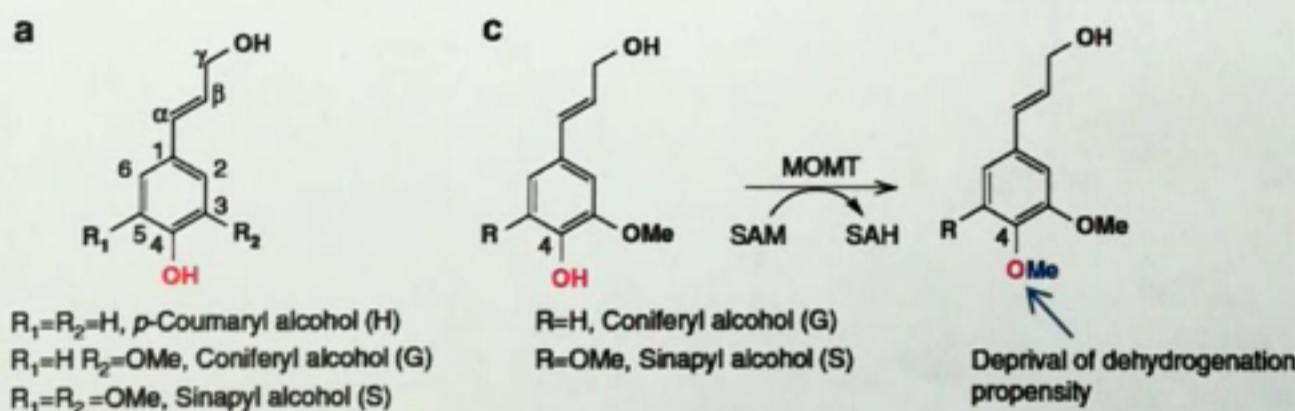


**Examen 12 Avril 2018 Biologie et Biotechnologie des Plantes.**  
**B4 salle A1 10H00-11h30**

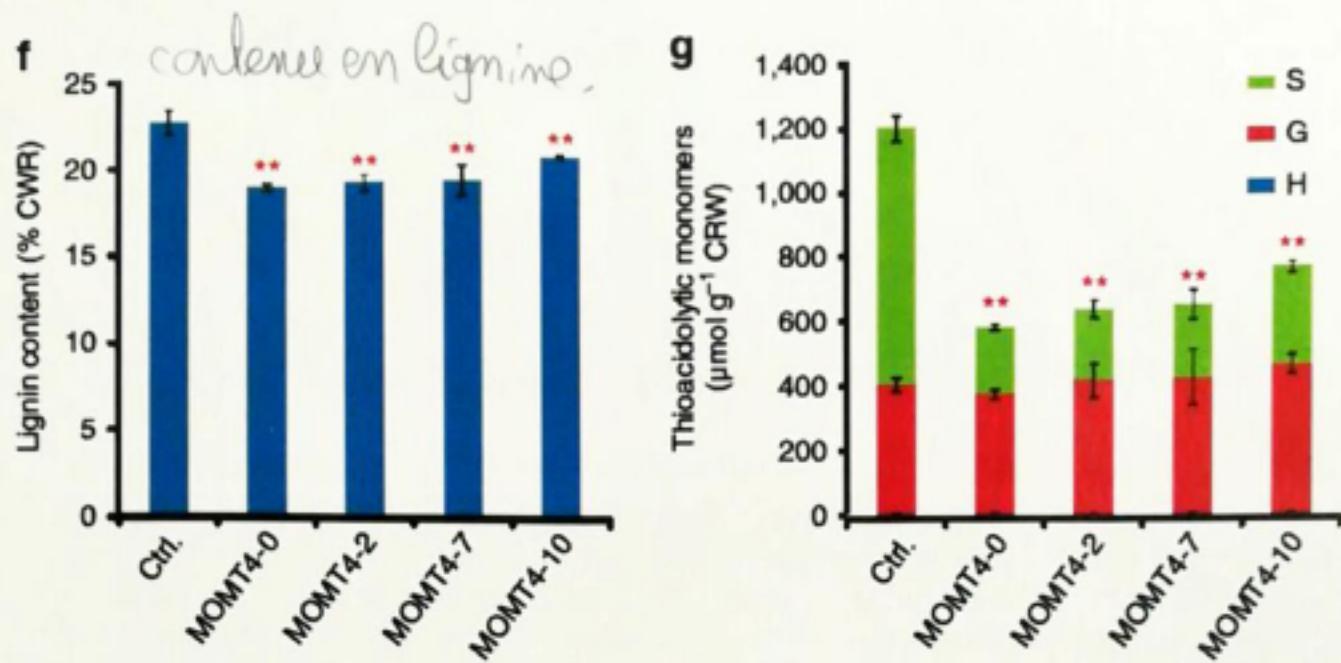
**Sujet de Mr Gomès (1h30) : Biosynthèse des lignines (Documents de cours et de TD autorisés)**

La production de **biocarburants** de seconde génération à partir de **cellulose** issue de la **biomasse végétale ligneuse butte actuellement sur la présence de forte quantité de lignine dans le bois**. Le degré de réticulation de la lignine est un facteur clé de l'extractibilité et de la thermolyse acide de la cellulose en sucres simples, fermentable en bioéthanol. Dans un article publié en 2016 dans la revue Nature Communications, Cai et collaborateurs étudient l'impact de l'expression dans des **peupliers transgéniques** d'une **monolignol 4-O-methyltransférase** modifiée (isogène *MOMT4*) sur la **digestibilité et le rendement** en éthanol obtenu à **partir du bois**. Les figures suivantes sont extraites de cet article.

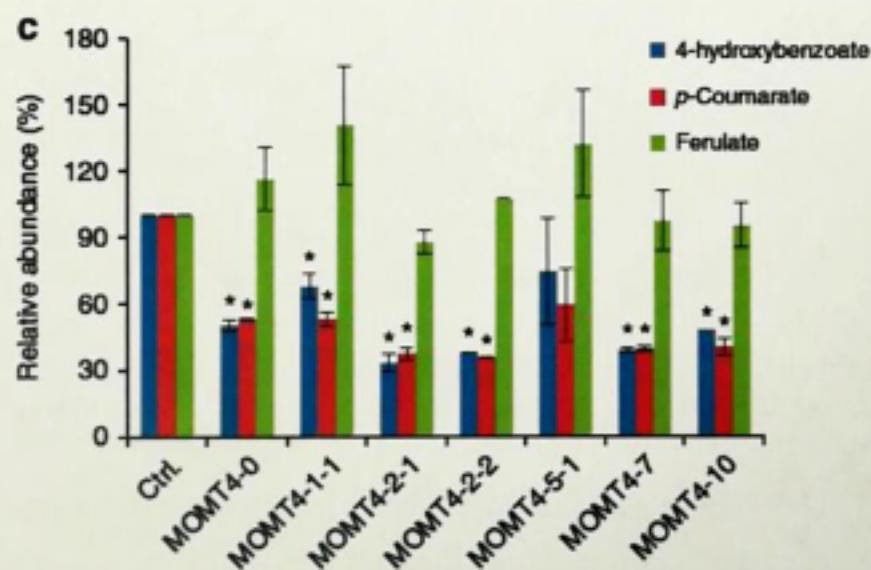
**Question :** Pour chaque figure, que vous analyserez, vous expliquerez, en 10-15 lignes maximum, les principales informations et messages que l'on peut en extraire. Puis, vous rédigerez une synthèse (2 pages maximum) expliquant comment les résultats obtenus pourraient faciliter dans un futur proche la production de bioéthanol à partir de bois.



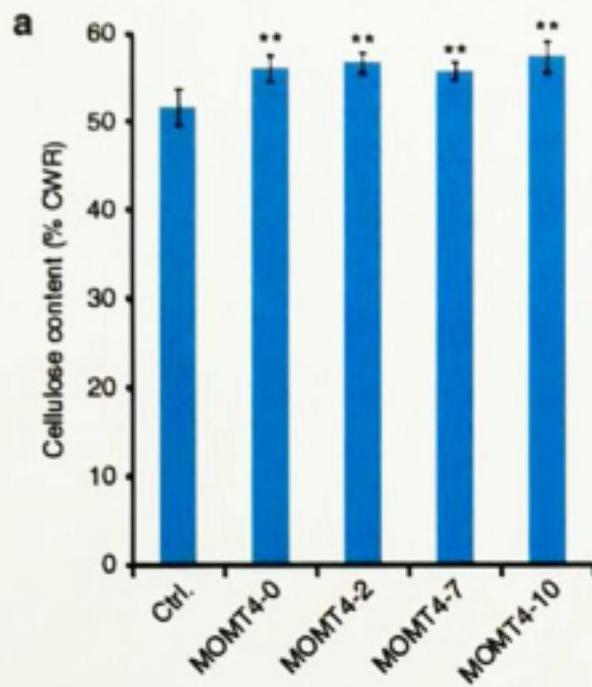
**Figure 1.** Illustration of MOMT4-mediated modification of monolignol structure.



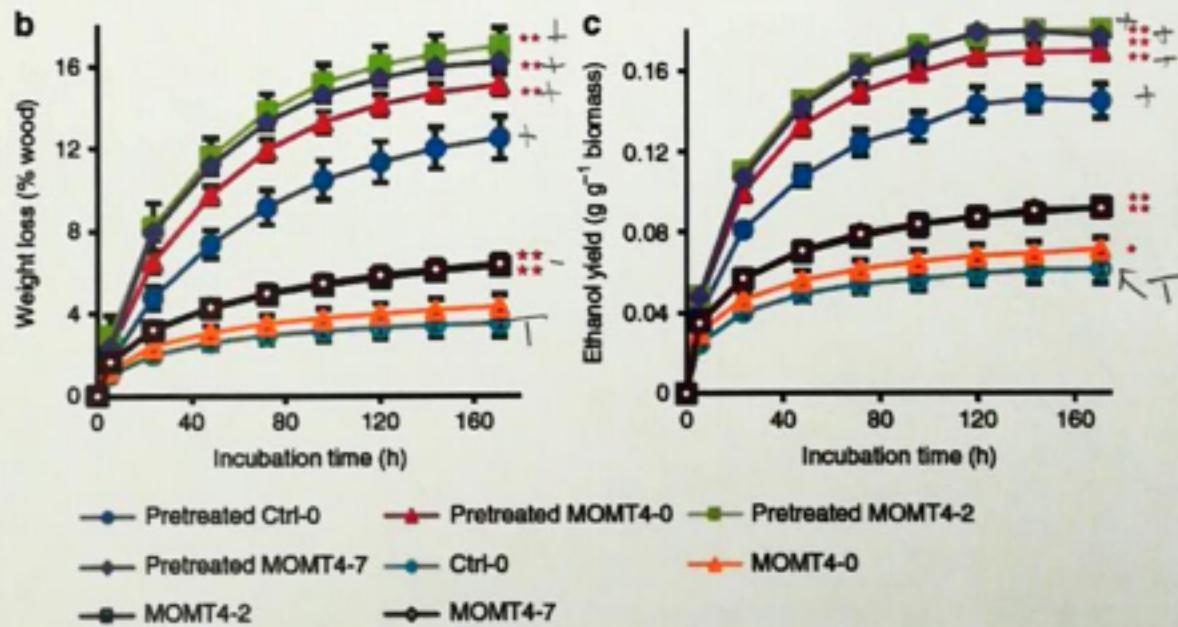
**Figure 2.** Effect of ectopic expression of MOMT4 on lignin content and composition in transgenic poplar. (f): total lignin content in the cell wall of control (Ctrl.) and MOMT4 transgenic lines poplar stems. (g): lignin monomers released by thioacidolysis of (Ctrl.) and MOMT4 transgenic lines poplar stems; S, syringyl; G, guaiacyl; H, p-hydroxyphenyl. Data are means $\pm$ SE of 3 biological replicates, \*\* indicates significant difference at P<0.01.



**Figure 3.** Alteration of the accumulated bound phenolics in stem cell walls of MOMT4 transgenic poplar. Ctrl: control. Data are means $\pm$ SE of 3 biological replicates, \* indicates significant difference at P<0.05.



**Figure 4.** Effect of *MOMT4* expression on cellulose content in control (Ctrl.) and transgenic poplar stems. Data are means $\pm$ SE of 3 biological replicates, \*\* indicates significant difference at  $P<0.01$ .



**Figure 5.** Enzymatic hydrolysis and bioconversion of control and *MOMT4* wood samples to ethanol. (b) digestive weight loss of wood samples treated with  $2\text{mg.ml}^{-1}$  of cellulase 72 h at  $37^\circ\text{C}$ . (c) ethanol yield during simultaneous cellulose digestion (saccharification) and sugar fermentation. Pretreated samples have been submitted to partial thermolysis before enzymatic digestion. Data are means $\pm$ SE of 3 biological replicates, \* indicates significant difference at  $P<0.05$ , \*\* indicates significant difference at  $P<0.01$ .

Examen Perl & Bases de données  
Le 2 Mai 2018, de 14h à 15h30  
Les documents ne sont pas autorisés  
Traiter chacune des deux parties sur une copie séparée

---

Perl

---

**EXERCICE 1 : PROGRAMME MYSTÈRE**

Le programme ci-dessous nécessite un fichier de données pour fonctionner. Nous appellerons ce fichier "ficenv.dat".

```
#!/usr/bin/perl
use strict;
use warnings;

#ouverture du fichier
...

my $nbve =0;
my @ve;
my @chems;
my %V_ENV;

while (<FENV>) {
    chomp;
    if (/([A-Z_]+)=(\d+\w+)/){
        $nbve++;
        my ($ve,$chems) = ($1,$2) ;
        $V_ENV{$ve} = $chems;
    }
}
print "nbve: $nbve \n";

for my $el (sort %V_ENV) {
    print $el," => ",$V_ENV{$el},"\n";
}

close(FENV);
```

1. Rajoutez la ligne nécessaire en début de programme pour ouvrir le fichier "ficenv.dat".
2. Après avoir bien lu le programme, écrivez 2 lignes qui pourraient se trouver dans "ficenv.dat" et qui valident le test du if.  
Quel est alors l'affichage produit en fin de programme pour ces deux lignes?
3. Décrivez ce que fait ce script d'une manière générale?

---

## Bases de données

---

Soit la base de données UNIVERSITÉ qui permet de gérer les données de ses activités. Une partie de cette base est composée des tables suivantes:

- Etudiant(NumEtu, Nom, Prénom, AnnéeNaissance)
- Enseignant(NumEns, Nom, Prénom, AnnéeNaissance)
- UE(NumUE, NomUE, Niveau) -
- Inscription(NumEtu, NumUE, Note) -
- AssuréePar(NumUE, NumEns)

Les attributs soulignés représentent la clé primaire de chacune des tables.

### EXERCICE 1 : SQL

Exprimer les requêtes suivantes en SQL

1. Afficher le nom des étudiants nés en 1996.
2. Afficher le numéro des enseignants qui ont un homonyme parmi les étudiants (même nom et même prénom).
3. Pour chaque UE, afficher son numéro ainsi que le nombre d'étudiants qui y sont inscrits.
4. Pour chaque étudiant dont la moyenne de toutes ses notes est inférieure à 10, afficher son numéro, son nom, et son prénom.
5. Pour l'étudiant(e) dont le numéro est 125, afficher le numéro des UE's auxquelles il (elle) est inscrit(e) mais où la note n'a pas été saisie.
6. Afficher la meilleure note à l'UE 183 ainsi que le nombre d'étudiants qui l'ont eue.

### EXERCICE 2 : CLÉS ET DÉPENDANCES

1. D'après la structure de cette base, il n'est pas possible d'avoir plusieurs enseignants pour la même UE. Expliquer pourquoi c'est le cas.
2. Comment peut on procéder pour qu'on ne puisse pas s'inscrire à une UE qui n'existe pas ?
3. Supposons que dans la table Enseignant, on a la dépendance fonctionnelle  
Nom, Prénom → AnnéeNaissance  
Cette table est-elle en 2FN ? Est-elle en 3FN ?

	<b>ANNEE UNIVERSITAIRE 2017 / 2018</b> <b>EXAMEN NGS</b>	<b>Collège</b> <b>Sciences et</b> <b>technologies</b>
	<b>PARCOURS / ETAPE :</b> Bioinformatique <b>Code UE :</b> <b>4TBI804U</b> <b>Epreuve : Bioinformatique</b> <b>Date :</b> 11/04/2018 <b>Heure :</b> 1h30 <b>Durée</b> : 1h30 <b>Documents :</b> non autorisés <b>Epreuve de M/Mme :</b> P Sirand Pugnet et P Thébault	

**Les documents ne sont pas autorisés.**

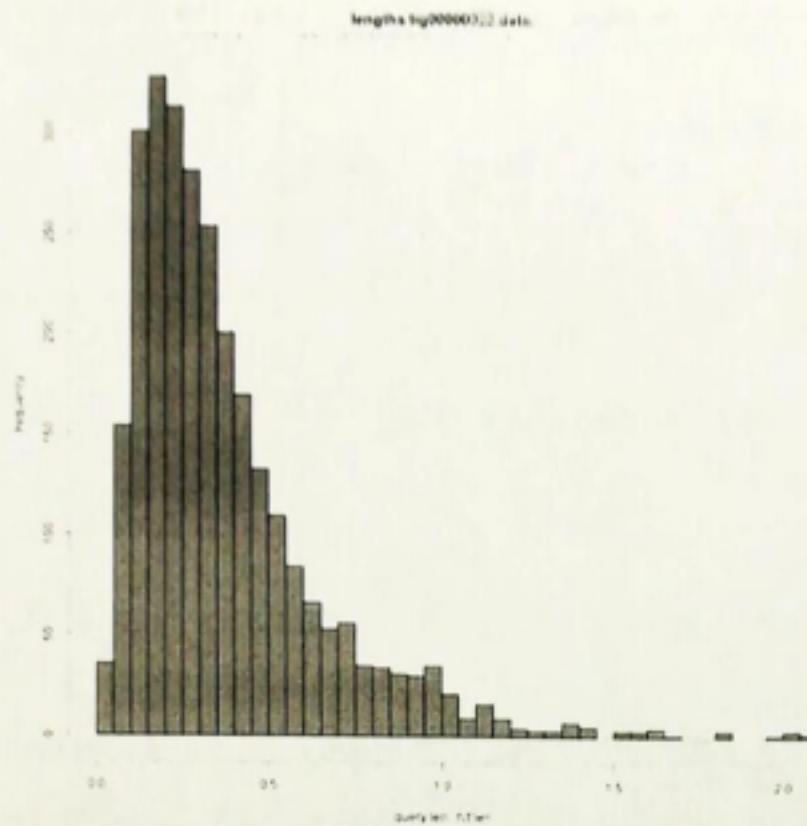
**Questions de cours : (12 points)**

- 1) La technologie de séquençage développée par Pacific Biosciences (PacBio) est actuellement en plein essor.
  - a. Expliquer le principe de cette technologie en vous appuyant sur des schémas (2 points)
  - b. Quels sont les avantages et inconvénients de cette technologie en comparaison avec les autres technologies NGS ? (2 points)
- 2) L'apparition des technologies NGS a bouleversé le domaine du séquençage depuis 10 ans. Pensez-vous que la technologie historique de Sanger est appelée à disparaître rapidement ? Justifiez votre réponse. (2 points)
- 3) Un séquençage a été réalisé sur une nouvelle souche *Bsu\_007* de la bactérie *Bacillus subtilis* à l'aide de la technologie PacBio. Après assemblage des reads avec le programme Canu, plusieurs contigs ont été obtenus. Les données présentées ci-dessous correspondent au contig principal de 3 Mb obtenu lors de cet assemblage. Les CDS présents sur ce contig ont été prédits avec le logiciel PROKKA et les protéines correspondantes ont été comparées par blastp aux protéines présentes dans la base de données Uniprot. Cette base de données contient notamment la séquence des protéines prédites à partir de la souche de référence de *B. subtilis*. Pour chaque protéine prédite de la souche *Bsu\_007*, le rapport entre la longueur de celle-ci et la longueur du meilleur hit identifié par blastp est calculé. Les résultats sont montrés figure 1.

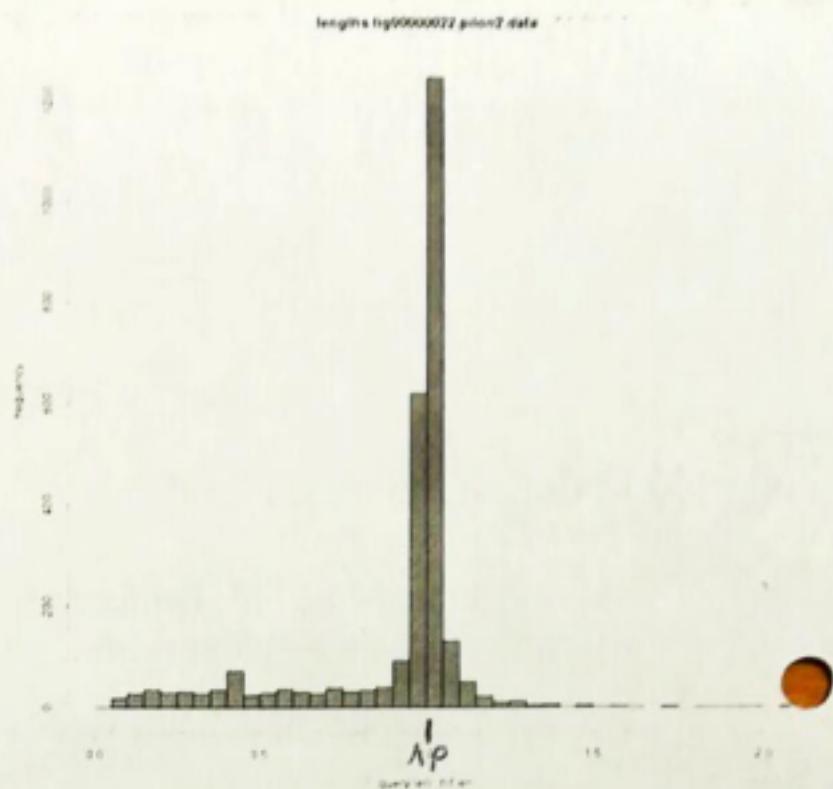
Sur 3233 protéines prédites, 2635 présentent une longueur de moins de 90% de la longueur du meilleur hit.

Analysez ce résultat. Quelles hypothèses pouvez-vous formuler pour expliquer cette distribution ? (2 points)

L'assemblage est retravaillé avec le logiciel de polishing Pilon afin de corriger un certain nombre d'erreurs sur la séquence de l'assemblage. La même analyse que précédemment est refaite et les résultats obtenus sont représentés figure 2.



*Figure 1. Histogramme du rapport de longueur des protéines prédictes de *Bsu\_007* et du meilleur hit identifié dans Uniprot par blastp.*



*Figure 2. Histogramme du rapport de longueur des protéines prédictes de *Bsu\_007* à partir de l'assemblage corrigé par Pilon et du meilleur hit identifié dans Uniprot par blastp.*

Sur 2842 protéines prédictes, 576 présentent une longueur de moins de 90% de la longueur du meilleur hit.

Analysez ce résultat. Que suggère-t-il quant au résultat de la figure 1 ? Justifiez votre réponse. (2 points)

Quelles recommandations pourriez-vous formuler concernant la détection des pseudogènes chez la souche *Bsu\_007* ? (2 points)