

Bio-informatique
Biologie, informatique ... et
gastronomie !

Sandrine Moreira Rousseau

Doctorante en Bio-informatique

Centre Robert Cedergren

Université de Montréal

Ressources Web



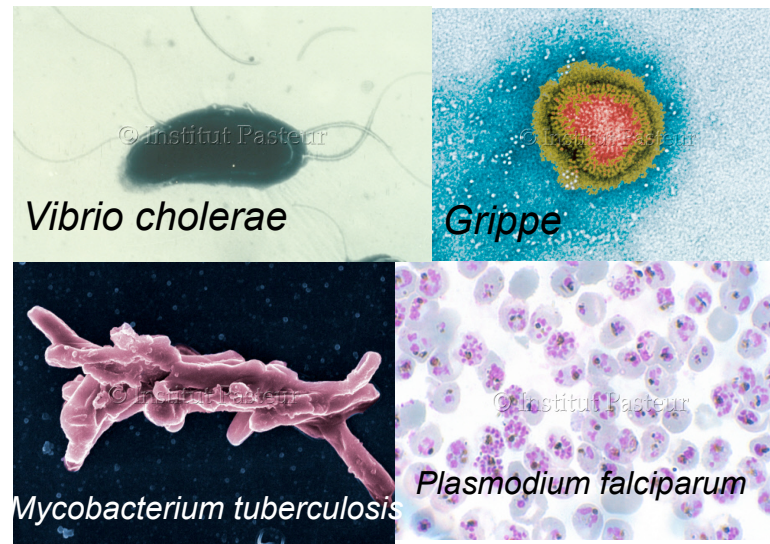
[http://sam.bioinfo.free.fr/animations_science/
SEUR_Listeria/index.html](http://sam.bioinfo.free.fr/animations_science/SEUR_Listeria/index.html)

Présentation



Pasteur : les maladies infectieuses

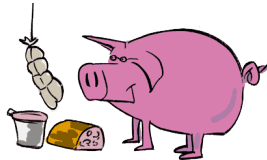
- Grippe
- Cholera
- Peste
- Paludisme
- Anthrax
- Tuberculose
- VIH



=> Des millions de morts par an

... et la listeriose !

- 200 à 300 cas par an
- C'est peu, mais cela touche :
 - la charcuterie



- le fromage



- le saumon fumé



... et en France, la gastronomie, c'est sacré !

Comment la bio-informatique peut sauver les fromages, la charcuterie et le saumon fumé ?

Vos
bio-
informaticiens
s'en occupent ?



Ne vous inquiétez
pas m'sieur : ces
gars-la sont des
champions !

Le bon, la brute et le truand

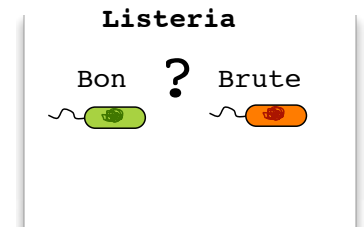


- Le problème : certaines espèces de *Listeria* sont dangereuses, d'autres non
- Comment les reconnaître ?
- A travers cet exemple concret, nous allons voir quelques aspects du travail du bio-informaticien.

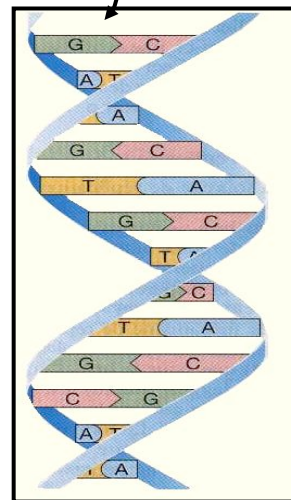
Notre question sera :

Comment reconnaître les *Listerias* inoffensives, les dangereuses et les truands !

Il faut y regarder de plus près ...



De beaucoup plus près !



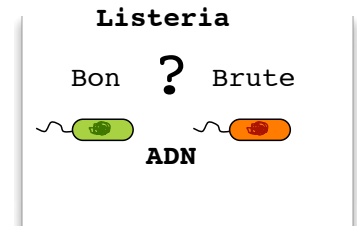
La bactérie a éclaté et laissé échappé son chromosome

*Schéma d'un morceau d'ADN.
On y voit la succession des 4 bases A, C, T et G*

Ressource web : chromosomes et ADN

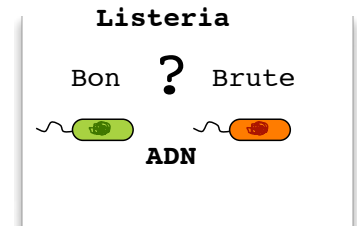
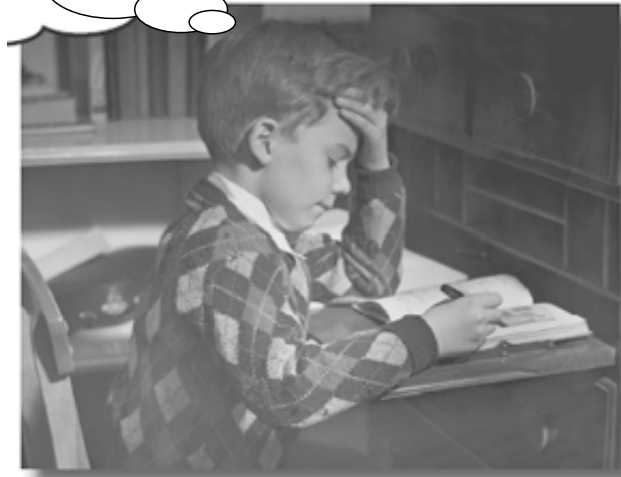
L'ADN

- Il existe dans le **programme génétique** de mes Listerias **toute l'information** dont elles ont besoin pour vivre et se multiplier
- Il doit y avoir aussi les informations qui les rendent **dangereuses** ; leurs "armes" bactériennes.



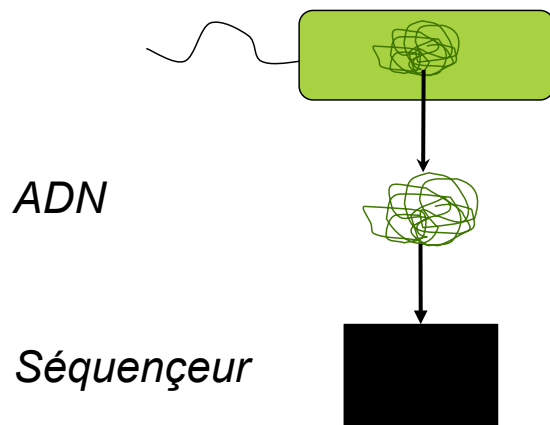
Qui sait lire l'ADN ?

Pff, lire l'ADN c'est
vraiment pas facile.
J'aurais bien besoin
d'un séquenceur !



Le séquençage de l'ADN

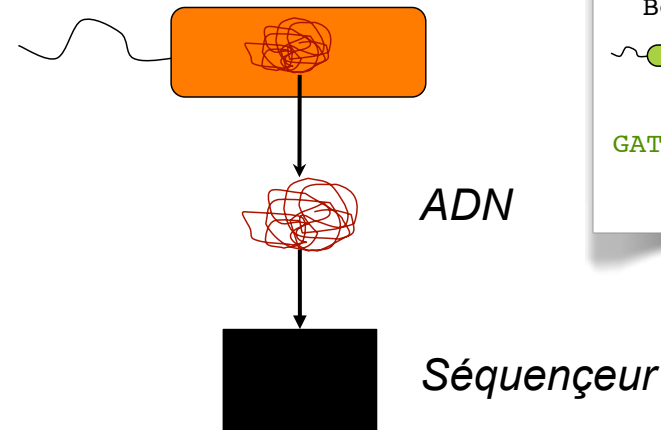
Listeria inoffensive



```
TAATTAATCTAACAATTCGTTACAGATTTCTTTACACACAAGTTATAC
ACAAGTTAACTGGCTGTGGACAACCGTTTTTCACATCTGGACAGTTTTGT
GGATAGATTTTGTAAAGTCCTTGCTATCAGAGTGATTTTCTGATATTATAA
TTCCTGTGCAATAGAAATATAGCTGGGGAAAACTAAAGTTATCCACAATA
CATTTTACTTTGTGGATAATTTTTTAACAGTGTTTGGATAACCTTATCC
ATGGCTTTTTCTATCTGTGGATAACTTTATAGCATCCATTACATTACAT
AAAAAAGGGGGGGTACTAGTGCAATCAATTGAAGACATCTGGCAGGAAAC
ACTGCAAATCGTAAAAAAAATATGAGTAAACCTAGTTACGACACATGGA
TGAAATCAACAACCGCTCATTCACTTGAAGGTAACACGTTTATTATTCA
GCGCCCAATAATTTTGTTCGCGATTGGTTAGAGAAGAGTTACACGCAATT
TATCGCTAACATTTTGAAGAAATAACTGGTCGCTTATTGATGTCGCT
TTATTGATGGCGAGCAGGAAGAAAACCTTTGAATACACTGTGATTAACCA
AATCCAGCGTTAGATGAAGATGGCGTTGAAATTGAAAAACATATGCTTAA
TCCGCGTTATGTTTTGATACTTTTGTCAATTGGTTGAGGAAACAGATTTG
CCCACGCAGCATCACTTGCAGTAGCCGAAGCACCAGCGAAAGCATATAAT
CCACTCTTCATTTATGGAGGAGTTGGCCTCGGTAAACACATTTAATGCA
CGCAGTTGGCCACTATGTTCAACAACATAAAGATAATGCGAAAGTAATGT
ACCTTTCCAGCGAAAAATTCACCAATGAGTTTATTAGCTCTATTGCGTAT
AATAAAACCGAAGAATTCCGTACAAAAATACCGGAATGTGGATGTCTTACT
```

X 3000

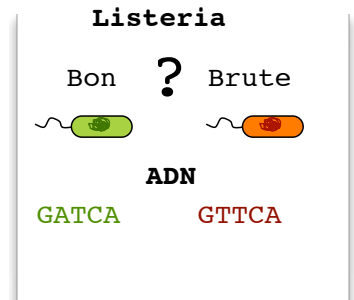
Listeria dangereuse



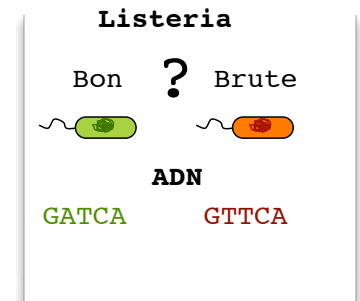
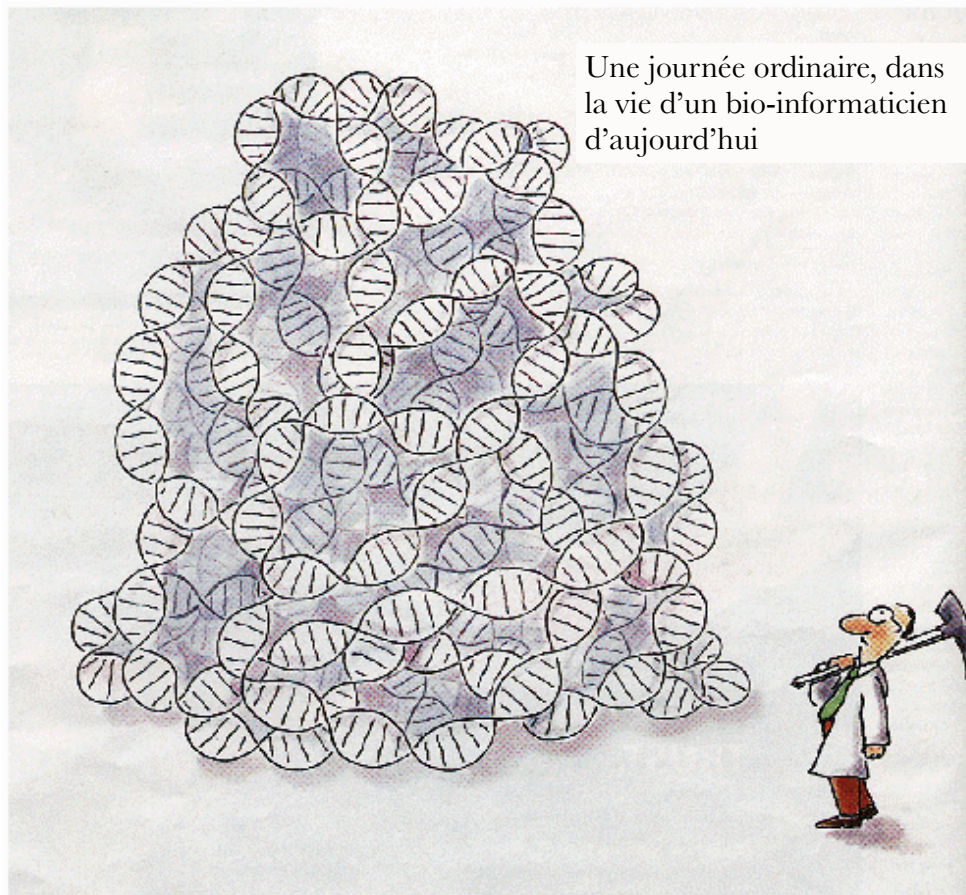
```
TAATTAATTCTAACAATTCGTTACAGATTTCTTTACACACAAGTTATAC
ACAAGTTAACTGGCTGTGGACAACCGTTTTTCACATCTGGACAGTTTTGT
GGATAGATTTTGTAAAGTCCTTGCTATCAGAGTGATTTTCTGATATTATAA
TTCCTGTGCAATAGAAATATAGCTGGGGAAAAATAAAGTTATCCACAATA
CATTTTACTTTGTGGATAATTTTTTAACAGTGTTTGGATAACCTTATCC
ATGGCTTTTTCTATCTGTGGATAACTTTAGAGCATCCATTACATTACAC
AAAAAAGGGGGGTACTAGTGCAATCAATTGAAGACATCTGGCAGGAAAC
ACTGCAAATCGTAAAAAAAATATGAGTAAACCTAGTTACGATACATGGA
TGAAATCAACAACCGCTCATTCACTTGAAGGTAACACGTTTATTATTTC
GCGCCCAATAATTTTGTACGCGATTGGTTAGAGAAGAGCTACACACAATT
TATCGCTAACATTTTGAAGAAATAACTGGTCGCTTATTGATGTCGCT
TTATTGATGGTGAGCAGGAAGAAAACCTTTGAATACACTGTGATTAACCA
AATCCAGCATTAGATGAAGATGGTATTGAAATTGAAAAACATATGCTTAA
TCCGCGTTATGTTTTGATACTTTTGTCAATTGGTTGAGGAAACAGATTTG
CCCACGCAGCATCACTTGCAGTAGCCGAAGCACCAGCGAAAGCGTATAAT
CCGCTTTTCATTTATGGCGGAGTTGGCCTTGGTAAACACATTTAATGCA
CGCAGTTGGCCATTACGTTCAACAACATAAAGATAATGCCAAAGTAATGT
ACCTTTCCAGCGAAAAATTCATAATGAGTTTATTAGCTCCATCCGAGAT
AATAAAACCGAAGAATTCGTACAAAAATACCGGAATGTTGACGTCTTAC
```

X 3000

- SAM -



Au travail !



C'est maintenant que commence le travail du bio-informaticien !

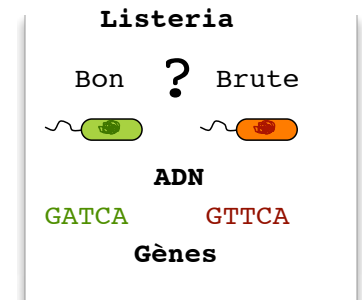
Rechercher du sens

l'origine des espèces je me propose de passer brièvement en revue les progrès de l'opinion relativement à l'origine des espèces jusqu'à tout récemment la plupart des naturalistes croyaient que les espèces sont des productions immuables créées séparément de nombreux savants ont habilement soutenu cette hypothèse quelques autres au contraire ont admis que les espèces se prouvent des modifications et que les formes actuelles descendent de formes préexistantes par voie de génération régulière

L'origine des espèces

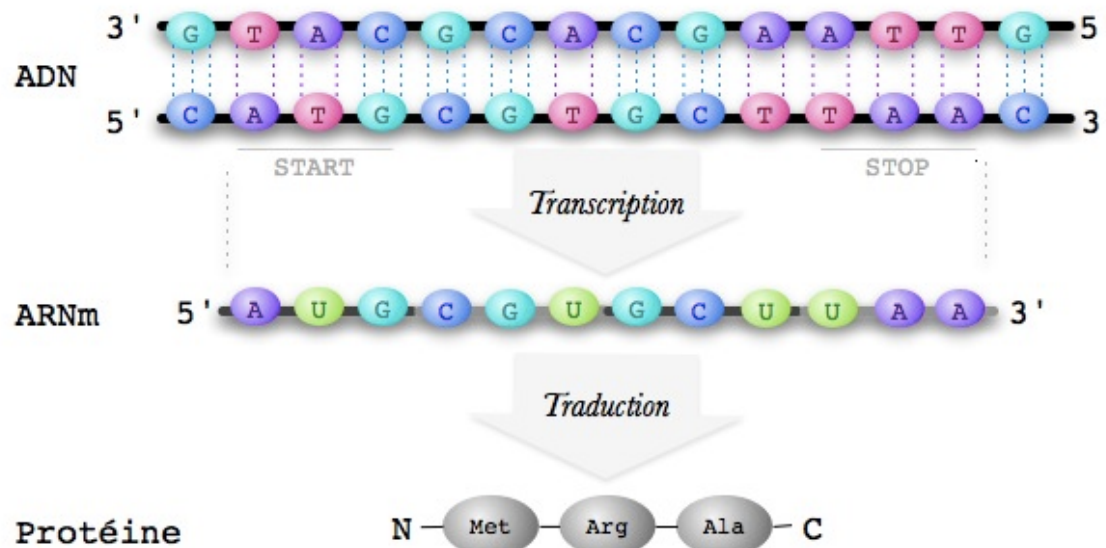
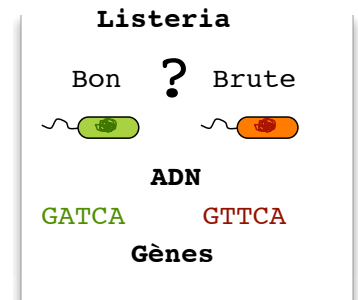
Je me propose de passer brièvement en revue les progrès de l'opinion relativement à l'origine des espèces. Jusque tout récemment, la plupart des naturalistes croyaient que les espèces sont des productions immuables créées séparément. De nombreux savants ont habilement soutenu cette hypothèse. Quelques autres, au contraire, ont admis que les espèces éprouvent des modifications et que les formes actuelles descendent de formes préexistantes par voie de génération régulière. ...

Charles Darwin



Rechercher les gènes

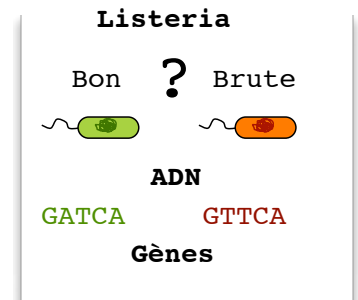
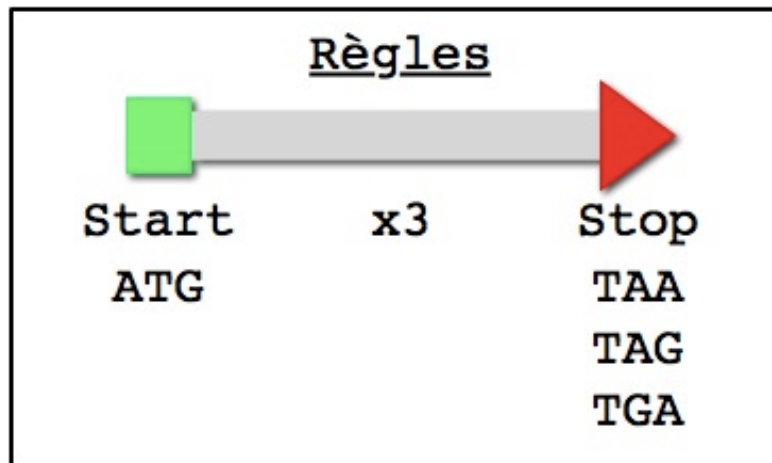
■ Sur l'ADN, le sens est donné par les **gènes**.
Il servent à fabriquer les **outils** moléculaires et sont associés à une **fonction** dans la bactérie



- SaM -

Ressource web : les gènes

A quoi ça ressemble un gène ?



Facile !



Le bio-informaticien écrit un programme qui va suivre ces règles pour trouver les gènes

Recherche de gènes ... à la main !

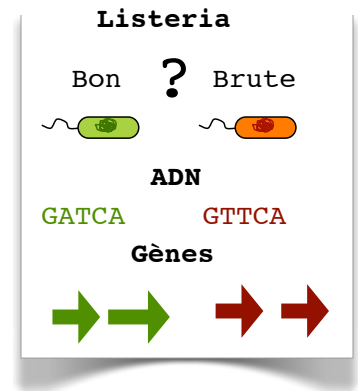
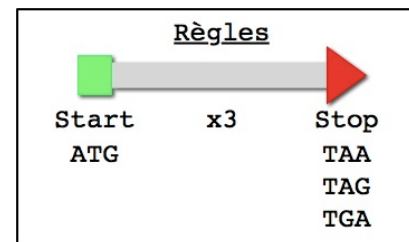
Prenons une partie de notre séquence de *Listeria* et appliquons les règles

A G T A T G T G G A A A C G T T G A G A G C G A T C C G A T G A A A

G A T G T G T T T G A G T C A C A G C G C A A C A A C A C G G A A A T

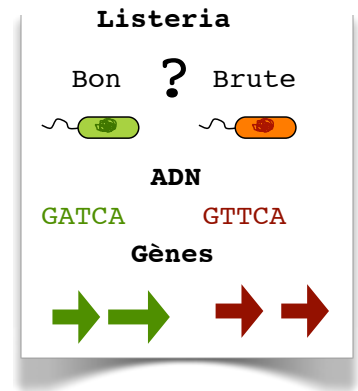
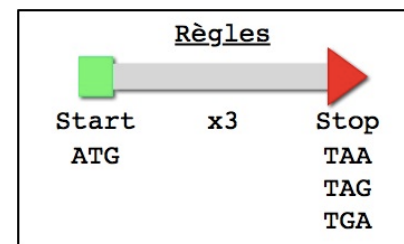
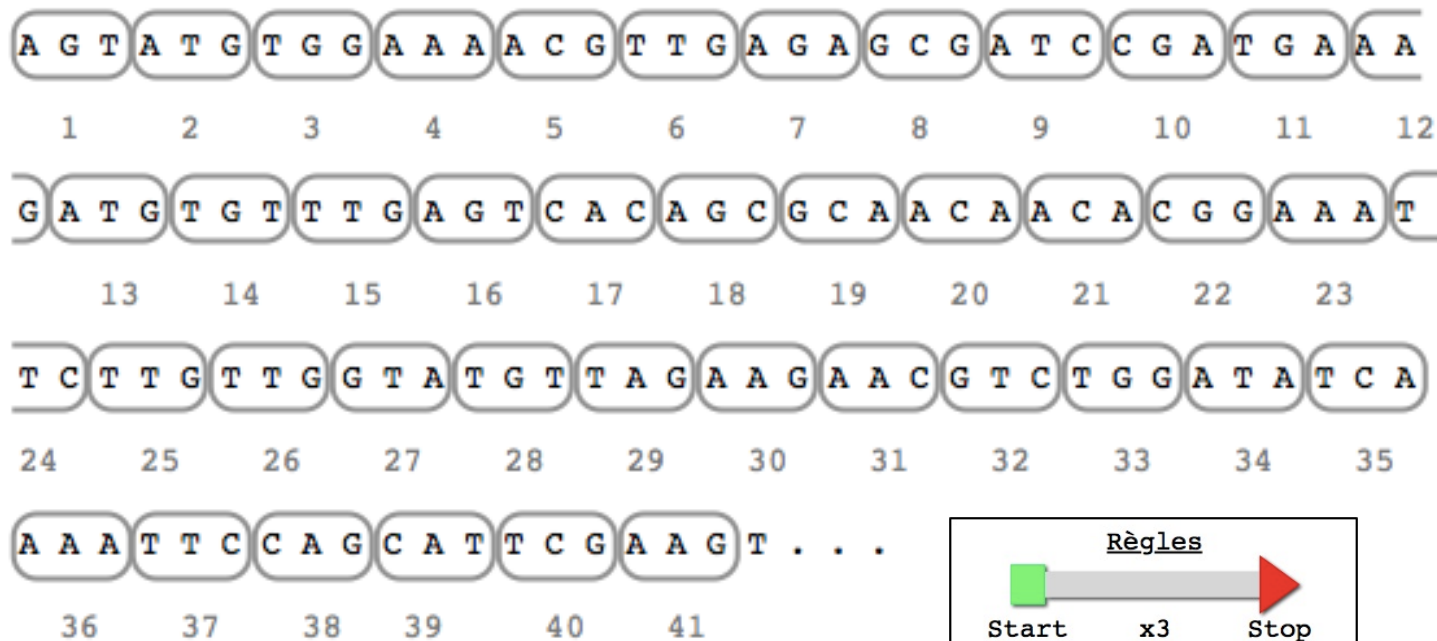
T C T T G T T G G T A T G T T A G A A G A A C G T C T G G A T A T C A

A A A T T C C A G C A T T C G A A G T . . .



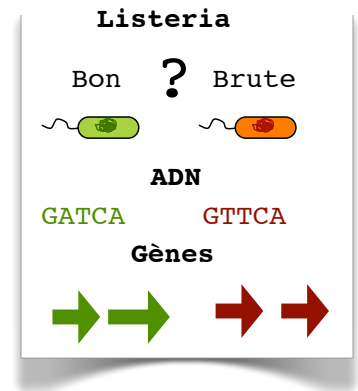
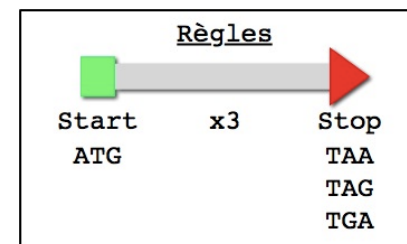
Recherche de gènes ... à la main !

1 - Diviser la séquence en triplets



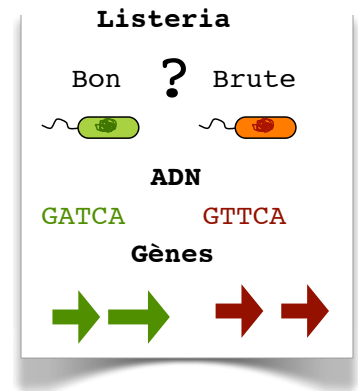
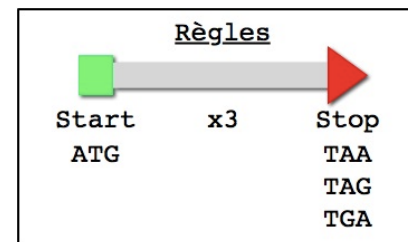
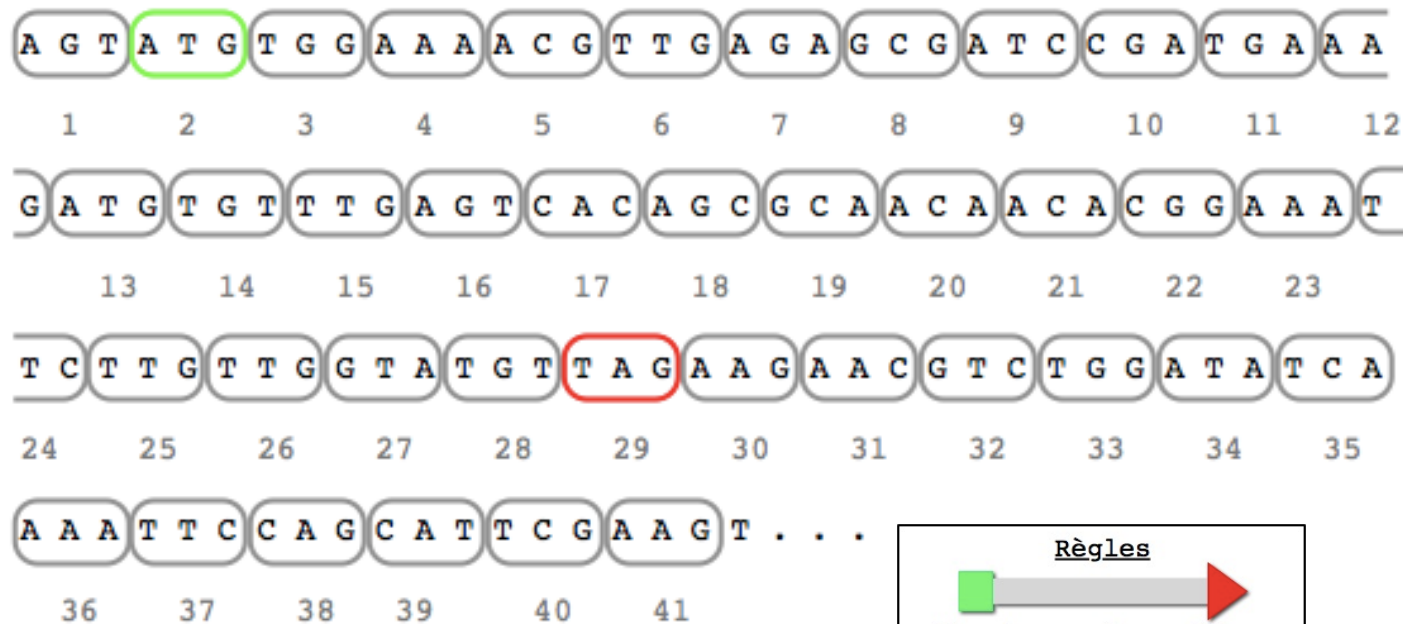
Recherche de gènes ... à la main !

2 - Rechercher les «START» et «STOP»



Recherche de gènes ... à la main !

2 - Rechercher les «START» et «STOP»



Recherche de gènes ... à la main !

3 - Repérer les gènes potentiels

A G T A T G T G G A A A C G T T G A G A G C G A T C C G A T G A A A

1 2 3 4 5 6 7 8 9 10 11 12

G A T G T G T T T G A G T C A C A G C G C A A C A A C A C G G A A A T

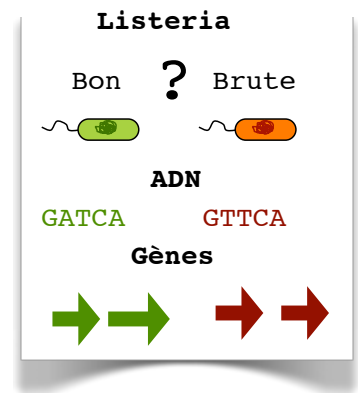
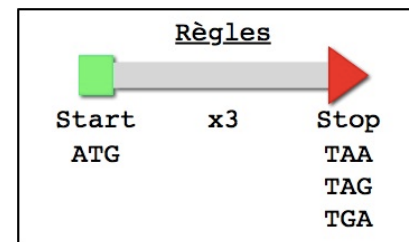
13 14 15 16 17 18 19 20 21 22 23

T C T T G T T G G T A T G T T A G A A G A A C G T C T G G A T A T C A

24 25 26 27 28 29 30 31 32 33 34 35

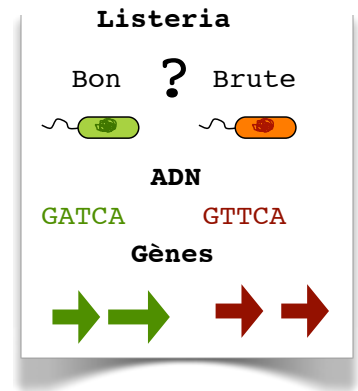
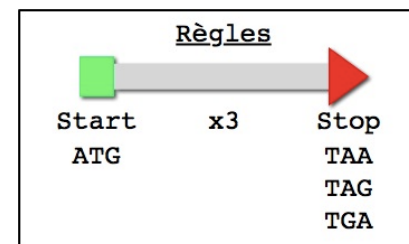
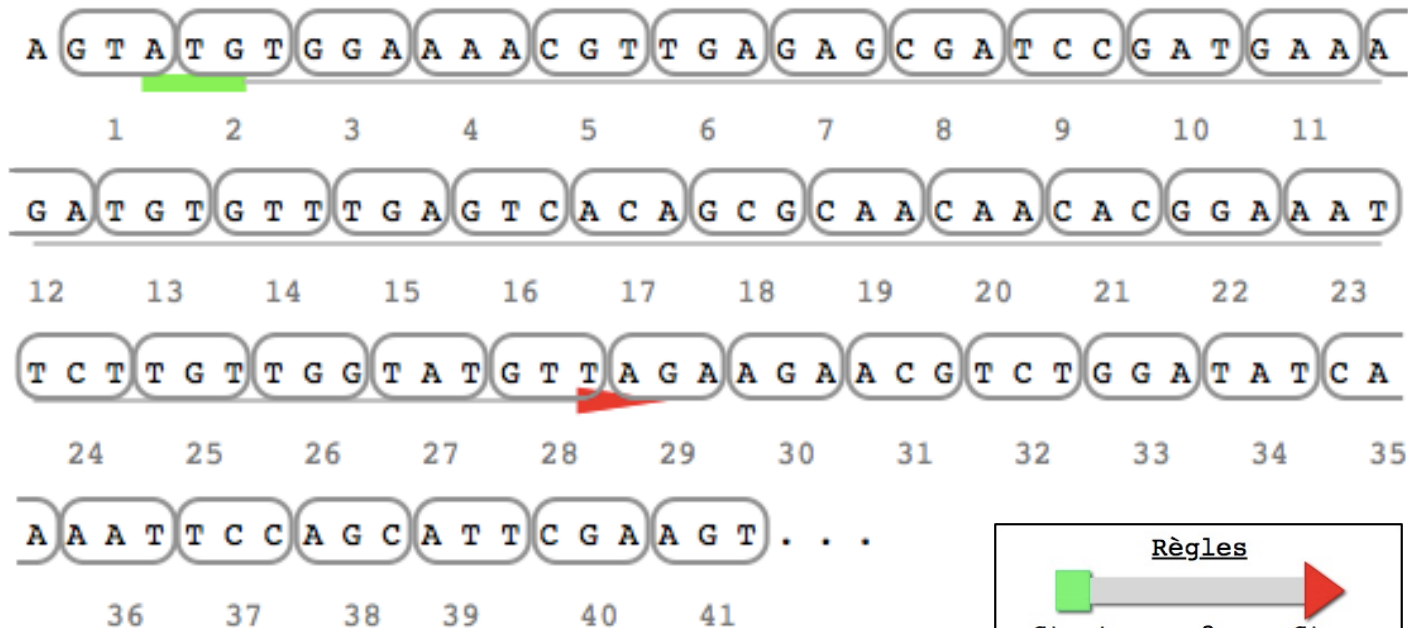
A A A T T C C A G C A T T C G A A G T . . .

36 37 38 39 40 41



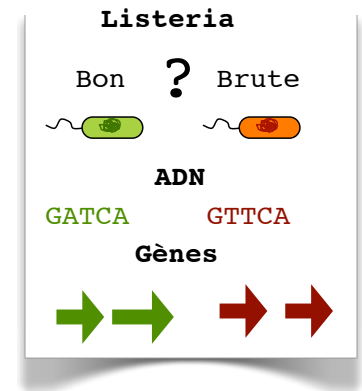
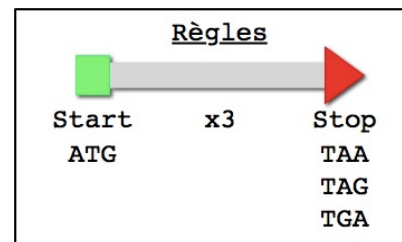
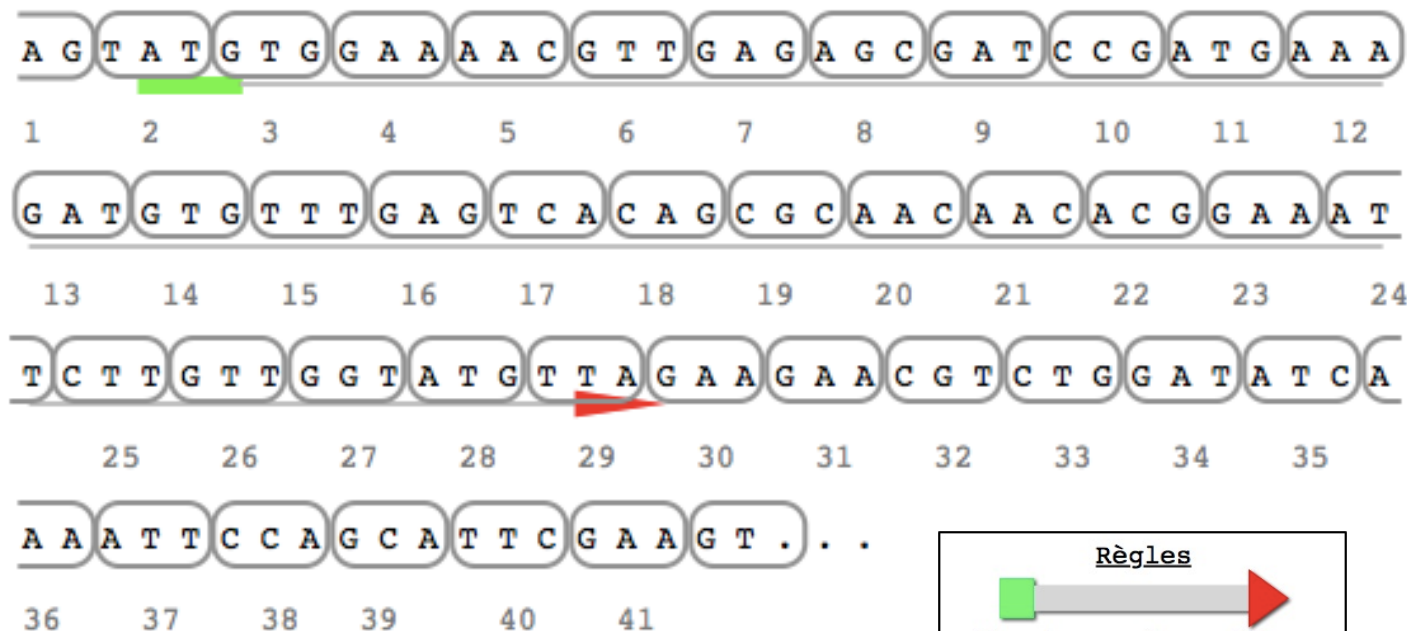
Recherche de gènes ... à la main !

4 - Décaler le cadre de lecture de 1 base



Recherche de gènes ... à la main !

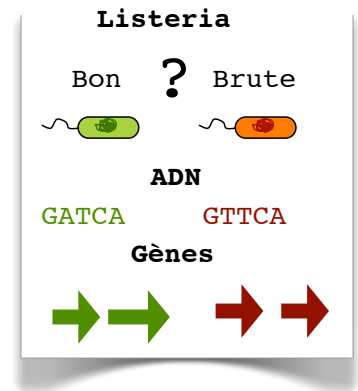
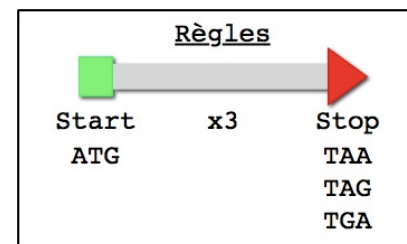
5 - Et recommencer !



... sur ce cadre, pas de gènes

Recherche de gènes ... à la main !

6 - Décaler le cadre de lecture de 1 base, recommencer ...



Recherche de gènes ... à la main !

7 - Repérer les gènes

A G T A T G T G G A A A A C G T T G A G A G C G A T C C G A T G A A A

1 2 3 4 5 6 7 8 9 10 11 12

G A T G T G T T T G A G T C A C A G C G C A A C A A C A C G G A A A T

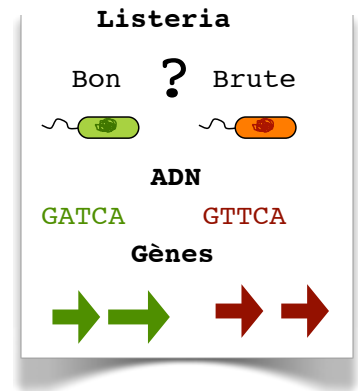
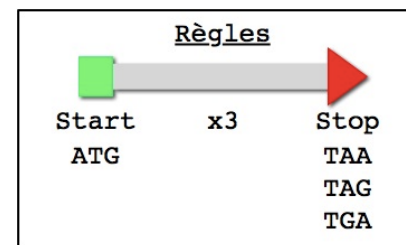
13 14 15 16 17 18 19 20 21 22 23 24

T C T T G T T G G T A T G T T A G A A G A A C G T C T G G A T A T C A

25 26 27 28 29 30 31 32 33 34 35

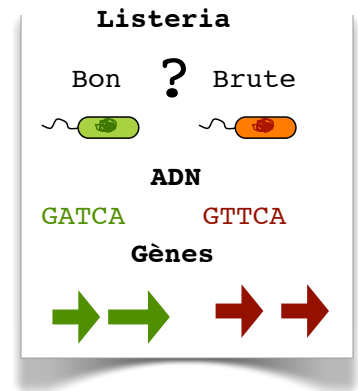
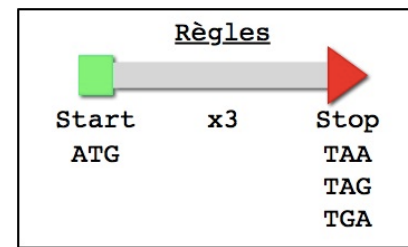
A A A T T C C A G C A T T C G A A G T . . .

36 37 38 39 40 41



Recherche de gènes ... à la main !

8 - Décaler le cadre de lecture de 1 base, et s'apercevoir qu'on revient au départ



Recherche de gènes ... à la main !

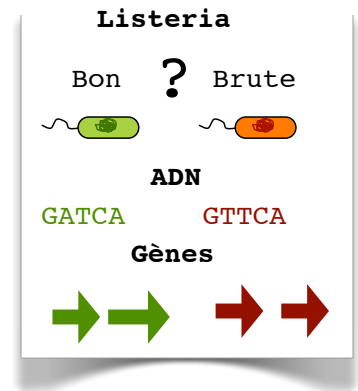
9 - Ne pas oublier qu'une séquence d'ADN a 2 brins !

```
A G T A T G T G G A A A C G T T G A G A G C G A T C C G A T G A A A
T C A T A C A C C T T T T G C A A C T C T C G C T A G G C T A C T T T
```

```
G A T G T G T T T G A G T C A C A G C G C A A C A A C A C G G A A A T
C T A C A C A A A C T C A G T G T C G C G T T G T T A T G C C T T T A
```

```
T C T T G T T G G T A T G T T A G A A G A A C G T C T G G A T A T C A
A G A A C A A C C A T A C A A T C T T C T T G C A G A C C T A T A G T
```

```
A A A T T C C A G C A T T C G A A G T . . .
T T T A A G G T C G T A A G C T T C A . . .
```



Plus que 4 000 000 de bases ...

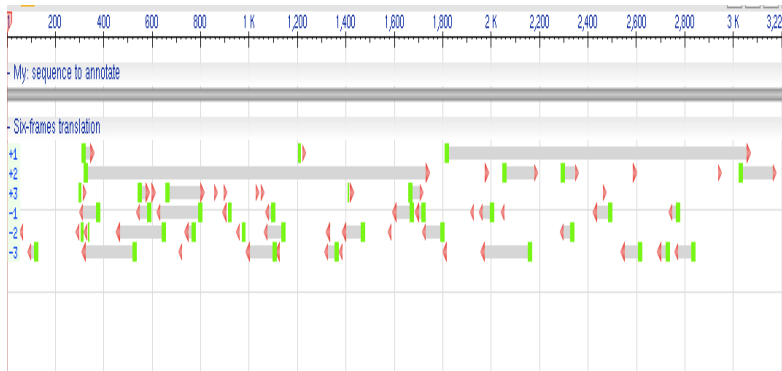
... et 2 brins ! Quel boulot ! Il est temps d'utiliser un matériel de pro !



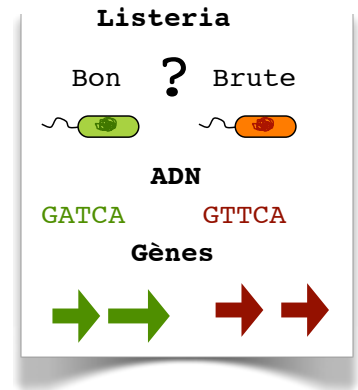
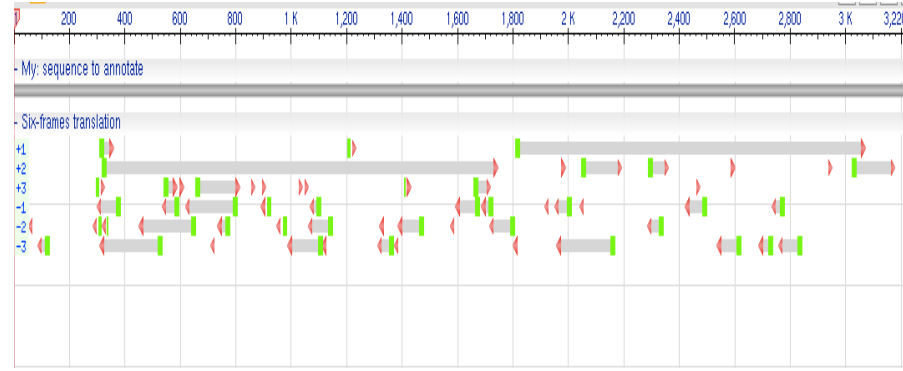
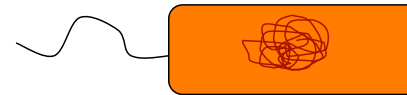
Ressource web : rechercher les gènes

Recherche de gènes

Listeria inoffensive



Listeria dangereuse

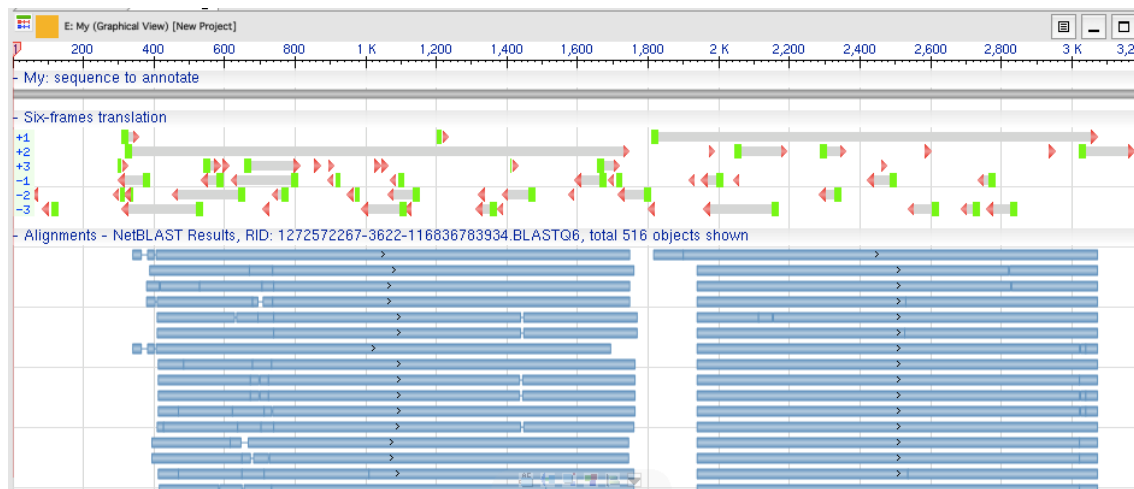


Tous ces gènes ! Mais comment savoir lesquels sont corrects ?

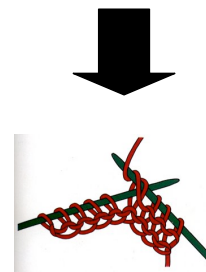


Le bio-informaticien recherche s'il existe déjà des séquences **similaires** à l'un de ces gènes

Recherche de similarités

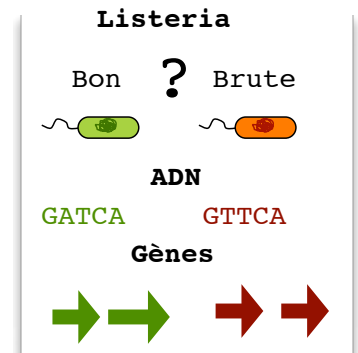


dnaA



dnaN

- SaM -

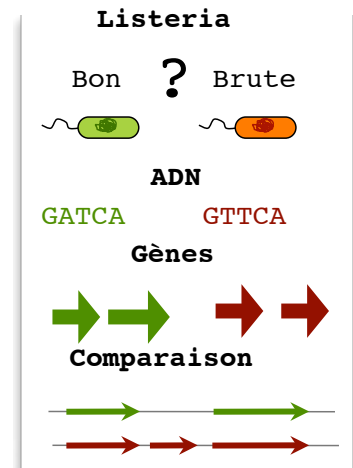
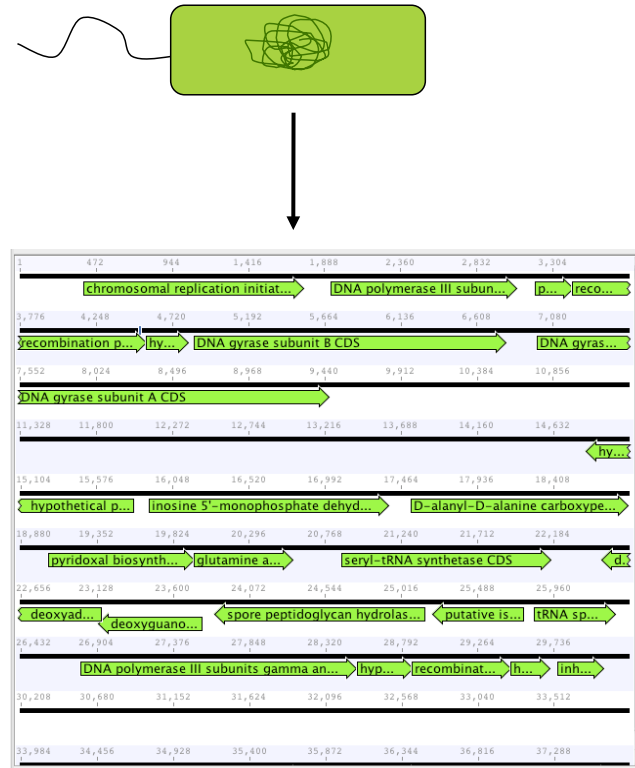


On en connaît beaucoup plus sur nos *Listerias* !

Listeria dangereuse



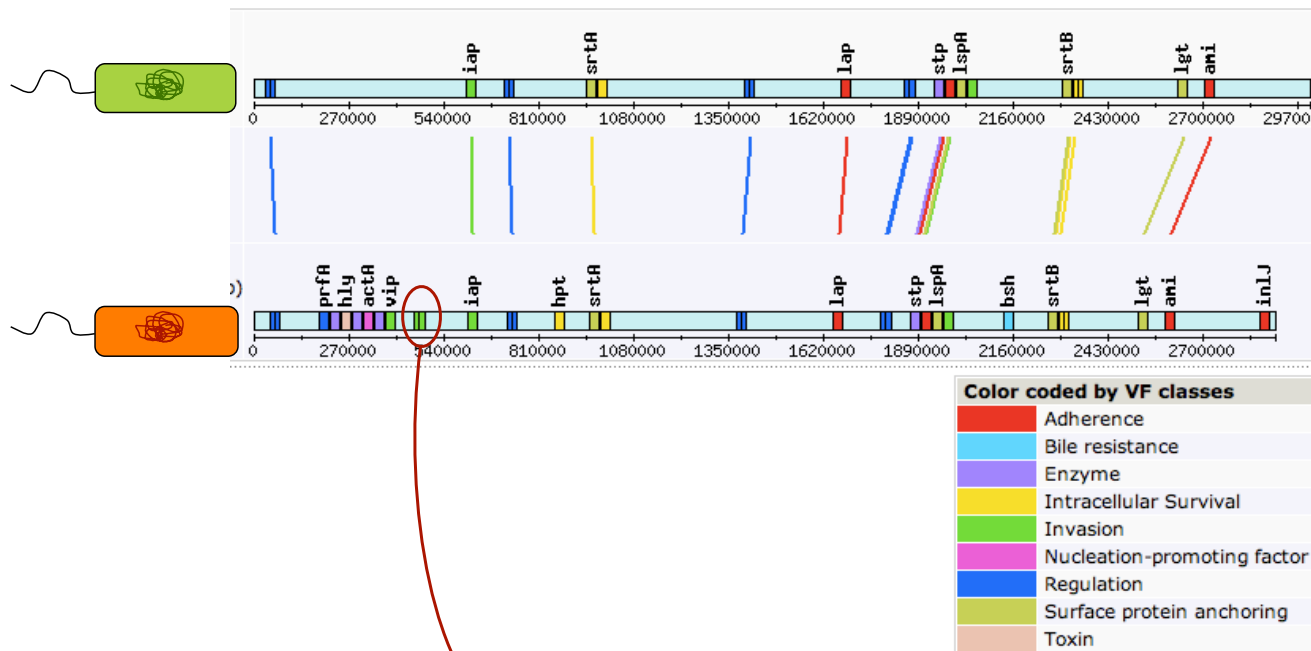
Listeria inoffensive



Cherchons les gènes “dangereux”, ceux qui existent ici mais pas la

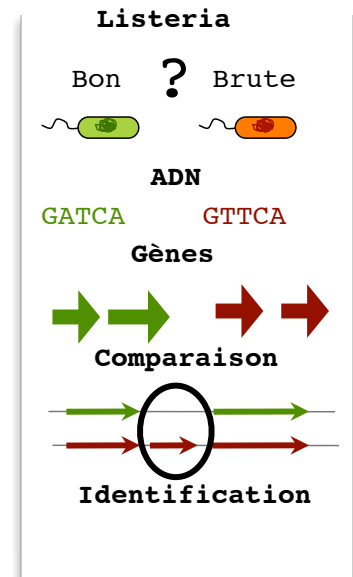
Le bon, la brute ... et les truands ?

 Le bio-informaticien utilise un programme qui compare les 2 *Listerias* et trouve les gènes spécifiques

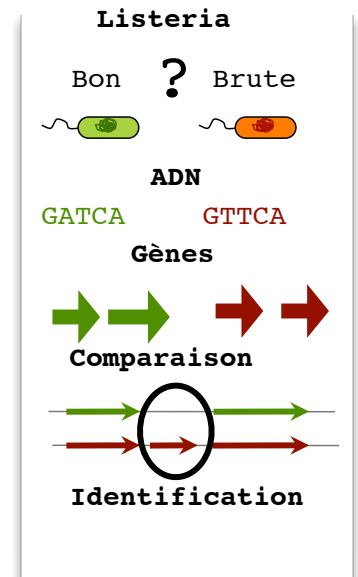
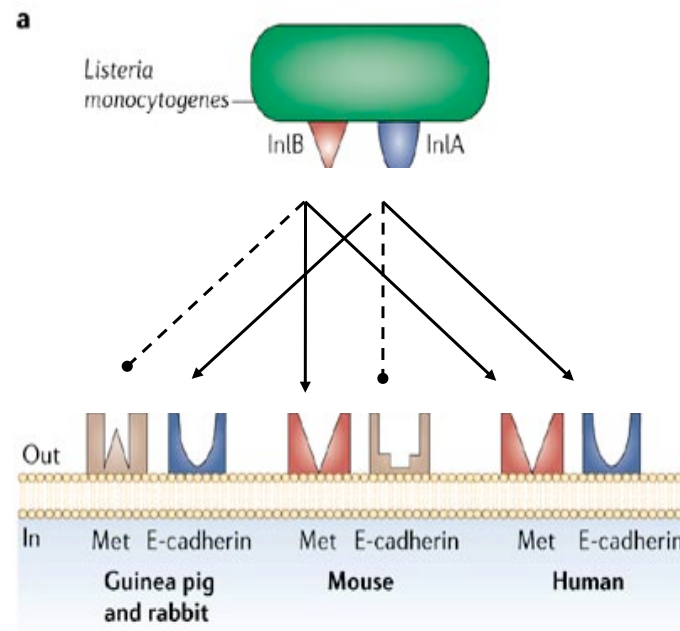


Les internalines

- SAM -



Les internalines

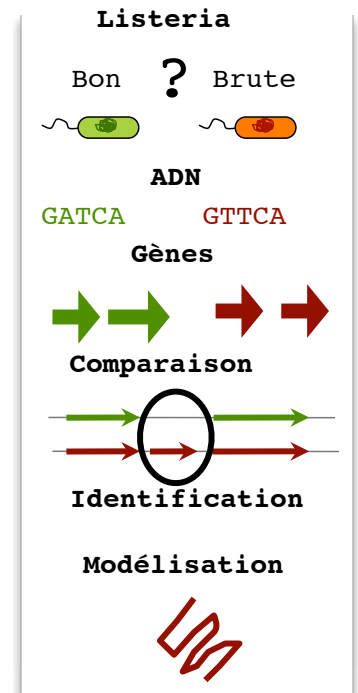
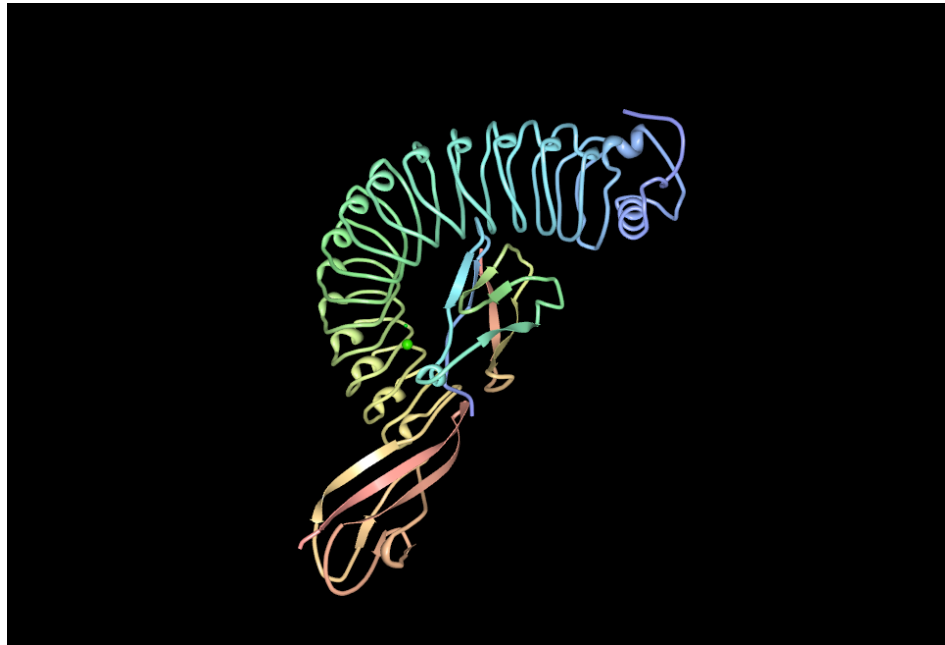


Nous avons donc trouvé 2 gènes, les internalines A et B qui permettraient d'expliquer pourquoi une des *Listerias* est dangereuse et pas l'autre.

Structure en 3D d'une internaline



Le bio-informaticien pourra modéliser la structure dans l'espace de l'internaline afin de comprendre comment elle interagit avec son récepteur.



Les bio-informaticiens

- Écrivent/utilisent des programmes informatiques pour répondre à des questions sur les génomes
 - Des **bases de données** pour stocker, organiser toutes ces informations
 - Des programmes pour comprendre **l'organisation des génomes** (statistiques, modèles mathématiques)
 - Des programmes pour comprendre **l'évolution des espèces**
 - Des programmes pour prédire la **fonction des gènes**
 - Des programmes pour prédire la **structure des protéines**
 -
- Métier exigeant
 - Formation universitaire **longue** (minimum master)
 - Bon niveau en biologie **ET** en informatique / statistique

Pourquoi j'ai choisi ce métier

- Parce que j'aime la **biologie**, mais pas avec une éprouvette
- Parce que l'informatique, je trouve ça **fun**
- Parce que les questions qui m'intéressent concernent l'évolution des génomes
- Parce que c'est un métier avec une **double compétence** ... donc 2x plus de chance de trouver du travail
- Parce que l'ADN est **universel** et l'on peut travailler un jour sur les plantes et le lendemain sur une bactérie
- Parce qu'au rythme où vont les découvertes, je ne m'ennuierai jamais ; c'est un métier de **passion** !

Merci !
Quel bon moment
nous venons de
passer !

