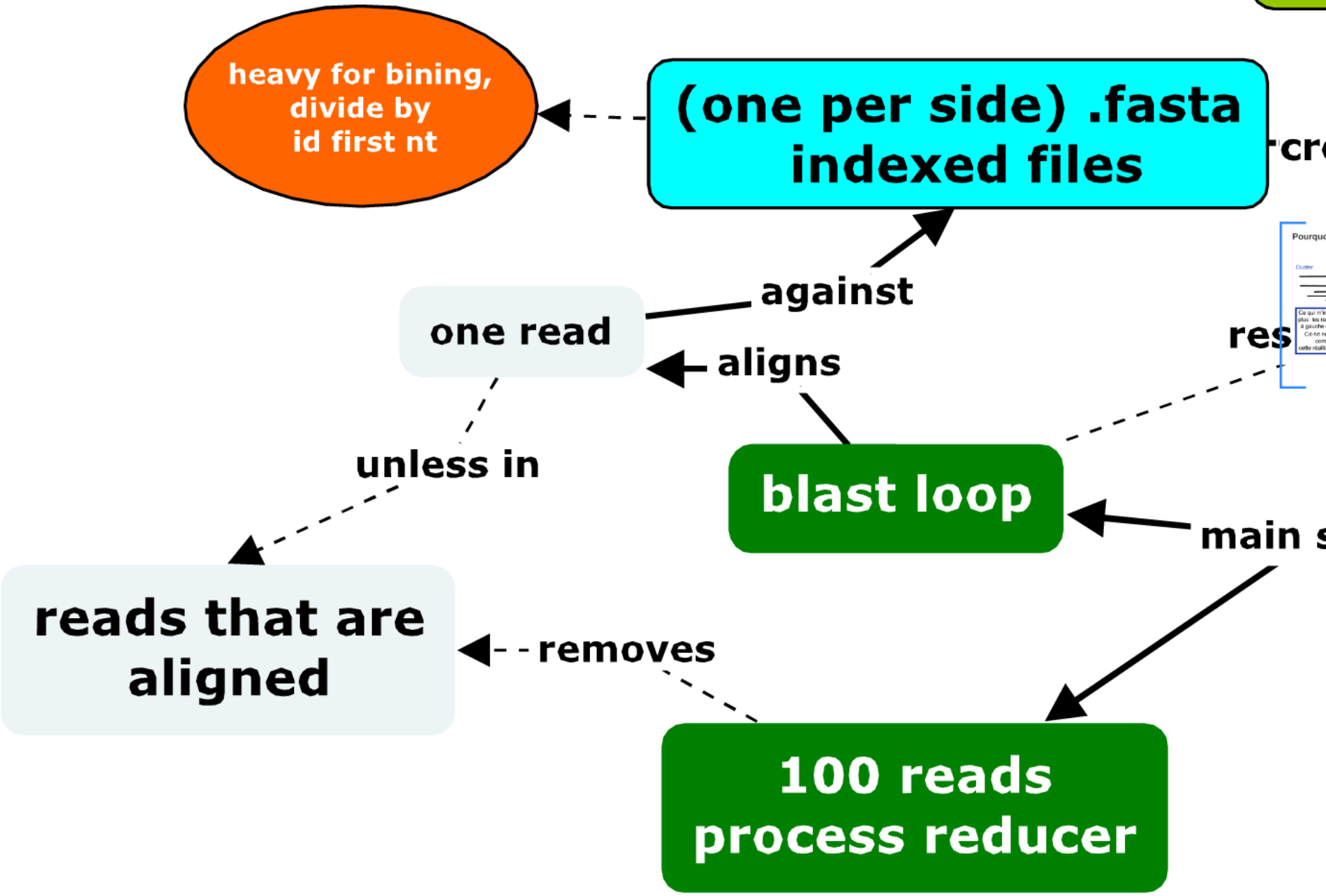


Cluster

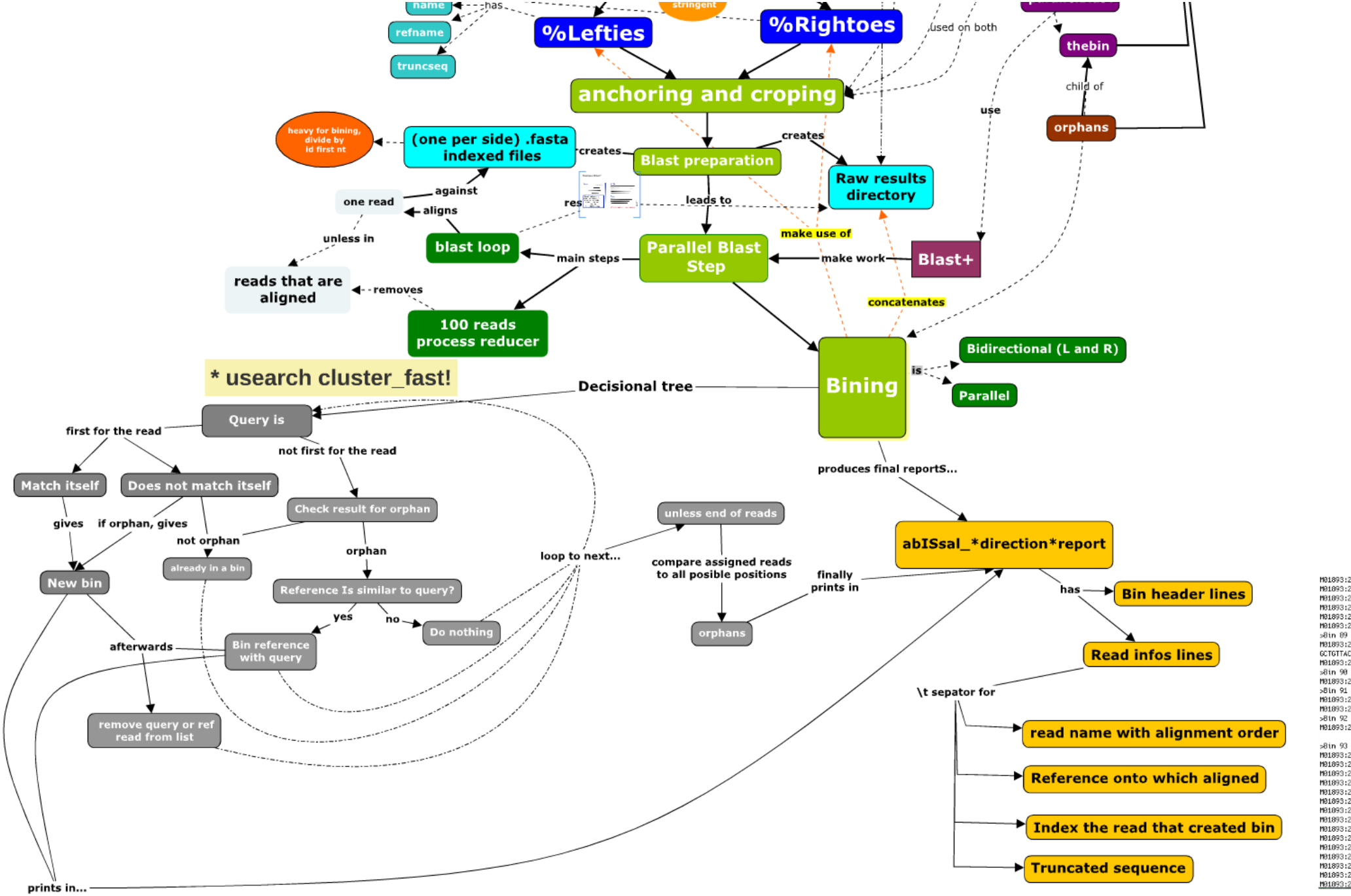


(À droite seulement, par exemple)


samread_Grumpy.pl
feeds
is made of
and
Chronological steps
Subroutines
bwa mem
Gives
samfile
first (partial) parse
second (full) parse
parallel
feeds parameter
count alignment
@append_name
used in tagging
core == 2
modulable cores
ambiguities_count_grumpy.pl
gives
Begin counting.
3218072
There are 0 ones (0.000%)
1602726 twos (99.757%)
3039 threes (0.189%)
and 787 fours (0.049%)
and 71 fives (0.004%).
Line parsing
Full_counter
Bad_counter
misc_counter
gives
==1 && XSYM
==end - Y +1
&& YMXS
get less stringent
used
used on Lefties only
used on both
used on both
printableTime
revCompAll
cutter
thesorter
parallelBlast
thebin
child of
orphans
use
name
refname
truncseq
has
%Lefties
%Rightoes
anchoring and croping
heavy for bining, divide by id first nt
(one per side) .fasta indexed files
creates
Blast preparation
creates
Raw results directory
one read
against
aligns
unless in
res
leads to
blast loop
main steps
Parallel Blast Step
make work
Blast+
make use of
concatenates
removes



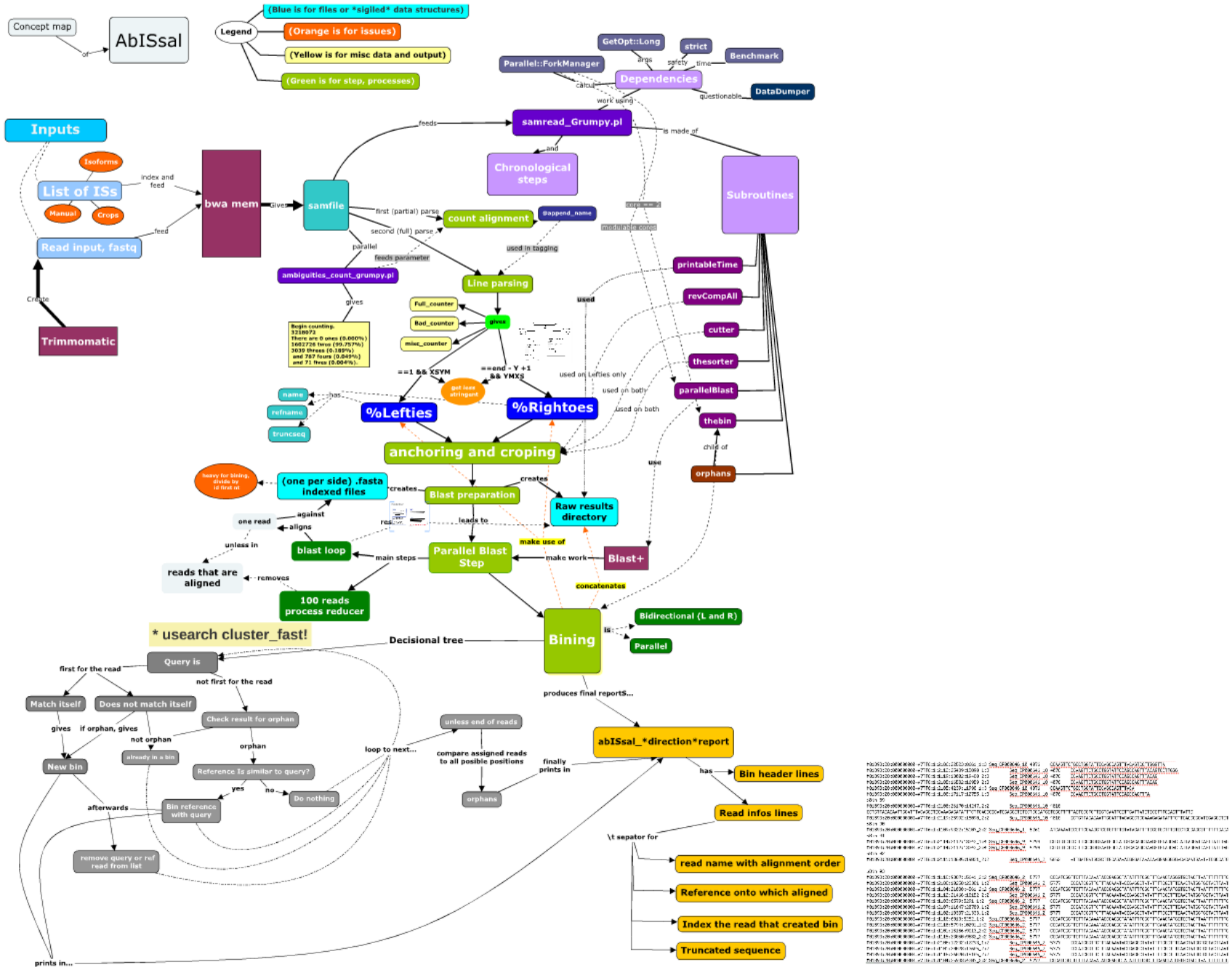
* usearch cluster_fast!


name
has
refname
truncseq
stringent
%Lefties
%Rightoes
used on both
thebin
child of
orphans
use
anchoring and croping
heavy for bining, divide by id first nt
(one per side) .fasta indexed files
creates
Blast preparation
creates
Raw results directory
one read
against
aligns
res
leads to
unless in
blast loop
make use of
main steps
Parallel Blast Step
make work
Blast+
reads that are aligned
removes
concatenates
100 reads process reducer
Bining
is
Bidirectional (L and R)
Parallel
* usearch cluster_fast!
Decisional tree
Query is
first for the read
not first for the read
Match itself
Does not match itself
Check result for orphan
produces final reportS...
gives
if orphan, gives
not orphan
orphan
unless end of reads
abISsal_*direction*report
New bin
already in a bin
loop to next...
compare assigned reads to all posible positions
finally prints in
has
Bin header lines
Reference Is similar to query?
yes
no
Do nothing
orphans
afterwards
Bin reference with query
Read infos lines
\t sepator for
remove query or ref read from list
read name with alignment order
Reference onto which aligned
Index the read that created bin
Truncated sequence
prints in...

```
M01893:20:000000000-A7TT6:1:2106:22523:8661_1:3 Seq_CP000646_10 4076    CCAAGTTCTGCCTGGTATTCCAGCCAGTTTACAGTCCTTGGGTTA
M01893:20:000000000-A7TT6:1:2112:25489:15890_1:3        Seq_CP000646_10 4076    CCAAGTTCTGCCTGGTATTCCAGCCAGTTTACAGTCCTTGGG
M01893:20:000000000-A7TT6:1:1119:13002:19480_2:3        Seq_CP000646_10 4076    CCAAGTTCTGCCTGGTATTCCAGCCAGTTTACAG
M01893:20:000000000-A7TT6:1:2105:16582:16959_2:3        Seq_CP000646_10 4076    CCAAGTTCTGCCTGGTATTCCAGCCAGTTTACAG
M01893:20:000000000-A7TT6:1:2105:4239:11998_1:3 Seq_CP000646_10 4076    CCAAGTTCTGCCTGGTATTCCAGCCAGTTTACA
M01893:20:000000000-A7TT6:1:1108:17117:12755_1:3        Seq_CP000646_10 4076    CCAAGTTCTGCCTGGTATTCCAGCCAGTTTA
>Bin 89
M01893:20:000000000-A7TT6:1:2108:26170:14247_2:2        Seq_CP000646_10 4810
GCTGTTACACAATTGGATTTACAGGCTCCAAAGAGATATTTCTTGACCCGCATGGAGGCTCTGCTCCCATGGTCGGTTTTTACTCCCTCTTGGTCAATTCCTTGATTATCTCCCTTTCCAGTTTATTC
M01893:20:000000000-A7TT6:1:2119:25992:15090_2:2        Seq_CP000646_10 4810    GCTGTTACACAATTGGATTTACAGGCTCCAAAGAGATATTTCTTGACCCGCATGGAGGCTCT
>Bin 90
M01893:20:000000000-A7TT6:1:1109:9322:15107_2:2 Seq_CP000646_1  5261    ATCAAAATCCCTTCCAAGGTCCTGTTTTGTATAGATTTTCGCTGCTTTGTCCTGCGAGCGTTTTTTGACA
>Bin 91
M01893:20:000000000-A7TT6:1:2114:3117:13343_1:4 Seq_CP000646_9  5294    CCGTGTCTCCTTTCCCGGGGAATGGCTATGGCACAGCGAAGGTGTAGCACTATTAGGATCAGTTATTTAG
M01893:20:000000000-A7TT6:1:2114:3117:13343_3:4 Seq_CP000646_9  5294    CCGTGTCTCCTTTCCCGGGGAATGGCTATGGCACAGCGAAGGTGTAGCACTATTAGGATCAGTTATTTAG
>Bin 92
M01893:20:000000000-A7TT6:1:2111:13605:19031_2:2        Seq_CP000646_2  5652    ATTGATGATGCGCTTGCAGAAACGGGATAAACAAGGAGGGGGACACAATGAATATCGCCATG

>Bin 93
M01893:20:000000000-A7TT6:1:1118:9387:15641_2:2 Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTAGGGTGCTACTTAATTTTTTTTC
M01893:20:000000000-A7TT6:1:2108:10258:23381_1:2        Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAAT
M01893:20:000000000-A7TT6:1:1104:21030:4561_2:2 Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAATTTTTTTTC
M01893:20:000000000-A7TT6:1:1112:21466:18152_2:2        Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAAT
M01893:20:000000000-A7TT6:1:1103:6579:5291_1:2  Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAATTTTTTTTC
M01893:20:000000000-A7TT6:1:2107:11647:20789_1:2        Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAAT
M01893:20:000000000-A7TT6:1:1102:10337:21333_1:2        Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAAT
M01893:20:000000000-A7TT6:1:1118:6010:5252_1:2  Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAATTTTTTTTC
M01893:20:000000000-A7TT6:1:2118:5794:10291_1:2 Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAATTTTTTTTC
M01893:20:000000000-A7TT6:1:1101:15156:9013_2:2 Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAATTTTTTTTC
M01893:20:000000000-A7TT6:1:1115:23050:9682_2:2 Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAATTTTTTTTC
M01893:20:000000000-A7TT6:1:2106:12232:12293_1:2        Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAAT
M01893:20:000000000-A7TT6:1:1101:20078:13605_2:2        Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAAT
M01893:20:000000000-A7TT6:1:1116:26070:15105_2:2        Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAAT
M01893:20:000000000-A7TT6:1:1104:25933:4207_2:2 Seq_CP000646_2  5777    CCCATCGGTTCTTTACAAATACCGAGGCTATATTTTCGCTTTCAACTATGGTGCTACTTAATTTTTTTTC
```

Les étapes d'AbISsal



Et après...

Détection d'insertions rares
Comparaisons

Analyses sur lecture
Comparaisons
"Transformable"

Transpositions
Connaissances
????

Orienter AbISsal vers la comparaison de jeux de données

- à la sortie
- à l'entrée

Incorporer des méthodes de clustering plus efficaces

Autres suggestions?

CRSNG
NSERC

Ressources Aquatiques Québec RAQ

Club Bioinfo IBIS