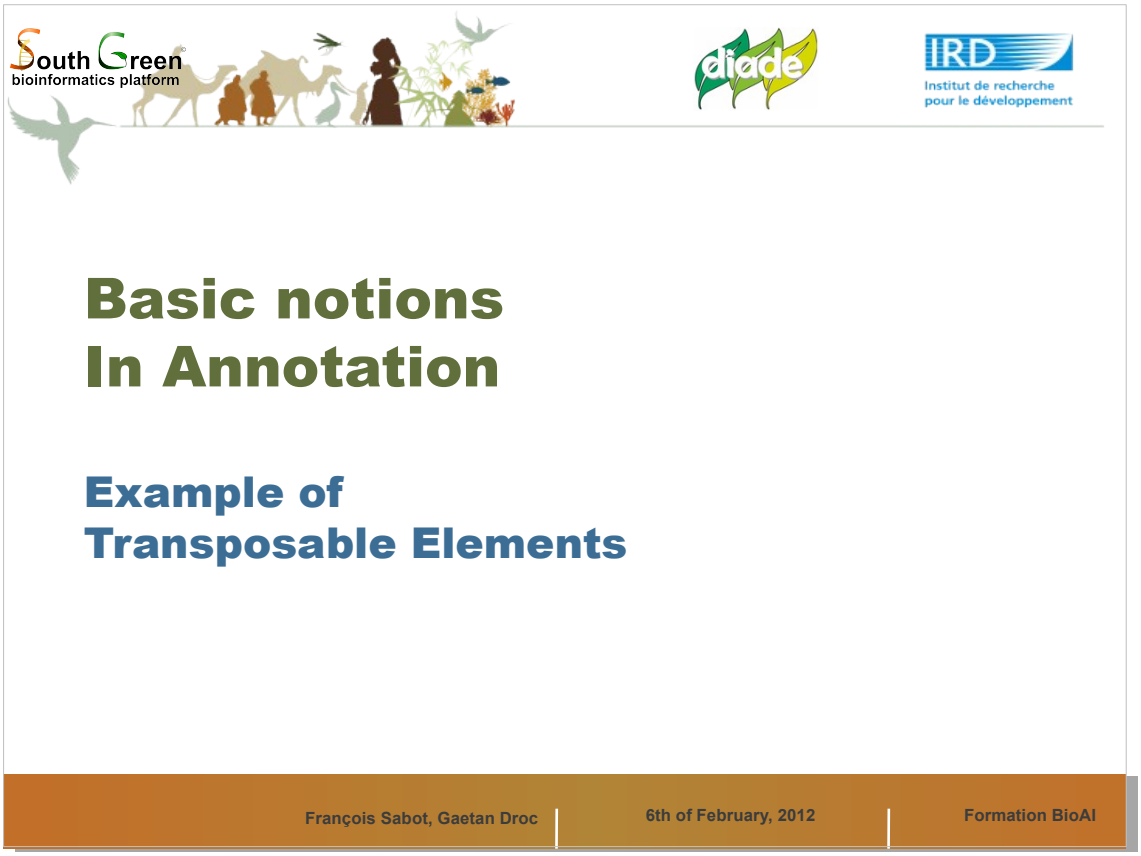
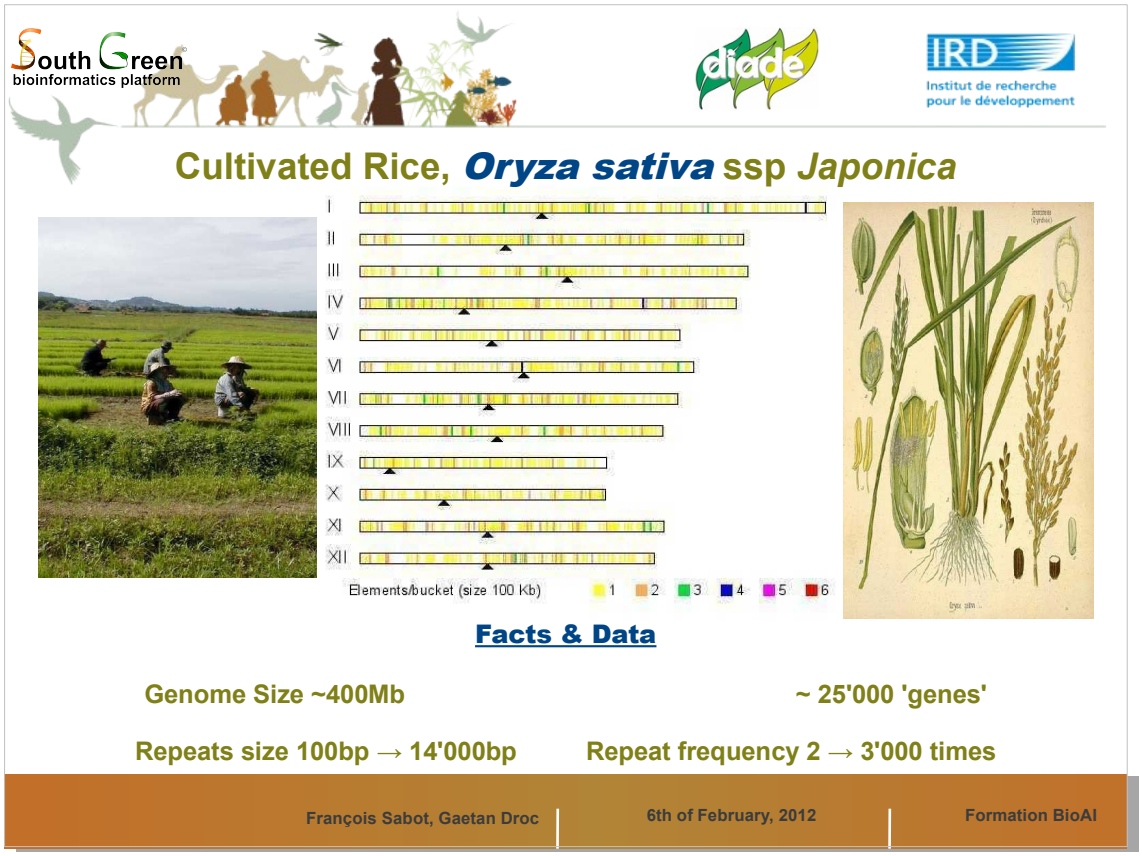


Le but du TP est de vous donner une idée rapide des méthodologies d'annotation des éléments transposables dans une grande région régomique, grâce à l'utilisation d'outils visuel, Gepard et Artemis.

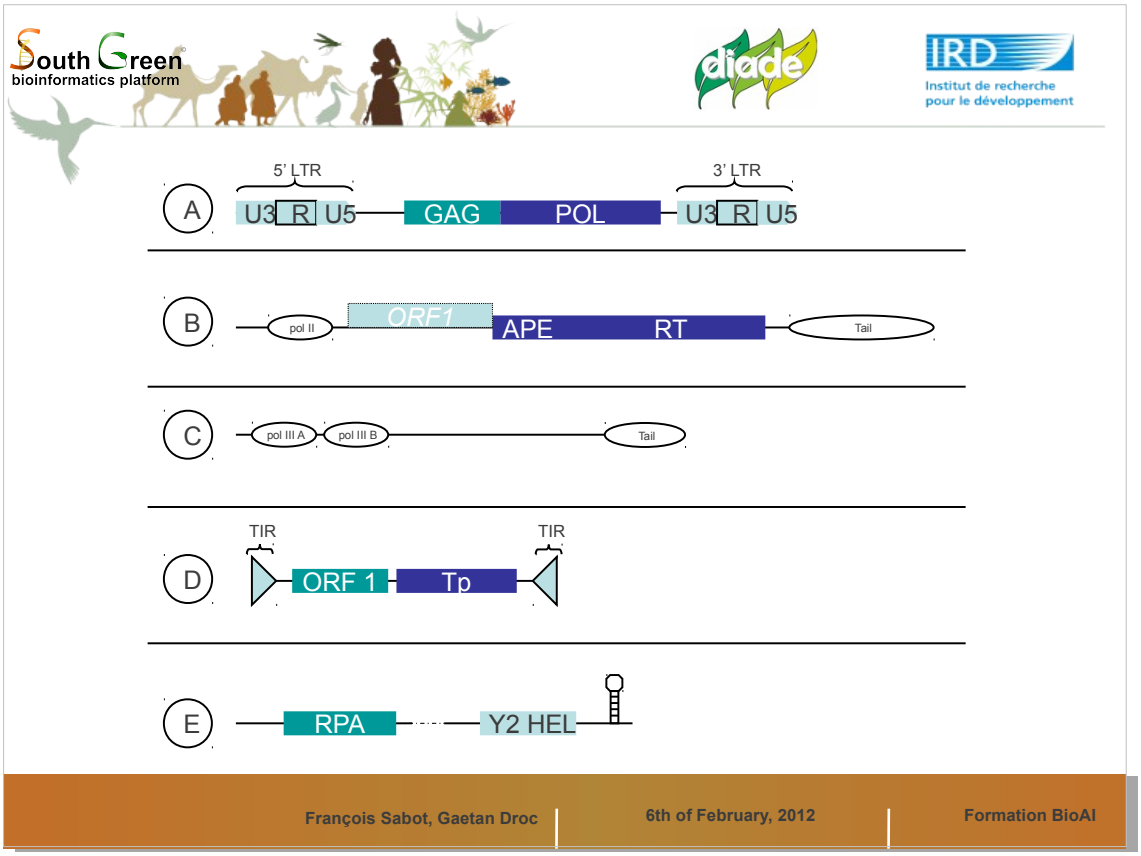
De plus, vous serez amené à manipuler Artemis, ce qui vous aidera fortement pour le prochain TD



Nous allons travailler sur le riz Asiatique, qui a un petit génome, et peu d'éléments transposables, comparativement à des plantes comme l'orge ou le blé.

Paradoxalement, il est parfois plus dur d'annoter un génome avec peu d'éléments.

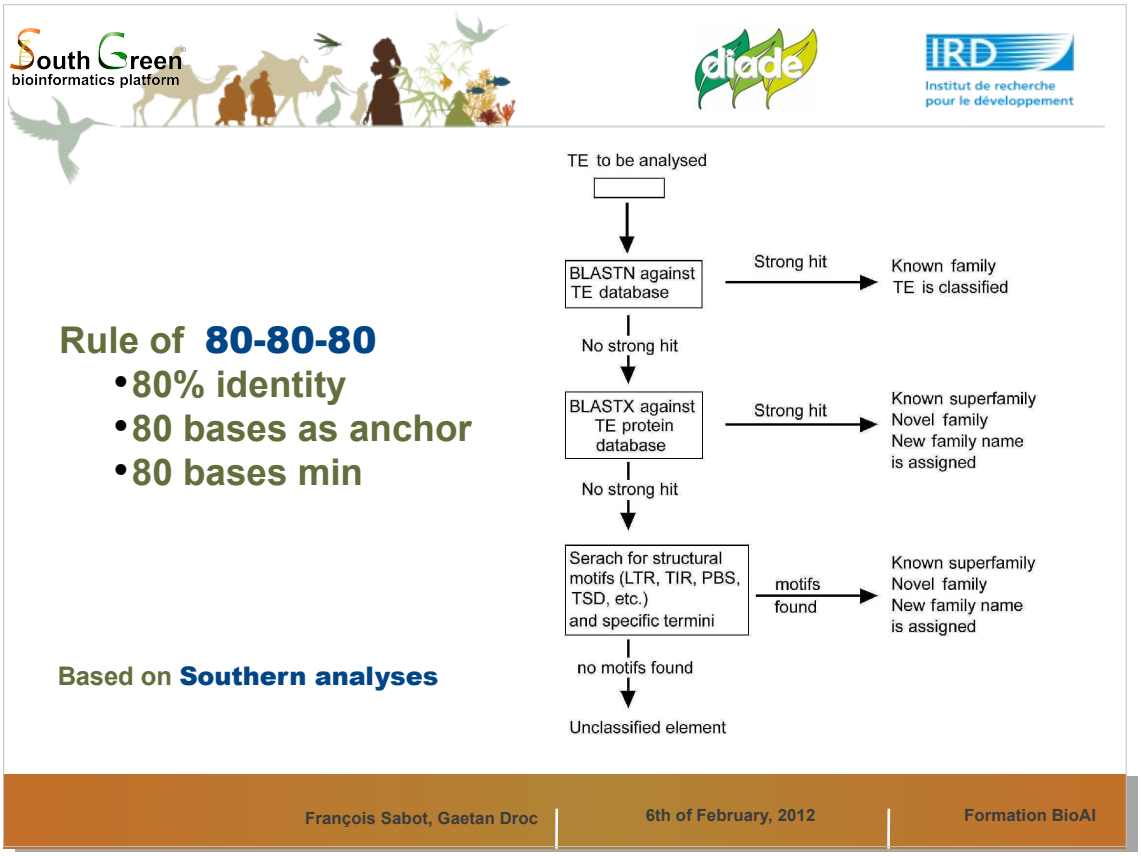
La classification des éléments transpables, ou TE, est un sujet complexe. Un bonne idée de leur diversité est présenté dans le papier cité ci-dessus



Une classification simplifiée des éléments présents dans les plantes.

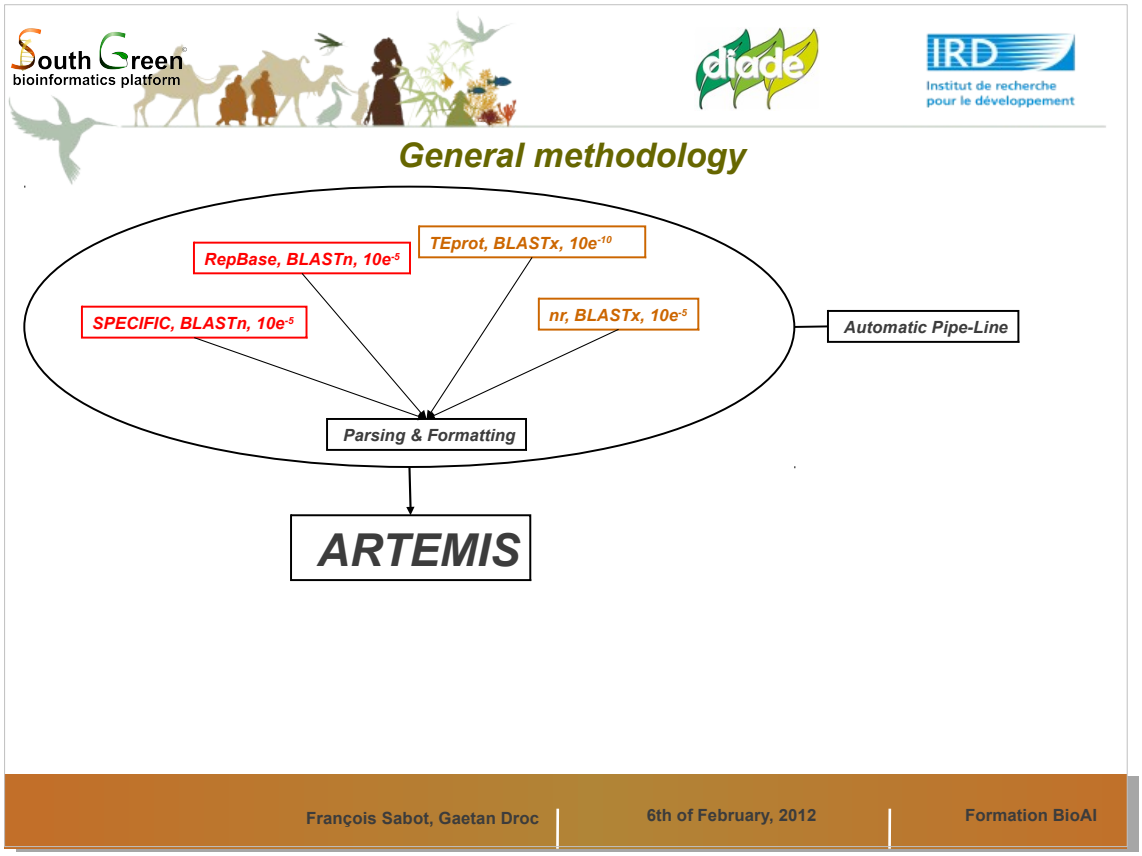
LTR = Long Terminal Repeats
TIR = Terminal Inverted Repeats

Les éléments sont en général bordés par des TSD, target site duplication, obtenus via une coupure asymetrique du double brin d'ADN au site d'insertion.



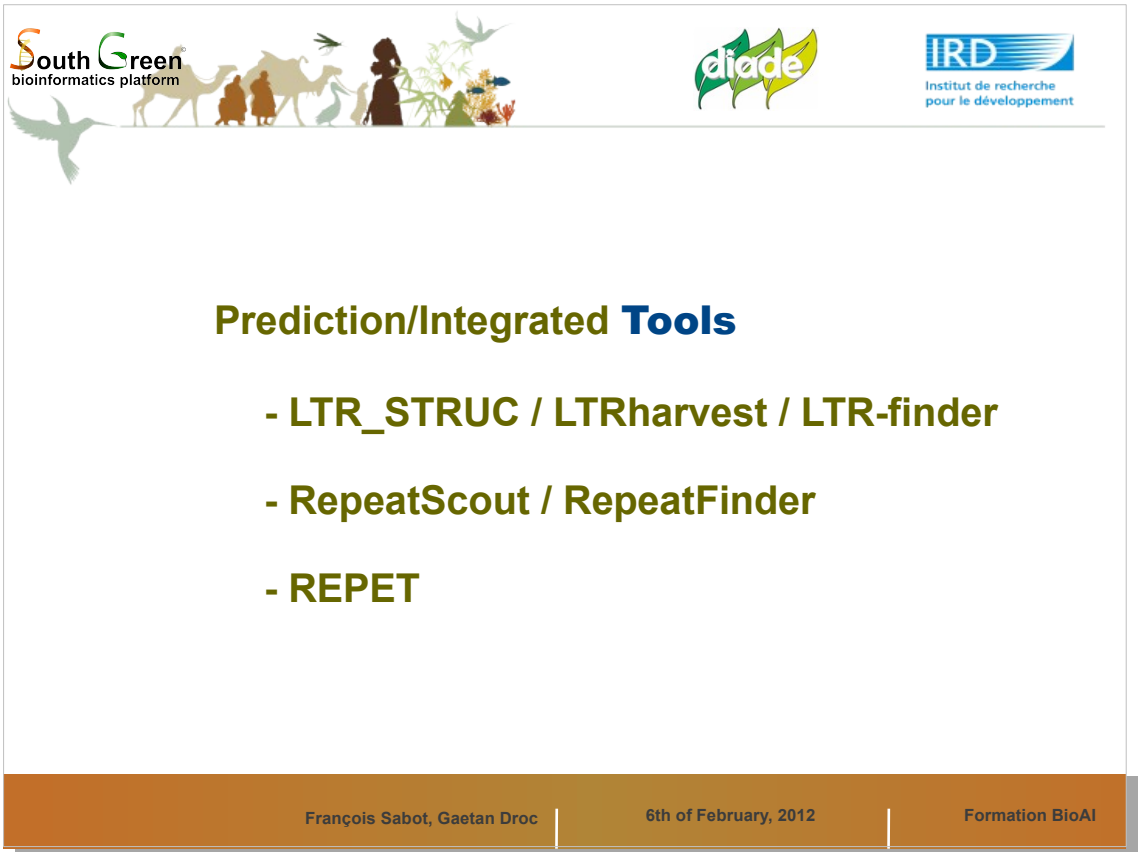
Ici un protocole basique d'annotation de TEs. La regles des 80 – 80 – 80 marche a peu pres bien en automatique, et évite de créer des faux positifs.

Par contre elle ne permet pas de récupérer toutes les répétitions...

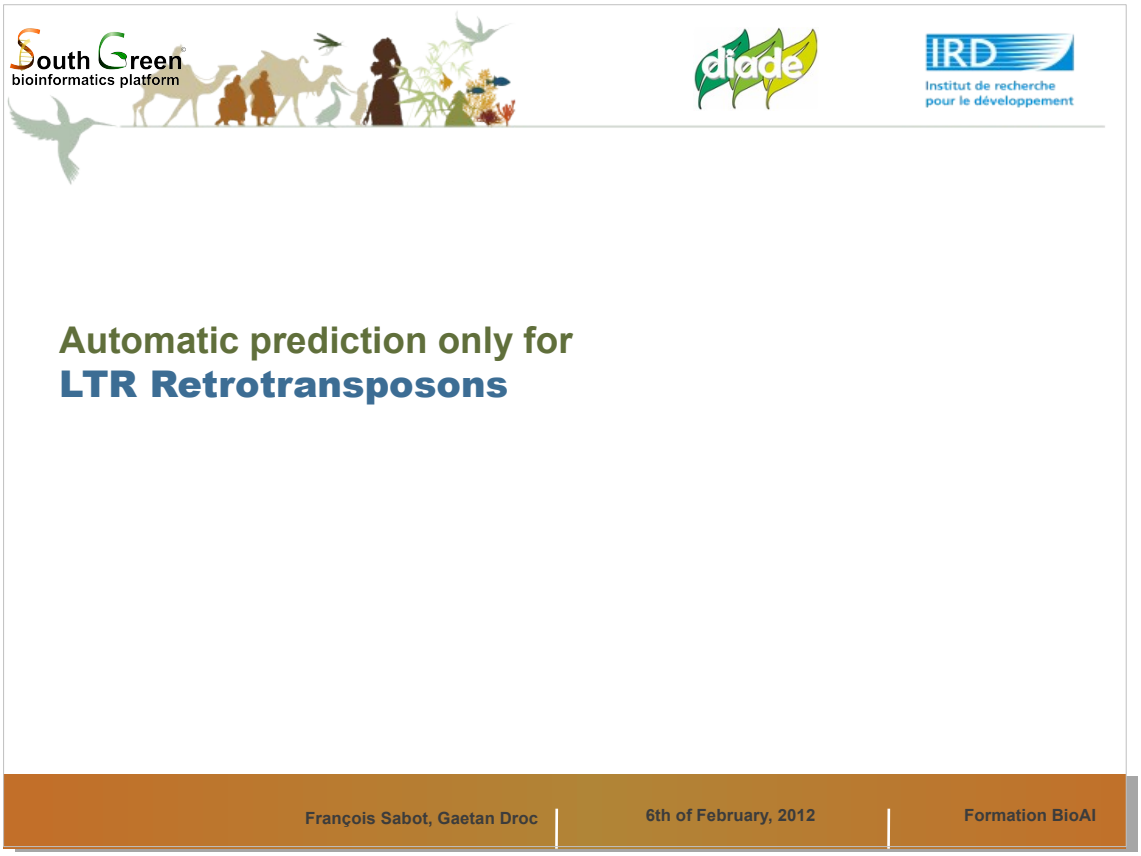


Voici un type de pipeline d'annotation automatique, qui fournit un set de données à utiliser ensuite dans Artemis pour finaliser le travail

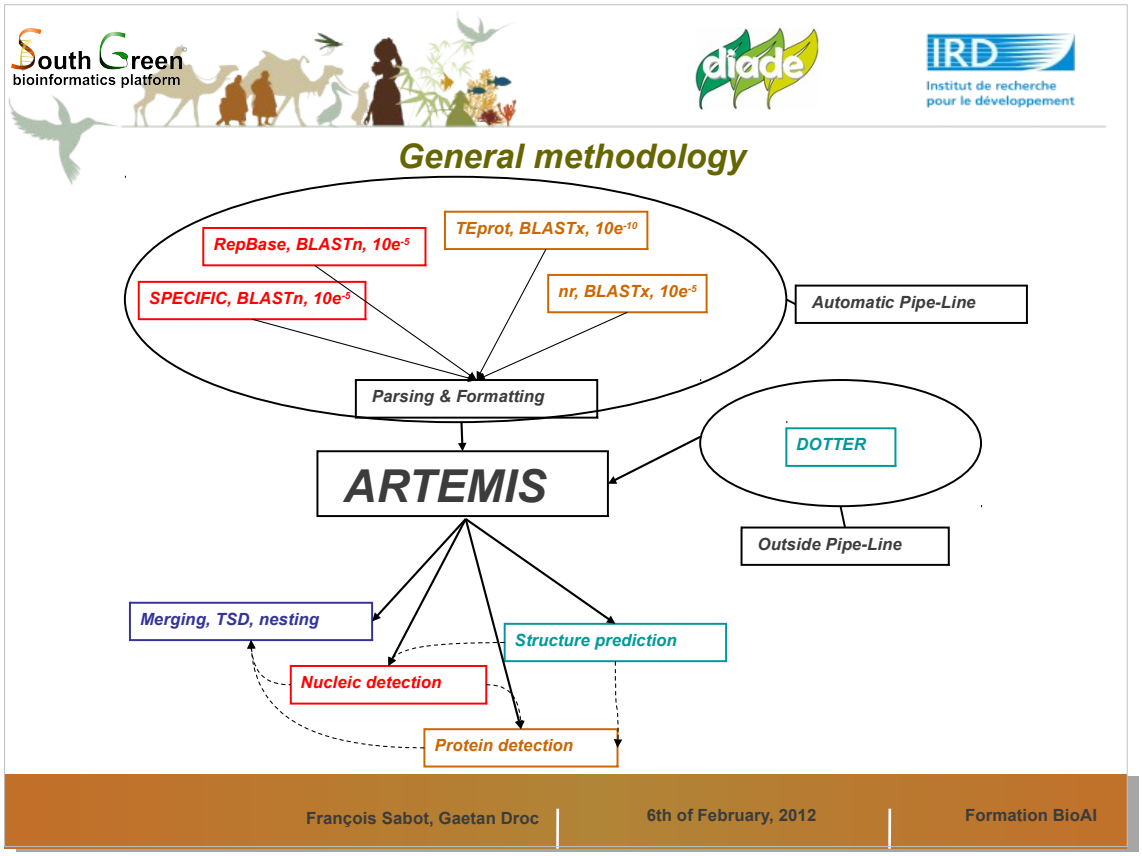
Différents types d'outils existent. Ceux de comparaisons réclament une base de données de séquences pré-existantes



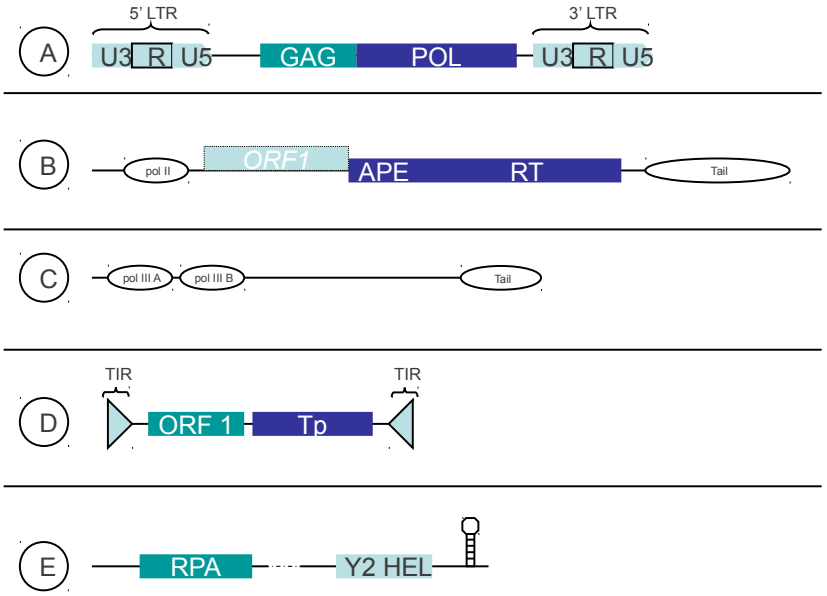
Certains outils permettent de prédire *ab initio*, donc sans connaissance préalable de séquences, simplement via des structures spécifiques des éléments transposables.

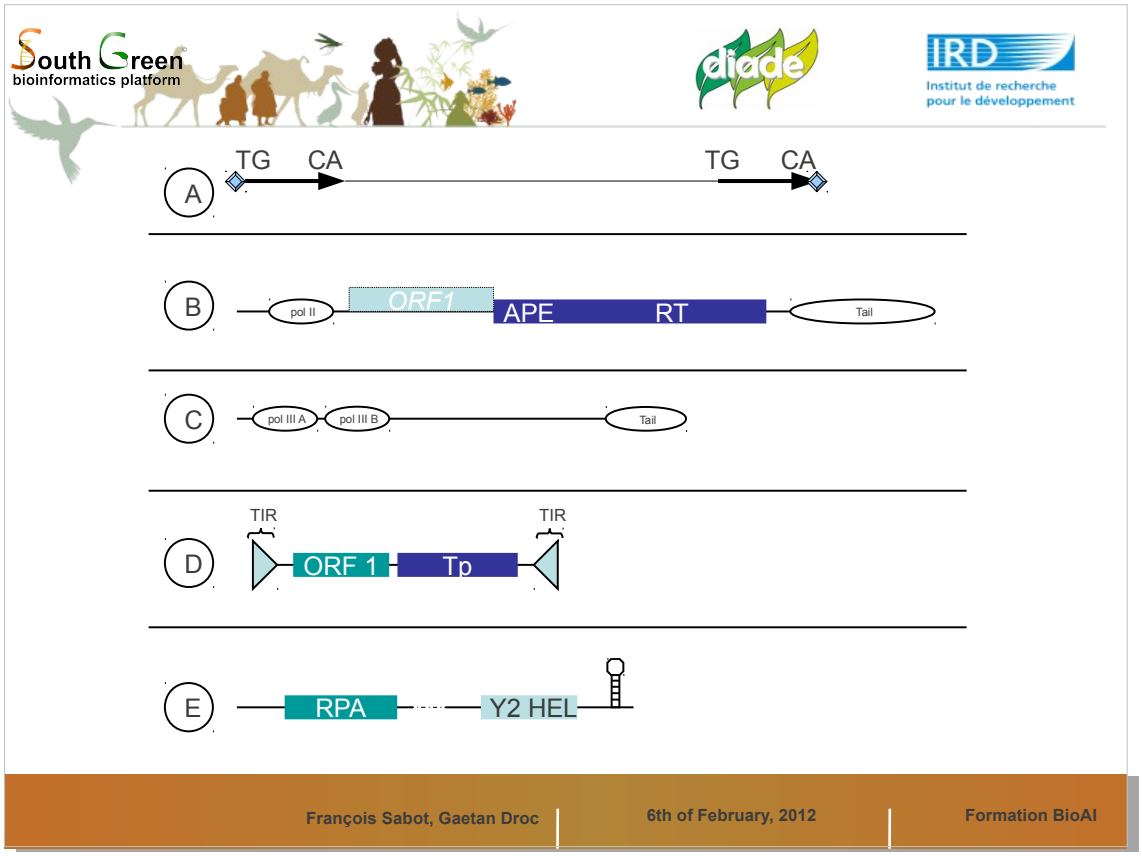


Malheureusement, la prédiction *ab initio* ne marche que pour les LTR rétrotransposons.



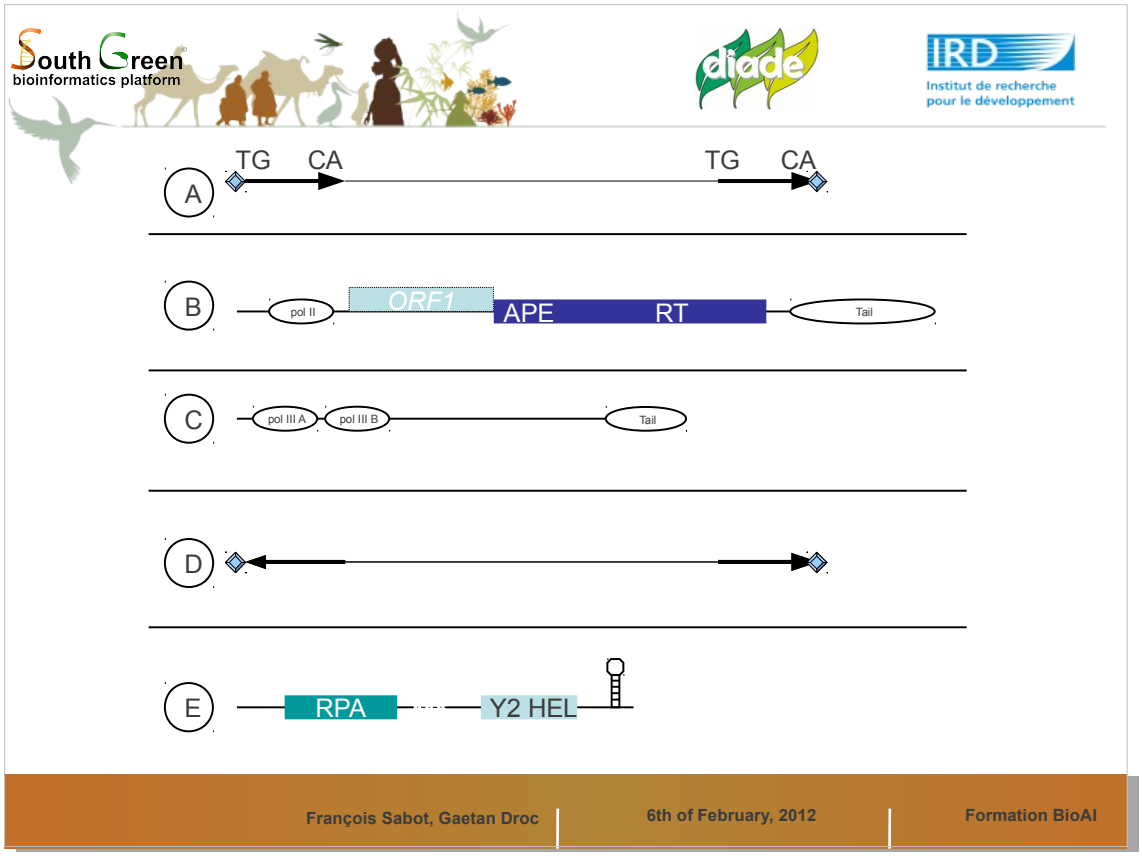
En plus du pipeline automatique, nous utiliserons des outils de type Dotter, puis grâce à Artemis nous allons identifier et annoter des structures d'éléments transposables.


A
5' LTR
U3
R
U5
GAG
POL
3' LTR
U3
R
U5
B
pol II
ORF1
APE
RT
Tail
C
pol III A
pol III B
Tail
D
TIR
ORF 1
Tp
TIR
E
RPA
Y2 HEL

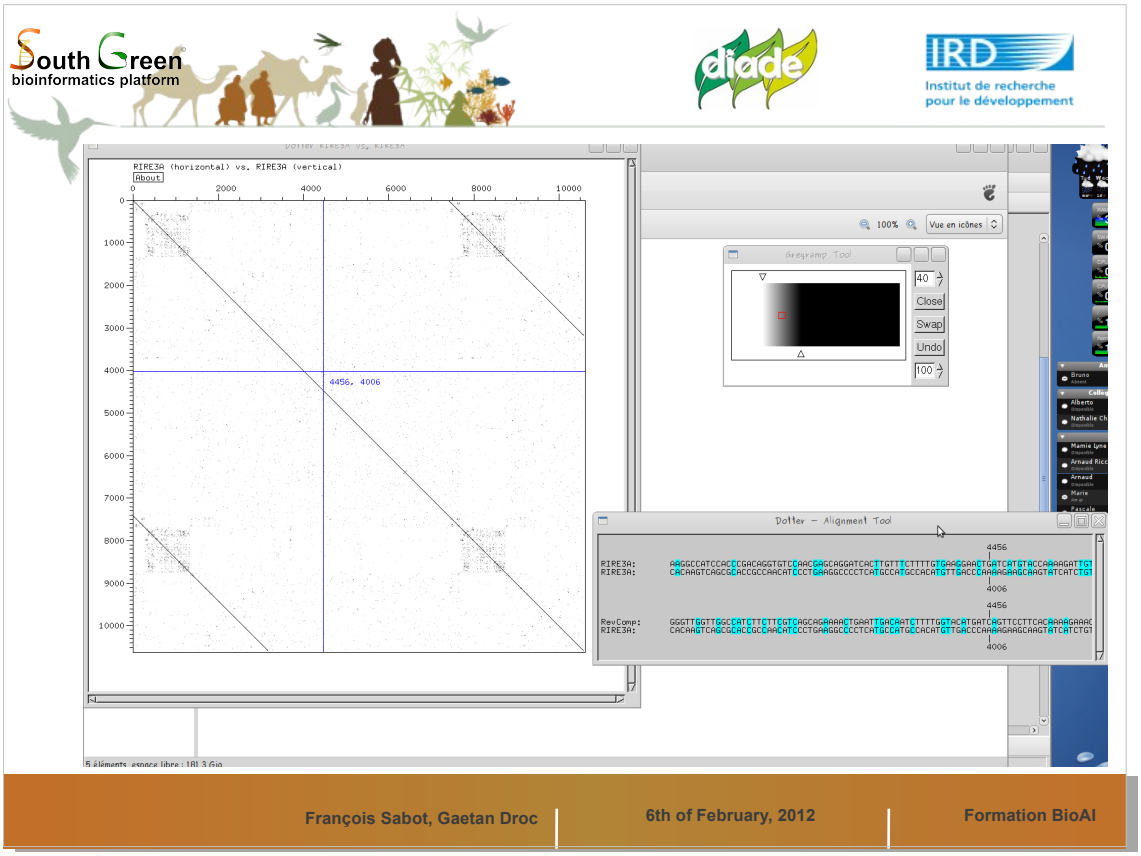


Les LTR commencent par TG et finissent par CA dans 99% des cas. Ils sont identiques entre eux au moment de l'insertions.

Les deux TSD sont de 5 bases

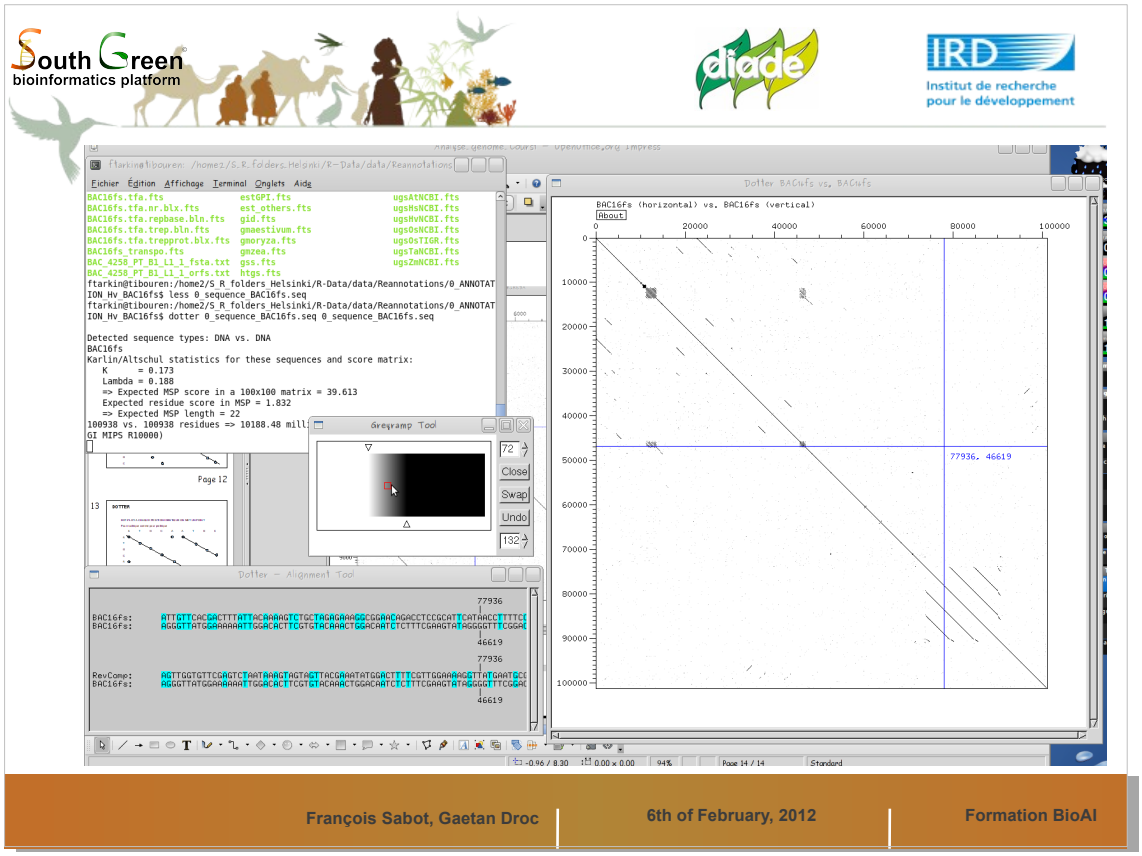


Les TIR sont très similaires, voire identiques, mais en orientation inverses. L'élement est aussi flanqué de TSD, de 2 à 19 bases

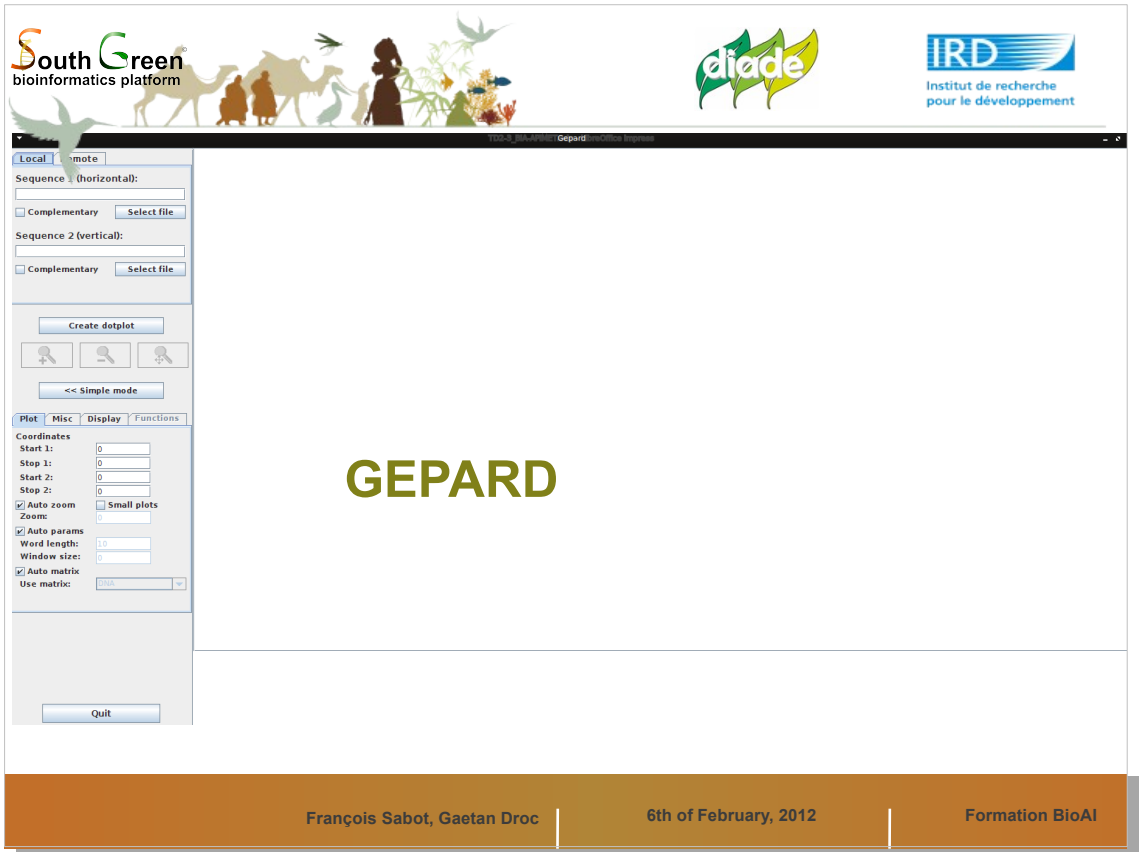


Pour déterminer les structures en répétitions directes ou indirectes, nous allons utiliser des outils de Dot-Plot.

Ici une image de Dotter, le plus souple. Mais il n'est plus disponible sur le web...

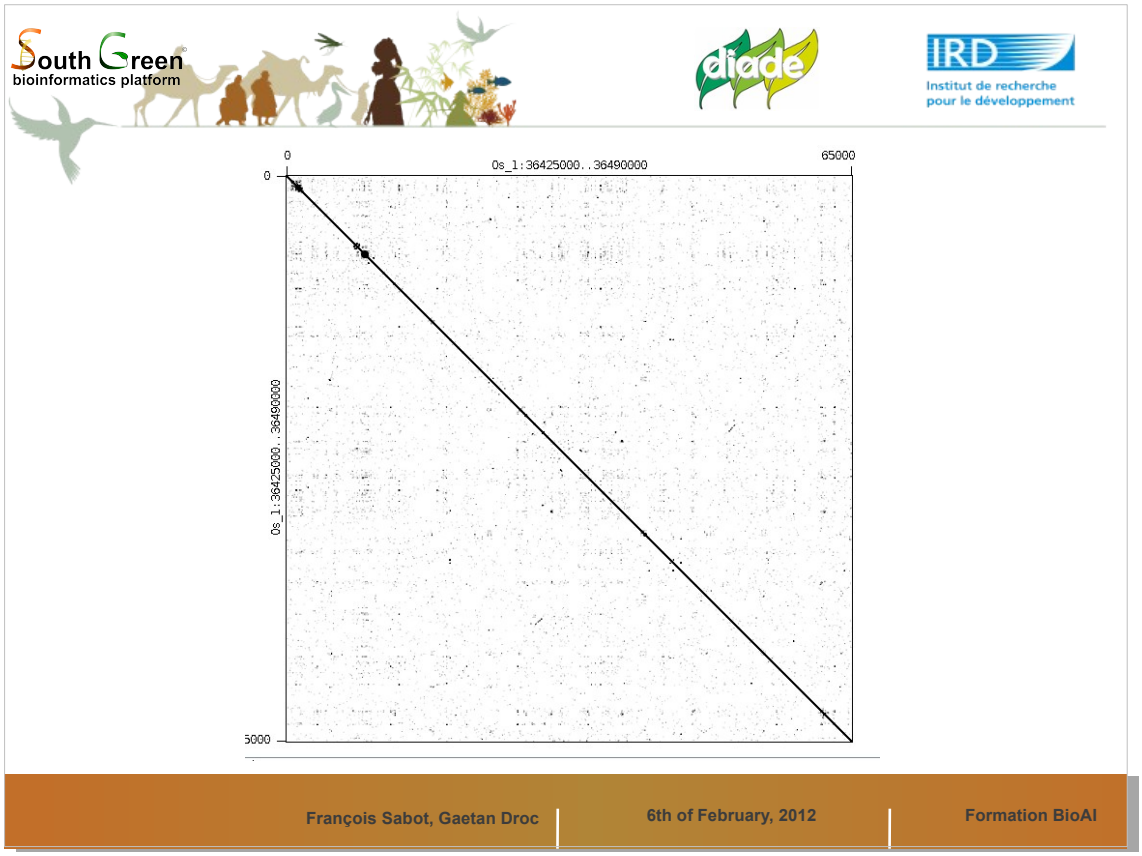


A chaque base identique le logiciel dessine un point. Donc, si une suite de bases est répétée plus loin, nous verrons deux lignes parallèles (direct) ou face-à-face (indirect, forme une croix avec la ligne centrale)

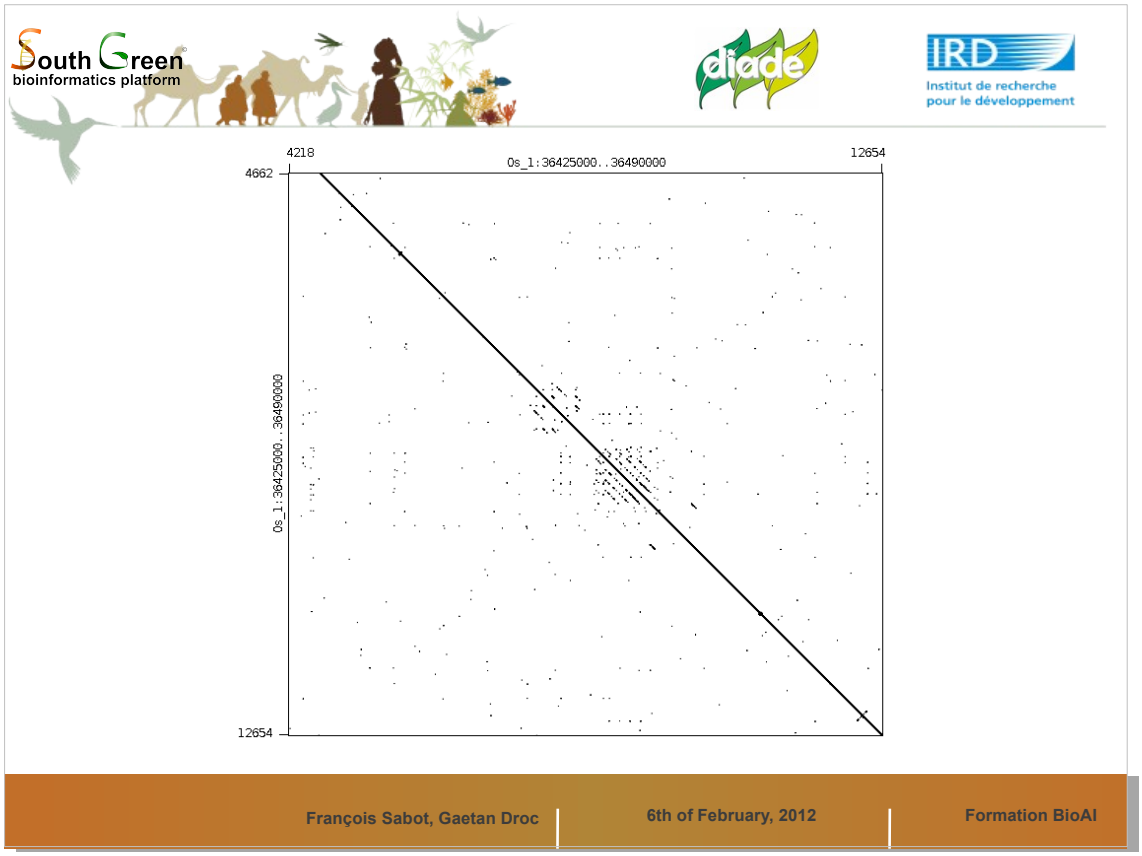


Nous allons utiliser Gepard, logiciel de dot-plot écrit en Java. Il est multiplateforme et plutot performant.

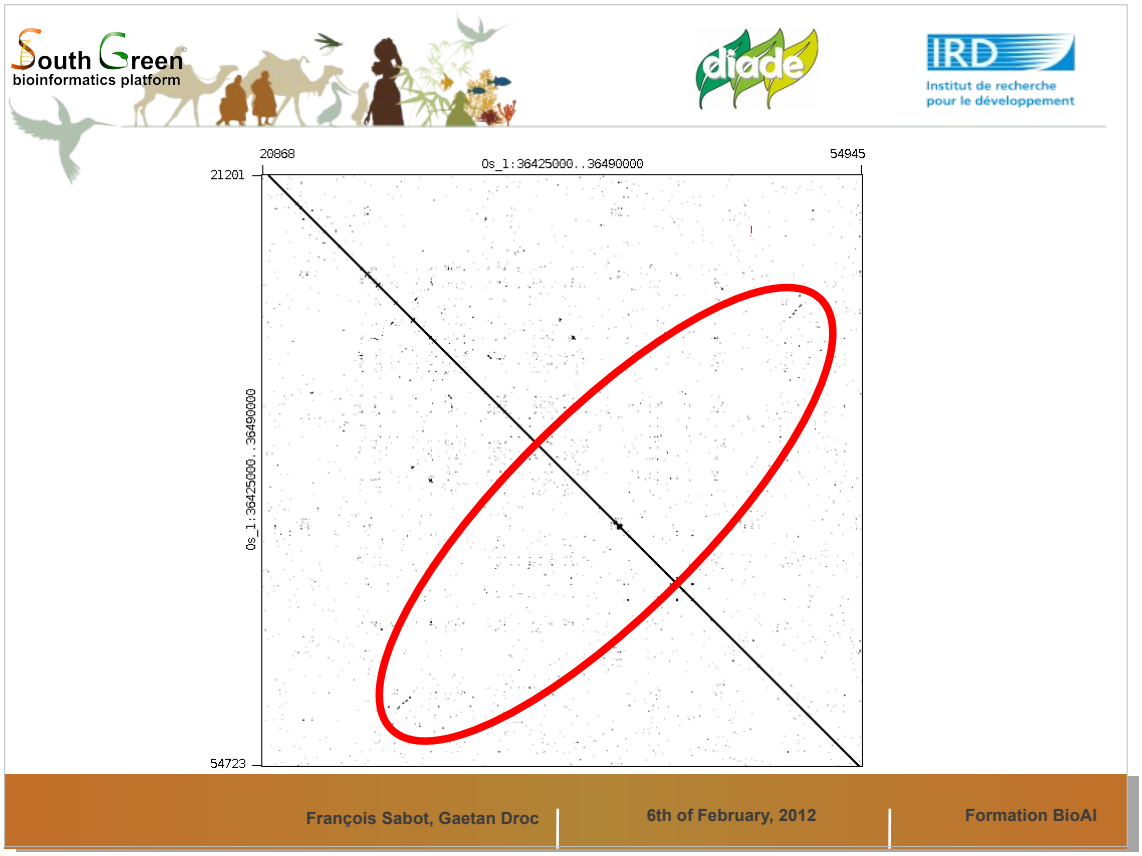
La séquence 1 et la séquence 2 sont les mêmes, il s'agit de la séquence de la Data library TE annotation récupérée ci avant. La séquence doit etre de préférence en fasta.



Voici l'image de Dot d'origine. La ligne centrale représente la séquence sur elle meme.

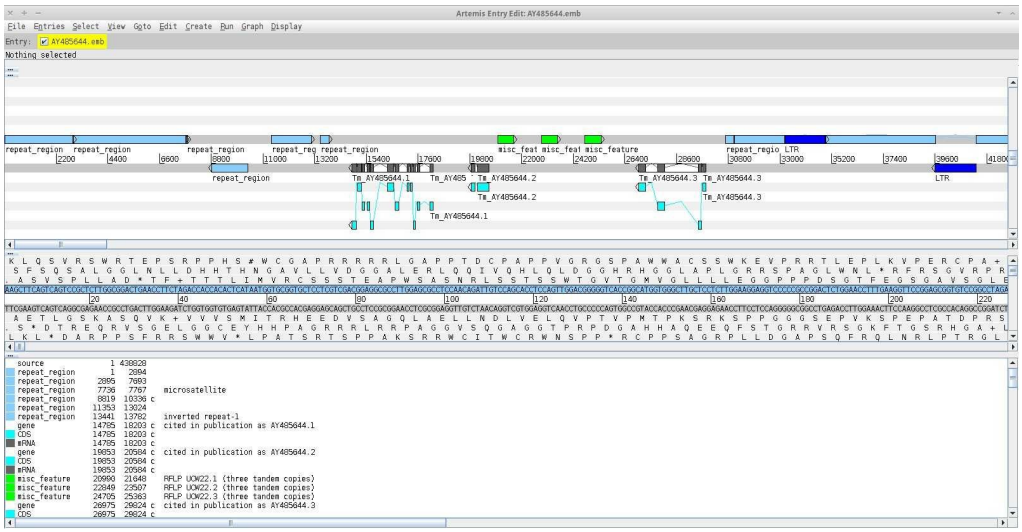


Ici une image de type microsatellite long, de type par exemple (TAGCGTA)n.
Cela peut etre aussi plusieurs répétitions courtes.

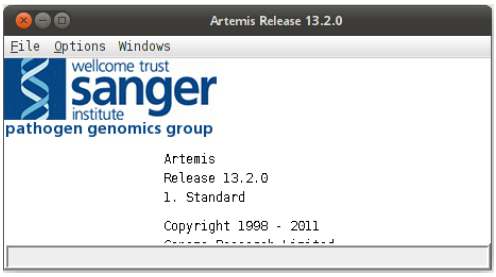


Notez ici les deux lignes en croix par rapport au centre, souvent un signe de TIR TE.

ARTEMIS: tool for viewing and annotating large sequences

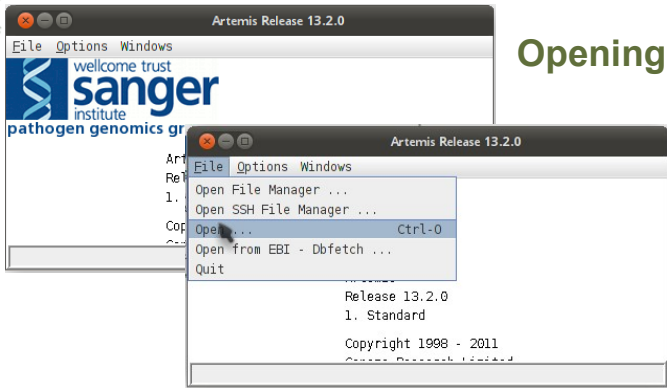


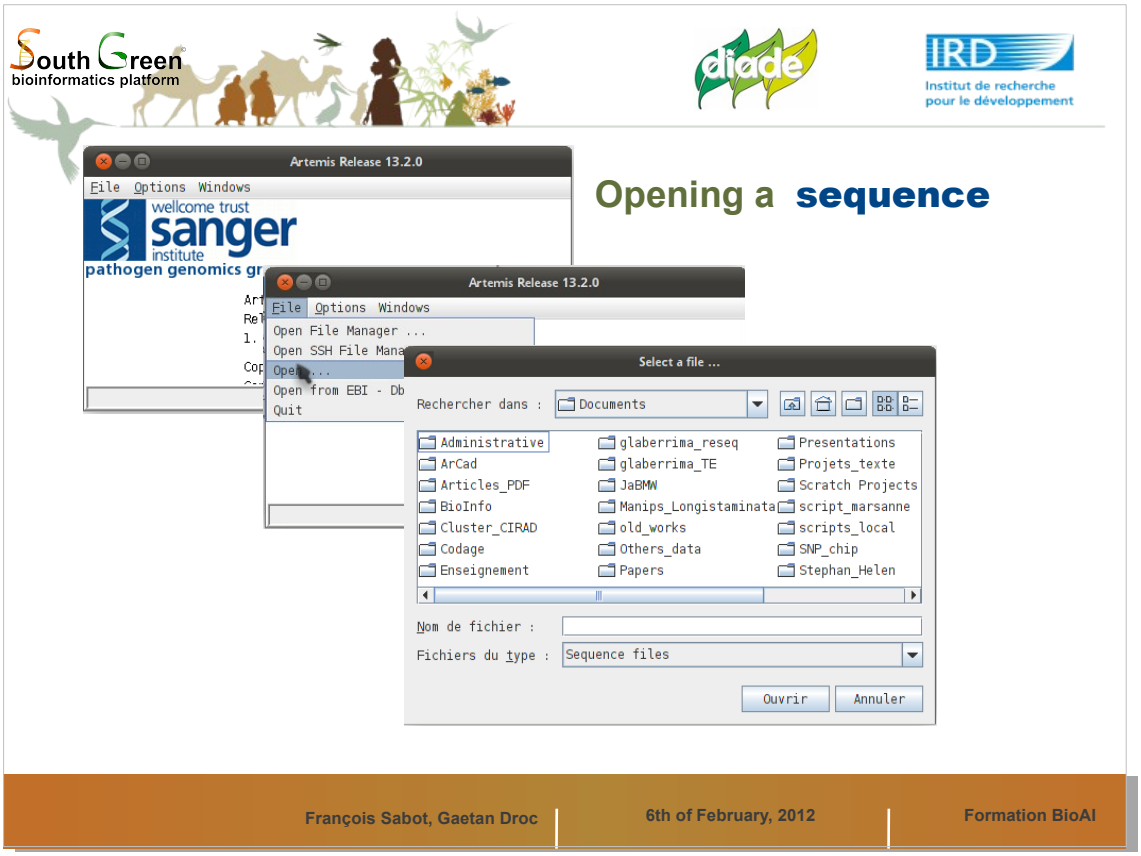
http://www.sanger.ac.uk/resources/software/artemis/



Opening a **sequence**

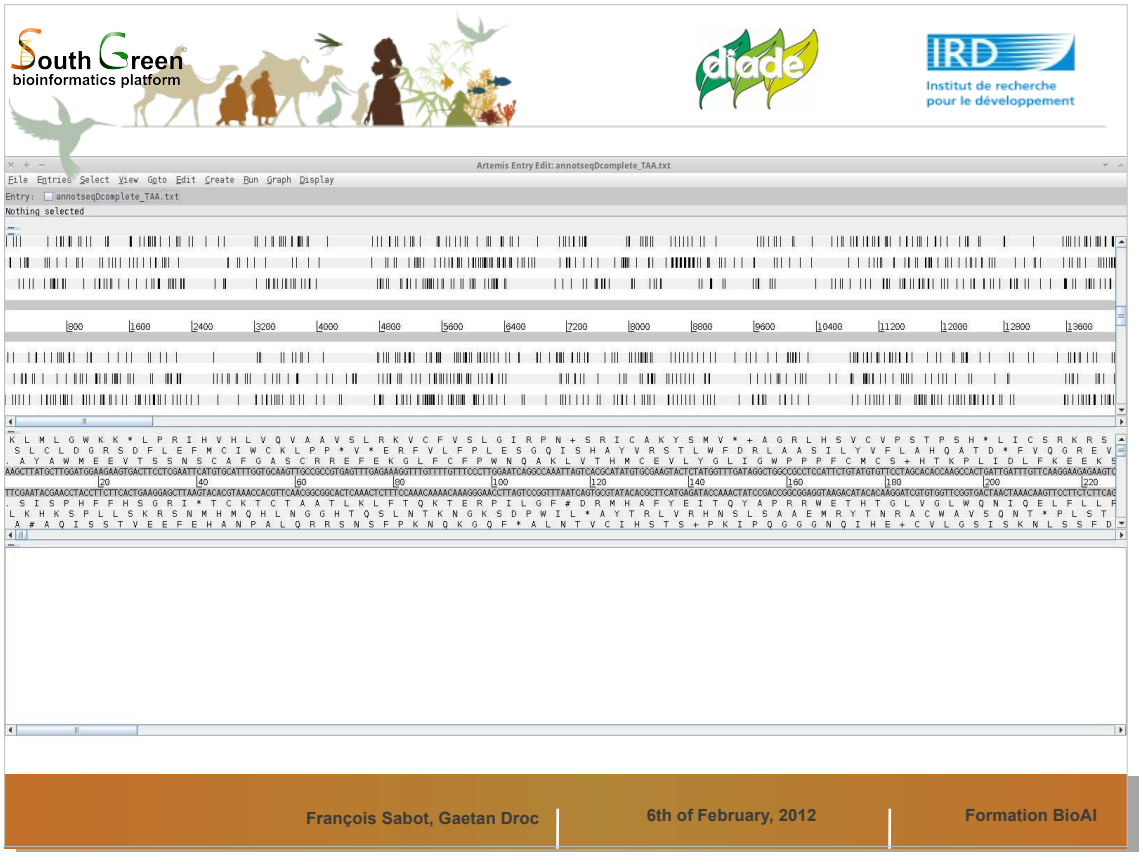
Opening a **sequence**



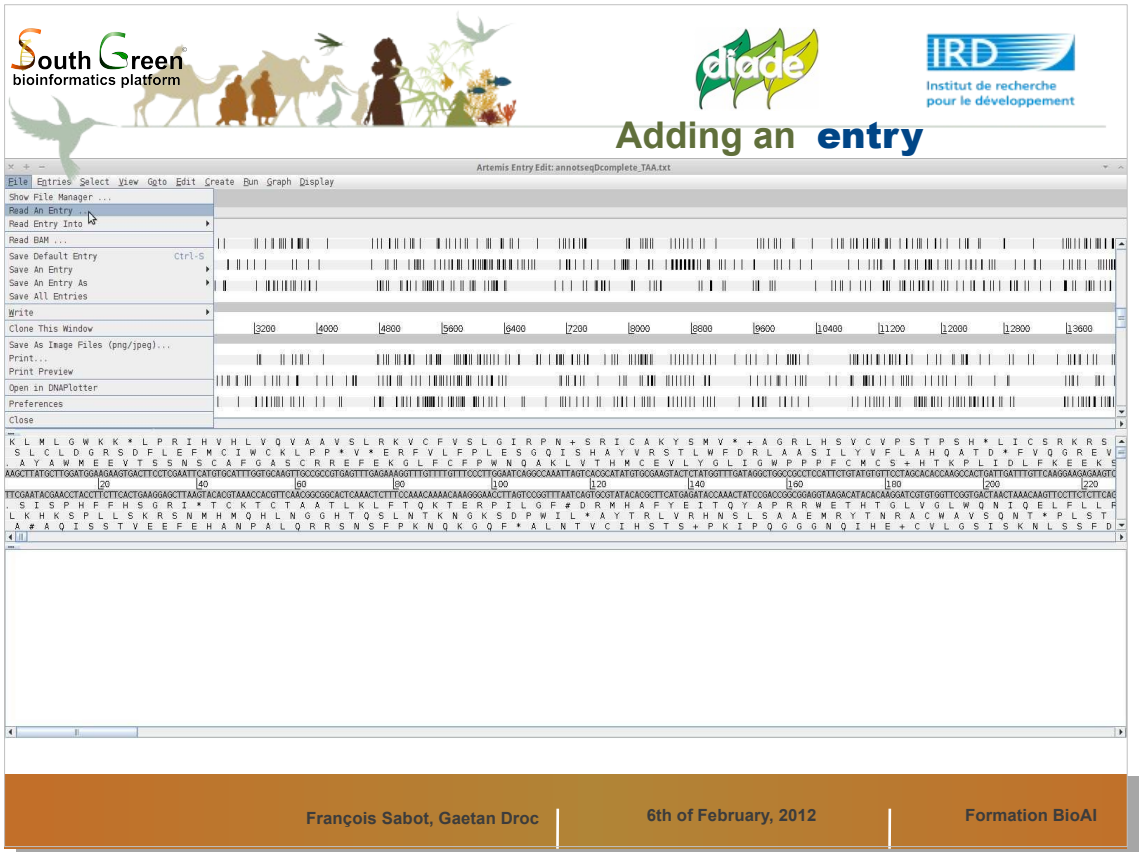


Artemis peut ouvrir des séquences Fasta, EMBL, GFF, GenBank, GTF, etc..

Si vous ne voyez pas votre fichier, demandez à voir 'Tous les fichiers' et pas seulement les 'Sequence files'.



Vue d'ensemble d'une séquence dans Artemis.
Les barres noires sont les codons stops



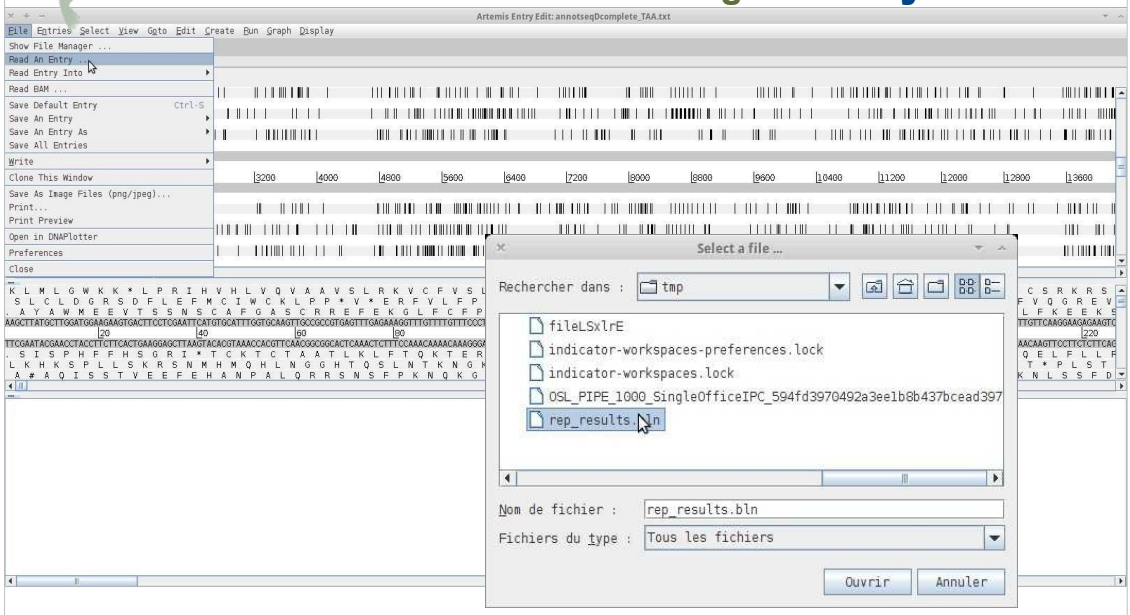
Pour ajouter une 'Entry', soit une information d'annotation/de résultat BLAST/autre, choisissez le menu comme décrit

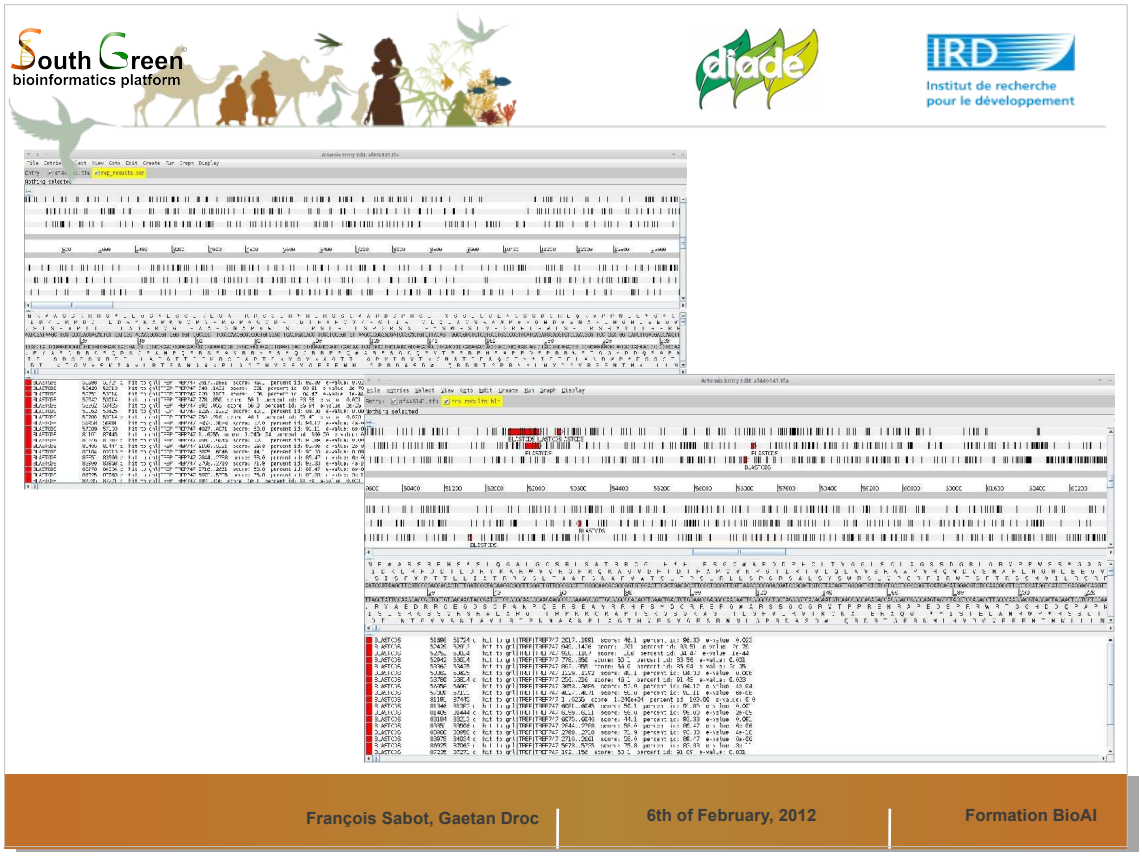
South Green
bioinformatics platform

diade

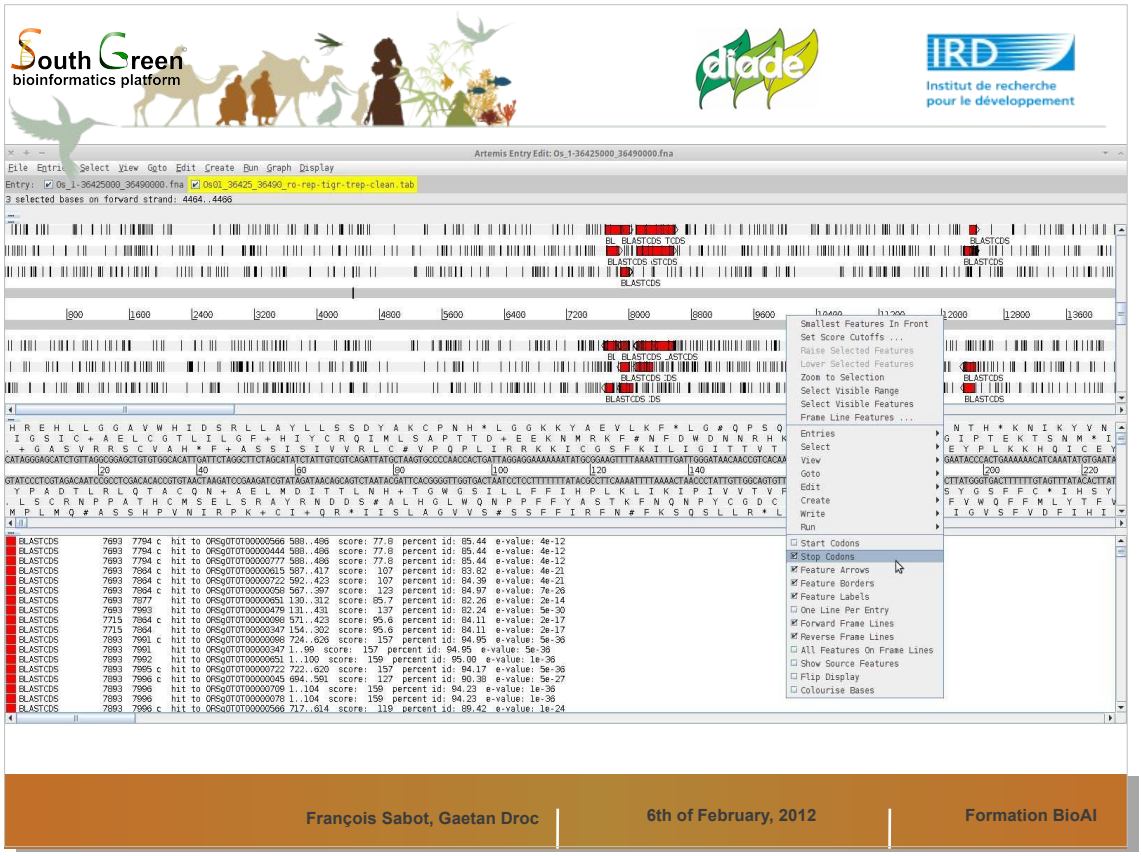
IRD
Institut de recherche pour le développement

Adding an entry

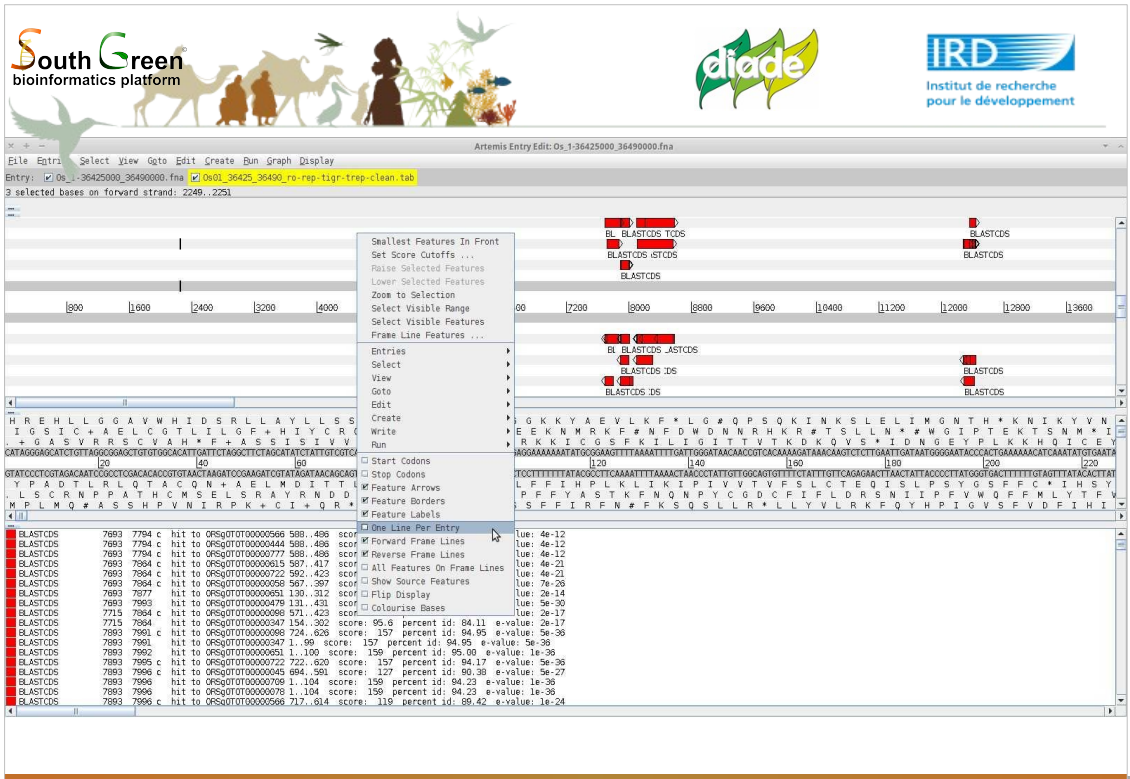


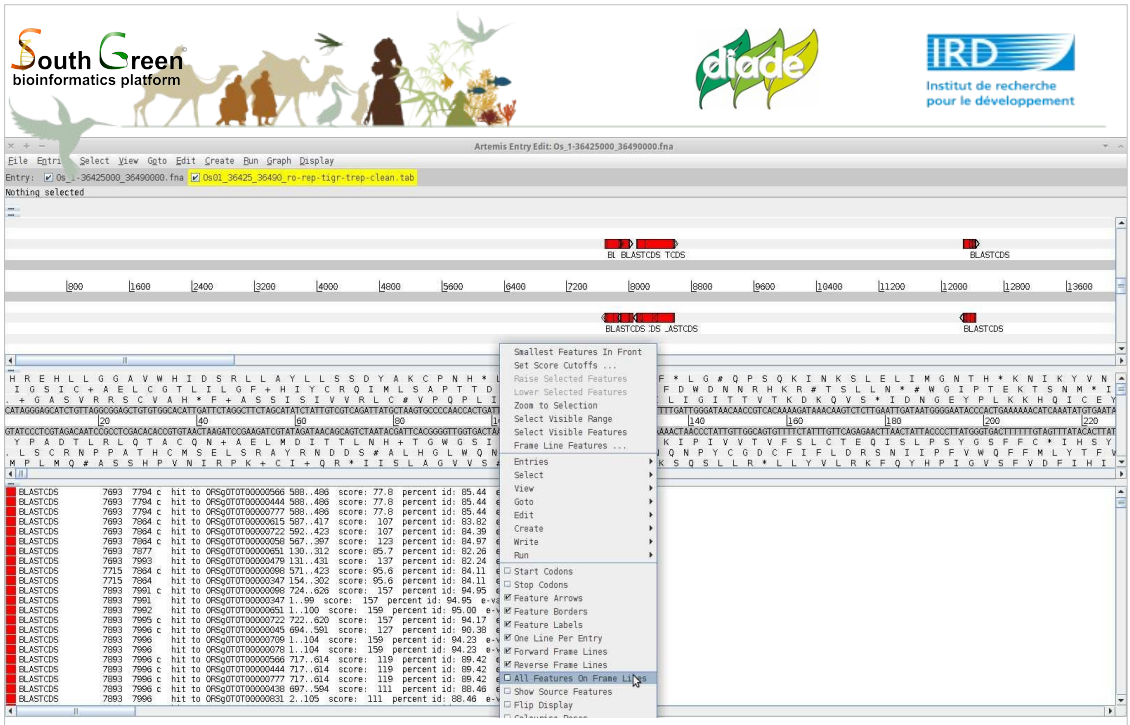


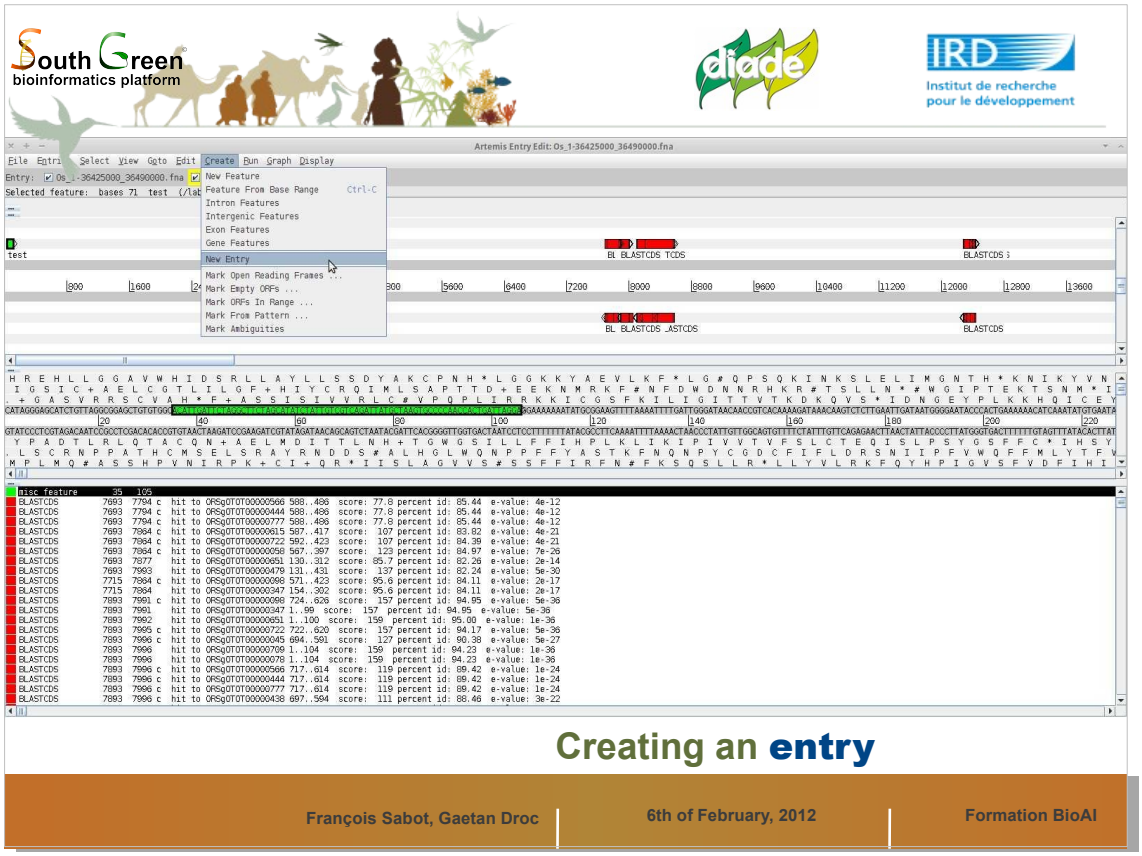
Dans le cas d'une entrée contenant un résultat BLAST (sous forme tabulée obligatoirement), les 'hits' sont notés en rouge comme BLASTCDS, meme si l'information vient d'un blast nucléique.



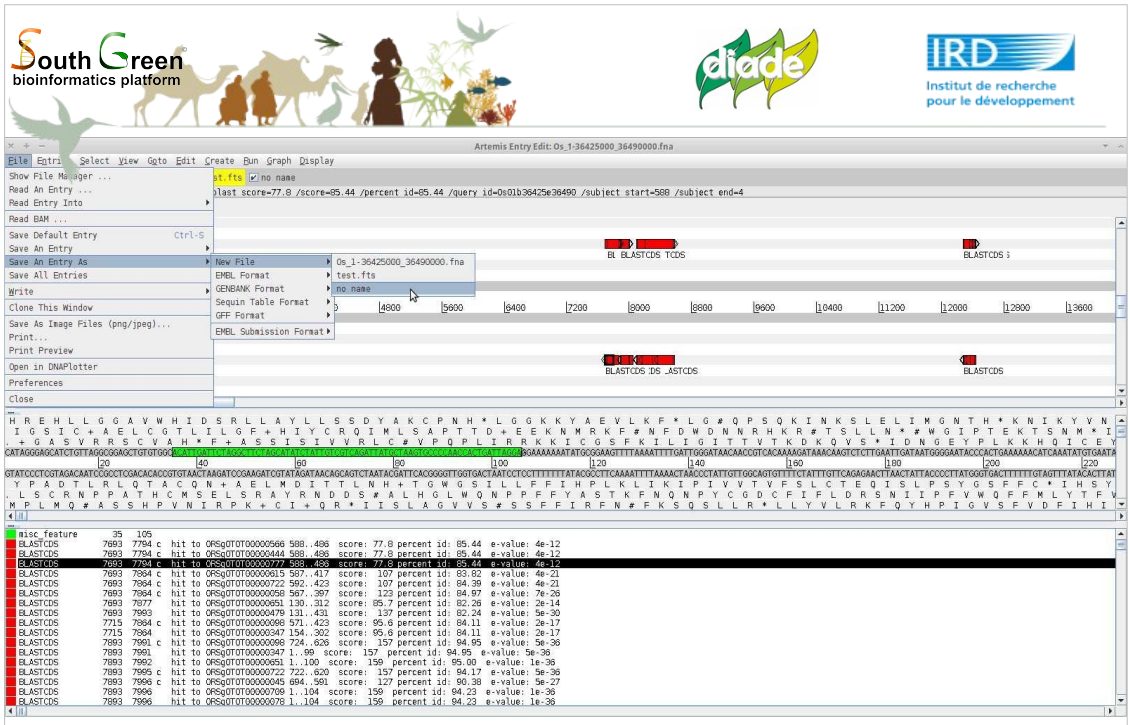
Commençons à améliorer notre visuel...


South Green
bioinformatics platform
diade
IRD
Institut de recherche
pour le développement
Artemis Entry Edit: Os_1-36425000_36490000.fna
File Entri Select View Goto Edit Create Run Graph Display
Smallest Features In Front
Set Score Cutoffs ...
Raise Selected Features
Lower Selected Features
Zoom to Selection
Select Visible Range
Select Visible Features
Frame Line Features ...
Entries
Select
View
Goto
Edit
Create
Write
Run
Start Codons
Stop Codons
Feature Arrows
Feature Borders
Feature Labels
One Line Per Entry
Forward Frame Lines
Reverse Frame Lines
All Features On Frame Lines
Show Source Features
Flip Display
Colourise Bases


South Green
bioinformatics platform
diade
IRD
Institut de recherche
pour le développement
Artemis Entry Edit: Os_1-36425000_36490000.fna
File Entri Select View Goto Edit Create Run Graph Display
Entry: Os_1-36425000_36490000.fna Os01_36425_36490_ro-rep-tigr-trep-clean.tab
Nothing selected
BLASTCDS
Smallest Features In Front
Set Score Cutoffs ...
Raise Selected Features
Lower Selected Features
Zoom to Selection
Select Visible Range
Select Visible Features
Frame Line Features ...
Entries
Select
View
Goto
Edit
Create
Write
Run
Start Codons
Stop Codons
Feature Arrows
Feature Borders
Feature Labels
One Line Per Entry
Forward Frame Lines
Reverse Frame Lines
All Features On Frame Lines
Show Source Features
Flip Display



Créez une entrée pour ajouter l'annotation que vous allez faire. Attention, on ne peut pas écrire dans l'entrée BLAST.

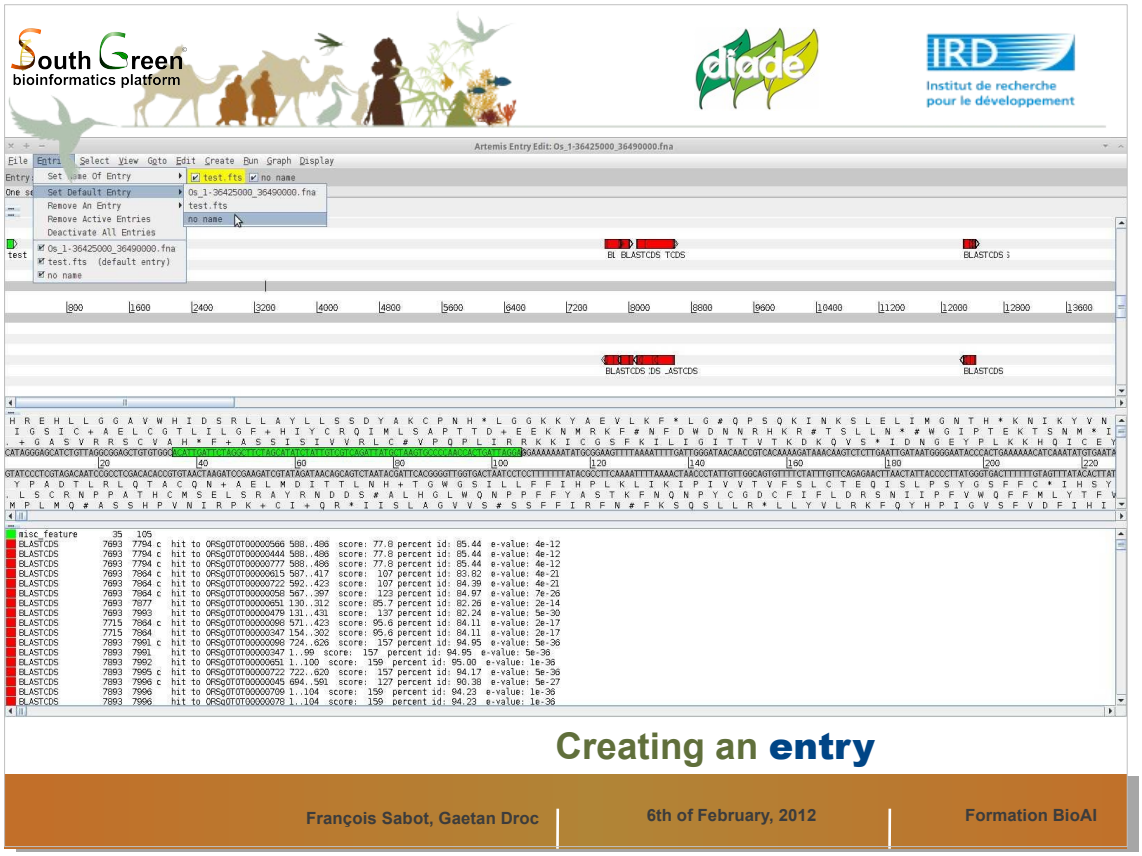


Creating an entry

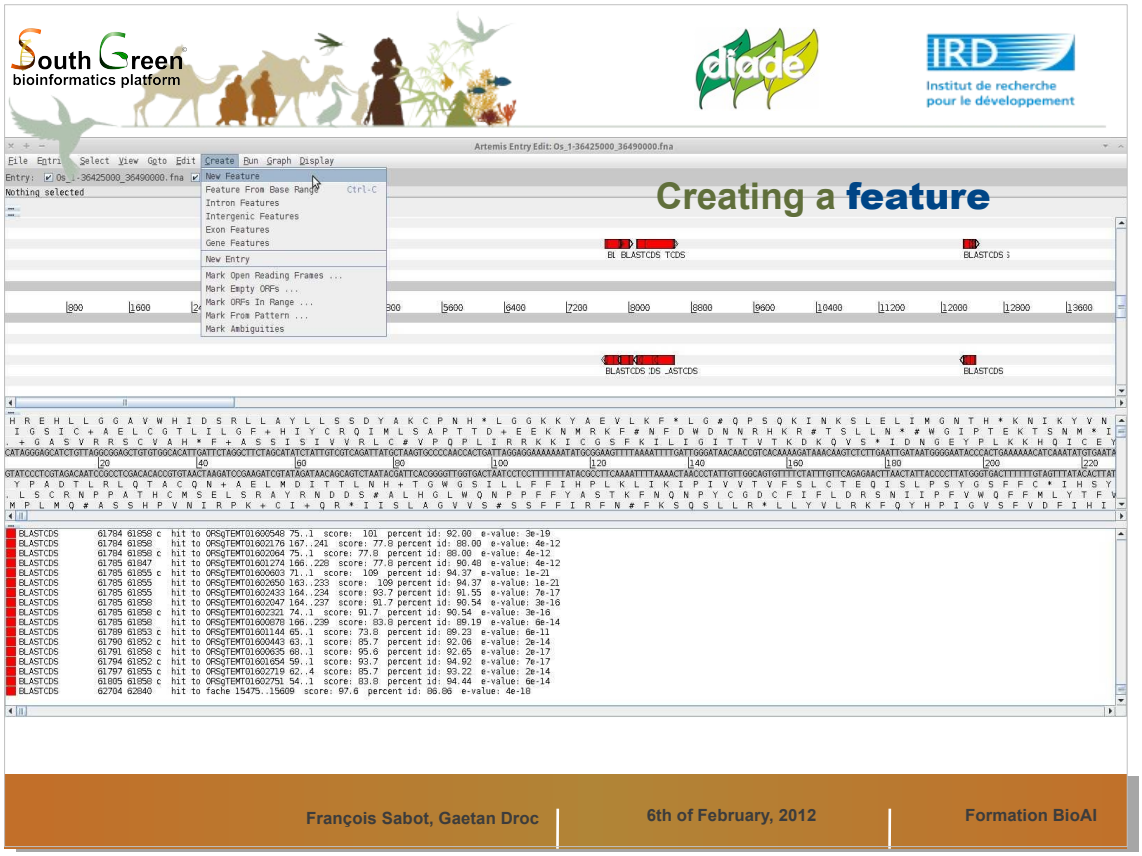
François Sabot, Gaetan Droc

6th of February, 2012

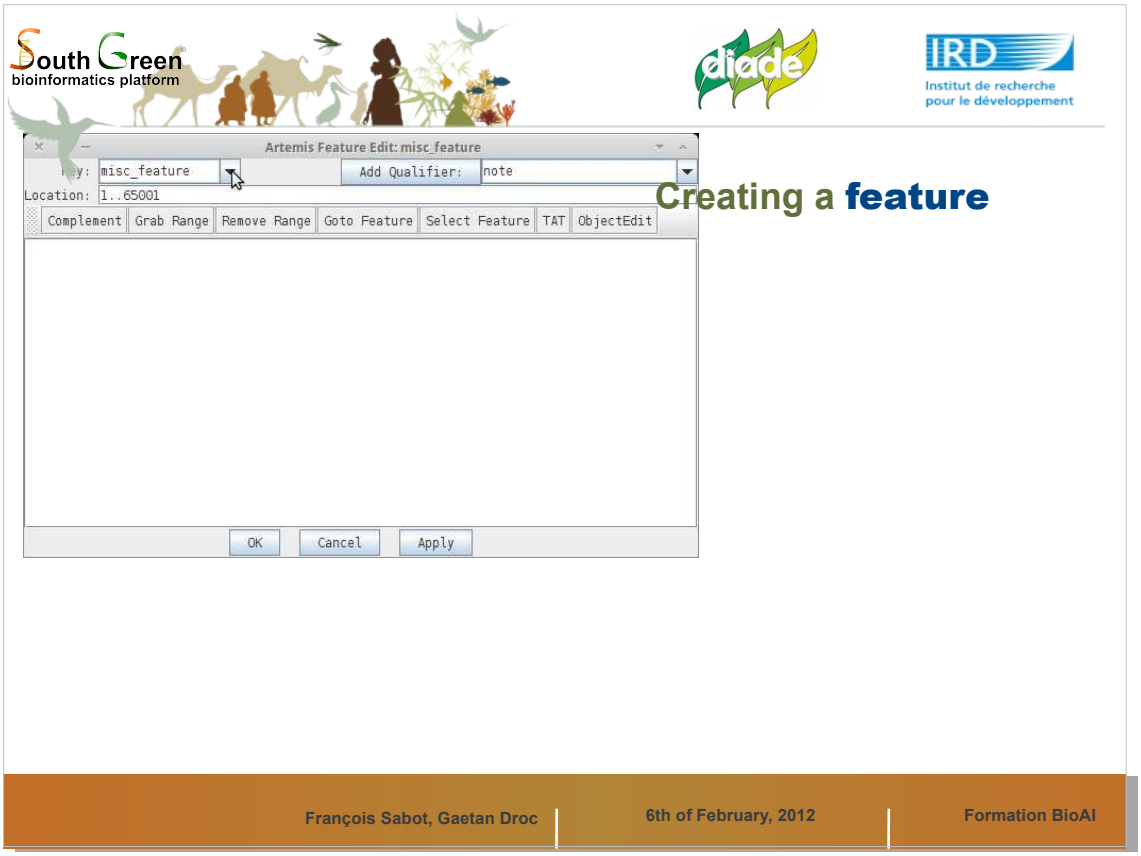
Formation BioAI



Sélectionnez votre entrée comme étant celle par défaut



Pour ajouter une information, ou Feature, il y a plusieurs possibilités

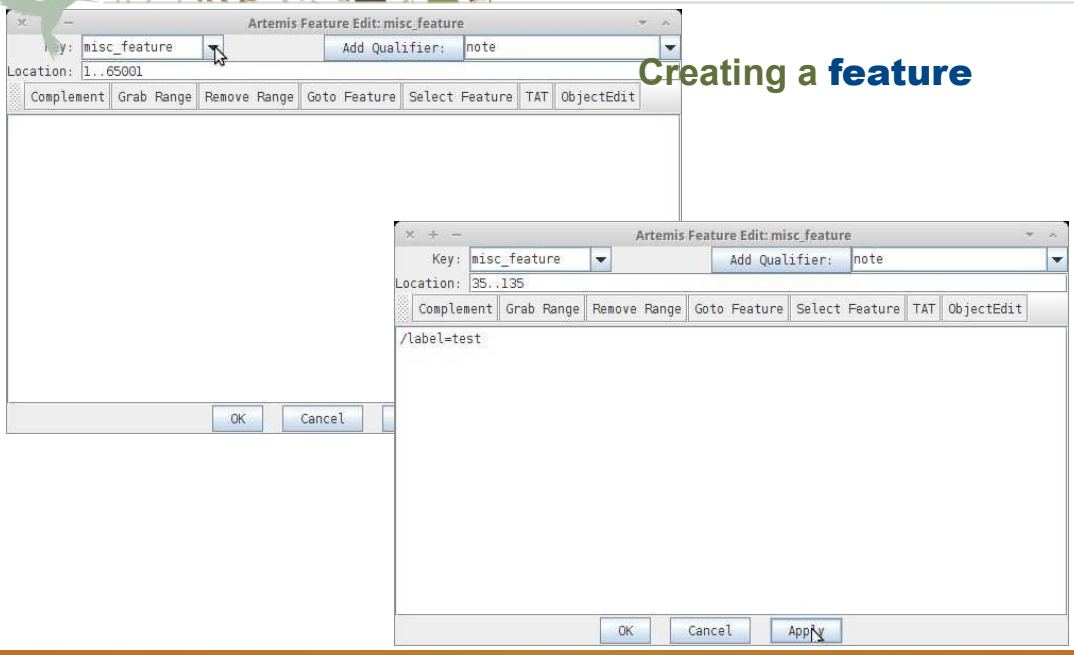


En ajoutant directement via le menu Feature une nouvelle information

South Green
bioinformatics platform

IRD
Institut de recherche pour le développement

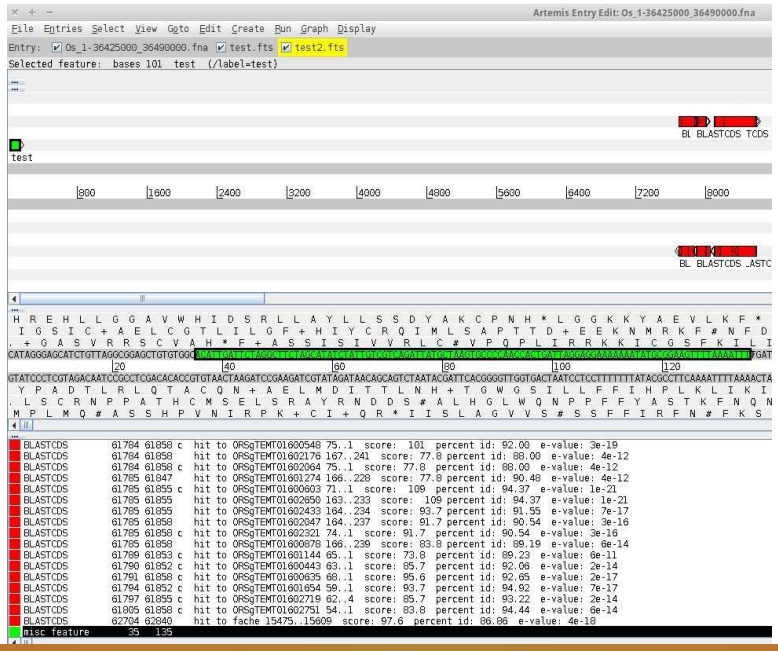
Creating a **feature**

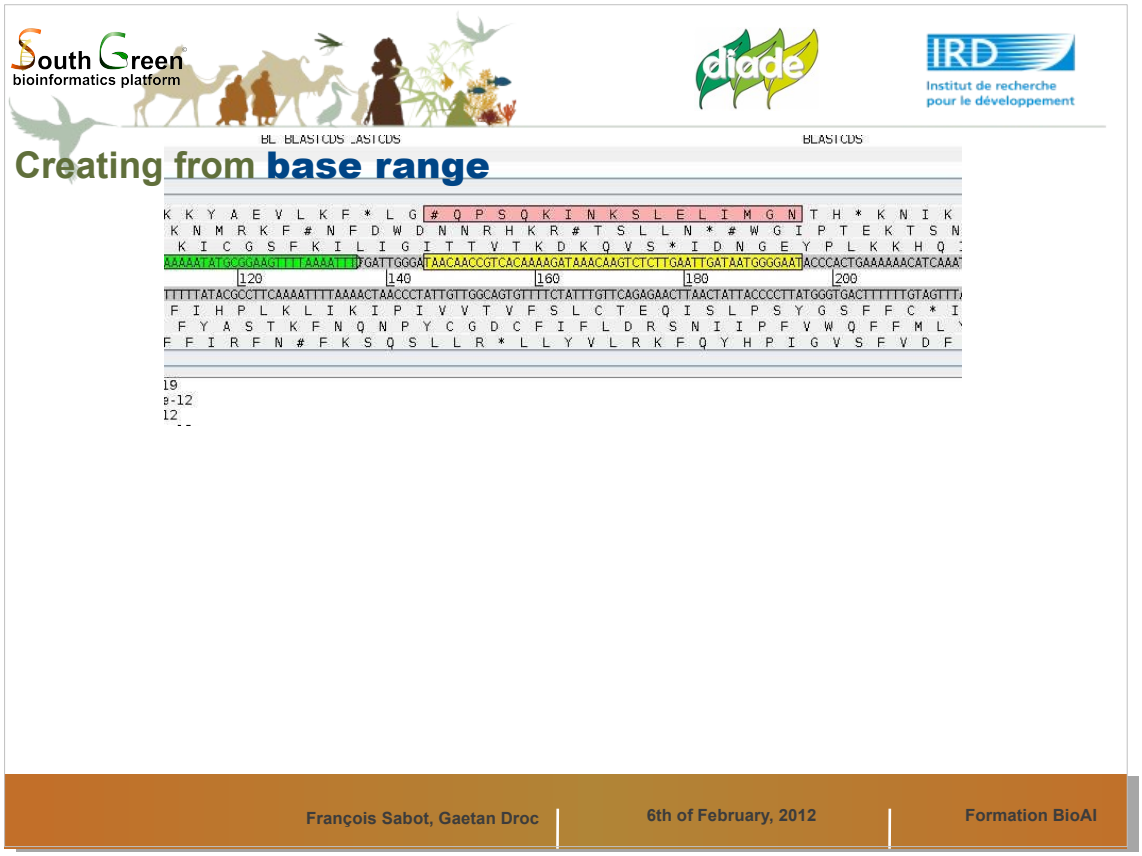


South Green bioinformatics platform

diade

IRD Institut de recherche pour le développement



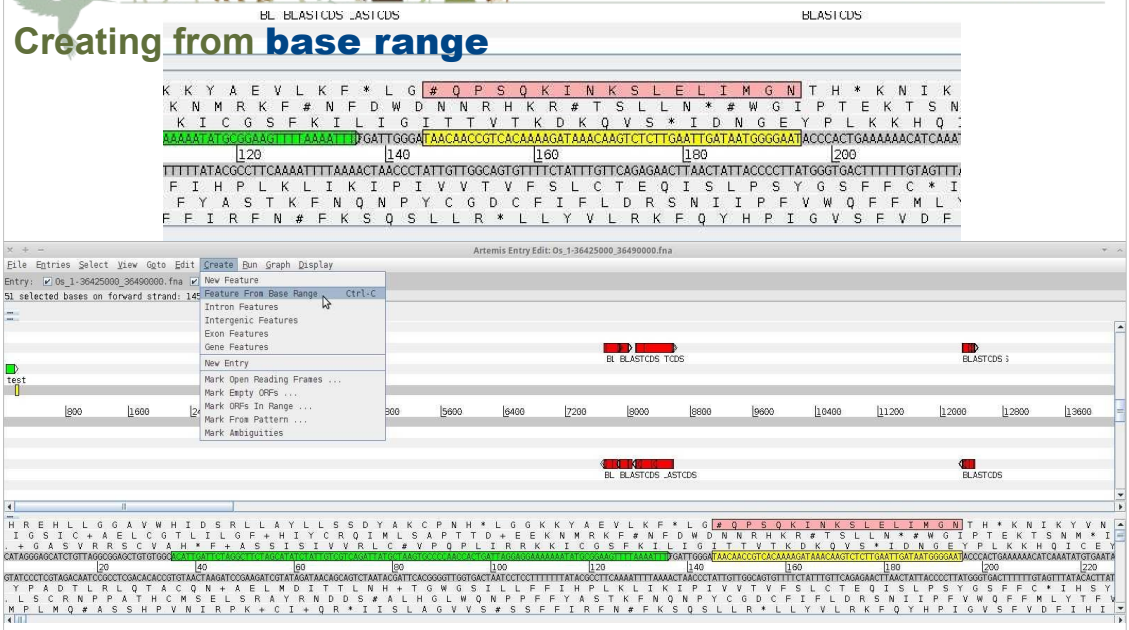


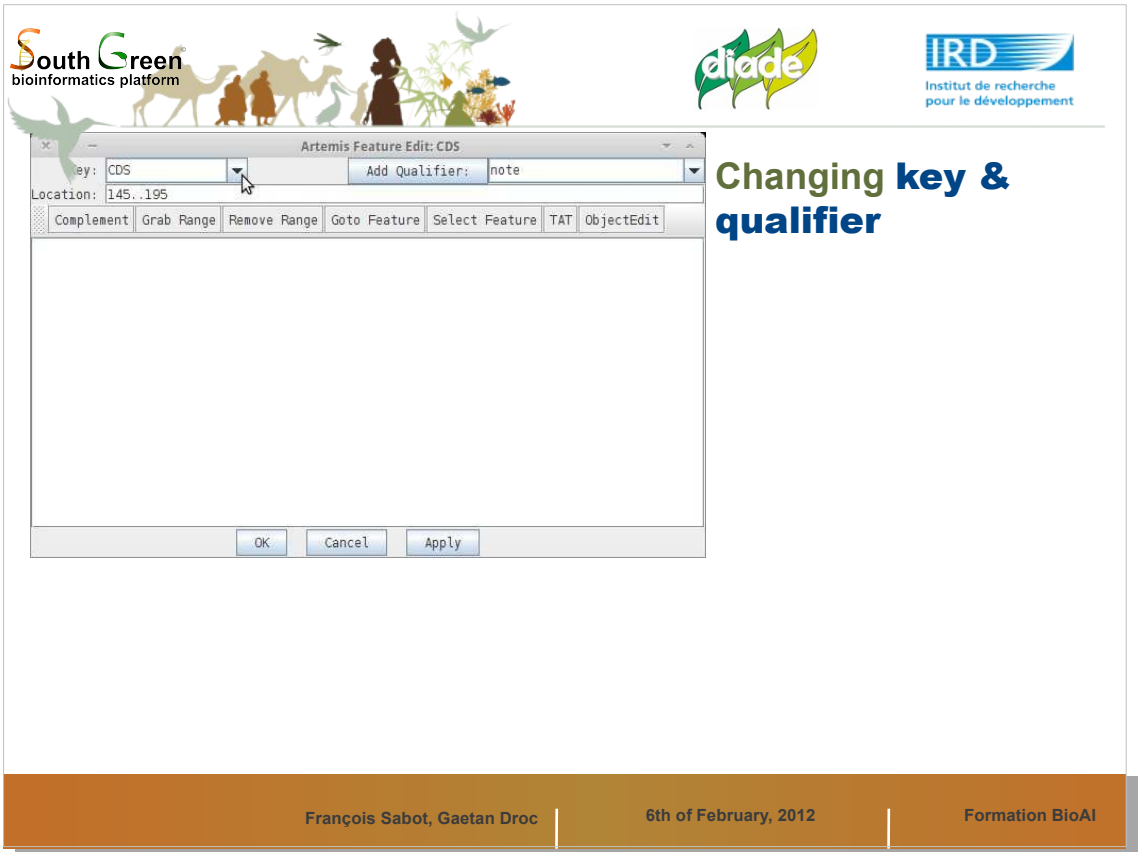
En sélectionnant une zone à la main

South Green bioinformatics platform

IRD Institut de recherche pour le développement

Creating from base range

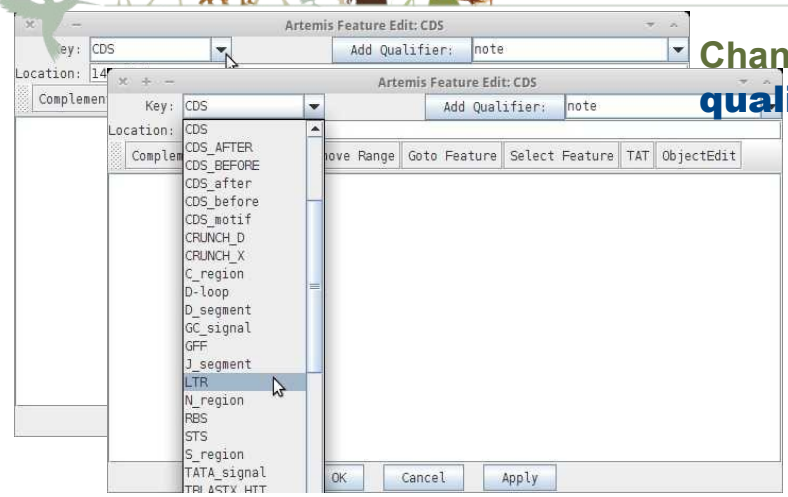




On peut aussi modifier le type de feature et lui apporter des quaificatifs.

On peut modifier une feature en la sélectionnant, et en tapant Ctrl + E, ou bien via le menu Feature/Edit selected feature in Editor

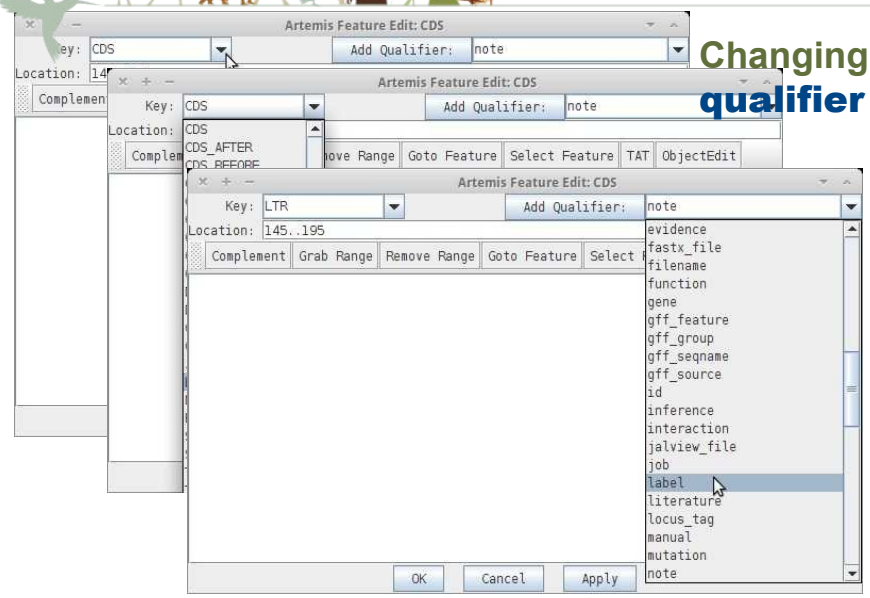
Changing key & qualifier



South Green
bioinformatics platform

IRD
Institut de recherche pour le développement

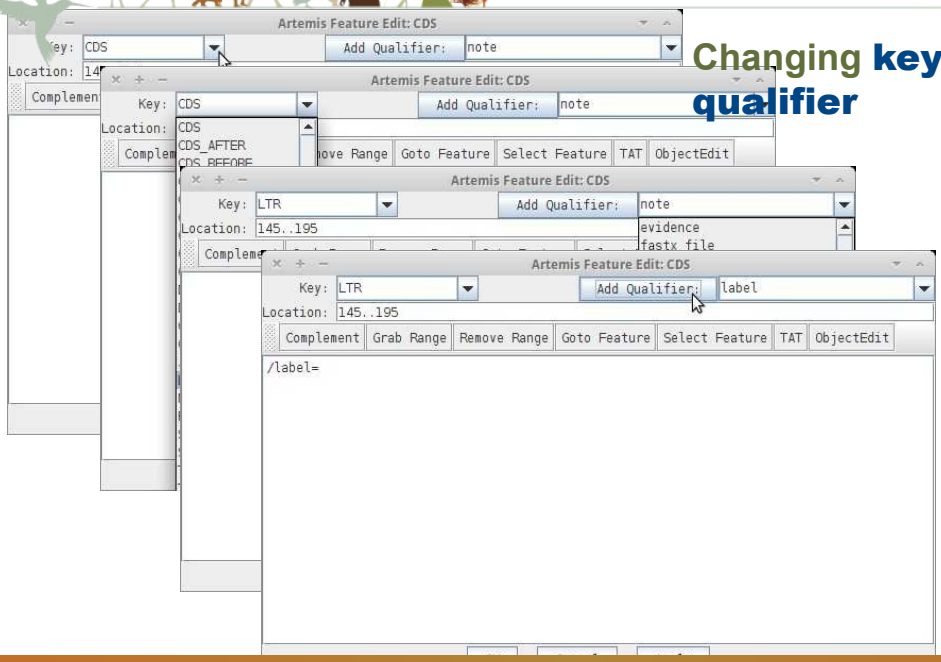
Changing key & qualifier



South Green
bioinformatics platform

IRD
Institut de recherche pour le développement

Changing key & qualifier

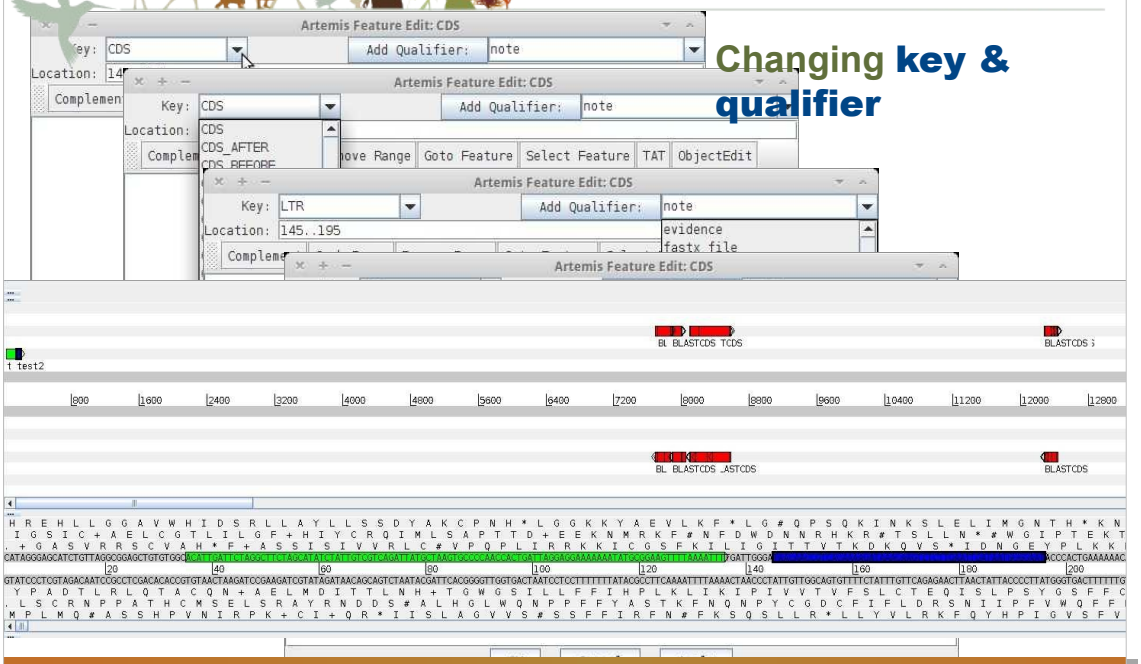


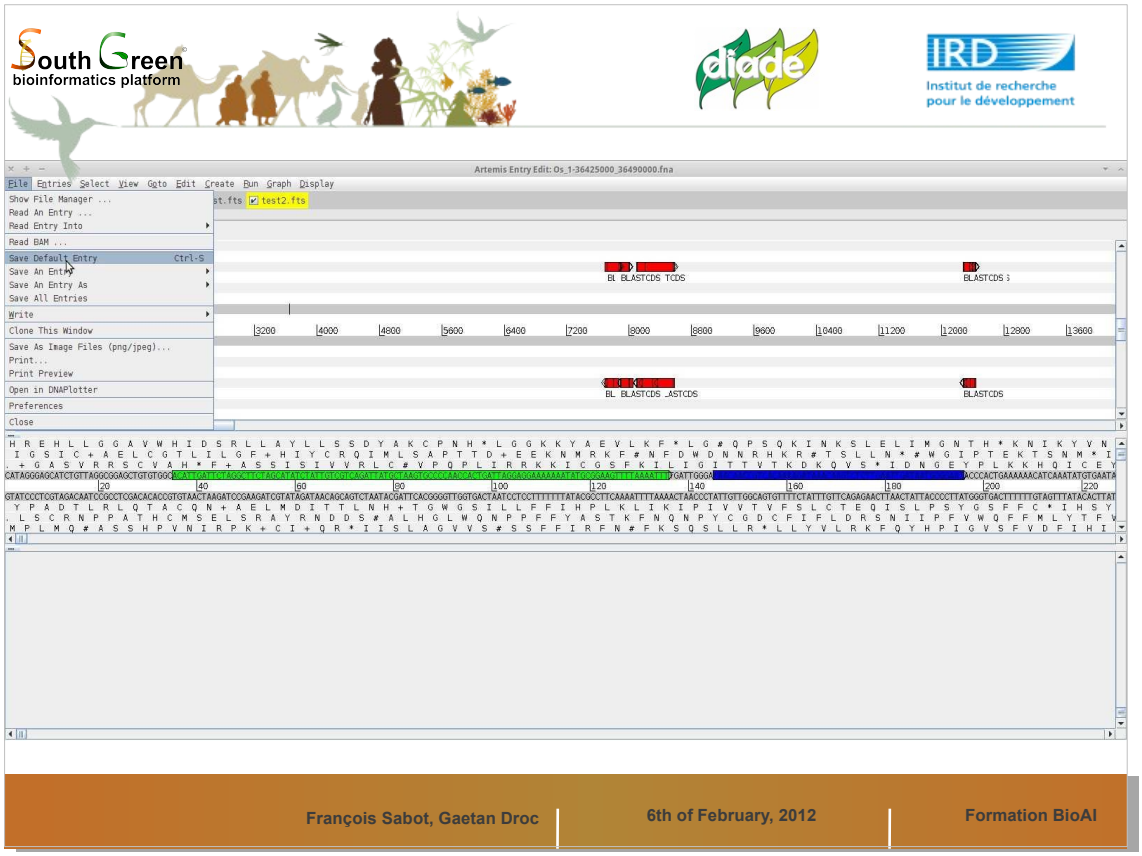
South Green
bioinformatics platform

diade

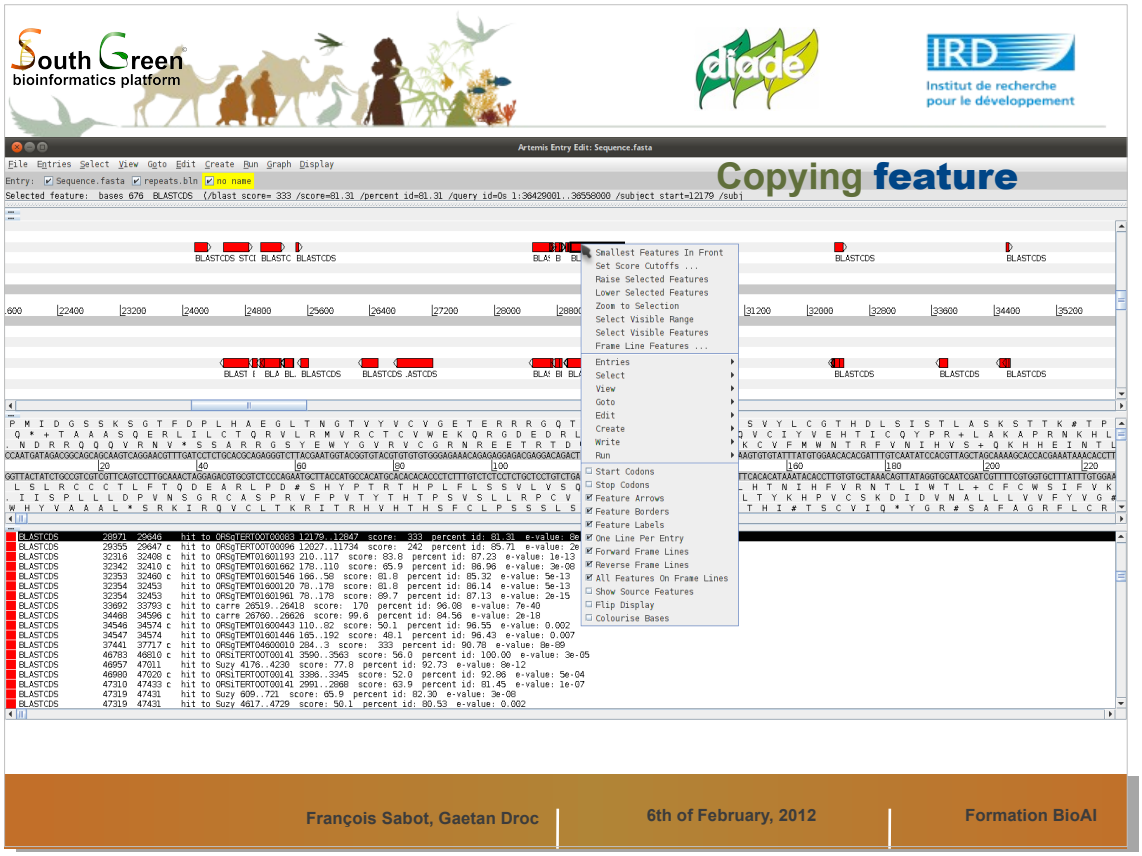
IRD
Institut de recherche pour le développement

Changing key & qualifier



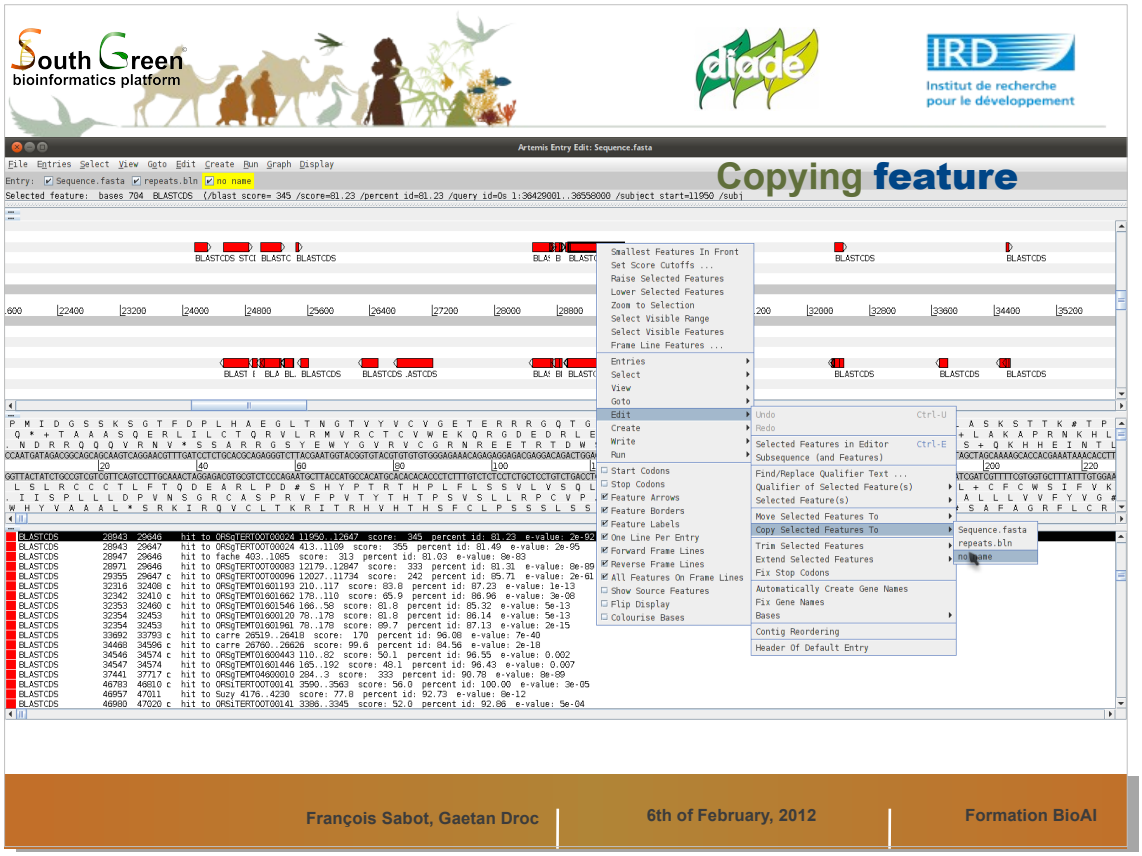


Pensez à sauvegarder régulièrement vos ajouts...



Il est possible de transferer une annotation d'une entrée à une autre.

Cliquez-droit sur une feature en rouge,



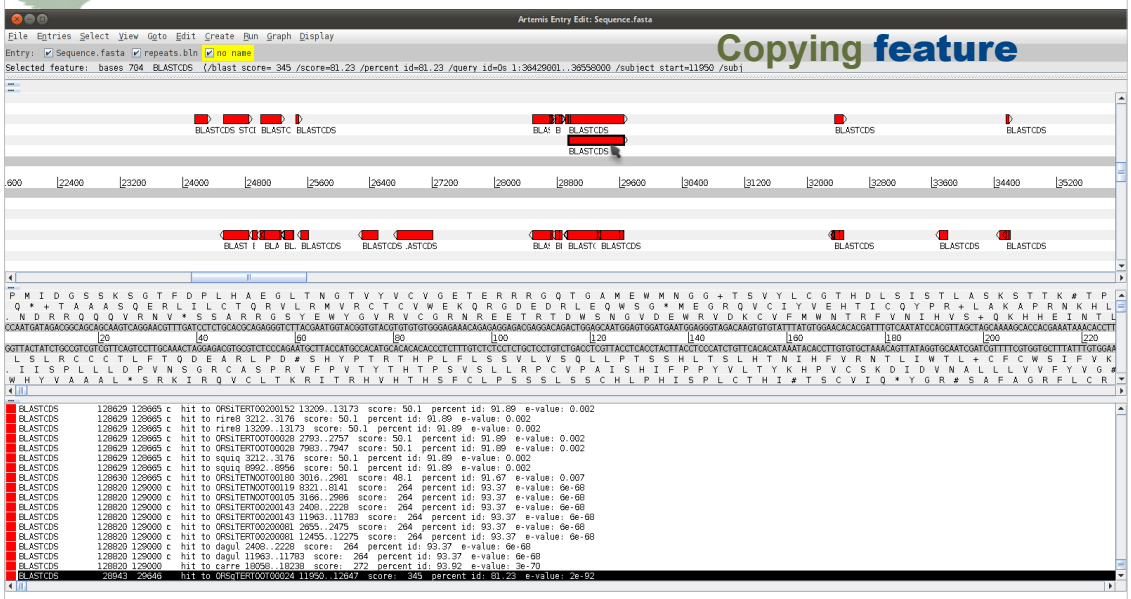
Puis sélectionnez Entry/Copy entry to
Choisissez le nom de votre entrée.

South Green bioinformatics platform

diade

IRD Institut de recherche pour le développement

Copying feature

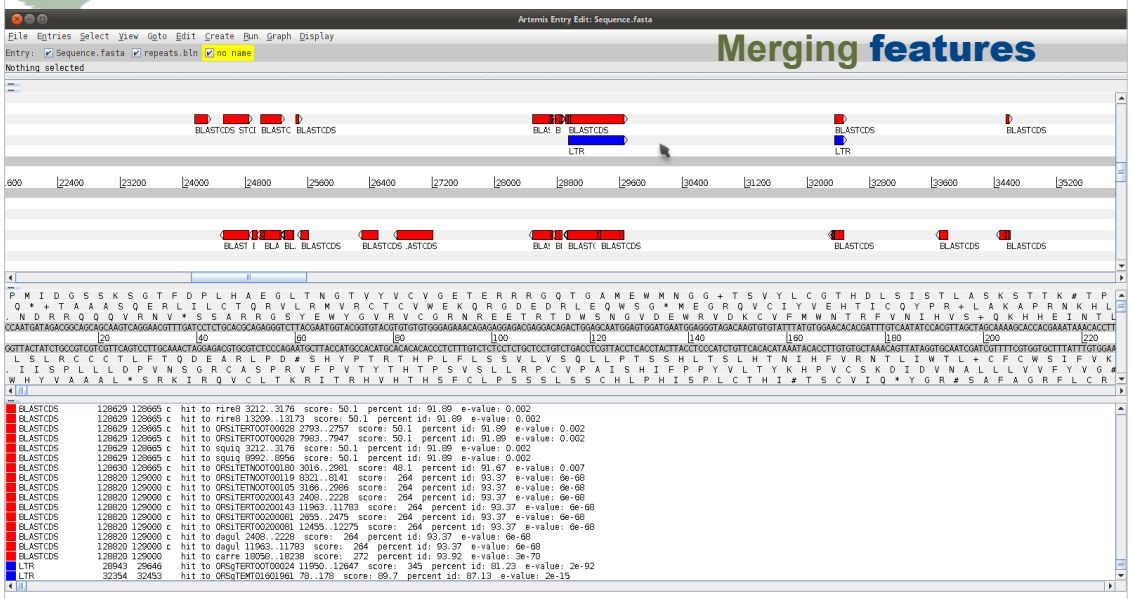


South Green bioinformatics platform

diade

IRD Institut de recherche pour le développement

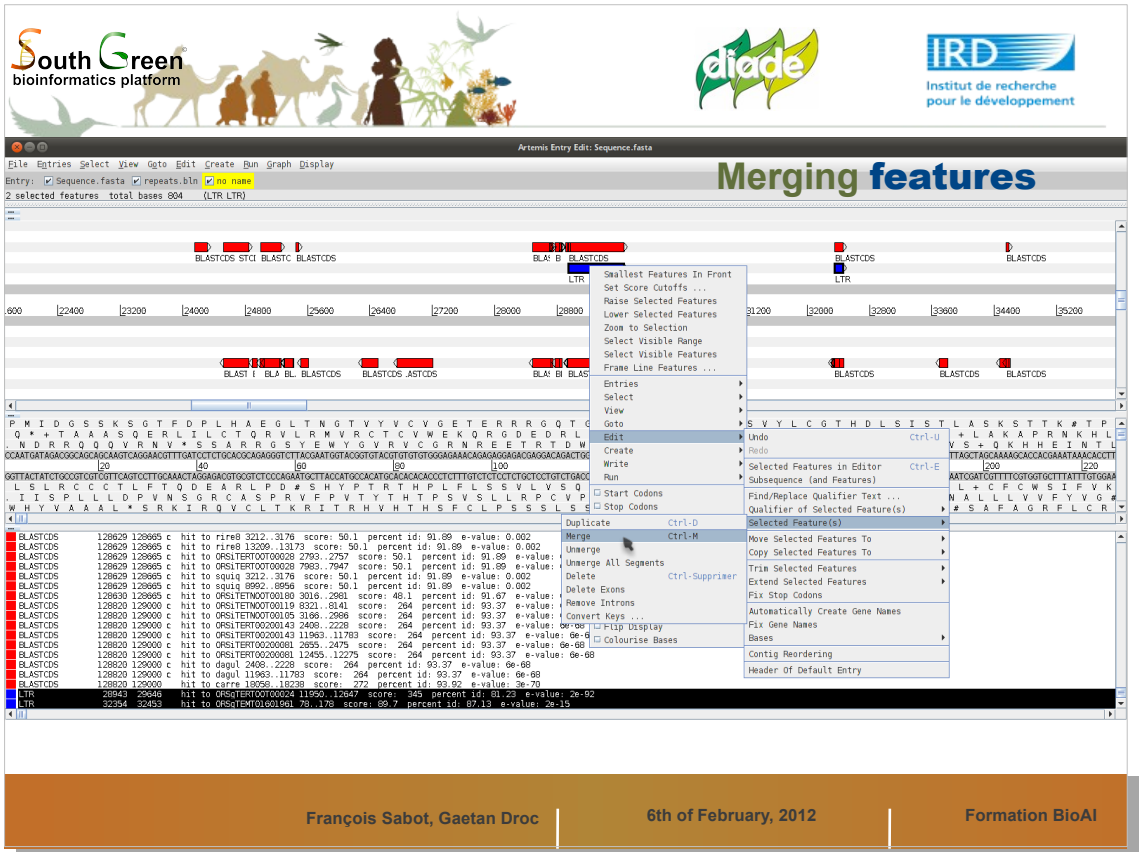
Merging features



François Sabot, Gaetan Droc

6th of February, 2012

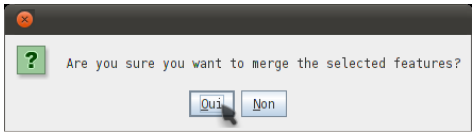
Formation BioAI



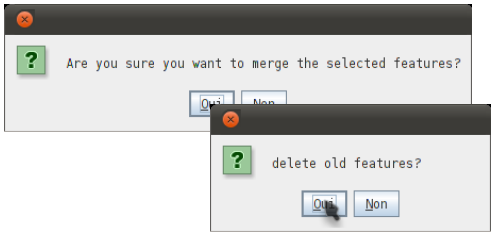
On peut aussi joindre ('Merge') deux features. Sélectionnez les deux (ou plus), puis clic-droit/Entry/Merge.

Ou bien sélectionnez les, puis Ctrl + M

Merging features



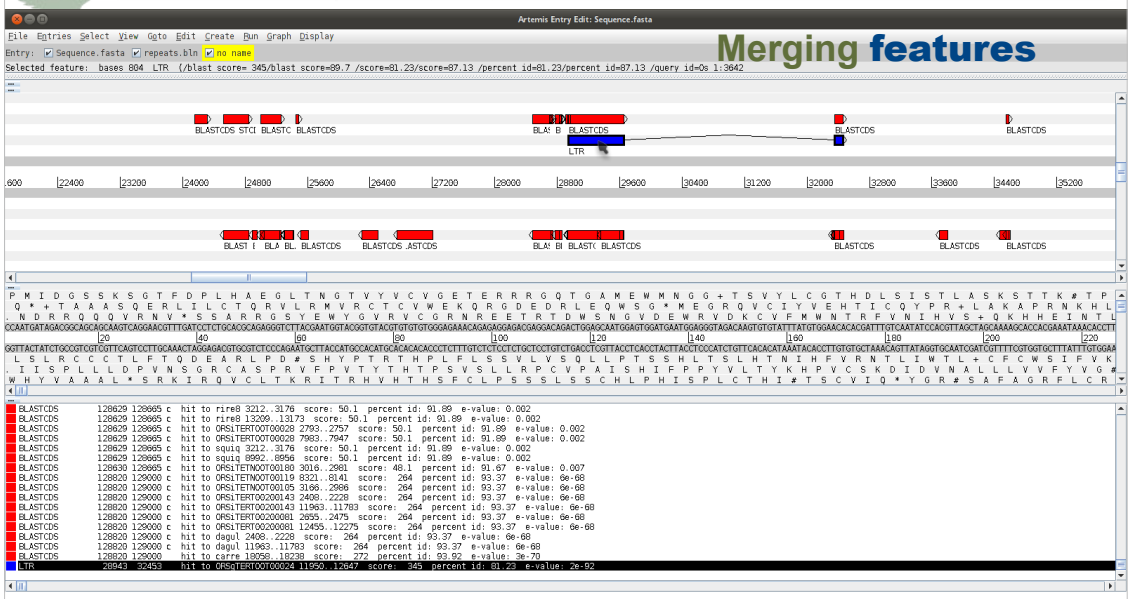
Merging features

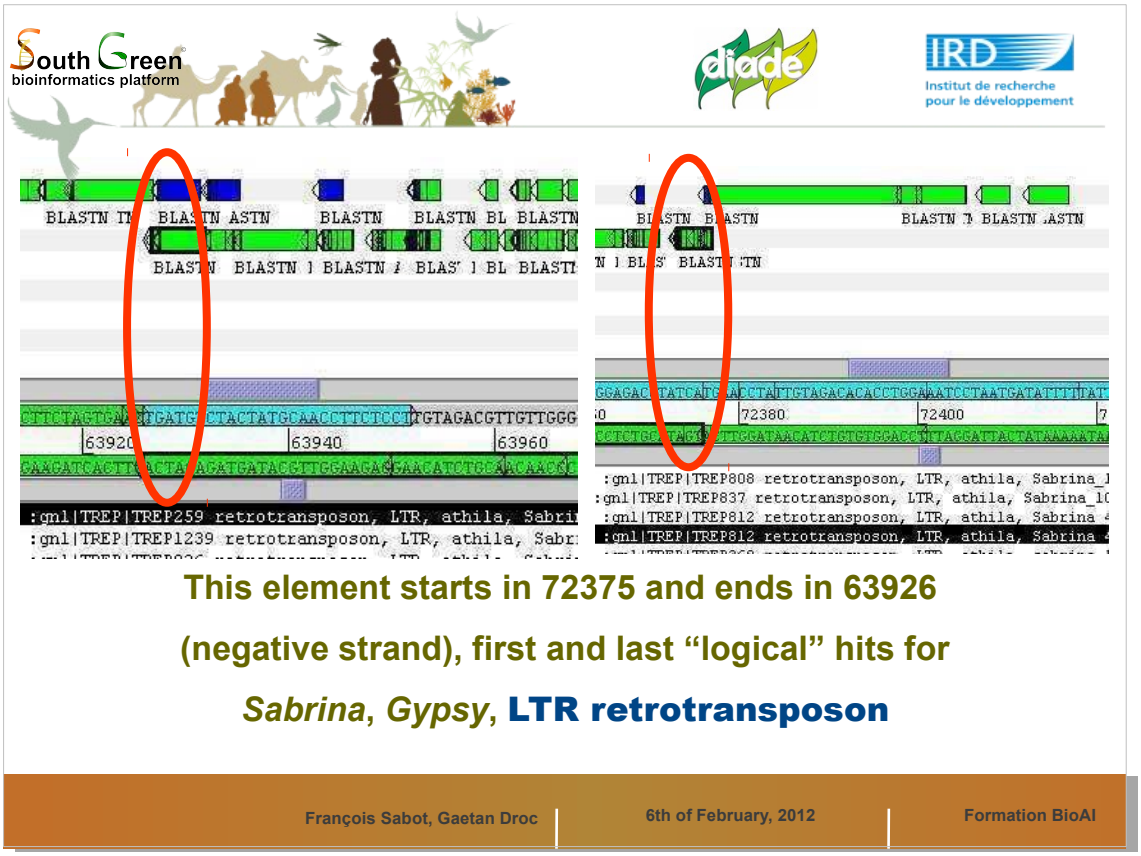


South Green bioinformatics platform

IRD Institut de recherche pour le développement

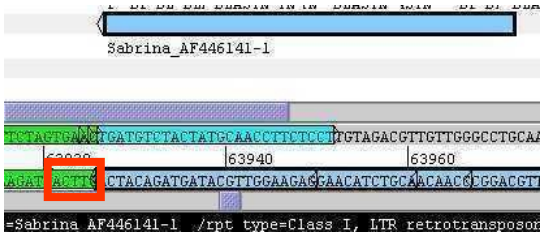
Merging features

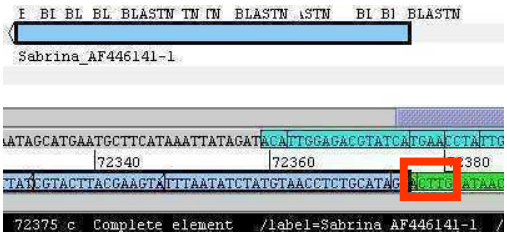




Quelques informations supplémentaires...

Manual identification of the **TSD**



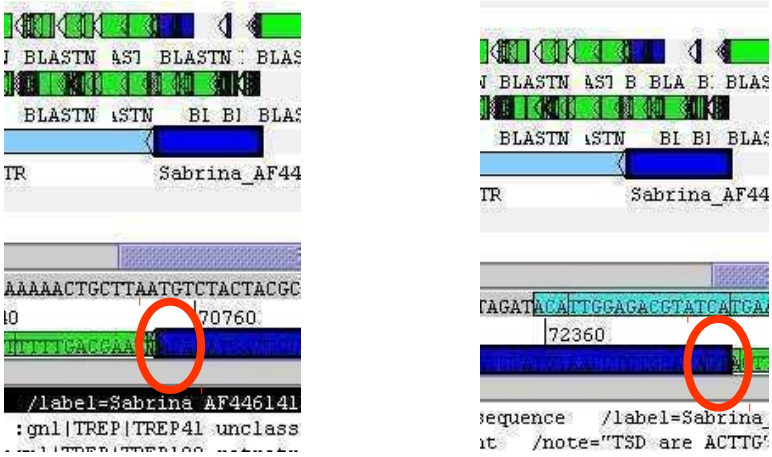


TSD are **ACTTG** here, and are added to the description

South Green bioinformatics platform

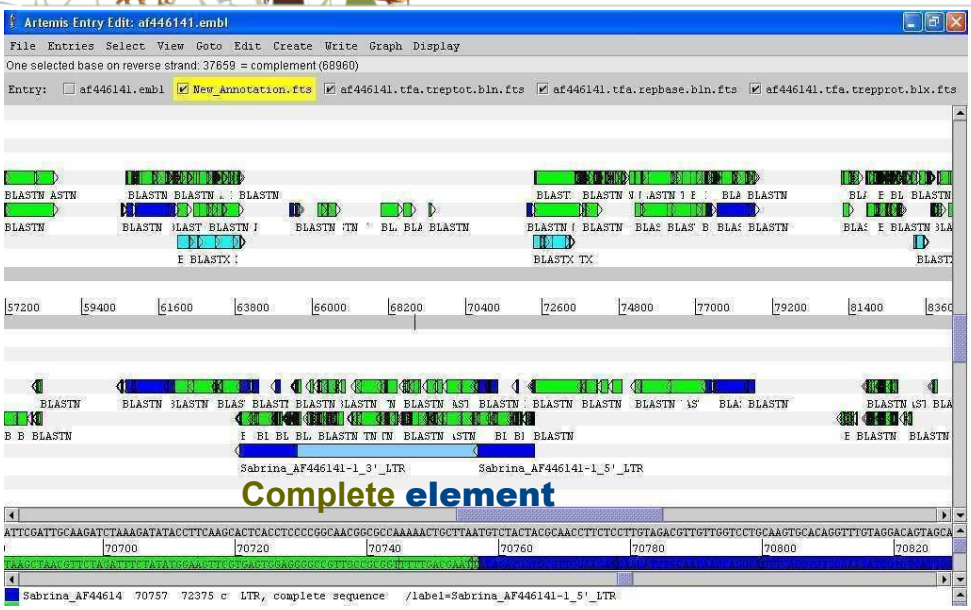
IRD Institut de recherche pour le développement

Identification of the 5' LTR TG..CA

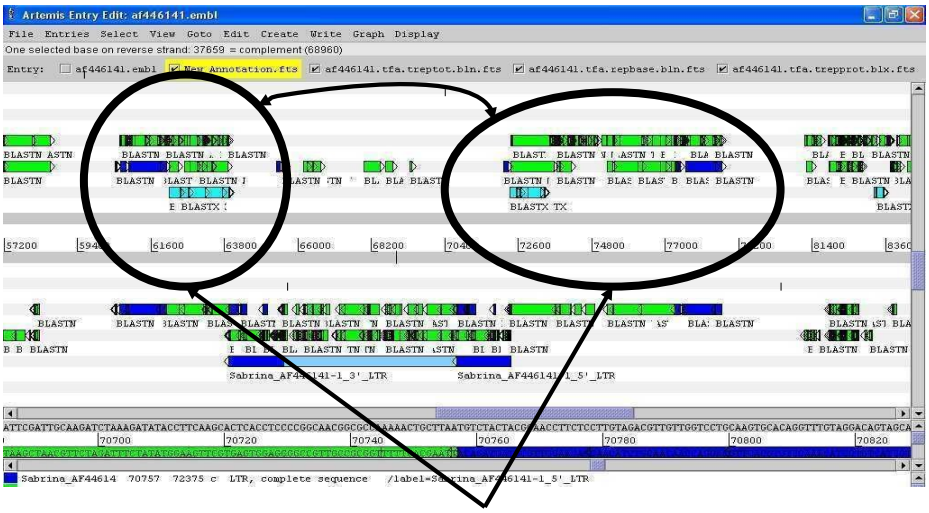


5' LTR from 70756..72375, new feature created

Key feature: LTR

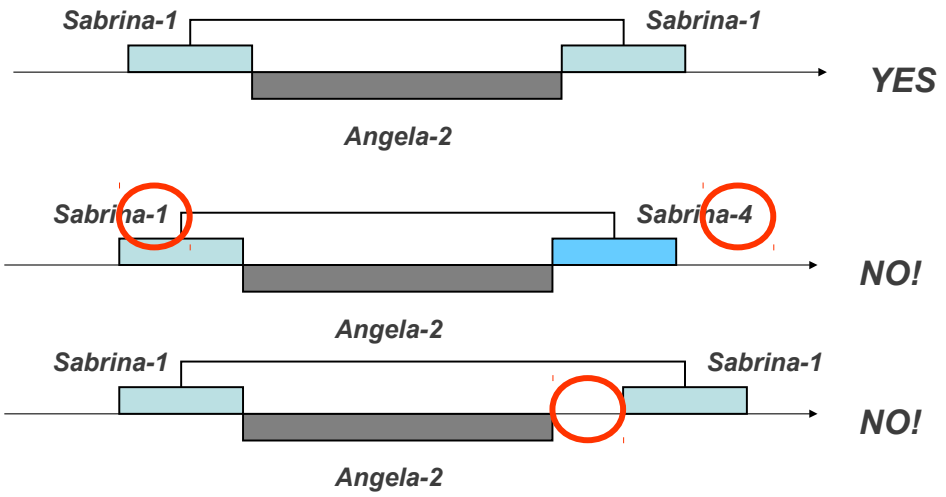


Complete element

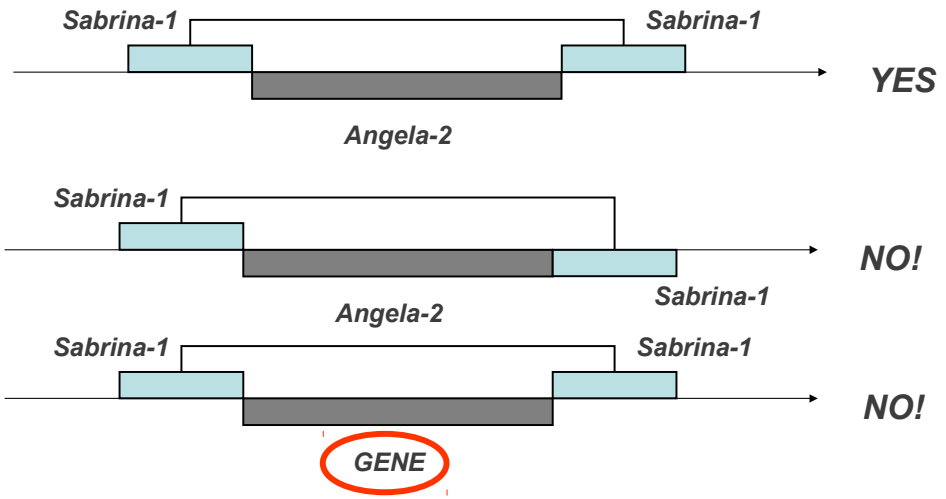


Classical aspect of a nested element: same hits on each side, Continuous and related (no breaks in the sequence)

Be careful for the **merging** of elements! Some examples



Be careful for the **merging** of elements! Some examples



Few additional **advices**

_ Always prefer the **BLASTn** to the BLASTx results

_ Always prefer the **Specific** db to the RepBase (more details)

_ Be careful **not to cross multiple** annotations at the same position (e.g., end of element X and start of element y at the same nucleotide)

_ Do not fear to put "**uncertain**" as a comment for an ambiguous annotation

_ An EST/cDNA is **not necessary** a gene