



**HAL**  
open science

# Étude phénologique du hêtre (*Fagus sylvatica* L., 1753) dans la réserve naturelle de la Massane

Lauren Clément

► **To cite this version:**

Lauren Clément. Étude phénologique du hêtre (*Fagus sylvatica* L., 1753) dans la réserve naturelle de la Massane. Sciences de l'environnement. 2022. dumas-03932287

**HAL Id: dumas-03932287**

**<https://dumas.ccsd.cnrs.fr/dumas-03932287v1>**

Submitted on 10 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Etude phénologique du hêtre (*Fagus sylvatica* L., 1753) dans la réserve naturelle de la Massane

Lauren CLEMENT



Année universitaire 2021 - 2022

Sous la direction de : M. Olivier PANAUD

Laboratoire Génome et Développement des Plantes (LGDP), Bâtiment T, 58 Avenue Paul Alduy -

66860 PERPIGNAN

Tél : 04.68.66.21.19 - Fax : 04.68.66.84.99

Photographie de couverture : hêtre en train de débourrer dans la Réserve Naturelle Nationale de La  
Massane, Pyrénées-Orientales (66)

## Remerciements

Je tiens à remercier dans un premier temps mon maître de stage M. Panaud, directeur du Laboratoire Génome et Développement des Plantes et enseignant chercheur, pour m'avoir donné l'opportunité de réaliser ce stage et de travailler avec lui. Grâce à lui et sa bienveillance, j'ai découvert dans les meilleures conditions ce que représente le travail en laboratoire et j'ai pu progresser au niveau technique en bio-informatique. Il m'a permis de développer mes connaissances, et pas seulement en bio-informatique. Je remercie aussi l'indépendance et la confiance qu'il m'a accordées au cours de ce stage. Je suis extrêmement reconnaissante du temps qu'il m'a consacré, malgré son emploi du temps fortement chargé.

J'aimerai aussi remercier M. Garrigue, conservateur de la Réserve Naturelle Nationale de la Massane, qui m'a accompagné au cours de ce stage pour les travaux sur le terrain. Il m'a permis de découvrir cette réserve et d'en apprendre plus sur cet écosystème forestier. J'ai pu développer avec lui mes compétences naturalistes. Avec sa bonne humeur débordante, j'ai pu travailler aussi dans les meilleures conditions. Je remercie aussi l'ensemble de l'équipe de la Réserve de La Massane pour m'avoir aidé au cours de mon stage.

Je voudrais aussi remercier Mme Chuine, directrice de recherche au CNRS, qui travaille elle aussi sur le projet sur lequel j'ai effectué mon stage (le projet FAGADAPT). Elle m'a appris les bases de la phénologie. Malgré le fait que mon stage ne portait pas sur sa partie (modélisation), je savais qu'elle serait présente pour m'aider en cas de soucis sur le recueil des données phénologiques.

Je remercie aussi l'ensemble de l'équipe dans laquelle j'ai été accueilli au sein du LGDP, l'équipe MANGO, ainsi que tout le staff administratif. Ils m'ont eux aussi aidé et apporté leur bienveillance. Je m'y suis sentie à ma place au sein de cette équipe.

Je remercie aussi les doctorants, et ingénieurs en bioinformatique qui ont travaillé dans le même bureau que moi et qui ont été présents lorsque j'avais besoin d'aide.

Enfin, je remercie ma famille, mon compagnon et mes amis qui m'ont conseillés lors de la rédaction de mon rapport de stage et relus.

## Liste des abréviations

INPN : Inventaire National du Patrimoine Naturel

UICN : Union Internationale pour la Conservation de la Nature

GWAS : Genome-Wide Association Studies

SNP : Single Nucleotide Polymorphism

CEFE : Centre d'Ecologie Fonctionnelle et Evolutive

BIOM : Biologie Intégrative des Organismes Marins

LGDP : Laboratoire Génome et Développement des Plantes

ZNIEFF : Zones Naturelles d'Intérêts Écologique, Faunistique et Floristique

BBCH : Biologische Bundesanstalt Bundessortenamt und CHEmische Industrie

ACP : Analyse en Composantes Principales

ADN : Acide désoxyribonucléique

BAM : Binary Alignment Map

VCF : Variant Call Format

MAF : Minor Allele Frequencies

csv : Comma-Separated values

INRAE : Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement

# Sommaire

<b>1. Introduction</b>	<b>1</b>
<b>2. Matériels et Méthodes</b>	<b>6</b>
2.1. Recueil des données phénologiques sur le terrain	6
2.2. Analyses statistiques des données phénologiques	7
2.3. Analyses bioinformatiques des données génomiques et phénotypiques	7
2.3.1. Filtre et tri des données génomiques	7
2.3.2. Recherche d'association entre le débourrement précoce et les polymorphismes nucléotidiques	8
<b>3. Résultats</b>	<b>9</b>
3.1. Trois phénotypes de débourrement distincts, répartis aléatoirement dans la réserve	9
3.2. Les phénotypes restent corrélés aux températures	11
3.3. La diversité chez le hêtre est en grande partie expliquée par le statut de débourrement	13
3.4. Pas de facteurs de génétiques associés aux phénotypes de débourrement pour le moment	14
<b>4. Discussion</b>	<b>16</b>
<b>5. Conclusion et perspectives</b>	<b>18</b>
<b>6. Bibliographie</b>	<b>19</b>
<b>7. Webographie</b>	<b>21</b>

## 1. Introduction

Dans le contexte du réchauffement climatique, les questions liées à la résilience des écosystèmes se posent de plus en plus. Les zones en climat méditerranéen sont beaucoup plus touchées par la sécheresse et les hausses de températures, et cela met en péril les espèces, végétales et animales, qui y sont localisées. Cela implique aussi que les zones de distribution des espèces végétales changent et engendrent par conséquent un dépérissement à grande échelle des forêts à travers le monde (Jump *et al.*, 2009). Il a été aussi observé que le changement climatique induit des modifications de la phénologie des espèces (Le collectif scientifique de l'Observatoire des Saisons, 2017).

En 100 ans, il y a eu une augmentation de température de 0.7°C en moyenne au niveau mondial (Zhang *et al.*, 2019). De nombreux modèles climatiques prévoient de manière générale une augmentation de la température de la surface terrestre de 3°C, jusqu'à 6°C pour les latitudes tempérées au cours de ce siècle (Zhang *et al.*, 2019, Saxe *et al.*, 2001). Cette hausse de température affecte de différentes façons tous les organismes vivants, y compris les plantes. En effet, il a été remarqué que les températures élevées au début du printemps arrivent de plus en plus tôt, ce qui leur permet d'avoir une plus grande période de production primaire (Zhang *et al.*, 2019). Il a ainsi été remarqué qu'en Europe depuis 1971, le débourrement des arbres avait été avancé de 2,5 jours par décennie dû à la hausse de température (Zhang *et al.*, 2019). Ceci accroît leur assimilation de la lumière et leur croissance (Saxe *et al.*, 2001). Cependant le fait de débourrer plus tôt augmente en parallèle le risque de dommages faits par le gel tardif printanier (Zhang *et al.*, 2019). Le réchauffement affecte aussi les aires de distribution des espèces. En effet, chaque espèce a des préférences de températures et de pluviométrie. Les différents modes d'adaptation possibles sont des changements génétiques ou non génétiques, comme des adaptations phénotypiques (Duputié *et al.*, 2015). En ce qui concerne les arbres, qui sont des organismes avec une durée de vie longue, il leur est plus compliqué de s'adapter, car l'adaptation génétique est lente (Duputié *et al.*, 2015).

Le monde terrestre a été subdivisé en 14 biomes et 8 royaumes biogéographiques ; et chacun de ces biomes a lui-même était divisé en écorégions (David M. Olson *et al.*, 2001). Un biome est une région biogéographique se trouvant dans un climat identique ; on peut aussi l'appeler un macro-écosystème (Webréférence 1). En France, trois biomes sont principalement présents regroupés en un même grand type qu'on appelle les forêts tempérées. On en distingue trois sortes : les forêts tempérées de conifères, les forêts méditerranéennes et les forêts tempérées caducifoliées mixtes (David M. Olson *et al.*, 2001). Les forêts tempérées caducifoliées mixtes sont le biome

majoritaire en France. On retrouve dans ces forêts des arbres caduques tels que le chêne et le hêtre (Webréférence 2). C'est un écosystème important, car il abrite des grands prédateurs terrestres (loups, ours etc.)(Webréférence 2). Or, ces forêts sont elles aussi touchées par le réchauffement climatique. Des évènements de dépérissement très prononcé ont été remarqués sur le hêtre européen (*Fagus sylvatica*) au Nord de l'Eurasie (Jump *et al.*, 2009).

Le hêtre de forêt, présent en Europe, est l'espèce *Fagus sylvatica* (L., 1753) (Le collectif scientifique de l'Observatoire des Saisons, 2017). C'est cette espèce que l'on retrouve en France métropolitaine. Il fait partie de la famille des *Fagaceae*, au même titre que le chêne et le châtaignier (Webréférence 3). D'après l'Inventaire National du Patrimoine Naturel (INPN), c'est une espèce ayant une préférence pour les températures fraîches et une pluviométrie relativement importante ; c'est pourquoi on la trouve généralement dans les zones montagneuses, en altitude (entre 800 et 1800 mètres d'altitude). Plus on se rapproche de la région méditerranéenne, moins le hêtre est présent, mis à part sur certains versants Nord possédant des températures plus froides (Webréférence 3). Le hêtre est une espèce dicline monoïque, c'est-à-dire qui possède des fleurs à sexes séparés mais possède les deux types de sexes sur le même arbre, et à feuilles caduques (Webréférence 3). Le hêtre a un feuillage qui peut être marcescent, c'est-à-dire avec des feuilles mortes qui restent accrochées aux branches pendant la période végétative. Cette espèce est notée sur la liste rouge de l'Union Internationale pour la Conservation de la Nature (UICN) en tant que préoccupation mineure depuis 2018 (Webréférence 3).

L'étude de la phénologie des individus à l'échelle populationnelle est importante pour comprendre comment cette espèce est affectée par le réchauffement climatique. Les stades phénologiques chez les arbres caduques sont le débourrement, c'est-à-dire à l'éclosion des premiers bourgeons, la floraison, la maturation des fruits et la sénescence des feuilles (Le collectif scientifique de l'Observatoire des Saisons, 2017). Cette dynamique va permettre de différencier deux grandes phases chez les plantes pérennes : la phase de dormance, ou repos hivernal, de l'automne au printemps, et la phase de croissance primaire et secondaire des pousses, du printemps à l'automne (Le collectif scientifique de l'Observatoire des Saisons, 2017). La période à laquelle la plante sort de la phase de dormance est un trait adaptatif et cela affecte grandement le taux de survie de la plante (McKown *et al.*, 2018). Le moment auquel la plante débourre va dépendre principalement de la température, mais aussi de la photopériode, c'est-à-dire le rapport entre la durée du jour et la durée de la nuit, pour certaines espèces telles que le hêtre (Le collectif scientifique de l'Observatoire des Saisons, 2017). La date précise à laquelle un arbre débourre va définir sa période de croissance.

Avec l'avancée des technologies dans le domaine du séquençage, il est dorénavant possible d'accéder relativement facilement à la diversité génétique de populations naturelles. Les approches Genome-Wide Association Studies, plus communément appelées par son acronyme GWAS, préalablement développées pour étudier les maladies humaines, sont maintenant utilisées à plus grande échelle (Atwell *et al.*, 2010). Cela consiste dans un premier temps à séquencer les génomes des individus étudiés, puis de les aligner à l'aide du génome de référence qui doit être de bonne qualité et correctement annoté (Atwell *et al.*, 2010). Ensuite, des modèles statistiques (ex: modèles linéaires mixtes) permettent de vérifier s'il y a une association entre des polymorphismes nucléotidiques (SNP), c'est-à-dire des différences de paires de base à un endroit donné dans le génome, et le phénotype prédéfini (Atwell *et al.*, 2010).

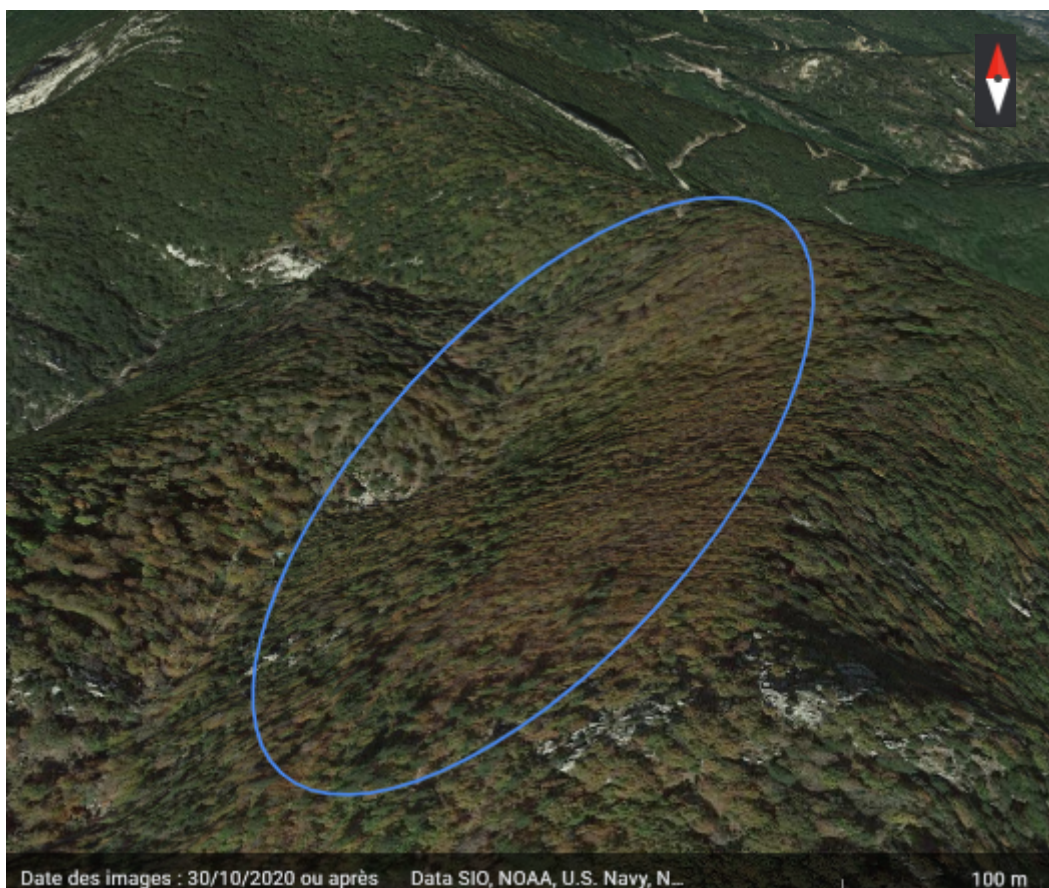
En parallèle de ces études génomiques sont aussi développés des modèles qui permettent de prédire la phénologie des arbres avec les conditions climatiques futures (Chuine *et al.*, 2016). Ceux-ci prennent en compte les dates de débourrement, la date de la fin de l'endodormance, qui est la première phase de la dormance, et la température (Chuine *et al.*, 2016). Les études ont été faites sur trois espèces d'arbres fruitiers mais sont adaptables pour les arbres forestiers (Chuine *et al.*, 2016). D'autres modèles permettent de prédire la distribution d'espèces d'arbres en rentrant les températures, les précipitations et la longueur de la nuit (Chuine et Beaubien, 2001).

C'est dans ce contexte que s'inscrit le projet FAGADAPT ([Projet pilote FAGADAPT – Biodivoc \(umontpellier.fr\)](#)). C'est un projet multidisciplinaire, subventionné par la région Occitanie, qui consiste à étudier l'adaptation des forêts tempérées, telles que les forêts de hêtres, dans le contexte du réchauffement climatique. Il a pour but de rechercher des stratégies pour aider ces forêts à s'adapter. Ce projet mélange la modélisation (CEFE Montpellier, BIOM Banyuls) et la génomique (LGDP Perpignan). L'hypothèse de départ est qu'il y a une origine génétique de la différence des dates de débourrement chez le hêtre, qui serait le trait d'une possible adaptation locale. Le but est de trouver dans un premier temps des facteurs génétiques qui induiraient un changement de date de débourrement chez le hêtre, avec des approches GWAS ; puis d'estimer à quel point ces changements affectent la valeur sélective de chaque type de phénotypes à l'aide de modèle de physiologie des arbres (PHENOFIT, Chuine & Beaubien, 2001). Cela permettra en finalité de connaître le devenir de chaque type de phénotypes avec le climat futur, et ainsi d'envisager ou non de faire de l'adaptation assistée pour les forêts tempérées de France. Pour cela, nous avons à notre disposition 150 hêtres, dans la Réserve Naturelle Nationale de la Massane, identifiés et séquencés avec la méthode Illumina dans le cadre du projet BeechGenome ([Beech genome resource](#)). Une



surveillance régulière de ces arbres est faite concernant la mortalité, la phénologie, la croissance et la vigueur.

La forêt de la Massane est localisée dans le département des Pyrénées Orientales, sur les communes d'Argelès-sur-Mer (66008) et de Sorède (66196), dans le massif des Albères. La forêt est traversée par un fleuve côtier appelé La Massane, qui prend sa source dans le massif des Albères et se jette dans la mer Méditerranée ; d'où le nom de cette forêt (Garrigue J., 2017). Cette forêt, inexploitée depuis plus d'un siècle, a une superficie de 358,9 hectares (Garrigue J., 2017). En effet, jusque dans les années 1880 la forêt a été dans un premier temps exploitée pour son bois dans le but de produire du charbon de bois pour les forges catalanes, qui elles-mêmes permettaient de convertir le minerai de fer (Garrigue *et al.*, 2010). Or après de nombreuses restrictions, les coupes de bois ont été suspendues, car l'exploitation était trop intensive (Garrigue *et al.*, 2010). Une grande partie de cette forêt a été classée en tant que Réserve Naturelle Nationale par arrêté ministériel depuis le 30 Juillet 1973 (Garrigue J., 2017). Cette réserve a une superficie de 336 hectares et s'étend de 600 à 1 253 mètres d'altitude (Garrigue *et al.*, 2010). La forêt de la Massane est dorénavant classée en tant que Zones Naturelles d'Intérêts Écologique, Faunistique et Floristique (ZNIEFF) de type 1, c'est-à-dire que c'est un espace homogène, au niveau écologique, possédant des espèces rares, remarquables ou notées en tant que patrimoine naturel régional (Webréférence 4). Il a été inventorié 8 119 espèces (faune et flore confondues) sur le site de La Massane en 2017, ce qui représente le record mondial en effort d'inventaire en nombre d'espèces inventoriées par hectare (Garrigue J., 2017). La forêt de la Massane a été choisie pour le projet FAGADAPT, car on y retrouve principalement du hêtre. Les 150 hêtres que j'ai étudiés se trouvent dans une zone bien spécifique : une réserve intégrale de 9,4 hectares entourée de barrières pour éviter que le bétail, encore quelque peu présent dans la réserve, ne puisse venir dégrader le peuplement de cette zone. Cette forêt fait partie des dernières hêtraies présentes sur le littoral, et plus précisément sur le littoral Méditerranéen (Garrigue J., 2017). Le site de la Massane a la particularité d'abriter cette espèce d'arbre malgré son rapprochement à la mer Méditerranée, car il est localisé en altitude ; les températures y sont plus fraîches qu'au bord du littoral. Mais il s'agit bien d'une hêtraie méditerranéenne de basse altitude (600 mètres) soumise à de fortes perturbations météorologiques, telles que le vent, la pluie, la sécheresse et le froid (Garrigue *et al.*, 2010). La forêt de la Massane est en particulier touchée par le réchauffement climatique, avec une augmentation de température de 1°C depuis 1980 (Garrigue J., 2018).



**Figure 1** : Carte en 3D de la réserve intégrale de La Massane à Argelès-sur-Mer

Au cours de mon stage, je me suis intéressée à la partie génomique et non celle de modélisation du projet FAGADAPT. J'ai donc cherché à répondre aux problématiques suivantes : Quelles sont les tendances de débourrement pour les 150 hêtres dans la Réserve Naturelle Nationale de la Massane ? Et surtout, existe-t-il des facteurs génétiques qui causeraient cette différence de date de débourrement ?

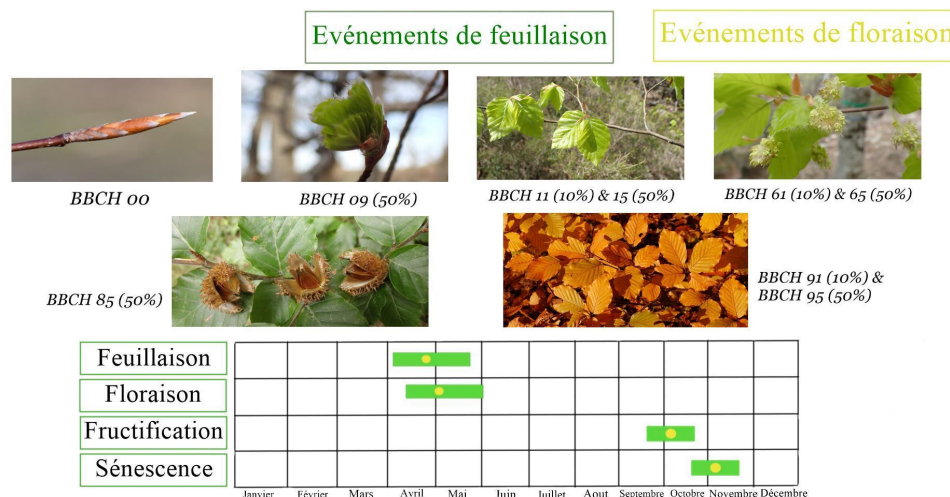
Pour répondre à cette problématique, j'ai dû diviser mon travail en deux temps : un temps sur le terrain pendant lequel j'ai noté la date précise de plusieurs stades phénologiques, dont le débourrement, pour les 150 hêtres étudiés, dans le but d'apporter un complément d'information au premier phénotypage effectué l'an dernier par l'équipe de la Massane ; et un temps pendant lequel j'ai fait des analyses statistiques, puis des analyses GWAS afin de déterminer les facteurs génétiques en lien avec ce phénotype.

## 2. Matériels et Méthodes

### 2.1. Recueil des données phénologiques sur le terrain

Mon stage s'est effectué de mi-février à début mai. Les 150 hêtres séquencés dans le cadre du projet BeechGenome ont été préalablement identifiés sur une carte de la réserve via l'application QField et identifiés visuellement à l'aide de rubalise plastique.

Par la suite, un peu avant et pendant la période de débourrement (de avril à mai), la visite des arbres s'est effectuée deux à trois fois par semaine pour surveiller à l'aide de jumelles (10x42 Leica) les bourgeons de chacun des arbres suivis ; puis la date des différents stades phénologiques a été notée sur des tablettes de terrain via l'application QField. Les observations se font en tournant autour de l'arbre observé, pour pouvoir estimer au mieux le pourcentage de feuilles en cours de développement ou étalées. Pour pouvoir faire les observations de manière correcte, j'ai dû suivre une formation en ligne sur le site de Tela Botanica : la formation Observatoire des Saisons. Cette formation et le protocole qui en découle, permettent de coder chaque événement phénologique selon un code BBCH à deux chiffres : les dizaines codant pour le type d'événement phénologique (0 : le développement des bourgeons, 1 : la feuillaison, 5 et 6 : la floraison, 7 et 8 : la fructification, et 9 : la sénescence foliaire) et les unités pour l'avancement de l'événement (ex: 65 correspond à 50% des fleurs ouvertes) (Le collectif scientifique de l'Observatoire des Saisons, 2017). Durant ce projet, nous nous intéresserons principalement au stade 09 (apparition pour au moins 50% du houppier le bout des premières feuilles), qui correspond au débourrement à proprement parlé, mais aussi aux stades 11 (10% de feuilles ouvertes étalées), et 15 (50% de feuilles ouvertes étalées), ainsi qu'aux stades phénologiques pour les fleurs : le stade BBCH 61 (10% du houppier avec des fleurs mâles épanouies) et 65 (50% de fleurs mâles épanouies). La date précise pour chacun de ces stades est nécessaire pour alimenter les modèles de Mme Chuine, du CEFV de Montpellier. L'étude GWAS se focalise uniquement sur les dates de débourrement (BBCH 09). Ces stades sont représentés sur la figure 2 ci-dessous.



**Figure 2** : Frise chronologique des événements phénologiques chez le hêtre

L'application QField m'a permis de noter d'autres informations lorsque j'étais sur le terrain : diamètre de l'arbre, la présence ou non de branches avec des feuilles marcescentes ainsi que la bonne santé de l'arbre. Les données ont été récupérées et mises sous la forme d'un fichier Excel (base de données en annexe), et il a pu être fait une carte via l'application QGIS version 3.16.11.

## 2.2. Analyses statistiques des données phénologiques

Les données phénologiques ont été dans un premier temps analysées grâce au logiciel Rstudio version 4.1.1 (script en annexe). Le seuil de significativité est fixé à 0,05. La normalité a été testée avec le test de Shapiro. Le test non-paramétrique de Kruskal-Wallis et le test de Tukey ont permis de comparer les moyennes des données de débournement. Une Analyse en Composantes Principales (ACP) a permis de savoir si certains paramètres étaient corrélés entre eux. Enfin un test de corrélation de Spearman a été fait pour les données de débournement et les températures.

## 2.3. Analyses bioinformatiques des données génomiques et phénotypiques

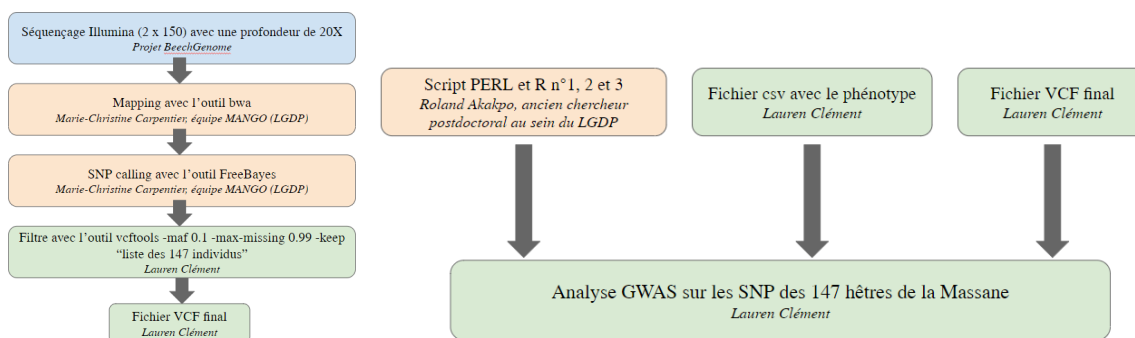
### 2.3.1. Filtre et tri des données génomiques

En ce qui concerne les données génomiques, celles-ci ont été obtenues par un séquençage en utilisant une plateforme Illumina (2 x 150), avec une profondeur de 20x. Par la suite, un mapping a été fait avec les fichiers FASTQ (format de fichier texte qui permet de stocker les séquences nucléiques et leurs scores de qualité) à l'aide de l'outil bwa qui permet de faire correspondre les "reads" (petits bouts de séquence d'ADN) de chaque individu contre le génome de référence. Il en

ressort des fichiers BAM (fichier stockant les séquences nucléiques alignées). Ensuite, le SNP calling a été fait à l'aide de l'outil FreeBayes. Cela permet de connaître en quelle position la base diffère de celle du génome de référence. Il en ressort alors un fichier VCF (format de fichier texte stockant les variations de séquences génétiques) utilisable pour les analyses. Il faut cependant le filtrer à l'aide de l'outil vcftools. Le paramètre -keep a permis de garder une liste d'individus. Le paramètre -maf (minor allele frequencies) est mis à 0,1. La fréquence allélique correspond au rapport du nombre de fois qu'un allèle apparaît à un site donné chez tous les individus étudiés sur le nombre total d'allèles non manquants dans ce site. Le paramètre -max-missing est mis à 0,99. Cela signifie que seuls les sites ayant moins de 1% de données manquantes sont gardés.

### 2.3.2. Recherche d'association entre le débourrement précoce et les polymorphismes nucléotidiques

Pour les analyses GWAS, il a été utilisé le fichier VCF final filtré comme précisé précédemment. Il a été nécessaire de convertir le fichier Excel avec les noms des arbres et les phénotypes en un fichier csv. Les dates de débourrement ont été codées en jours juliens, car cela sera plus pratique pour les analyses futures. J'ai ensuite utilisé des scripts PERL et R déjà existants, créés par Roland Akakpo qui a travaillé en tant que chercheur postdoctoral au LGDP jusqu'en Septembre 2019 (Annexe). Pour utiliser ces scripts, il est nécessaire d'avoir le logiciel R, ainsi qu'un ordinateur sous Linux. Les scripts PERL permettent d'utiliser les scripts R correspondants avec n'importe quel fichier : cela ne nécessite pas de noter les chemins d'accès ou de réadapter le script R pour qu'il corresponde au jeu de données. Le premier script permet d'étudier la structure de la population. Le second permet d'imputer des valeurs aux données manquantes qui restent, sinon l'analyse GWAS n'est pas possible. Le troisième permet de faire l'analyse GWAS à proprement parlée, c'est-à-dire que pour chaque SNP il est calculé une p-value démontrant l'association ou non avec le phénotype fourni par le fichier texte en amont. Il en ressort un Manhattan plot et un fichier csv avec l'ensemble des p-values pour chaque SNP. Le schéma de mon implication lors de ces analyses bioinformatiques est représenté en figure 3.

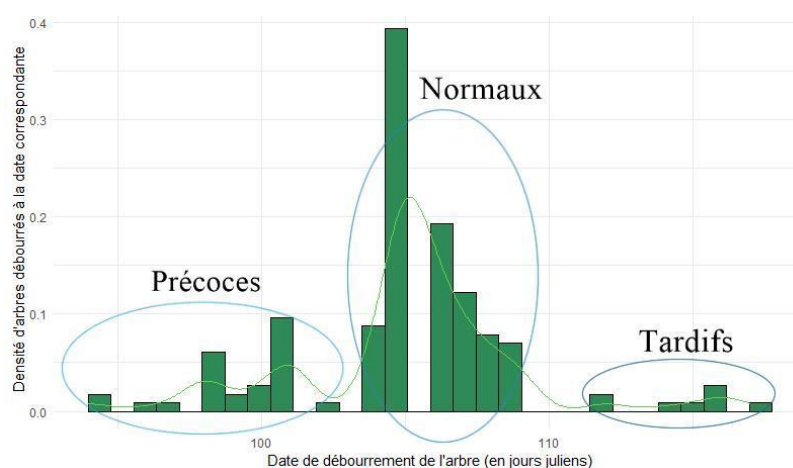


**Figure 3 :** Schéma de mon implication personnelle dans les analyses bioinformatiques

### 3. Résultats

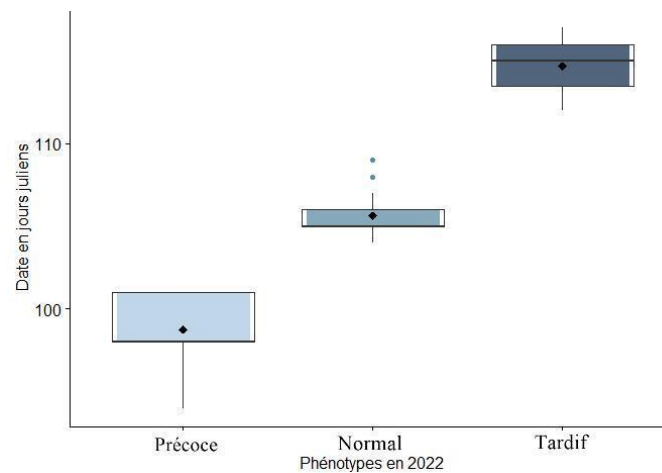
#### 3.1. Trois phénotypes de débourrement distincts, répartis aléatoirement dans la réserve

Suite aux observations sur le terrain, trois phénotypes de débourrement ont été trouvés : les individus précoces, ceux normaux et les tardifs (Figure 4). Les individus précoces représentent environ 20% de la population des 150 arbres séquencés. Les individus normaux représentent 75% de la population et les tardifs, seulement 5%. Un individu est noté comme précoce lorsque sa date de débourrement se situe entre le 4/04/2022 et le 13/04/2022 ; un individu normal a une date de débourrement qui se situe entre le 14/04/2022 et le 19/04/2022. Enfin un individu est noté tardif lorsqu'il a débourré entre le 20/04/2022 et le 27/04/2022. Cette notation a été déterminée à l'aide de la figure n°4 : en effet, on peut voir qu'il y a trois pics bien distincts, correspondant donc à chacune des classes définies juste avant.



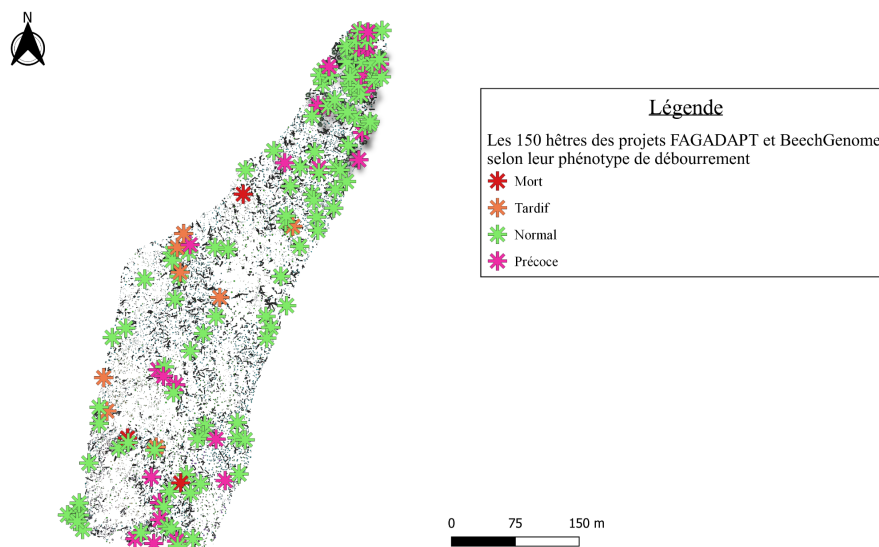
**Figure 4 :** Histogramme de la densité d'arbres débourrés (stade BBCH 09) dans notre population en fonction de la date en jours juliens

La moyenne de la date de débourrement pour chaque phénotype a été calculée et mise sous la forme d'un boxplot (Figure 5). La répartition des données n'est pas normale. De plus, il y a un défaut de normalité sur les résidus et ils ne sont pas homogènes. C'est pourquoi le test non-paramétrique de Kruskal-Wallis a été effectué. La p-value obtenue est inférieure à 0,05 donc l'hypothèse d'égalité des moyennes est rejetée. Un test de Tukey a ensuite été fait pour comparer les moyennes deux à deux. Les p-values sont inférieures au seuil de 0,05, donc les trois phénotypes de débourrement sont significativement distincts.



**Figure 5 :** Boxplot des trois phénotypes de débourrement en fonction de la date (jours juliens)

Sur la figure 6, nous pouvons voir que les hêtres précoces sont répartis aléatoirement dans la réserve, et non pas dans une seule et même zone spécifique. Ce repérage a permis de remarquer trois arbres morts (L15-014, BQ1-007 et BW8-003).



**Figure 6 :** Carte QGIS de la répartition des 150 arbres séquencés dans la réserve de la Massane

Les phénotypes sont corrélés en partie à l'état sanitaire de l'arbre. Trois tests de corrélation ont été fait : le test de Spearman, celui de Pearson et celui de Kendall. Ce sont des tests non-paramétriques qui permettent de mesurer le degré d'association entre les variables. Les p-values sont dans les trois cas inférieures à 0,05 et les indices de corrélation sont supérieurs à 0 mais éloignés de 1 (tableau 1). Donc les variables sont significativement corrélées, mais faiblement corrélées positivement. Cela signifie que l'état sanitaire de l'arbre, principalement lorsqu'il est dépérissant, a donc un impact sur le débourrement ; et l'arbre aura tendance à débourrer plus tardivement lorsqu'il est dépérissant.

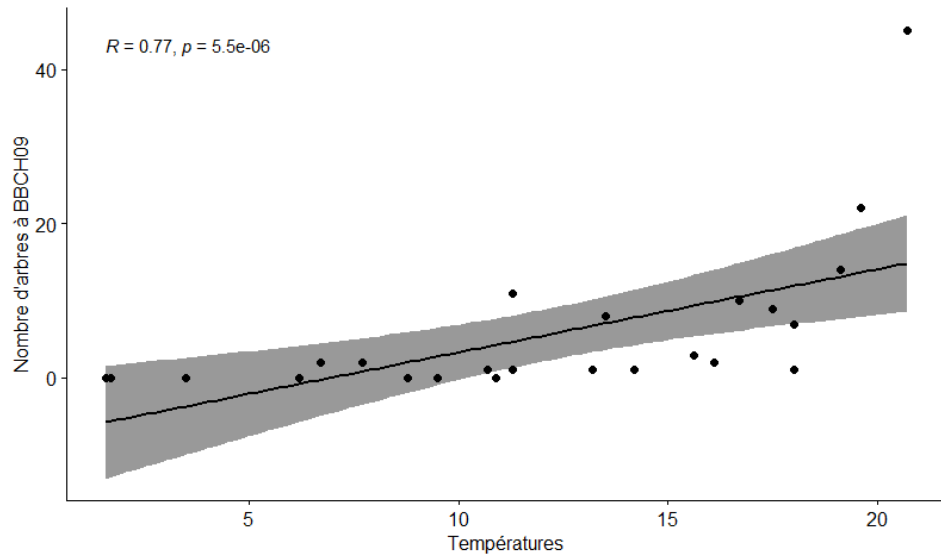
**Tableau 1 :** Tableau des p-values et indices pour les trois tests de corrélation (Spearman, Pearson et Kendall) sur les variables "Phénotypes" et "Dépérissant"

	<b>P-values</b>	<b>Indices de corrélation</b>
<b>Test de Spearman</b>	0.002348	rho : 0.2507632
<b>Test de Pearson</b>	0.001282	cor : 0.2649123
<b>Test de Kendall</b>	0.00262	tau : 0.2449451

### 3.2. Les phénotypes restent corrélés aux températures

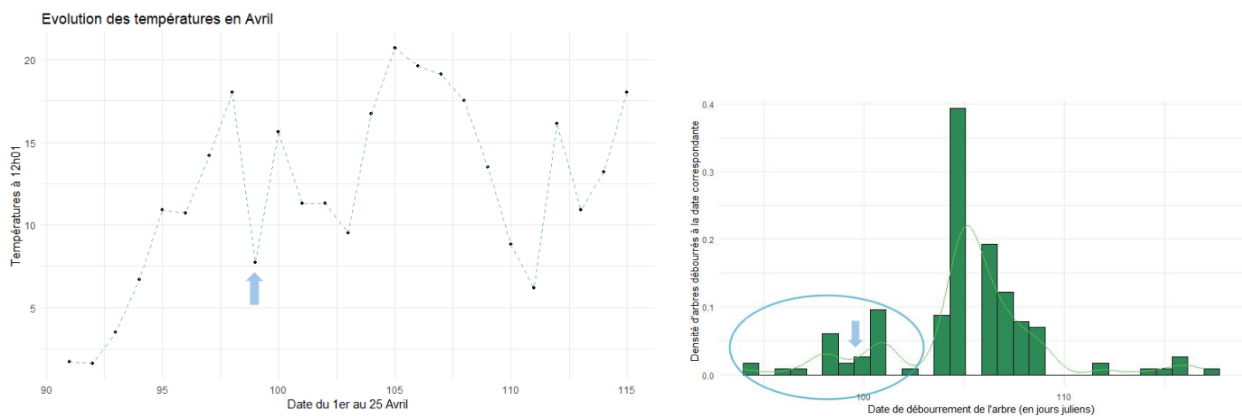
La réserve de La Massane est dotée d'une station météo, enregistrant la pluviométrie et les températures. Une corrélation de Spearman a été faite entre les températures de chaque jour pendant le mois d'Avril et le nombre d'arbres débourrés à la même date. La p-value est inférieure à 0,05 et le rho tend vers 1. Cela signifie qu'une corrélation existe entre ces deux variables et celle-ci est positive, comme en témoigne la figure 7. Chaque phénotype de débourrement est corrélé à un pic d'augmentation de la température locale, comme nous pouvons le voir sur la figure 8.





**Figure 7 :** Graphique de corrélation de Spearman entre les températures et le nombre d'arbres débourrés, aux mêmes dates

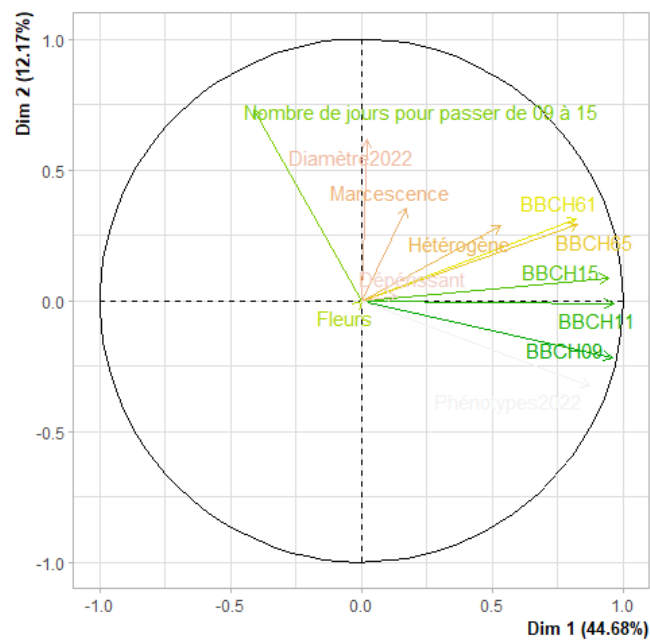
Sur la figure 4, pour le phénotype précoce, une tendance à deux phénotypes se dessine. Cependant ces individus font partie du même phénotype. Cette distribution anormale est probablement due au fait, qu'à la date représentée par la flèche bleue sur la figure 8, il y a eu une baisse des températures. Or il a été vu juste au-dessus que le débourrement est fortement corrélé à la température, et plus exactement à des températures relativement chaudes.



**Figure 8 :** Graphique des températures à la Massane pendant le mois d'Avril avec le graphique de la figure 4

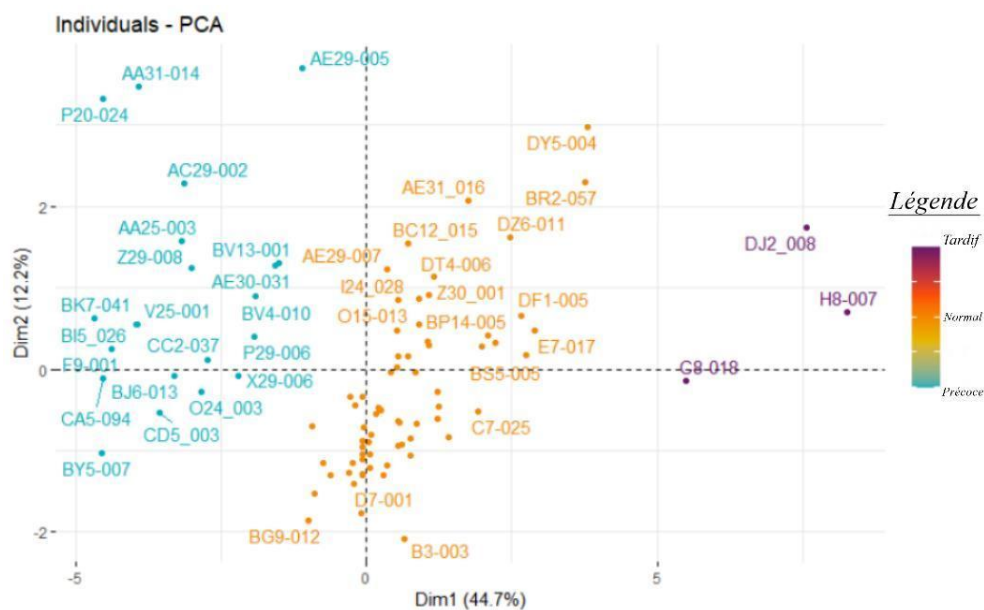
### 3.3. La diversité chez le hêtre est en grande partie expliquée par le statut de débourrement

L'ACP a permis de voir que plusieurs variables étaient corrélées. Sur la figure 9, nous pouvons voir que le premier axe (Dim 1) est principalement expliqué par les stades phénologiques concernant la feuillaison. Cet axe explique 44% de la diversité de la population. Donc 44% de la diversité de cette population de hêtres est expliquée par les différents phénotypes de débourrement et feuillaison. Le second axe est expliqué par le nombre de jours pour passer du stade BBCH 09 au stade BBCH 15, et par le diamètre de l'arbre.



**Figure 9** : Représentation graphique de l'ACP

La figure 10 permet de visualiser plus facilement le fait que la population est divisée en trois grands groupes : en bleu les individus précoces, en orange ceux qui débourrent à une date "normale" et en violet les individus tardifs. Cela confirme le fait que la population est divisée par le phénotype de débourrement.



**Figure 10** : Représentation graphique des coordonnées des individus en ACP

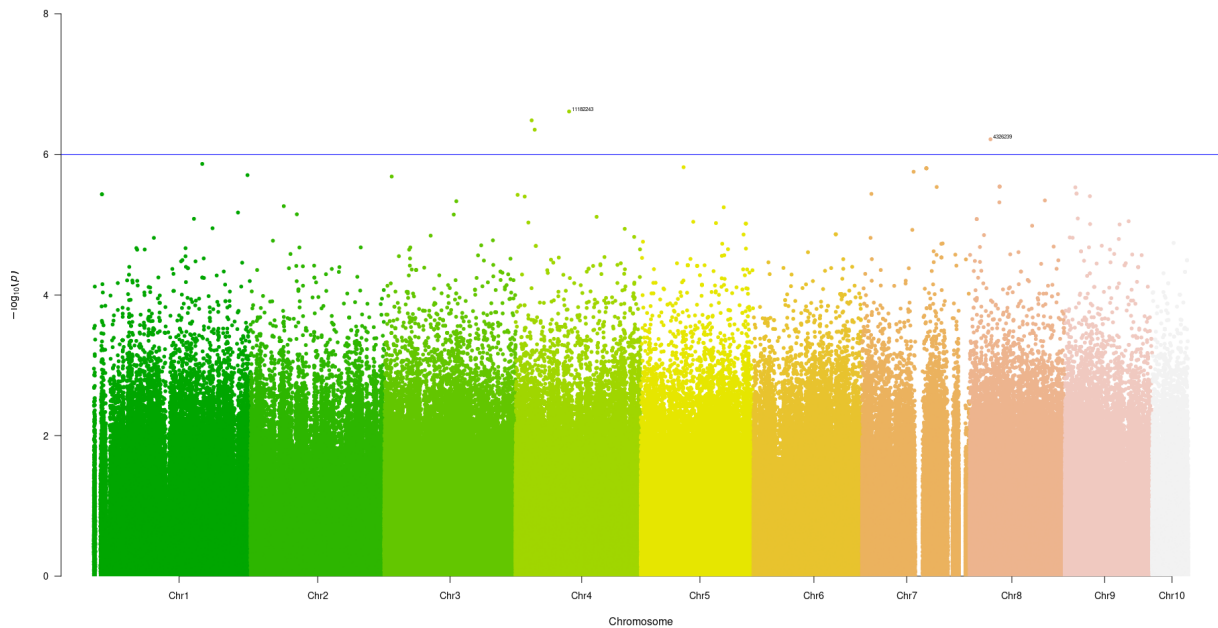
### 3.4. Pas de facteurs de génétiques associés aux phénotypes de débourrement pour le moment

Avant de commencer toute analyse GWAS, il faut connaître la structure de la population. La population de la Massane avait déjà été étudiée à l'aide de marqueurs moléculaires (microsatellites) qui avaient révélé qu'il s'agit d'une seule population (au sens de Hardy-Weinberg)(G. Delafontaine *et al.*, 2013). Une valeur de  $k = 1$  a donc été utilisée pour la suite de mes analyses.

Au début de mon stage, un fichier vcf avait été fourni par les collaborateurs de l'INRAE Avignon. Cependant, il s'est avéré que celui-ci contenait beaucoup trop de données manquantes pour pouvoir mener à bien des analyses GWAS correctes. Nous avons donc dû refaire le SNP calling sur le cluster du laboratoire, ce qui a nécessité une mise au point assez longue. Le SNP calling est en train de tourner et je n'ai donc pas pu réaliser mon étude sur l'ensemble du génome. Néanmoins, afin de pouvoir fournir un résultat pour cette partie de mon stage, j'ai réalisé une analyse GWAS sur les 10 premiers scaffolds des génomes des 150 arbres, qui avaient été analysés.

Le résultat de l'analyse GWAS a été mis sous la forme d'un graphique appelé Manhattan plot (figure 10). Chaque point représente un SNP et chaque couleur correspond à un chromosome. L'axe des ordonnées correspond à la force d'association d'un haplotype (=groupe d'allèles) dans une région génomique qui affecte le phénotype étudié ; cela correspond à la p-value des tests statistiques de l'analyse GWAS. La ligne bleue horizontale représente le seuil de significativité qui

est de  $-\log(10^{-6})$ , donc 6. Plus un pic est haut, plus celui-ci est fortement associé au phénotype. Nous pouvons voir sur la figure 11 que 4 SNP (trois provenant du chromosome 4, et un du chromosome 8) semblent associés aux différents phénotypes du débourrement. Cependant, pour que ces SNP soient significativement corrélés au phénotype, il faudrait qu'ils dépassent un seuil de significativité de  $-\log(10^{-8})$ . Ces données préliminaires ne montrent pas d'association significative entre un SNP et la date de débourrement. Des analyses sur le génome complet seront nécessaires pour conclure sur les bases génétiques du débourrement.



**Figure 10 :** Manhattan plot avec le seuil de significativité du SNP mis à  $-\log(10^{-6})$

#### 4. Discussion

La répartition géographique aléatoire des différents phénotypes de débourrement confirme le fait que les différences de date de débourrement n'ont pas un déterminisme écologique, car si cela avait été le cas, les individus précoces par exemple auraient été regroupés dans une unique zone avec des conditions écologiques particulières.

Une des principales difficultés sur le terrain a été l'estimation de feuilles ouvertes dans le cas d'arbres hétérogènes. En effet, certains arbres avaient des branches considérées comme "précoces", car les feuilles étaient déjà étalées sur celles-ci alors que le reste de l'arbre avait des bourgeons encore fermés. Par conséquent, le stade BBCH 09 n'était pas atteint alors que parfois le stade BBCH 11 l'était déjà.

Trois différents phénotypes pour le débourrement ont été trouvés. Cependant les phénotypes précoces et normaux représentent la plus grande partie de la population (environ 95% à eux deux). Les individus tardifs sont généralement des individus étant dépérissants ; or il n'y a pas assez d'individus pour faire un test statistique pertinent. C'est pourquoi, par la suite, les tests de corrélation ont été faits sur l'ensemble des phénotypes. Les trois tests de corrélation ont été fait pour pouvoir confirmer avec certitude la robustesse de la corrélation.

Nous avons vu que le débourrement est fortement corrélé à l'augmentation des températures. En effet, dans la littérature, il est noté que le débourrement est en partie dû à la température : il faut une température constante au-dessus d'un certain seuil pour que les arbres débourrent (Saxe *et al.*, 2001). De plus, selon Zhang *et al.* (2019), les températures plus chaudes affectent de manière significative les dates de débourrement, en les avançant, chez des espèces de conifères. Cependant dans notre cas, chaque phénotype va débourrer à un pic différent de température. Et c'est pour expliquer cette différence de réponse que l'on recherche des facteurs génétiques.

Pour l'ACP, des individus ont été enlevés pour l'analyse, car il y avait des données manquantes (notées NA). Or il n'est pas possible de faire une ACP avec des données manquantes. Ces données manquantes étaient dues au fait que certains arbres n'avaient pas de dates de floraison. En effet, seuls 75% des arbres étudiés avaient des fleurs, et parmi ceux-là 83% ont atteint le stade BBCH 65. La floraison chez le hêtre, en tout cas dans la forêt de La Massane, n'a pas lieu chaque année et pas chez tous les arbres.

Les premières données génomiques fournies comprenaient énormément de données manquantes. Or il faut une matrice avec un minimum de 5 millions de SNP pour avoir la possibilité

de trouver un SNP associé avec le phénotype, ce qui n'était pas le cas avec le premier fichier. Ce minimum a été calculé avec la taille du génome du hêtre (535 Mb) divisée par le déséquilibre de liaison de celui-ci (~100 kb)(Pfenninger *et al.*, 2021). Le déséquilibre de liaison correspond à une association non aléatoire d'allèles de gènes distincts dans une population. C'est pourquoi un SNP calling a été refait. Cependant c'est un processus lourd en calculs et par conséquent très long. Les données génomiques utilisées pour l'analyse GWAS dans ce rapport sont de très bonnes qualités (moins de 1% de données manquantes) mais il n'y a qu'environ 2,5 millions de SNP. C'est pourquoi une autre analyse GWAS sera nécessaire, avec cette fois-ci l'ensemble du génome de chacun des 150 hêtres séquencés, pour pouvoir conclure sur la présence ou non d'un facteur génétique associé au débourrement précoce chez le hêtre. Le déterminisme génétique pour le débourrement précoce chez le hêtre reste envisageable, car il a été trouvé, chez le peuplier, deux gènes régulant positivement le débourrement : le gène EBB1 (EARLY BUD BREAK 1), régulateur positif du débourrement, qui code un facteur de transcription de la famille des AP2/ERF (ayant un homologue chez *Arabidopsis* régulant le débourrement (régulateur d'activité DORNROSCHEN)) ; et le gène EBB3 (EARLY BUD BREAK 3), sensible à la température, qui est un facteur de transcription aussi de la famille AP2/ERF qui régule directement et positivement le gène CYCLIND3.1 (qui est un promoteur de la progression G1/S du cycle cellulaire dont l'expression est corrélée à la réactivation de la croissance des bourgeons au débourrement)(Azeez *et al.*, 2021).

Enfin, les SNP ne sont pas les seuls facteurs génétiques existants. Les polymorphismes d'insertions d'éléments transposables (TIPs) sont responsables de nombreuses avancées biologiques et peuvent contribuer à la biodiversité des plantes (Akakpo *et al.*, 2019). Les éléments transposables sont des séquences d'ADN capables d'auto-réplication, et ubiquitaires dans les génomes eucaryotes (Akakpo *et al.*, 2019). Il est possible d'étudier l'impact d'insertion d'éléments transposables avec une adaptation du GWAS, qu'on appelle TE-GWAS (Akakpo *et al.*, 2019).

## **5. Conclusion et perspectives**

Trois phénotypes de débourrement ont été trouvés chez le hêtre dans la Réserve de La Massane. Ces phénotypes sont répartis aléatoirement dans la Réserve, donc le déterminisme écologique est écarté. Ces phénotypes sont faiblement corrélés à l'état sanitaire de l'arbre, mais fortement corrélés aux températures locales. La grande majorité de la diversité chez le hêtre est expliquée par le statut de débourrement de l'arbre. La population étudiée est bien une seule et même population, génétiquement parlant. Enfin, la première analyse GWAS, avec une partie des données génomiques, n'a pas révélé de facteurs génétiques associés au débourrement précoce chez le hêtre.

Une autre analyse GWAS avec l'ensemble des données génomiques est nécessaire pour conclure sur les résultats. Il serait intéressant aussi de faire plusieurs TE-GWAS avec différentes familles d'éléments transposables présents chez le hêtre. Enfin, les données récoltées sur le terrain ont été partagées avec le CEFV de Montpellier pour obtenir prochainement des résultats pour la partie modélisation du projet FAGADAPT.

## 6. Bibliographic

- Akakpo, R., Carpentier, M., Ie Hsing, Y., and Panaud, O. (2020). The impact of transposable elements on the structure, evolution and function of the rice genome. *New Phytol* 226, 44–49.
- Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631.
- Azeez, A., Zhao, Y.C., Singh, R.K., Yordanov, Y.S., Dash, M., Miskolczi, P., Stojkovič, K., Strauss, S.H., Bhalerao, R.P., and Busov, V.B. (2021). EARLY BUD-BREAK 1 and EARLY BUD-BREAK 3 control resumption of poplar growth after winter dormancy. *Nature Communications* 12, 1123.
- Chuine, I., and Beaubien, E.G. (2001). Phenology is a major determinant of tree species range. *Ecology Letters* 4, 500–510.
- Chuine, I., Bonhomme, M., Legave, J.-M., García de Cortázar-Atauri, I., Charrier, G., Lacoite, A., and Améglio, T. (2016). Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. *Glob Change Biol* 22, 3444–3460.
- Duputié, A., Rutschmann, A., Ronce, O., and Chuine, I. (2015). Phenological plasticity will not help all species adapt to climate change. *Global Change Biology* 21, 3062–3073.
- Garrigue, J. (2017). Plan de gestion 2017-2027 de la Réserve Naturelle Nationale de La Massane.
- Garrigue, J., Magdalou, J.-A., and Sorel, D. (2018). Climatologie, analyse des données de 2017, Travaux de la Massane Tome 112.
- Garrigue, J., Magdalou, J.-A., and Grel, A. Plan de gestion 2010-2014 de la Réserve Naturelle Nationale de La Massane.
- Jump, A.S., Mátyás, C., and Peñuelas, J. (2009). The altitude-for-latitude disparity in the range retractions of woody species. *Trends in Ecology & Evolution* 24, 694–701.



de Lafontaine, G., Ducouso, A., Lefèvre, S., Magnanou, E., and Petit, R.J. (2013). Stronger spatial genetic structure in recolonized areas than in refugia in the European beech. *Mol Ecol* 22, 4397–4412.

Le collectif scientifique de l'Observatoire des Saisons (2017). *Les plantes au rythme des saisons, Guide d'observation phénologique* (Biotope Editions).

McKown, A.D., Klápště, J., Guy, R.D., El-Kassaby, Y.A., and Mansfield, S.D. (2018). Ecological genomics of variation in bud-break phenology and mechanisms of response to climate warming in *Populus trichocarpa*. *New Phytol* 220, 300–316.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., et al. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* 51, 933.

Pfenninger, M., Reuss, F., Kiebler, A., Schönnenbeck, P., Caliendo, C., Gerber, S., Cocchiara, B., Reuter, S., Blüthgen, N., Mody, K., et al. (2021). Genomic basis for drought resistance in European beech forests threatened by climate change. *ELife* 10, e65532.

Saxe, H., Cannell, M.G.R., Johnsen, Ø., Ryan, M.G., and Vourlitis, G. (2001). Tree and forest functioning in response to global warming. *New Phytologist* 149, 369–399.

Zhang, S., Isabel, N., Huang, J.-G., Ren, H., and Rossi, S. (2019). Responses of bud-break phenology to daily-asymmetric warming: daytime warming intensifies the advancement of bud break. *International Journal of Biometeorology* 63, 1631–1640.

## 7. Webographie

**Webréférence 1** : site internet Larousse : [biome - LAROUSSE](#), dernière consultation le 12/05/2022

**Webréférence 2** : site internet de l'Université du Muséum de Paléontologie de Californie : [The forest biome \(berkeley.edu\)](#), dernière consultation le 12/05/2022

**Webréférence 3** : site de l'Inventaire National du Patrimoine Naturel pour la fiche espèce du Hêtre des forêts : [Fagus sylvatica L., 1753 - Hêtre des forêts, Hêtre, Fayard, Hêtre commun, Fouteau-Présentation \(mnhn.fr\)](#), dernière consultation le 12/05/2022

**Webréférence 4** : site de l'Inventaire National du Patrimoine Naturel pour l'inventaire ZNIEFF : <https://inpn.mnhn.fr/programme/inventaire-znieff/presentation>, dernière consultation le 12/05/2022

## **ANNEXE**

Dossier Drive avec (1) la base de données de terrain, (2) les scripts personnels pour les analyses statistiques et (3) les scripts de Roland Akakpo pour le GWAS :

<https://drive.google.com/drive/folders/17adCccJOzmwelJAchEyf3Jj8C4QUjHjE?usp=sharing>

## Abstract

With the increase in temperatures due to climate change, questions related to the resilience of ecosystems are increasingly being asked. The distribution zones of plant species change, and the phenology of species is modified. Temperate broadleaf and mixed forests, such as beech forests (*Fagus sylvatica L.*), are the majority biome in France and are also strongly affected by this climate change. Trees have cyclic events differentiated into phenological stages, such as budbreak. Differences in budbreak date have been observed in beech in the Massane Reserve, which could be the result of local adaptation. We studied this beech forest with statistical analysis and GWAS approaches, with the aim of finding a genetic factor that could explain this difference in budbreak date. We were able to confirm the presence of “early” and “normal” budbreakers. This difference in budbreak is not due to an ecological determinism, but there is a correlation with temperatures. However, preliminary results show that there is no genetic factor associated with different budbreak dates.

**Key words :** Genome-Wide Association Studies (GWAS), phenology, climate change, adaptation, European beech (*Fagus sylvatica L.*)

## Résumé

Avec l'augmentation des températures due au changement climatique, les questions liées à la résilience des écosystèmes se posent de plus en plus. Les zones de distribution des espèces végétales changent, et la phénologie des espèces est modifiée. Les forêts tempérées caducifoliées mixtes, telles que les forêts de hêtres (*Fagus sylvatica L.*), sont le biome majoritaire en France et sont elles aussi fortement touchées par ce changement climatique. Les arbres ont des événements cycliques différenciés en stades phénologiques, tels que le débourrement. Il a été remarqué des différences de date de débourrement chez le hêtre dans la Réserve de La Massane, qui pourraient être le fruit d'une adaptation locale. Nous avons étudié cette hêtraie avec des analyses statistiques et des approches GWAS, dans le but de trouver un facteur génétique pouvant expliquer cette différence de date de débourrement. Nous avons pu confirmer la présence d'individus “précoces” et “normaux” en ce qui concerne le débourrement. Cette différence de débourrement n'est pas due à un déterminisme écologique, mais il existe une corrélation avec les températures. Cependant, les résultats préliminaires montrent qu'il n'y a pas de facteur génétique associé aux différentes dates de débourrement.

**Mots-clés :** Genome-Wide Association Studies (GWAS), phénologie, réchauffement climatique, adaptation, hêtre Européen (*Fagus sylvatica L.*)