PROM ANALYZER

MERCI POUR VOTRE TÉLÉCHARGEMENT !

Bienvenue sur **PromAnalyzer** !

Cette application a pour but de vous **faciliter le traitement** de vos données *PlantPAN 3.0* obtenues avec l'outil *Promoter Analysis.*

Codé en python, cet exécutable peut être **lancé après téléchargement** sans que vous ayez besoin de télécharger les packages nécessaires à son fonctionnement.

Vous trouverez ci-dessous **un manuel d'aide** récapitulant les grandes étapes du programme et les fichiers obtenus.

Sur ce, bon courage ! :-)

SOMMAIRE

1 | DONNÉES À UTILISÉES

2 | PLACER VOS DONNÉES DANS LE BON DOSSIER

3 | LANCEMENT DU PROGRAMME

4 | CHOIX DU TRAITEMENT

5 | SAISIE PAR L'UTILISATEUR

6 | PENDANT ET À LA FIN DU TRAITEMENT

7 | FICHIERS RÉSULTATS

Cette application a été développé dans le cadre d'un stage par deux étudiants L3, Achil Michel et Solenne Rouillard.

1 | DONNÉES À UTILISÉES

PlantPAN 3.0

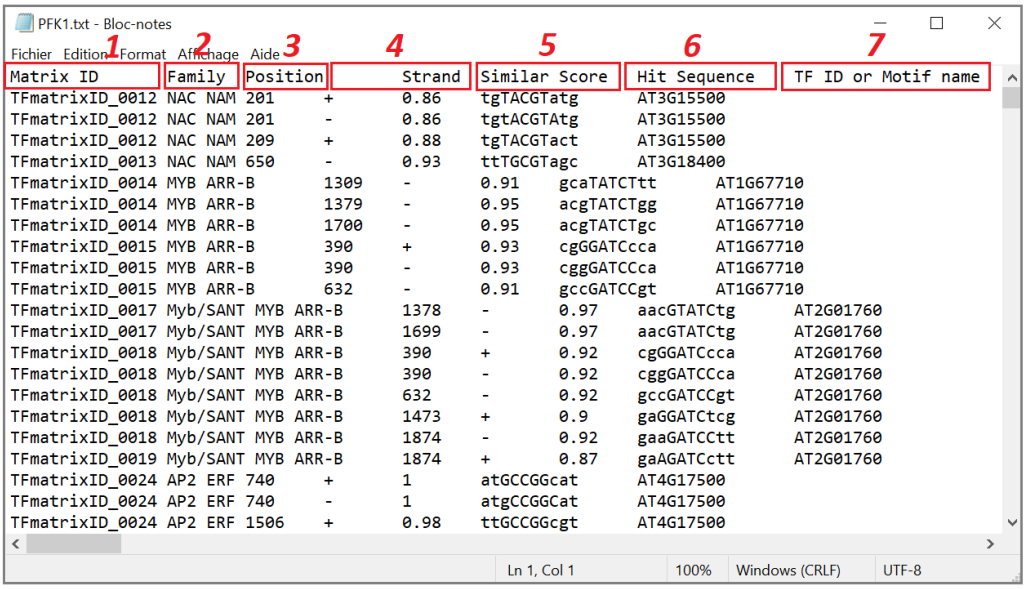
Le type de fichier accepté par ce programme sont les **fichiers texte** obtenus avec l'outil **Promoter Analysis de PlantPAN 3.0.**

Ces fichiers regroupent l'ensemble des TF/motifs retrouvés sur votre ou vos promoteurs et sont téléchargeables après l'analyse directement sur le site.

Lien vers PlantPAN : http://plantpan.itps.ncku.edu.tw/promoter.php

---

Pour le bon fonctionnement du programme, la première ligne de vos fichiers doit être identique à celle présentée ci-dessous *(cela est fait automatiquement par PlantPAN normalement)* :



PFK1.txt - Bloc-notes

Fichier Edition Format Affichage Aide

1 Matrix ID	2 Family	3 Position	4 Strand	5 Similar Score	6 Hit Sequence	7 TF ID or Motif name
TFmatrixID_0012	NAC NAM	201	+	0.86	tgTACGTatg	AT3G15500
TFmatrixID_0012	NAC NAM	201	-	0.86	tgtACGTAtg	AT3G15500
TFmatrixID_0012	NAC NAM	209	+	0.88	tgTACGTact	AT3G15500
TFmatrixID_0013	NAC NAM	650	-	0.93	ttTGCGTagc	AT3G18400
TFmatrixID_0014	MYB ARR-B	1309	-	0.91	gcaTATCTtt	AT1G67710
TFmatrixID_0014	MYB ARR-B	1379	-	0.95	acgTATCTgg	AT1G67710
TFmatrixID_0014	MYB ARR-B	1700	-	0.95	acgTATCTgc	AT1G67710
TFmatrixID_0015	MYB ARR-B	390	+	0.93	cgGGATCcca	AT1G67710
TFmatrixID_0015	MYB ARR-B	390	-	0.93	cggGATCCca	AT1G67710
TFmatrixID_0015	MYB ARR-B	632	-	0.91	gccGATCCgt	AT1G67710
TFmatrixID_0017	Myb/SANT MYB ARR-B	1378	-	0.97	aacGTATCtg	AT2G01760
TFmatrixID_0017	Myb/SANT MYB ARR-B	1699	-	0.97	aacGTATCtg	AT2G01760
TFmatrixID_0018	Myb/SANT MYB ARR-B	390	+	0.92	cgGGATCcca	AT2G01760
TFmatrixID_0018	Myb/SANT MYB ARR-B	390	-	0.92	cggGATCCca	AT2G01760
TFmatrixID_0018	Myb/SANT MYB ARR-B	632	-	0.92	gccGATCCgt	AT2G01760
TFmatrixID_0018	Myb/SANT MYB ARR-B	1473	+	0.9	gaGGATCtcg	AT2G01760
TFmatrixID_0018	Myb/SANT MYB ARR-B	1874	-	0.92	gaaGATCCtt	AT2G01760
TFmatrixID_0019	Myb/SANT MYB ARR-B	1874	+	0.87	gaAGATCctt	AT2G01760
TFmatrixID_0024	AP2 ERF 740	+	1	atGCCGGcat	AT4G17500	
TFmatrixID_0024	AP2 ERF 740	-	1	atgCCGGCat	AT4G17500	
TFmatrixID_0024	AP2 ERF 1506	+	0.98	ttGCCGGcgt	AT4G17500	

Ln 1, Col 1 | 100% | Windows (CRLF) | UTF-8

ATTENTION

Si vos fichiers n'ont pas l'ensemble de ces sept colonnes ou si l'ordre n'est pas correct, PromAnalyzer aura de fortes chances **de se bloquer** et **le traitement ne pourra pas se faire**.

Cependant, si vos **fichiers sont bien sous cette forme**, PromAnalyzer fonctionnera sans problème ✔

CONSEIL

Pour chacun de vos fichiers, essayer de mettre **un nom qui vous parle** *(par exemple le nom du gène auquel appartient ce promoteur)* car par la suite, ce nom sera utilisé en **en-tête** de la colonne dans le fichier résultat bilan.

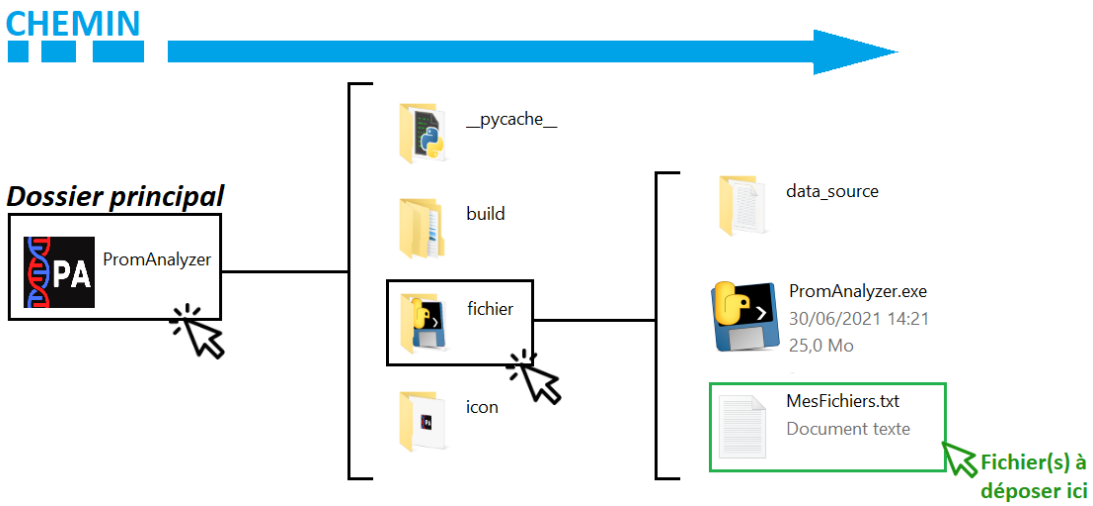
Ainsi un nom tel que « PlantPAN-rzf2926 » est beaucoup moins informatif que « HXK » pour l'hexokinase ou encore « PK1 » pour la pyruvate kinase.

2 | PLACER VOS DONNÉES DANS LE BON DOSSIER

Vos fichiers sont corrects, c'est parfait ! Mais attention, s'ils ne sont pas dans le **même dossier que l'exécutable** du programme *(fichier avec l'extension .exe)*, ils ne seront malheureusement pas lus par l'application.

---------- Or comment faire ? ----------

Il suffit de mettre vos fichiers dans **le dossier « fichier »,** présent dans **le dossier principal** du programme :



INFOS

Le dossier **data_source** contient un fichier texte utilisé pour le programme. Il s'agit d'une base de données de tous les TFs d'*Arabidopsis qui a été récupérée* sur le **site Agris** *(https://agris-knowledgebase.org/downloads.html).*

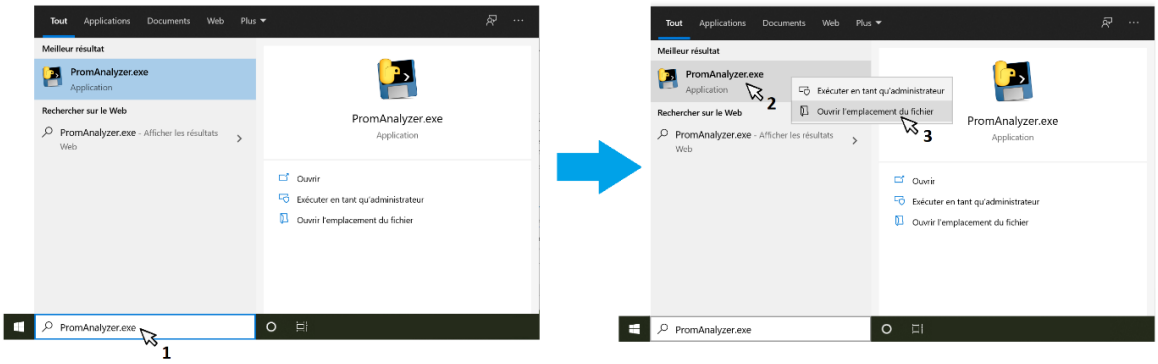
Si vous souhaitez **modifier ce fichier** *(pour mettre à jour les informations)*, faites attention à ce que le nouveau fichier ait la **même forme/nom** que celui-là. Sinon, il pourrait y avoir des soucis au niveau des fichiers résultats.

De plus, à la fin de votre analyse, veuillez ne **pas supprimer le dossier data_source et le fichier exécutable *(PromAnalyzer.exe).*** Car le programme ne fonctionnerait plus ! ☹

Au secours : Je ne sais plus de tout où j'ai mis le dossier du programme !

Pas de panique, tout se retrouve sur un ordinateur ! *(Enfin généralement)*

Pour le retrouver, essayer de **rechercher le nom du programme sur votre ordinateur** (PromAnalyzer.exe). Par exemple, sur Windows, vous pouvez le faire grâce à la **barre de recherche** comme suit :



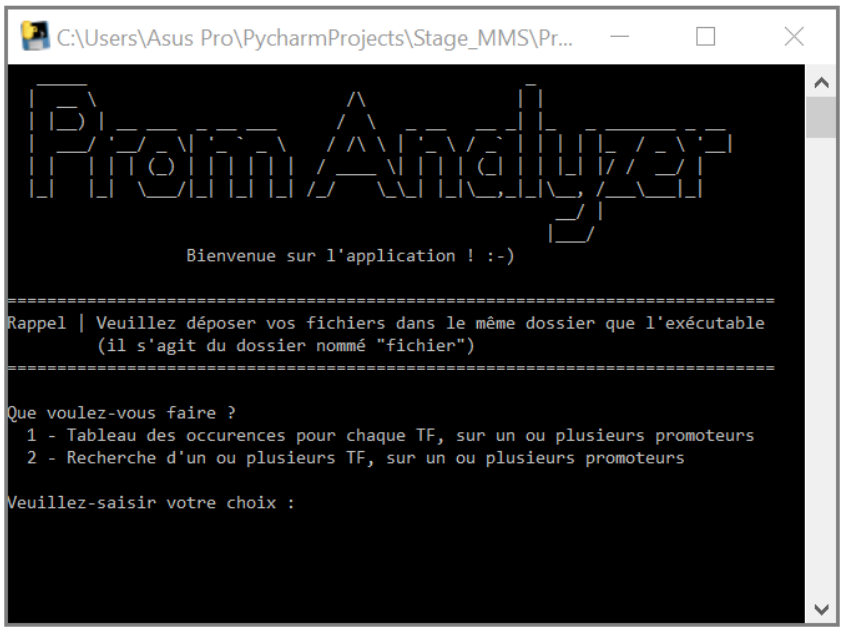
Windows vous ouvrera ensuite la fenêtre du dossier où se trouve le programme *(normalement dans « fichier »)*

Si vous ne le retrouvez toujours pas, vous pouvez toujours re-télécharger l'application sur GitHub grâce à ce lien : *https://github.com/SoR-72/PromAnalyzer*

3 | LANCEMENT DU PROGRAMME

Si vous avez bien suivi les instructions jusqu'à présent, bravo ! Vous êtes désormais prêt à lancer l'application ! 🎉

Pour cela, rien de plus simple, faites **un double-clic sur le fichier exécutable** *(PromAnalyzer.exe)*. L'application devrait se lancer sans problème et le menu suivant doit s'afficher :



ET SI … ÇA NE MARCHE PAS ?

Si ce n'est pas le cas ou que le programme **ne se lance absolument pas**, c'est assez problématique.

Ce souci peut avoir **plusieurs raisons** et peut venir soit du programme en lui-même, soit de votre système d'exploitation.

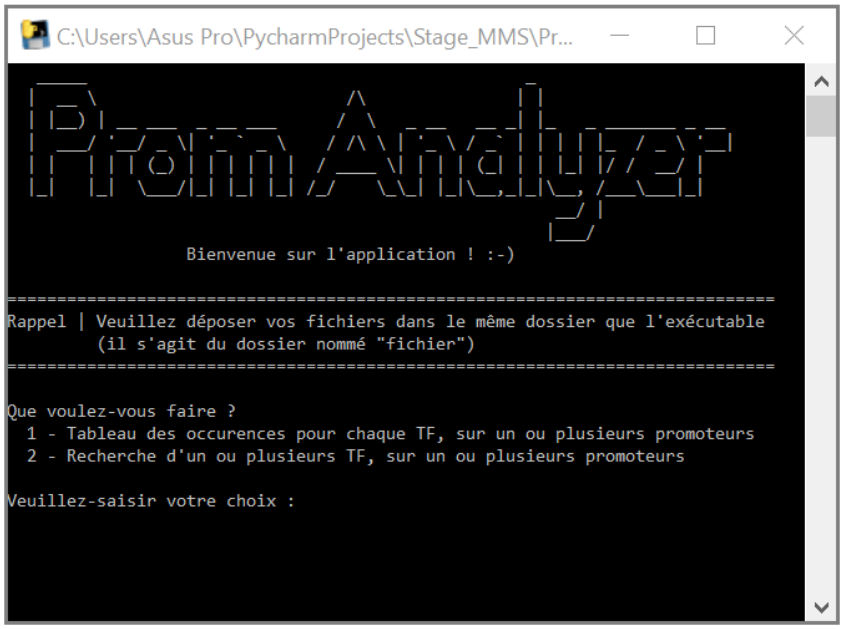
Tout d'abord, vous pouvez essayer de **re-télécharger le logiciel** à partir de GitHub pour voir si cela résout le problème : *https://github.com/SoR-72/PromAnalyzer*

De plus, s'il y a une fenêtre pop-up avec **une erreur qui s'affiche**, essayer de rechercher sur internet ce qu'elle signifie pour localiser le problème.

Si malgré tout, l'application ne se lance toujours pas, vous ne pourrez malheureusement pas utiliser PromAnalyzer.. À notre grand regret ☹

4 | CHOIX DU TRAITEMENT

Ainsi, voici donc **l'écran d'accueil** de **PromAnalyzer** :



Et comme vous pouvez le voir, deux choix de traitement sont possibles :

CHOIX – 1 Tableau des occurrences pour chaque TF, *sur un ou plusieurs promoteurs*	**CHOIX - 2** Recherche d'un ou plusieurs TF, *sur un ou plusieurs promoteurs*
Permets de **visualiser le nombre d'occurrences** *(sur combien de gènes et le total de répétition)* pour chaque TF sur les promoteurs donnés par l'utilisateur Très utile pour faire une **analyse d'ordre globale** ou pour faire **le repérage des TF intéressants** présents sur vos promoteurs et leur fréquence ☑ **Il y a des informations complémentaires sur le TF (nom de la protéine, similarité, ...) ainsi que des liens vers le site TAIR et UniProt**	Permets de **trouver les positions et les motifs** d'un ou plusieurs **TF** sur tous les promoteurs donnés par l'utilisateur Très utile pour faire la **cartographie** des TF d'intérêts sur vos promoteurs

Lorsque votre choix est fait, il suffit de **saisir votre choix** sur la fenêtre de l'application (soit 1 ou 2).

A NOTER que tant que vous n'aurez pas **saisi le chiffre 1 ou 2**, le programme vous redemandera votre choix *(permets d'éviter de bloquer le programme si vous faites une faute de frappe)*.

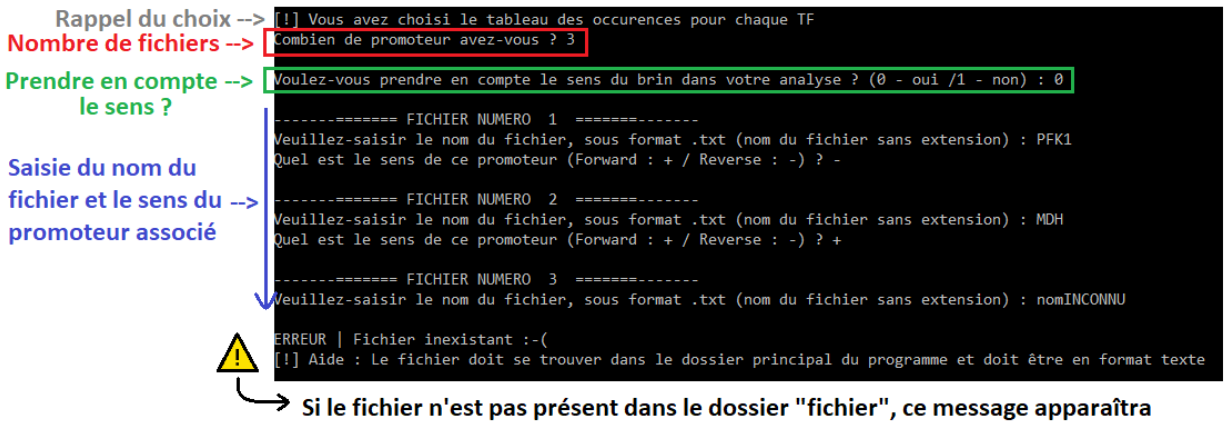
5 | SAISIE PAR L'UTILISATEUR

Après avoir fait votre choix, vous allez devoir **saisir plusieurs choses** en fonction des traitements :

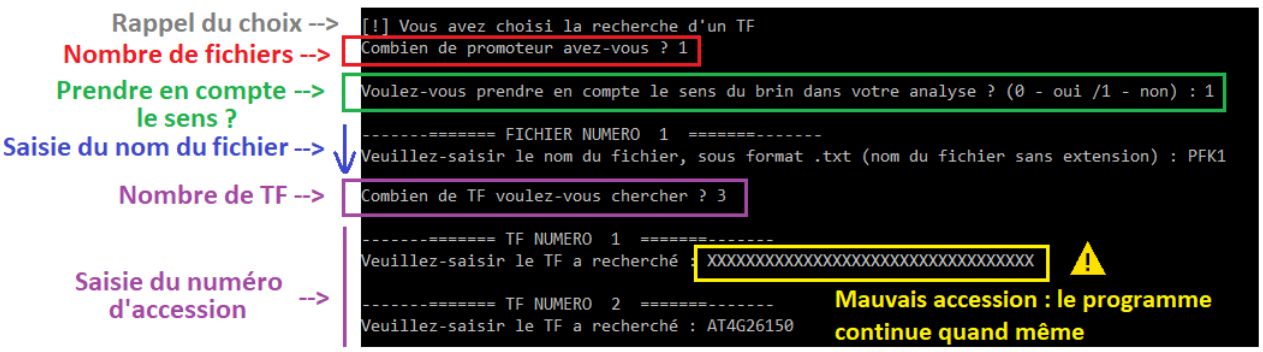
1	CHOIX 1 CHOIX 2	**Saisie du nombre de fichiers à analyser**	Il faut saisir un nombre positif non nul pour qu'il soit valide
2	CHOIX 1 CHOIX 2	**Si vous voulez prendre en compte le sens du brin de vos promoteurs**	Entrer 0 si vous voulez prendre en compte le sens de vos promoteurs *(un message vous demandera alors le sens pour chaque fichier)* Sinon, entrer 1 pour ne pas le prendre en compte *(vous aurez seulement à saisir le nom du fichier et l'analyse prendra en compte les deux sens + /-)*
3	CHOIX 1 CHOIX 2	**Saisie des infos pour chaque fichier** *(son nom et son sens si vous vouliez le prendre en compte)*	Ici, il faut saisir le nom du fichier sans son extension. Par exemple, si vous avez un fichier HXK.txt, vous allez devoir saisir HXK. A noter que si le fichier saisi n'est pas présent dans le dossier de l'exécutable, un message d'erreur s'affichera. Ensuite, si vous avez choisi de prendre en compte le sens du promoteur, veuillez saisir le sens + ou – en fonction de vos fichiers.
4	CHOIX 2	**Saisie du nombre de TF à rechercher**	Il faut saisir un nombre positif non nul (comme pour les fichiers)
5	CHOIX 2	**Saisie du numéro d'accession pour chaque TF**	Il faut saisir le numéro d'accession du TF à rechercher *(par exemple : AT5G66320, AT3G50870, …)*. A noter que si vous faites une faute de frappe (donc un numéro erroné), le traitement se lancera quand même et le tableau de recherche sera vide.

A la fin de la saisie, le traitement sera lancé et il ne restera plus qu'à attendre :-)

Voici un exemple de saisie pour le choix 1 (avec prise en compte du sens du brin) :



Et voici un exemple de saisie pour le choix 2 (pas de prise en compte du sens du brin) :



6 | PENDANT ET À LA FIN DU TRAITEMENT

Comme dit précédemment, le **traitement se lance à la fin de vos saisies** et vous pouvez le voir grâce au message suivant qui s'affiche :

```
EN COURS | Le traitement vient d'être lancé ! :-)
A noter que le programme peut prendre plus ou moins de temps selon la taille des fichiers (généralement 20-30s)
```

Sachez qu'il met généralement 20 à 30 secondes pour effectuer le traitement. Toutefois ce temps peut varier en fonction du nombre et de la taille de vos fichiers. Donc soyez patient ! :-D

C'EST NORMAL LES CHOSES BIZARRES QUI APPARAISSENT PENDANT LE TRAITEMENT ?

```
EN COURS | Le traitement vient d'être lancé ! :-)
A noter que le programme peut prendre plus ou moins de temps selon la taille des fichiers (généralement 20-30s)

C:\Users\ASUSPR~1\AppData\Local\Temp\_MEI108802\PromAnalyzer.py:47: FutureWarning: Your version of xlrd is 1.2.0. In xlr
d >= 2.0, only the xls format is supported. As a result, the openpyxl engine will be used if it is installed and the eng
ine argument is not specified. Install openpyxl instead.
C:\Users\ASUSPR~1\AppData\Local\Temp\_MEI108802\PromAnalyzer.py:183: FutureWarning: Your version of xlrd is 1.2.0. In xl
rd >= 2.0, only the xls format is supported. As a result, the openpyxl engine will be used if it is installed and the en
gine argument is not specified. Install openpyxl instead.
C:\Users\ASUSPR~1\AppData\Local\Temp\_MEI108802\PromAnalyzer.py:232: FutureWarning: Your version of xlrd is 1.2.0. In xl
rd >= 2.0, only the xls format is supported. As a result, the openpyxl engine will be used if it is installed and the en
gine argument is not specified. Install openpyxl instead.
```

Oui ! Ne vous inquiétez pas, c'est tout à fait normal ! Pour vous expliquer, PromAnalyzer utilise un package pour son bon fonctionnement *(xlrd)* mais ce dernier est sous une version antérieure *(1.2.0)* car c'est la seule à permettre d'ouvrir des fichiers excel comme nous le souhaitions. De ce fait, ce message s'ouvre à chaque fois que l'on ouvre un fichier Excel dans le traitement pour nous rappeler que les futures versions de ce package ne le permettront pas.

ET LES FICHIERS QUI APPARAISSENT ET DISPARAISSENT DU DOSSIER « FICHIER » ?

C'est tout à fait normal également ! Ce sont les fichiers créés à partir de ceux que vous nous avez fournis pour faciliter leur traitement. Ils seront ensuite supprimés à la fin du traitement pour ne pas encombrer votre dossier.

Après quelque temps, le **traitement est fini** lorsque vous voyez ce message :

```
SUCCES | Le traitement vient de se terminer, vous pouvez désormais récupérer le fichier Bilan_analyse.xlsx

[!] Appuyer sur la touche enter pour terminer le programme

 Achil MICHEL, Solenne ROUILLARD - Stage 2021
```

Vous pouvez alors fermer l'application : soit en tapant sur **la touche Enter** ou en **cliquant sur la croix** pour fermer la fenêtre). Puis vous pouvez récupérer votre fichier résultat dans le dossier « fichier » et le déplacer où vous voulez.

ATTENTION !	Si vous **ne déplacer pas** ou **ne renommer pas** ce fichier, il est possible qu'il soit **écrasé** si vous faites une nouvelle analyse.

7 | FICHIERS RÉSULTATS

Votre traitement est terminé et vous avez le fichier résultat, mais qu'est-ce qu'il contient précisément ?

CHOIX – 1 Tableau des occurrences pour chaque TF

Voici un exemple du fichier résultat obtenu après avoir choisi le **traitement « Tableau des occurrences pour chaque TF »** :

Informations diverses sur le TF — Liens — Nombre d'occurrences

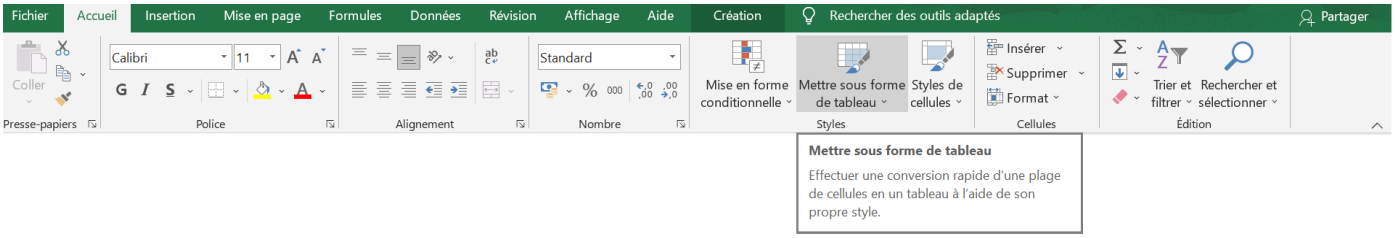
Sur tous les TF

	A	B	C	D	E	F	G	H	I	J	K
	PlantPAN Matrix ID	Motif Family	TF ID or Motif name	Name	Info	Others	TAIR Link	Uniprot Link	Found on	Sum of repeats	PFK1
	TFmatrixID_0013	NAC NAM	AT3G18400	not_available	organ separation protein	NA	https://www.arabi	https://www.uniprot	1	4	4
	TFmatrixID_0014	MYB ARR-B	AT1G67710	ARR11 ARP3	F12A21.15	hypothetical protein	https://www.arabi	https://www.uniprot	1	15	15
	TFmatrixID_0017	Myb/SANT MYB ARR-B	AT2G01760	ARR14	two-component response regulator pro	supported by full-length cDNA: Ceres:41875.	https://www.arabi	https://www.uniprot	1	16	16
	TFmatrixID_0024	AP2 ERF	AT4G17500	AtERF1	ethylene responsive element binding fa	identical to ethylene responsive element binding	https://www.arabi	https://www.uniprot	1	3	3
	TFmatrixID_0025	Homeodomain bZIP HD-ZIP	AT3G61890	ATHB-12	homeobox-leucine zipper protein ATHB-	supported by full-length cDNA: Ceres:32615.	https://www.arabi	https://www.uniprot	1	1	1
	TFmatrixID_0028	bZIP	AT1G42990	AtbZIP60	bZIP family transcription factor	contains Pfam profile: PF00170: bZIP transcripti	https://www.arabi	https://www.uniprot	1	1	1
	TFmatrixID_0030	MYB-related	AT2G46830	not_available	MYB-related transcription factor (CCA1	not_available	https://www.arabi	https://www.uniprot	1	4	4
	TFmatrixID_0037	AP2	AT2G40340	not_available	AP2 domain transcription factor	NA	https://www.arabi	https://www.uniprot	1	4	4
	TFmatrixID_0040	B3 ARF	AT2G33860	ARF3 ETTIN	auxin response transcription factor 3 (I	ARF3)	https://www.arabi	https://www.uniprot	1	3	3
	TFmatrixID_0042	GATA tify	AT5G25830	GATA-12	GATA zinc finger protein	GATA transcription factor Arabidopsis thaliana (	https://www.arabi	https://www.uniprot	1	49	49
	TFmatrixID_0044	MYB G2-like	AT2G20570	GLK1	expressed protein	NA	https://www.arabi	https://www.uniprot	1	1	1
	TFmatrixID_0048	Myb/SANT MYB G2-like	AT5G16560	KAN	protein	similar to unknown protein (gb|AAD20098.1) su	https://www.arabi	https://www.uniprot	1	2	2
	TFmatrixID_0059	AP2 ERF	AT1G03800	not_available	ethylene responsive element binding fa	identical to ERF domain protein 10 GI:11414990	https://www.arabi	https://www.uniprot	1	3	3

⚠ Le lien TAIR ne fonctionne que pour des TF d'*Arabidopsis Thaliana*

Présence sur combien de promoteurs — Total de répétition — Par promoteur

A noter que ce tableau a été mis en forme grâce à la fonctionnalité *« Mettre sous forme de tableau »* d'Excel :



Cette mise en forme vous permettra de trier ou de filtrer vos colonnes en cliquant sur cet élément dans vos en-têtes de colonnes.

CHOIX - 2 Recherche d'un ou plusieurs TF

Voici un exemple du fichier résultat obtenu après avoir choisi le **traitement « Recherche d'un ou plusieurs TF »** :

	A	B	C	D	E
1	TF recherché :	AT3G54620			
2	Sur promoteur du gène :	PFK1			
3					
4					
5	Hit Sequence	Positio	Similar Scor	Strand	
6	TACGT	3	1	-	
7	TACGT	82	1	-	
8	TACGT	211	1	-	
9	CACGT	472	1	-	
10	CACGT	940	1	-	
11	gccagCGTGG	1125	0,7	-	
12	AACGT	1378	1	-	
13	GACGT	1433	1	-	
14	AACGT	1699	1	-	
15					

PFK1 | GLK | PK5 | PK4A | OGDH | MDH | G6PDH | 6PGDH_1

Rappel du TF recherché et le promoteur

Informations sur la localisation des motifs en lien avec ce TF

Promoteur(s) analysé(s)