

# Diversité phylogénétique de l'assiette

**Stagiaire:**

Tommaso  
BARBERIS

**Encadrant:**

Bastien  
BOUSSAU

Avril-Juin 2021



## Résumé

Connaît-on vraiment tous les aspects de ce que nous mangeons ? Dans ce stage nous avons exploré une façon originale d'évaluer notre alimentation : la **diversité phylogénétique** !

Ce point de vue nous permet d'évaluer une idée répandue : une alimentation variée, est-elle vraiment synonyme d'un régime **sain** et **équilibré** ?

Il nous permet aussi de vulgariser des concepts de **biodiversité**, en illustrant comment l'Homme est relié à l'histoire du **Vivant** au travers de l'alimentation.

Dans le même objectif, nous avons également voulu attirer l'attention du grand public sur l'**arbre de la Vie** avec des visualisations intéressantes de nos plats.

Nous vous montrerons les liens entre l'**écologie** du Vivant et la **cuisine** avec l'outil que nous avons conçu pour répondre à ces demandes.

# Table des matières

<b>Résumé</b>	<b>1</b>
<b>Abréviations</b>	<b>3</b>
<b>1 Introduction</b>	<b>4</b>
1.1 État de l'art . . . . .	5
1.1.1 Naissance de la question scientifique . . . . .	7
<b>2 Matériel et méthodes</b>	<b>8</b>
2.1 Bibliothèques Python . . . . .	9
2.2 Bibliothèques R . . . . .	9
<b>3 Résultats obtenus</b>	<b>9</b>
3.1 Command Line Interface (CLI) . . . . .	10
3.2 Graphic User Interface (GUI) . . . . .	11
3.3 Analyse des données . . . . .	13
3.3.1 Calculs de diversité phylogénétique . . . . .	14
3.4 Fusion des trois groupes . . . . .	15
3.4.1 Corrélations avec la diversité phylogénétique . . . . .	17
3.5 Modèle de régression linéaire multiple . . . . .	18
3.5.1 Modèle uniquement avec le type de recette . . . . .	19
<b>4 Discussion</b>	<b>20</b>
4.1 Description des données . . . . .	20
4.2 Distribution de la diversité phylogénétique pondérée . . . . .	20
4.3 Corrélation entre la diversité phylogénétique pondérée et les valeurs nutritionnelles	21
4.4 Modèle des données . . . . .	21
4.5 Limites techniques . . . . .	22
4.5.1 Collecter les informations . . . . .	22
4.5.2 Taxonomie et diversité phylogénétique . . . . .	22
<b>5 Conclusion</b>	<b>23</b>
<b>Table des figures</b>	<b>24</b>
<b>Références</b>	<b>25</b>

## Abréviations

**CD** Continuous Delivery and Continuous Deployment. 10

**CI** Continuous Integration. 10

**CLI** Command Line Interface. 7, 9–12, 23

**DP** Diversité Phylogénétique. 4–7, 14, 15, 22–24

**DPP** Diversité Phylogénétique Pondérée. 14, 15, 17–21, 24

**GUI** Graphic User Interface. 7, 9, 11, 13, 22, 23

**TPA** Travail Projet d’Automne. 7, 8, 10, 11, 22

# 1 Introduction

L'alimentation humaine dépend principalement de ressources **animales** et **végétales** et possède une composante hautement **culturelle**. Nous mangeons différemment selon le pays de provenance et selon l'époque dans laquelle nous vivons [12]. Néanmoins, réfléchissons-nous assez au contenu de nos assiettes ? Dans ce projet nous avons creusé une façon originale de penser au contenu de nos assiettes, en l'analysant à travers la biodiversité qui y est contenue.

**La biodiversité** définit généralement la **variabilité biologique** d'un ensemble d'êtres vivants dans un contexte bien décrit qui est communément appelé environnement ou **écosystème**.

Plusieurs métriques existent pour pouvoir la mesurer et, sa quantification est particulièrement utile dans des domaines comme l'agriculture où elle est calculée pour suivre l'évolution d'une parcelle au cours du temps [2], pour renseigner sur l'efficacité de traitements chimiques et pratiques de culture [14], ou encore plus simplement pour pouvoir comparer deux écosystèmes différents [17]. Dans ce stage nous n'avons pas étudié la biodiversité d'un vrai écosystème, mais d'un plat, ce qui fait un usage inhabituel de cette grandeur.

**Pourquoi nous sommes-nous intéressés à la biodiversité d'un plat ?** L'idée de penser qu'une **alimentation diversifiée** soit synonyme d'un **régime sain** [18][1] est une opinion assez répandue dans la société d'aujourd'hui [7]. Pour cette raison nous nous intéressons à la biodiversité du plat, pour pouvoir explorer l'existence d'une corrélation entre la biodiversité et sa qualité nutritive.

**Diversité phylogénétique** Une des mesures qui peut être utilisée pour quantifier la biodiversité est la **Diversité Phylogénétique (DP)**, car elle permet de traiter différemment deux espèces proches contre deux espèces éloignées, ce qui nous intéresse beaucoup dans notre contexte culinaire.

Mais pourquoi utiliser la diversité phylogénétique et pas une autre métrique, comme la *richesse spécifique*<sup>1</sup>, pour étudier la biodiversité ? La DP est une mesure qui se base sur les branches d'un **arbre phylogénétique**, et qui permet donc d'intégrer un aspect évolutive que la richesse spécifique ne permet pas [16].

Or, ce stage vise à développer un outil qui permet de quantifier cette mesure de DP pour une recette trouvée sur le site de cuisine [Marmiton](#) [13] pour ensuite essayer de répondre à la question centrale de ce stage : « *Est-ce que la diversité phylogénétique est liée à la composition nutritionnelle du plat ?* ».

Le travail réalisé pourra être réutilisé par d'autres personnes afin de répondre à des autres questions de type culturel comme :

- ☛ Est-ce que la diversité phylogénétique est liée à l'origine géographique de la recette ?
- ☛ Est-elle influencée par l'époque dans laquelle on vit ? (Est-ce que la diversité phylogénétique des plats du moyen-âge est différente de celle des plats actuels ?)
- ☛ Est-ce qu'elle varie selon le régime alimentaire ?

---

1. La richesse spécifique représente le nombre d'espèces comptées dans un environnement donné.

Le travail réalisé cherche à aller au-delà de ces questions, auxquelles des réponses tranchantes et objectives peuvent s'avérer compliquées à trouver. En fait, indépendamment de cela, le stage vise à sensibiliser à la réflexion sur le vivant avec un point de vue original, la **nourriture**. Notamment, il permet de faire réfléchir sur l'*arbre du Vivant*, de façon particulière aux relations qui peuvent s'instaurer entre le règne **animal** et le règne **végétal**, composantes principales de notre alimentation.

D'autre part, il permet de faire comprendre comment la biodiversité est impliquée dans la vie de tous les jours et comment l'*Homme* est inévitablement lié à la structure intrinsèque de la **Terre** et à l'histoire du **Vivant**.

## 1.1 État de l'art

Ce projet s'appuie sur des concepts issus de l'écologie, notamment la **diversité phylogénétique**, pour porter un regard original sur notre alimentation, afin de réfléchir à celle-ci et de pouvoir disséminer des concepts de la biologie de l'évolution au grand public.

D'autres personnes se sont déjà intéressées à la nourriture avec un point de vue différent de celui typiquement culinaire ou culturelle ; c'est le cas du travail effectué par Ruben MILLA qui s'intéresse à l'aspect évolutif de l'alimentation [15]. Dans ses travaux, il a constitué une base de données d'espèces cultivées avec un but alimentaire, incluant des données descriptives comme les organes récoltés, les différentes formes lors de la croissance et leurs répartition géographique. En se basant sur ces données, il a pu construire une **phylogénie moléculaire** avec une échelle **temporelle** des cultures (mais qui donc ne contient pas d'autres organismes comestibles, notamment les animaux).

D'ailleurs, comme anticipé dans l'introduction, l'utilisation de la diversité phylogénétique peut s'étendre du simple concept de mesure de la biodiversité à des concepts plus complexes : par exemple, elle peut exprimer la stabilité des écosystèmes, lesquels dépendent de la diversité des espèces présentes. Cela fait de la DP une vraie métrique qui peut être utilisée pour décrire les différences écologiques [4]. Un article proposé par *Cadotte et al.* [3] montrerait comme des plantes produisent plus de **biomasse** en présence d'espèces **phylogénétiquement distantes** par rapport à une **mono-culture** ; inversement elles auraient la tendance à produire la quantité de biomasse prédite par la mono-culture dans le cas d'un système d'espèces plus proches phylogénétiquement.

La DP est une mesure utilisée en nombreuses applications, comme dans l'étude de la préservation de l'évolution potentielle [9] ou encore plus généralement dans la caractérisation de la biodiversité d'un environnement.

**Mais plus précisément, c'est quoi la diversité phylogénétique ?** La diversité phylogénétique est une mesure de la biodiversité basée sur la phylogénie. En 1992, Daniel **Faith** [8] décrivait la DP comme suit : "la DP d'un sous-ensemble d'espèces (ou d'autres niveaux taxonomiques) à partir d'un arbre phylogénétique est la somme des longueurs des branches couvrant ce sous-ensemble."

**Comment peut-on la calculer ?** La DP des recettes peut être calculée en utilisant la métrique proposé par *Webb et al.* [23], qui considère la DP comme la somme des distances moyennes de chaque espèce par rapport aux autres.

Il est tout à fait possible de la calculer sur un arbre phylogénétique, en considérant les distances entre espèces en nombre de branches.

**En quoi diffère la diversité phylogénétique des autres mesures de biodiversité ?** La DP se distingue des autres types de biodiversité (spécifique, taxonomique et fonctionnelle) car permet de refléter l'histoire évolutive d'un groupe d'espèces et bien plus que ça, elle permet de représenter cette diversité avec un intérêt en biologie de la conservation.

Si nous comparons DP et **richesse spécifique**, nous pouvons constater que la DP est moins dépendante des biais taxonomiques, car quand on se contente de compter les taxons ou les espèces, la richesse est biaisée par le fait que certains groupes d'organismes sont mieux échantillonnés que d'autres (difficultés et erreurs d'échantillonnage, etc.).

Parmi d'autres métriques intéressantes et très utilisées, on peut citer l'**indice** de **Shannon** [20], qui permet de quantifier la diversité **spécifique** (il s'agit d'une mesure d'entropie), et l'**indice** de **Simpson** [19]. L'indice de **Simpson** nous donne la possibilité de décrire la probabilité que deux individus pris au hasard appartiennent au même groupe taxonomique.

D'autre part, ce stage cherche à fournir une description évolutive des recettes en calculant la DP des espèces contenues dans l'assiette. La première étape est donc de récupérer les données d'une recette : ingrédients, quantités et unités, titre de la recette etc. Pour faire cela, il est nécessaire de parser la page web à partir de laquelle on veut récupérer la recette. Des outils pour réaliser cette étape existent déjà, comme on peut voir un exemple [ici](#), mais lors du projet d'automne (comme on verra plus tard dans le paragraphe 1.1.1 (*Naissance de la question scientifique*)) nous avons évalué qu'aucun de ceux-ci n'était adapté à notre but car soit ils étaient orientés pour des sites anglophones soit ils ne permettaient pas de récupérer des informations importantes comme la quantité de chaque ingrédient. Alternativement nous aurons aimé utiliser une **API** de recettes (comme en est un exemple [Spoonacular](#) [21]) française, mais à ce jour, aucun outil de ce type est disponible en ligne.

**Correspondance nom vernaculaire - nom scientifique** Pour ce qui concerne la correspondance entre ingrédient et espèce nous avons trouvé plusieurs bases de données permettant le passage de nom vernaculaire à nom scientifique, cependant cela amenait à des fausses correspondances dus à la non-unicité de noms vernaculaires (le nom vernaculaire d'un ingrédient peut être le même de celui d'une autre espèce qui quant à elle n'a rien à voir avec l'ingrédient). Le projet *Community Garden Book* [6] propose entre autres options, un outil numérique qui permet de faire la correspondance entre nom vernaculaire et scientifique d'une espèce. Malheureusement, à l'heure actuelle, le projet est encore en cours de développement et donc nous n'avons pas pu en profiter.

**Visualisation des espèces** Pour la représentation graphique des espèces nous nous sommes appuyés sur la représentation de l'arbre de la vie proposé par [LifeMap](#) [22].

**Manipulation des arbres phylogénétiques** Étant donné que les calculs de DP peuvent se baser sur les arbres phylogénétiques, nous avons utilisé la librairie Python *ete3* [11] pour la manipulation/visualisation de ces derniers.

**Valeurs nutritionnelles** Vu qu'il s'agit d'aliments, il était intéressant de proposer les résultats avec les valeurs nutritionnelles de chaque ingrédient, pour cela nous avons opté pour l'utilisation

de la base de données [Ciqua](#) [5].

### 1.1.1 Naissance de la question scientifique

Le sujet de stage naît comme suite d'un projet réalisé au cours du semestre d'automne de l'année universitaire en cours encadré par M. BOUSSAU Bastien et Mme PARIS Mathilde intitulé « *Diversité phylogénétique de l'alimentation* ». Le travail réalisé par moi-même et deux autres étudiants (DEGUISE Victor et FROUTÉ Timothée) peut se trouver [ici](#) (pour les personnes possédant un compte GitLab sur la plate-forme **pedago-service** de l'**Université Lyon 1**). Pour simplicité, dans la suite du document, nous nous référons au travail effectué pendant le projet d'automne avec l'acronyme **TPA** (Travail Projet d'Automne).

Donc ce stage pars où le projet d'automne s'était arrêté, c'est-à-dire avec un outil écrit en Python qui se présente déjà avec une version en ligne de commande et avec une version ayant une interface graphique incomplète.

La version en ligne de commande proposait à l'utilisateur de rentrer l'URL d'une recette provenant d'un site web parmi *Marmiton*, *CuisineAZ* et *750g*, et qui retournait donc un tableau contenant les ingrédients avec les espèces correspondantes ainsi que les valeurs nutritionnelles et le calcul de DP (comme la somme des distances (en nombre de branches) entre la racine et chacune des feuilles). La DP était calculé sur l'arbre phylogénétique de **Newick** (construit à partir de la taxonomie proposé par le **NCBI**) contenant toutes les espèces trouvées pour les ingrédients et avec tous les nœuds intermédiaires. La DP était aussi calculée avec pondération par la quantité de **matière sèche** de chaque ingrédient.

Au même temps, le programme sauvegardait l'arbre de Newick dans un fichier texte et le tableau dans le format *.tsv* dans le même répertoire à partir duquel le programme avait été lancé et il ouvrait un navigateur web pour afficher les espèces répertoriées sur l'arbre LifeMap.

L'interface graphique, en plus de ça, proposait déjà des liens vers les sites utiles comme ceux de cuisine ainsi que vers le répertoire distant sur GitLab contenant le projet. Une fenêtre d'erreur dans le cas où l'utilisateur aurait inséré un URL non valide (provenant d'un site non supporté ou URL corrompue) se serait ouverte. Ensuite comme en ligne de commande, elle listait les ingrédients pour lesquels l'espèce manquait, les ingrédients pas trouvés dans la table Ciqua, elle affichait le même tableau avec les résultats, la diversité phylogénétique pondérée et non.

La plus grande différence entre les deux versions consistait dans le donner le **choix à l'utilisateur** de faire certaines actions via l'utilisation de boutons, dans le cas de l'interface graphique. Ainsi l'utilisateur se voyait devant lui le choix de pouvoir : télécharger le tableau dans un fichier au format *.tsv*, d'ouvrir l'arbre avec LifeMap dans un navigateur, d'ouvrir l'arbre avec l'outil de navigation des arbres phylogénétiques proposé par *ete3* ou encore de copier dans le presse-papier l'arbre de Newick.

Dans la TABLE 1 nous montrons les différences de la Command Line Interface (CLI) et de la Graphic User Interface (GUI) entre la version du TPA et celle du stage.

	Projet d'automne	Stage
CLI	oui	oui
📄 Rapport html	non	oui



☛ Arbre de Newick (en output)	oui	oui
☛ Visualisation sur Life-Map	oui (ouverture d'un navigateur web pour l'affichage)	oui (intégré dans le rapport HTML via un iframe)
☛ Type de données en entrée	URL	URL, fichier avec plusieurs URL, fichier tsv
☛ Installation automatique des dépendances	non	oui
<b>GUI</b>	oui	oui
☛ Option multi-recette	non	oui
☛ Option fichier .tsv	non	oui
☛ Possibilité de rajouter des informations manquantes (espèces, quantités)	non	oui

TABLE 1 – Différences entre la version produite lors du projet d'automne et la version produite pendant le stage.

## 2 Matériel et méthodes

Nous avons choisi d'utiliser le site [Marmiton](#) comme source de recettes car, il est parmi les sites de cuisine les plus utilisés par les français et parce qu'il compte plus de **70 000** recettes.

Le projet est développé en **Python 3.8** et lors du TPA les dépendances étaient gérées avec **Conda**. Avec la prise en main du projet avec le stage, nous avons choisi de gérer les dépendances avec le paquet **Pipenv** *v. 2021.5.29* pour la raison suivante : pour construire l'interface graphique du programme, nous étions partis avec la librairie Python **Tkinter** qui sous Conda (ou Anaconda) résultait peu agréable visuellement (en particulière pour ce qui concerne les polices). Cela, car Anaconda n'inclut pas la librairie **Freetype** dans son interpréteur **Tk DLL** (c'est le moteur qui permet d'interpréter le code de l'interface graphique).

Or, Pipenv permet de remédier à ce problème en utilisant l'interpréteur du système d'exploitation, qui généralement supporte la librairie Freetype.

## 2.1 Librairies Python

Voici la liste des dépendances qui sont nécessaire à Pipenv pour créer un environnement virtuel qui permet de faire tourner le programme (☛ pour les librairies déjà présentes en automne, ☞ pour celles ajoutées avec le stage) :

☛ beautifulsoup4==4.9.0 ;	☛ pyperclip ;
☛ ete3==3.1.2 ;	☞ pytest==6.1.1 ;
☛ numpy==1.20.0 ;	☞ coverage==5.3 ;
☛ requests==2.24.0 ;	☞ customtkinter ;
☛ selenium==3.141.0 ;	☞ Pillow==8.2.0 ;
☛ xlrd==1.2.0 ;	☞ plotly==4.14.3 ;
☛ six==1.15.0 ;	☞ kaleido==0.2.1.
☛ webdriver-manager==3.4.1 ;	

## 2.2 Librairies R

Pour analyser les données obtenus on a utilisé **RStudio** utilisant **R** *v. 3.6.3* avec les librairies suivantes :

- ☛ naniar ;
- ☛ tidyr ;
- ☛ dplyr.

## 3 Résultats obtenus

Les objectifs techniques atteints dans ce stage sont les suivants :

- ☛ Mise en place de tests unitaires permettant de vérifier l'intégrité du code à chaque dépôt sur le répertoire distant (GitLab) ;
- ☛ Facilitation du débogage avec la mise en place d'un fichier log ;
- ☛ Modularisation du code pour la réutilisation ;
- ☛ Re-look de la GUI ;
- ☛ Modification structurale de la GUI (voir TABLE 1) ;
- ☛ Modification structurale de la CLI (voir TABLE 1) ;
- ☛ Développement d'une sous-version de **LifeMap** (voir paragraphe 3.1 (*Développement LifeMap*)).

Les modifications communes aux deux versions (CLI et GUI) effectués sur le projet préexistant, consistent dans la mise en place d'un fichier de *log* pour permettre un débogage plus simple et,

dans un système de tests unitaires permettant de tester des petites portions de code à chaque *commit* de façon automatique grâce aux **pipelines CI/CD** mis à disposition par **Gitlab** [10].

### 3.1 Command Line Interface (CLI)

Pendant le stage, la version CLI du programme a été modifiée sur plusieurs aspects, qui seront montré dans cette section.

Dans le TPA, le fichier exécutable était un exécutable Python *.py*, maintenant c'est un fichier *.sh* (nommé *divAlim.sh*). Passer à travers un script **bash** permet la vérification des dépendances sur le système utilisé. En fait, il permet de vérifier de façon automatique l'installation de *Pipenv* ainsi que l'existence ou pas d'un environnement virtuel ayant les dépendances nécessaires pour faire fonctionner le programme (l'environnement virtuel est créé automatiquement après la première utilisation de la CLI). L'utilisation d'un script **bash** a permis d'augmenter les fonctionnalités du programme, en donnant la possibilité de passer des options avec les paramètres associés, lors de l'exécution.

À différence de la version TPA, lancer le programme sans aucun paramètre, permet d'afficher l'aide en ligne de commande (il est possible aboutir au même résultat en exécutant le programme avec le flag *-help*), qui renseigne les informations principales du programme ainsi que sur les possibles options utilisables. Ici la liste des options possibles, avec les paramètres nécessaires et des exemples d'utilisation :

- ☛ **-u** : cette option permet de passer en paramètre directement un URL d'une recette.  
Exemple d'utilisation :

```
1 ./divAlim.sh -u https://www.marmiton.org/recettes/recette_  
2 charlotte-aux-fruits-d-ete_20546.aspx
```

- ☛ **-f** : cette option permet de passer en paramètre un fichier texte contenant en chaque ligne un URL d'une recette. Elle permet donc d'analyser plusieurs recettes au même temps.  
Exemple d'utilisation :

```
1 ./divAlim.sh -t file.txt
```

- ☛ **-t** : cette option permet de passer en paramètre un fichier au format *.tsv* contenant un tableau avec une ou plusieurs compositions de recettes.  
Exemple d'utilisation :

```
1 ./divAlim.sh -t file.tsv
```

Dans chaque cas, le programme génère en sortie un répertoire intitulé avec le nom de la recette (option **-u**) ou du fichier d'entrée (**-f** et **-t**) contenant un rapport dans un fichier au format **.html** ainsi que d'autres fichiers :

- **data.tsv** : tableau au format *.tsv* qui contient les résultats ainsi que les informations de départ ;
- **Tree.txt** : fichier texte contenant le plus petit sous-arbre de **Newick** contenant toutes les espèces identifiées dans la recette.

D'ailleurs, un exemple du report *.html* est montré en FIGURE 1 mais il est aussi présent en ligne [ici](#) pour pouvoir explorer l'interactivité avec l'arbre phylogénétique.

**Développement LifeMap** L'intégration d'un *iframe*<sup>2</sup> avec LifeMap dans le rapport HTML, a nécessité de développer une version ad-hoc du homonyme site à laquelle j'ai contribué. Cela a été possible grâce à l'aide du créateur de LifeMap, Damien De VIENNE. Le travail réalisé peut se trouver [ici](#). Cette version pourra être davantage réutilisée par d'autres projets qui s'intéressent à l'affichage d'un ensemble d'espèces sur l'arbre de la Vie, c'est notamment le cas du projet *Community Garden Book*, qui a déjà exprimé intérêt pour ce travail, notamment pour pouvoir afficher les espèces recensées sur une parcelle.

Il a été nécessaire de développer une sous-version de LifeMap car pour pouvoir afficher les espèces souhaitées, il aurait fallu passer des paramètres à travers l'iframe pour pouvoir "rentrer" les informations nécessaires dans le LifeMap originel. Or, cela n'est pas possible pour des raisons de sécurité (ex. *phishing*), donc nous avons contourné le problème en développant une sous-version de LifeMap qui permet de passer des paramètres directement via l'URL (qui requête la base de données) sans que l'iframe ait rien à faire.

## 3.2 Graphic User Interface (GUI)

La version GUI se base toujours sur la librairie Python **Tkinter** comme pendant le TPA, mais maintenant elle vient interprétée par l'installation Python de défaut du système (au lieu d'une installation Python sous Conda).

Comme montré dans le TABLE 1, la version développée lors du stage permet de compléter les informations tirées de la recette Marmiton, à savoir les **espèces** pour les ingrédients qui n'ont pas trouvé de correspondance et les quantités que le programme n'a pas pu trouver sur la page de la recette (FIGURE 2a et FIGURE 2b).

Nous avons également modifié le style du tableau contenant les informations, qui dans le TPA était fait nativement avec la librairie **Tkinter**, en utilisant la librairie **Plotly**. Nous générons donc le tableau sous-forme d'image au format *.png* (nous avons appliqué le même style pour le tableau contenu dans le rapport HTML de la CLI, donc il est possible de voir un exemple de ce tableau en FIGURE 1).

En plus de ça, l'utilisation de la GUI génère automatiquement un fichier *tsv* (possédant un nom commençant par "\_") qui permet de sauvegarder les informations rentrées pour les réutiliser dans un deuxième temps si nécessaire (possibilité de l'utiliser comme fichier d'entrée).

---

2. *iframe* ou *inline frame* : il s'agit d'une balise *html* qui permet d'afficher un autre document *html* à l'intérieur du document courant.

## Diversité phylogénétique de l'assiette



Name of the recipe: Ratatouille

URL of the recipe: [https://www.marmiton.org/recettes/recette\\_ratatouille\\_23223.aspx](https://www.marmiton.org/recettes/recette_ratatouille_23223.aspx)

Number of ingredients: 11

Number of specie found for the ingredients: 10

Ingredients that haven't match with a species: Sel

Ingrédient	Espèce	Quantité	Qté de matière sèche (g)	Eau (%)	Glucides (%)	Lipides (%)	Protéines (%)
Aubergine	Solanum melongena	350 g	24.85 g	92,9	2,39	0,14	1,12
Courgette	Cucurbita pepo	350 g	18.55 g	94,7	1,8	0,26	1,23
Poivron	Capsicum annuum	350 g	27.65 g	92,1	4,55	0,27	0,8
Tomate	Solanum lycopersicum	500 g	59.0 g	88,2	8,53	0,2	2,05
Ail	Allium sativum	120 g	38.64 g	67,8	18,6	0,5	5,31
Huile d'olive	Olea europaea	90 g	89.91 g	0,1	0	99,9	0,5
Thym	Thymus vulgaris	5 g	1.75 g	65,1	10,5	1,68	5,56
Poivre	Piper nigrum	1 g	0.91 g	9	39,5	7,5	13,3
Sel	NA	1 g	1.0 g	0,027	0	0	0
Oignon	Allium cepa	350 g	36.4 g	89,6	6,25	0,62	1,1
Feuille de laurier	Laurus nobilis	3 g	2.84 g	5,44	48,6	8,36	7,61

Phylogenetic diversity: 144.22

Weighted phylogenetic diversity: 14.27

Shannon's index: 2.37

Simpson's index: 1.61

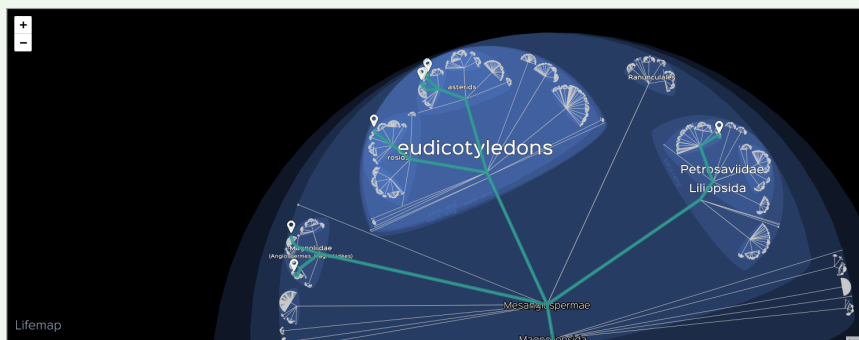


FIGURE 1 – Exemple de rapport au format *.html*

Maintenant, autre que l'URL, en entrée peut aussi prendre un fichier texte avec une liste d'URLs et, un fichier *.tsv* (tout comme la CLI).

Risotto Milanais

Si c'est possible, renseigner les espèces pour les ingrédients suivants:

parmesan râpé

bouillon de volaille

riz pour risotto

Test

Suivant

Risotto Milanais

Si c'est possible, renseigner les quantités pour les ingrédients suivants ainsi que leurs unités de mesure:

parmesan râpé

safran

riz pour risotto

huile d'olive

Avant

Terminer

(a) *Frame* permettant de compléter les espèces manquantes (b) *Frame* permettant de compléter les quantités manquantes

FIGURE 2 – Compléter les informations manquantes.

### 3.3 Analyse des données

L'outil développé a donc permis de générer des données (qui se trouvent [ici](#)) et, dans cette partie nous avons le but des les explorer et de les analyser.

**Les données** Les données sont générées en utilisant la GUI, de façon à compléter les informations manquantes, sur trois groupes de recettes tirées du site **Marmiton**, choisies arbitrairement :

- ☛ groupe **Vegetarian** : contient **30** recettes végétariennes sélectionnées à partir de la section [Végétarienne](#) . Il s'agit donc de recettes privées de viande mais avec la possibilité d'avoir des aliments d'origine animale (lait, œufs, etc.). Desserts exclus ;
- ☛ groupe **Top** : contient **30** recettes sélectionnées à partir de la section [Top des internautes](#). Il s'agit de recettes ayant au moins un ingrédient de viande ou de poisson. Desserts exclus ;
- ☛ groupe **Dessert** : contient **30** recettes sélectionnées à partir de la section [Top Dessert](#). Elles peuvent avoir des ingrédients d'origine animale ou pas.

Le tableau TABLE 2 montre les statistiques générales sur les ingrédients qui composent les recettes.

	Vegetarian	Top	Dessert
<b>Ingrédient le plus présent</b> (nombre)	Poivre (21)	Sel (21)	Oeuf (20)
<b>Espèce la plus représentée</b> (nombre)	Bos taurus (30)	Bos taurus (42)	Bos taurus (33)
<b>Énergie moyenne</b> (kJ/100g)	1026	985.4	1190
<b>Taux eau moyen</b> (g/100g)	51.56	50.736	38.83
<b>Taux glucides moyen</b> (g/100g)	15.01	13.10	29.54

<b>Taux lipides moyen</b> (g/100g)	14.93	14.31	15.99
<b>Taux protéines moyen</b> (g/100g)	5.672	6.641	5.8949

TABLE 2 – Informations sur les ingrédients composant les recettes.

### 3.3.1 Calculs de diversité phylogénétique

**Diversité phylogénétique** Comme vu dans l'état de l'art, nous avons choisi d'utiliser la métrique proposé par *Webb et al.* [23] pour calculer la DP.

Nous avons donc calculé la DP sur le plus petit arbre phylogénétique contenant toutes les espèces d'une recette en considérant les distances entre espèces le nombre de branches les séparant, avec la formule suivante :

$$DP = \sum_{i=0}^S \frac{\sum_{j \neq i}^{S-1} \mathbf{distance}_{j,i}}{S-1} \quad (1)$$

Explications des termes impliqués dans les formules :

- **S** : nombre total d'espèces au sein de la recette ;
- **distance<sub>j,i</sub>** : distance phylogénétique en nombre de branches entre l'espèce *j* et l'espèce *i* ;
- **q<sub>i</sub>** : quantité de matière sèche pour l'ingrédient correspondant à l'espèce *i*.

**Diversité phylogénétique pondérée** Pour avoir une mesure de la biodiversité plus cohérente, nous avons choisi de pondérer la diversité phylogénétique (diversité phylogénétique pondérée - **DPP**) avec la quantité de matière sèche (notée *q<sub>i</sub>*) des ingrédients, avec la formule suivante :

$$DPP = \frac{\sum_{i=0}^S \left( \frac{\sum_{j \neq i}^{S-1} \mathbf{distance}_{j,i}}{S-1} \times \mathbf{q}_i \right)}{\sum_{i=0}^S \mathbf{q}_i} \quad (2)$$

Pour la suite des analyses, nous avons choisi d'éliminer toutes les recettes ayant moins de 3 espèces identifiées, car la DPP de deux espèces revient au nombre de branches entre les deux. Les distributions de diversité phylogénétique, diversité phylogénétique pondérée et richesse, sont montrées en FIGURE 3 (en vert les recettes **végétariennes**, en rouge les **top** et en bleu le groupe **dessert**).

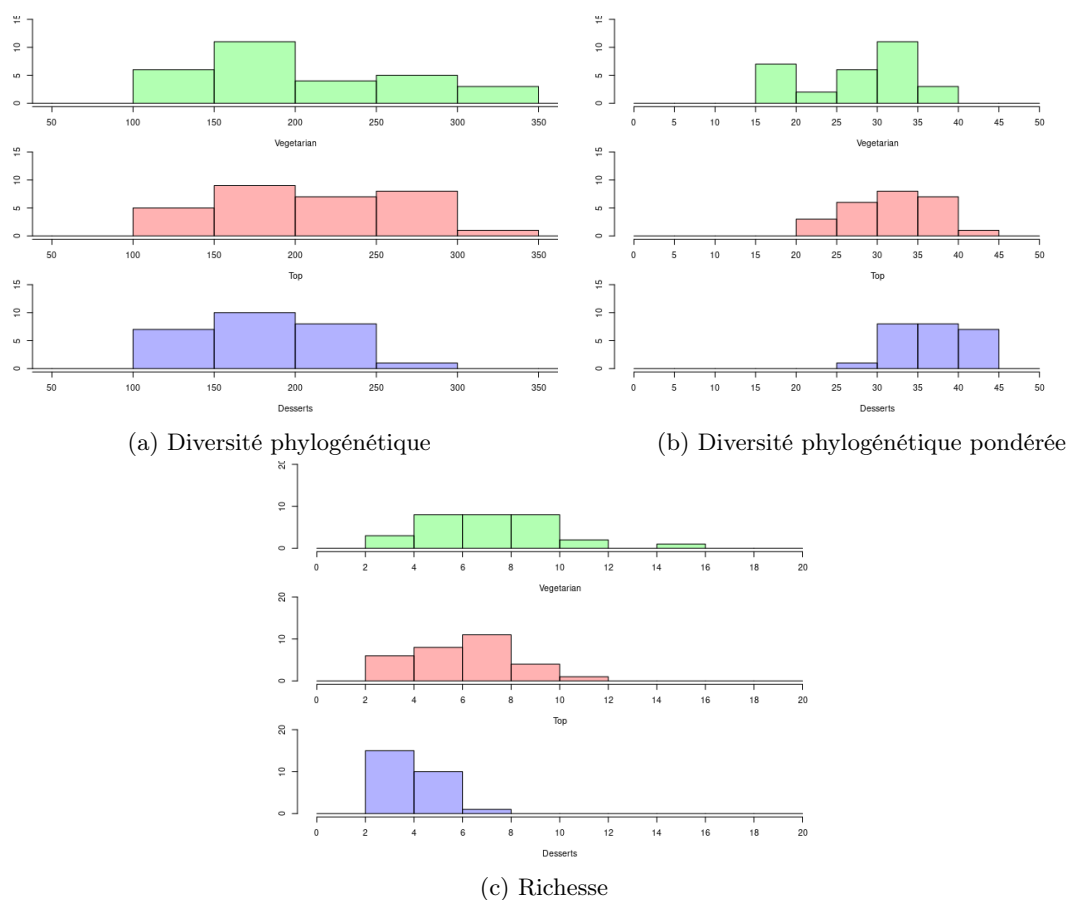


FIGURE 3 – Distribution de DP, DPP et richesse.

Le programme calcule aussi les indices de **Shannon** et de **Simpson**, mais suite à l'analyse des données nous avons évalué que ces deux indices se prêtent pas à l'analyse de recettes car on mesure pas l'abondance des espèces en nombre d'individu mais en quantité de matière sèche.

### 3.4 Fusion des trois groupes

Pour la suite des analyses, nous avons fusionné les données des trois groupes, mais en conservant l'information du groupe d'origine. Dans la TABLE 3 un résumé de ces données fusionnées.

Taux eau moyen (g/100g)	47.176
Taux glucides moyen (g/100g)	21.111



<b>Taux lipides moyen (g/100g)</b>	21.034
<b>Taux protéines moyen (g/100g)</b>	8.092
<b>Énergie moyenne (kJ/100g)</b>	1328.7
<b>Diversité phylogénétique moyenne</b>	200.2
<b>Diversité phylogénétique pondérée moyenne</b>	31.74
<b>Richesse moyenne</b>	6.349

TABLE 3 – Informations sur les recettes fusionnées.

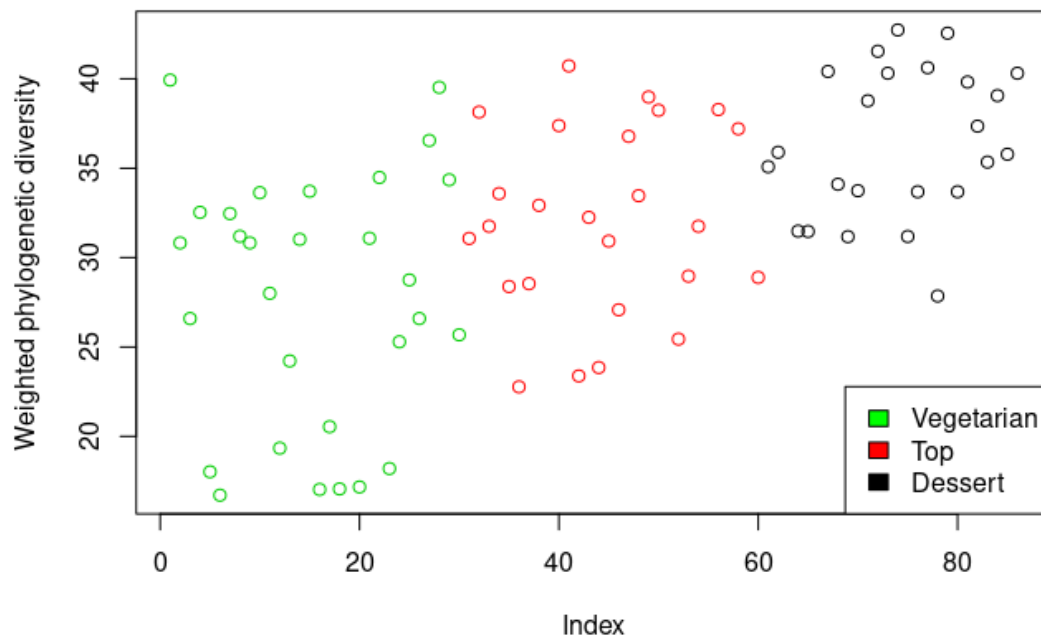


FIGURE 4 – Distribution de la diversité phylogénétique pondérée au sein des trois groupes.

### 3.4.1 Corrélations avec la diversité phylogénétique

Nous avons voulu explorer l'existence d'une corrélation entre la DPP et les différentes valeurs nutritionnelles d'une recette (énergie, eau, glucides, lipides et protéines), graphiquement en FIGURE 5 et numériquement avec le test de corrélation de **Pearson** (TABLE 4).

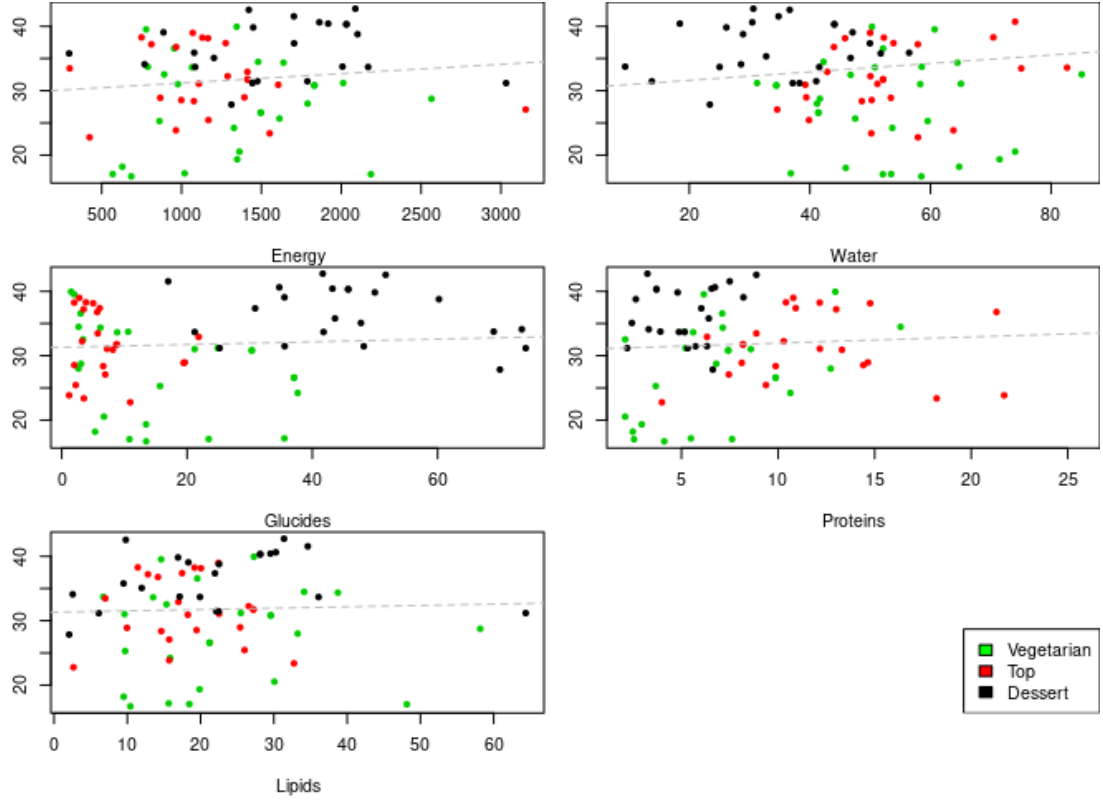


FIGURE 5 – Diversité phylogénétique pondérée en fonction des valeurs nutritionnelles.

	$r(df)$	p-value
<b>Énergie</b>	$r(70) = 0.1215734$	0.309
<b>Eau</b>	$r(76) = -0.1981784$	0.08198
<b>Glucides</b>	$r(68) = 0.1969102$	0.1023
<b>Lipides</b>	$r(68) = 0.03586738$	0.7682

<b>Protéines</b>	$r(68) = 0.06206846$	0.6097
------------------	----------------------	--------

TABLE 4 – Tests de corrélation

### 3.5 Modèle de régression linéaire multiple

Enfin, nous avons essayé d'expliquer la DPP avec un modèle de régression linéaire multiple. L'avantage d'utiliser un modèle linéaire par rapport à des simples corrélations (vues auparavant dans le rapport), est que le modèle permet d'expliquer notre variable (diversité phylogénétique pondérée) en fonction de plusieurs variables explicatives. Ici on a donc essayé de construire un modèle qui cherche d'expliquer la variabilité de la DPP en fonction des suivant facteurs : taux d'eau, taux de glucides, taux de protéines, taux de lipides, contenu énergétique et groupe d'origine (**Vegetarian**, **Top** et **Dessert**).

```

1 linear_model <- lm(Weighted_phylogenetic_diversity ~ Water +
2   Glucides + Proteins + Lipids + Energy + type,
3   data=all_simple)
4 summary(linear_model)
5
6 ##
7 ## Residuals:
8 ##      Min       1Q   Median       3Q      Max
9 ## -11.957  -4.421   0.744   4.775   9.709
10 ##
11 ## Coefficients:
12 ##              Estimate Std. Error t value Pr(>|t|)
13 ## (Intercept)    43.965547    7.547805   5.825 2.20e-07 ***
14 ## Water          -0.031896    0.083224  -0.383 0.702841
15 ## Glucides       -0.176164    0.077170  -2.283 0.025884 *
16 ## Proteins        0.356276    0.210972   1.689 0.096293 .
17 ## Lipids         -0.115770    0.106973  -1.082 0.283339
18 ## Energy          0.001455    0.002243   0.649 0.518853
19 ## typeTop        -13.233486    3.282596  -4.031 0.000154 ***
20 ## typeVegetarian -13.505177    2.495194  -5.412 1.07e-06 ***
21 ## ---
22 ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
23 ##
24 ## Residual standard error: 5.614 on 62 degrees of freedom

```

```

25 ## (16 observations deleted due to missingness)
26 ## Multiple R-squared: 0.3815, Adjusted R-squared: 0.3117
27 ## F-statistic: 5.464 on 7 and 62 DF, p-value: 6.404e-05

```

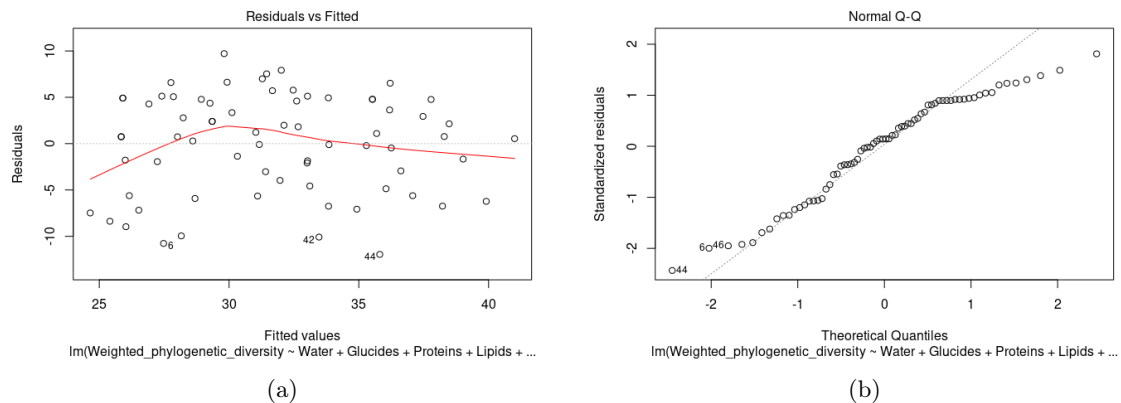


FIGURE 6 – Qualité du modèle linéaire.

### 3.5.1 Modèle uniquement avec le type de recette

Modèle de régression linéaire qui cherche à expliquer la DPP en fonction uniquement du groupe d'origine :

```

1 linear_model <- lm(Weighted_phylogenetic_diversity ~ type,
2   data=all_simple)
3 summary(linear_model)
4
5 ##
6 ## Residuals:
7 ##      Min       1Q   Median       3Q      Max
8 ## -10.9203  -3.4689  -0.0284   4.6156  12.2997
9 ##
10 ## Coefficients:
11 ##              Estimate Std. Error t value Pr(>|t|)
12 ## (Intercept)    36.413     1.175   30.996 < 2e-16 ***
13 ## typeTop        -4.384     1.645   -2.666  0.0094 **
14 ## typeVegetarian -8.782     1.588  -5.530 4.46e-07 ***
15 ## ---
16 ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
17 ##
18 ## Residual standard error: 5.755 on 75 degrees of freedom

```

```

19 ## (8 observations deleted due to missingness)
20 ## Multiple R-squared:  0.2902, Adjusted R-squared:  0.2713
21 ## F-statistic: 15.34 on 2 and 75 DF,  p-value: 2.609e-06

```

## 4 Discussion

### 4.1 Description des données

De manière plutôt étonnante, l'espèce la plus répandue dans chaque groupe est *Bos taurus* (nom scientifique utilisé pour définir l'ensemble des bovins), particulièrement pour ce qui concerne le groupe **Vegetarian**, qui devrait à priori présenter un nombre assez limité d'espèces animales dû à sa définition même (paragraphe 3.3 (*Les données*)).

Si on observe d'un peu plus près les données du groupe **Vegetarian**, on s'aperçoit que cela est dû, sans trop surprises, à l'utilisation de produits laitiers comme : crème fraîche, beurre, fromage et le lait.

**Forme des données après fusion** La fusion des trois groupes semblerait avoir modifié la structure des données, on peut le voir plutôt bien en regardant la quantité moyenne de protéines après fusion (**8.092** g/100g) par rapport aux moyennes des trois groupes avant (**5.672**, **6.641** et **5.8949** g/100g).

On voit comme le taux de protéines a augmenté inexplicablement. Cela peut être dû à la manière avec laquelle on a manipulé les données ; les moyennes des trois groupes (TABLE 2) a été calculé sur l'ensemble d'ingrédients (au sein d'un groupe), alors que la moyenne après fusion a été calculé sur les quantités moyennes de chaque recette.

On observe le même phénomène pour l'énergie et pour les lipides.

### 4.2 Distribution de la diversité phylogénétique pondérée

Dans la FIGURE 3c on peut voir comme les recettes végétariennes sont les plus riches en espèces mais qui pourtant ne sont pas les recettes les plus diverses si on observe la DPP.

Si on regarde la figure FIGURE 3a on a l'impression que soient bien les groupes **Vegetarian** et **Top** à être les plus divers, mais si on observe la DPP en figure FIGURE 3b on constate de façon inattendue une inversion de tendance, maintenant sont les desserts les plus divers (beaucoup de recettes entre **40** et **45**).

On peut observer cette tendance avec le graphique en FIGURE 4.

**Effet de la pondération sur la diversité phylogénétique d'un plat** Comme on vient de le voir, selon la diversité phylogénétique pondérée, sont les desserts à avoir les valeurs les plus élevées de DPP. Cela est probablement dû au fait que les desserts ont des compositions pauvres

en espèces (la recette la plus riche du groupe **Vegetarian** possède bien **15** espèces, alors que la plus riche du groupe **Dessert** possède que **7** espèces) mais que celles-ci proviennent à la fois du règne *animal* (notamment grâce au lait et au beurre) et à la fois du règne *végétal* (notamment cacao et sucre).

Du coup, on se retrouve avec peu d'espèces, relativement distantes, qui constituent la grande partie du plat à des fortes doses en quantité de matière sèche. D'ailleurs, dans les desserts on retrouve très peu d'ingrédients comme herbes aromatiques et/ou épices qui, permettent si d'augmenter la richesse, mais pas la DPP vu leur faible quantitatif utilisé dans les recettes.

### 4.3 Corrélation entre la diversité phylogénétique pondérée et les valeurs nutritionnelles

En (FIGURE 5) nous avons essayé de voir s'il existait une proportionnalité directe entre la DPP et les différentes valeurs nutritionnelles mais, ils semblerait pas être le cas.

Pour les trois groupes et pour tous les types de valeurs nutritionnelles, les recettes semblent être distribuées de façon hétérogène. Cela ne fait pas penser à une proportionnalité directe.

Ce résultat se confirme avec les tests de corrélation dans la TABLE 4, où aucune des valeurs nutritionnelles semble être corrélée avec la DPP. En observant les valeurs de *p-value* pour les différents tests on s'aperçoit que les résultats ne sont pas significatifs ( $p - value > 5\%$ ).

### 4.4 Modèle des données

Dans la sous-section 3.5 (*Modèle de régression linéaire multiple*), nous avons essayé d'expliquer la DPP en fonction des différents *paramètres* des recettes (énergie, valeurs nutritionnelles, type du groupe d'origine) avec un **modèle de régression linéaire multiple**.

On peut noter avec la valeur de **Adjusted R-squared** que, le modèle obtenu permet d'expliquer **31.2%** de la DPP avec les variables explicatives fournies et avec un degré de confiance assez élevé, le test a une *p-value* de **6.4e-05**. Autrement dit, le modèle explique 31.2% de la variabilité des données.

Mais le modèle obtenu peut être critiqué sous différents points de vue :

1. Dans le cas d'une régression linéaire parfaite on s'attend à avoir une droite nette et l'absence de pattern dans la distribution des points dans le graphique en FIGURE 6a. Dans notre cas, les points semblent avoir aucune distribution particulière (qui est donc plutôt bien pour expliquer le modèle) mais, la ligne rouge ressemble plutôt à une parabole aplatie qu'à une droite nette.
2. Dans le **QQ-plot** en FIGURE 6b, on devrait pouvoir observer les points pas trop loin de la ligne pointillée, laquelle représente une distribution *normale* des données. Si c'est pas trop le cas, alors le modèle n'est pas adéquat pour représenter les données. Ici, on peut voir comme les points dévient de la droite pour des valeurs élevées de quantiles théoriques.
3. Dans le modèle (section 3.5.1 - *Modèle uniquement avec le type de recette*) nous avons cherché à expliquer la DPP juste en fonction du type (groupe d'origine de la recette). Nous pouvons constater que le type de recette peut expliquer tout seul, bien le **27.1%** de la variabilité des données, en ayant toujours une *p-value* significative ( $2.6e - 06$ ).

D'après nos résultats préliminaires, il semblerait pas y avoir une corrélation entre la **diversité phylogénétique pondérée** et les **nutriments** d'un plat, comme on a vu dans la TABLE 4 en dans la FIGURE 4.

Cependant, nous pouvons pas exclure totalement une corrélation entre diversité phylogénétique et qualité nutritive, qui pourrait dépendre d'autres facteurs que nous avons pas analysé ici, comme la composition vitaminique, les oligo-éléments ou, de façon encore plus approfondie de facteurs comme la composition en acides aminés.

## 4.5 Limites techniques

### 4.5.1 Collecter les informations

Nous tenons à faire remarquer que la version GUI essaye de contourner l'absence d'une **API** de recettes françaises qui puisse être requêtée sur la composition en ingrédients d'une recette, les quantités à utiliser, les valeurs nutritionnelles ou alors pourquoi pas des informations supplémentaires comme le coût moyen d'un ingrédient ou les espèces qu'il contient avec les proportions de chacune.

Cela peut donc être un point de réflexion sur un développement futur, qui pourrait entraîner à la conception d'une interface web s'appuyant sur l'API qu'on vient de décrire et qui proposerait pour commencer les mêmes fonctionnalités que nous proposons ici dans cette application de bureau.

### 4.5.2 Taxonomie et diversité phylogénétique

Un des premiers problèmes rencontrés, déjà lors du TPA, est l'absence d'une phylogénie entière du Vivant datée. Pour pallier ce problème, nous avons choisi d'utiliser donc la taxonomie du **NCBI**. Quant à elle est "moins bien", car permet de raisonner uniquement en nombre de branches (comme décrit lors du document) au lieu d'un raisonnement en longueurs de branches correspondant à des temps de divergences entre espèces.

Nous sommes donc conscients d'avoir utilisé dans ce cadre, un calcul de diversité phylogénétique inhabituel par rapport à celui utilisé plus couramment.

Cela a aussi un impact assez fort sur la manière avec laquelle on interprète la DP : avec un arbre daté nous aurons pu qualifier, par exemple, un plat comme datant de 2 milliards d'années d'évolution, alors qu'en comptant le nombre de branches nous pouvons qualifier le plat comme le résultat de 200 événements de spéciation.

Pour tenter d'avoir des temps de divergence pour revenir à une définition plus classique de la diversité phylogénétique, nous aurons pu faire recours à une phylogénie datée comme [timetree of life](#).

Cependant, tous les temps de divergence n'y sont pas et donc il n'aurait pas résolu intégralement le problème. En perspective, il serait intéressant (et utile au sein de la communauté scientifique) de trouver un moyen pour obtenir des temps de divergence pour tous les nœuds taxonomiques.

## 5 Conclusion

L’approfondissement du développement de l’outil réalisé lors du ce stage a permis d’explorer la problématique scientifique posée en section 1 (*Introduction*) sous différents points de vue.

En premier lieu, nous avons mis à disposition un vrai outil permettant de calculer la diversité phylogénétique d’un plat. Ce programme a été rendu disponible en deux versions (GUI et CLI) pour pouvoir répondre à plusieurs besoins : la possibilité d’insérer des données manuellement (GUI) et la possibilité d’automatiser le programme (CLI) dans des pipelines.

Deuxièmement, nous avons essayé de répondre à certaines des questions posées dans l’introduction. On s’était posés la question sur l’existence d’une corrélation entre un régime diversifié et une alimentation saine. Nos analyses n’ont pas permis de détecter de corrélation entre la DP et qualité nutritive des plats, mais gagneraient à être répétées avec des mesures plus fines de la qualité nutritive.

En perspective, il pourrait donc être intéressant de faire des autres analyses pour étudier le lien entre diversité phylogénétique et qualité nutritive.



## Table des figures

1	Exemple de rapport au format <i>.html</i> . . . . .	12
2	Compléter les informations manquantes. . . . .	13
3	Distribution de DP, DPP et richesse. . . . .	15
4	Distribution de la diversité phylogénétique pondérée au sein des trois groupes. . .	16
5	Diversité phylogénétique pondérée en fonction des valeurs nutritionnelles. . . .	17
6	Qualité du modèle linéaire. . . . .	19

## Références

- [1] *Alimentation saine*. URL : <https://www.who.int/fr/news-room/fact-sheets/detail/healthy-diet> (visité le 23/06/2021).
- [2] Péter BATÁRY et al. « Biologia Futura : landscape perspectives on farmland biodiversity conservation ». In : *Biologia Futura* 71.1 (juin 2020), p. 9-18. ISSN : 2676-8615, 2676-8607. DOI : [10.1007/s42977-020-00015-7](https://doi.org/10.1007/s42977-020-00015-7). URL : <https://link.springer.com/10.1007/s42977-020-00015-7> (visité le 23/06/2021).
- [3] M. W. CADOTTE. « Experimental evidence that evolutionarily diverse assemblages result in higher productivity ». In : *Proceedings of the National Academy of Sciences* 110.22 (28 mai 2013), p. 8996-9000. ISSN : 0027-8424, 1091-6490. DOI : [10.1073/pnas.1301685110](https://doi.org/10.1073/pnas.1301685110). URL : <http://www.pnas.org/cgi/doi/10.1073/pnas.1301685110> (visité le 18/06/2021).
- [4] Marc W. CADOTTE, Russell DINNAGE et David TILMAN. « Phylogenetic diversity promotes ecosystem stability ». In : *Ecology* 93 (sp8 août 2012), S223-S233. ISSN : 0012-9658. DOI : [10.1890/11-0426.1](https://doi.org/10.1890/11-0426.1). URL : <http://doi.wiley.com/10.1890/11-0426.1> (visité le 18/06/2021).
- [5] *Ciqual Table de composition nutritionnelle des aliments*. URL : <https://ciqual.anses.fr/> (visité le 10/06/2021).
- [6] *Community Garden Book – Inria action exploratoire, agro-écologie*. URL : <https://project.inria.fr/polyculture/fr/> (visité le 11/06/2021).
- [7] A. I. CONKLIN et H. MOZAFFARI. *Varier son alimentation diminue le risque de maladie. Mais qu'entend-on par «variété» ?* The Conversation. URL : <http://theconversation.com/varier-son-alimentation-diminue-le-risque-de-maladie-mais-quentend-on-par-variete-124223> (visité le 23/06/2021).
- [8] Daniel P. FAITH. « Conservation evaluation and phylogenetic diversity ». In : *Biological Conservation* 61.1 (1992), p. 1-10. ISSN : 0006-3207. DOI : [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3). URL : <https://www.sciencedirect.com/science/article/pii/0006320792912013>.
- [9] Félix FOREST et al. « Preserving the evolutionary potential of floras in biodiversity hotspots ». In : *Nature* 445.7129 (fév. 2007), p. 757-760. ISSN : 0028-0836, 1476-4687. DOI : [10.1038/nature05587](https://doi.org/10.1038/nature05587). URL : <http://www.nature.com/articles/nature05587> (visité le 17/06/2021).
- [10] *GitLab CI/CD | GitLab*. URL : <https://docs.gitlab.com/ee/ci/> (visité le 15/06/2021).
- [11] Jaime HUERTA-CEPAS, François SERRA et Peer BORK. « ETE 3 : Reconstruction, Analysis, and Visualization of Phylogenomic Data ». In : *Molecular Biology and Evolution* 33.6 (juin 2016), p. 1635-1638. ISSN : 0737-4038, 1537-1719. DOI : [10.1093/molbev/msw046](https://doi.org/10.1093/molbev/msw046). URL : <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw046> (visité le 10/06/2021).
- [12] *Mangeons le monde ! Pour une géographie de l'alimentation. Pr Gilles Fumey*. URL : <https://www.diploweb.com/Mangeons-le-monde-Pour-une-geographie-de-l-alimentation.html> (visité le 23/06/2021).
- [13] MARMITON-AUFEMININ. *Marmiton.org*. marmiton.org. URL : <https://www.marmiton.org> (visité le 10/06/2021).
- [14] Amanda E. MARTIN et al. « Effects of farmland heterogeneity on biodiversity are similar to—or even larger than—the effects of farming practices ». In : *Agriculture, Ecosystems & Environment* 288 (2020), p. 106698. ISSN : 0167-8809. DOI : <https://doi.org/10.1016/j.agee.2019.106698>. URL : <https://www.sciencedirect.com/science/article/pii/S0167880919303147>.

- [15] Ruben MILLA. « Crop Origins and Phylo Food : A database and a phylogenetic tree to stimulate comparative analyses on the origins of food crops ». In : *Global Ecology and Biogeography* 29 (2020), p. 606-614. DOI : <https://doi.org/10.1111/geb.13057>.
- [16] Joseph T. MILLER et al. « Phylogenetic diversity is a better measure of biodiversity than taxon counting ». In : *Journal of Systematics and Evolution* 56.6 (nov. 2018), p. 663-667. ISSN : 1674-4918, 1759-6831. DOI : [10.1111/jse.12436](https://doi.org/10.1111/jse.12436). URL : <https://onlinelibrary.wiley.com/doi/abs/10.1111/jse.12436> (visité le 18/06/2021).
- [17] M. ORTEGA, R. ELENA-ROSELLO et J. M. GARCÍA DEL BARRIO. « Estimation of Plant Diversity at Landscape Level : A Methodological Approach Applied to Three Spanish Rural Areas ». In : *Environmental Monitoring and Assessment* 95.1 (1<sup>er</sup> juil. 2004), p. 97-116. ISSN : 1573-2959. DOI : [10.1023/B:EMAS.0000029893.27432.1c](https://doi.org/10.1023/B:EMAS.0000029893.27432.1c). URL : <https://doi.org/10.1023/B:EMAS.0000029893.27432.1c>.
- [18] Ministère des Solidarités et de la SANTÉ et Ministère des Solidarités et de la SANTÉ. *La santé par l'alimentation*. Ministère des Solidarités et de la Santé. 23 juin 2021. URL : <https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/strategie-nationale-de-sante/priorite-prevention-rester-en-bonne-sante-tout-au-long-de-sa-vie-11031/priorite-prevention-les-mesures-phares-detaillees/article/la-sante-par-l-alimentation> (visité le 23/06/2021).
- [19] E. H. SIMPSON. « Measurement of Diversity ». In : *Nature* 163.4148 (avr. 1949), p. 688. DOI : [10.1038/163688a0](https://doi.org/10.1038/163688a0).
- [20] Ian F. SPELLERBERG et Peter J. FEDOR. « A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the ‘Shannon–Wiener’ Index ». In : *Global Ecology and Biogeography* 12.3 (2003), p. 177-179. DOI : <https://doi.org/10.1046/j.1466-822X.2003.00015.x>. eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1466-822X.2003.00015.x>. URL : <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1466-822X.2003.00015.x>.
- [21] *spoonacular recipe and food API*. URL : <https://spoonacular.com/food-api> (visité le 10/06/2021).
- [22] Damien De VIENNE. « Lifemap : Exploring the Entire Tree of Life. » In : *PLOS Biology* 14.12 (2016). DOI : [10.1371/journal.pbio.2001624](https://doi.org/10.1371/journal.pbio.2001624).
- [23] Campbell O. WEBB et al. « Phylogenies and Community Ecology ». In : *Annual Review of Ecology and Systematics* 33.1 (2002), p. 475-505. DOI : [10.1146/annurev.ecolsys.33.010802.150448](https://doi.org/10.1146/annurev.ecolsys.33.010802.150448). eprint : <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>. URL : <https://doi.org/10.1146/annurev.ecolsys.33.010802.150448>.