



**Analyse du polymorphisme nucléotidique du gène
Phytochrome B2 chez le peuplier noir (*Populus nigra* L.),
espèce de la forêt alluviale :
Comparaison de deux populations sauvages.**

Crépin
Manon

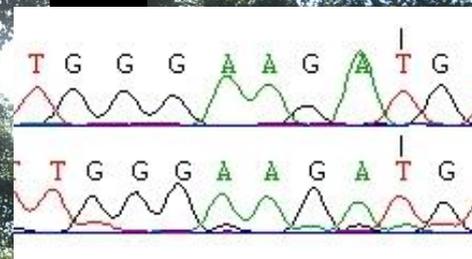
Da Silva
Théo

Fazillaud
Rachel

Galinat
Cédric

Jarry Anaïs

1^{ère} Science et technique de
laboratoire Option :
BIOTECHNOLOGIE
2018



LGT Jay de Beaufort – Périgueux



Remerciements

Nous remercions :

Mme Patricia FAIVRE-RAMPANT (patricia.faiivre-rampant@inra.fr) Chercheur scientifique –EPGV Etude des polymorphisme de génomes de végétaux Evry – France,

M. Marc VILLAR, (marc.villar@inra.fr) : Directeur de Recherche à l'Unité de Recherche Amélioration, Génétique et Physiologie Forestières, rattachée au département Ecologie des Forêts, Prairies et milieux Aquatiques (EFPA),

qui ont accepté d'être nos référents scientifiques dans cette étude. Merci pour leur conseils avisés.

Mme Corinne CRUAUD et M. Frédéric Gavory du CEA–Institut de Génomique, GENOSCOPE (Evry), pour le séquençage et l'envoi des séquences.

M.GALLET, notre professeur de biotechnologies qui a cru en nous pour la réalisation de cette activité. Il a su nous motiver pour participer à cette étude scientifique dans le cadre de l'atelier scientifique et technique de notre lycée.

M. Lapeyre, professeur de physique, qui intervient dans le fonctionnement de cet atelier scientifique.

M. Laurent SUAOU, professeur de biotechnologies, et ses élèves du lycée Raoul Dautry de Limoges. Ils ont bien voulu nous donner leurs séquences d'ADN à analyser pour réaliser cette étude, ainsi que les documents précisant la localisation de leurs arbres.

Nous remercions aussi "Science à l'école" qui permet de réaliser ces manipulations par le prêt de matériel et l'achat de consommable.

Sommaire

Résumé.....	4
Abstract	4
Introduction	5
La problématique.....	7
Place de notre travail dans cette expérimentation	8
Les problèmes rencontrés.....	8
RESULTATS	10
A) Qualité et longueur des séquences analysées	10
B) Etude des SNPs	10
Cas des peupliers DDG	11
Cas des peupliers RD_ND_Pn	11
C) Comparaison des deux populations.....	13
D) Spécificité de la technique	13
E) Conséquence de ces SNPs sur la protéine phytochrome B2	14
Conclusion.....	15
Bibliographie.....	16
Annexe 1 : Description botanique de <i>Populus nigra</i>	17
Annexe 2 : Les Phytochromes et leurs rôles : quelques données.....	18
Annexe 3 : Structure et séquences complètes du gène Phy B2 chez <i>P. trichocarpa</i>	19
étudié dans le projet <i>Populus</i> (extrait du document génome à l'école)	19
Annexe 4 :Les marqueurs du polymorphisme génétique	21
Annexe 5 : Le séquençage.....	22
Annexe 6 : Lecture et correction des électrophorègrammes	23
Annexe 7 : Analyse des séquences par le logiciel ‘GENALYS™’ : le code IUPAC.....	24

Résumé

Longtemps confondu et mis à l'écart par le peuplier hybride de culture, le peuplier noir (*Populus nigra* L.) suscite un regain d'intérêt à la fois écologique et génétique. Cette essence est en danger. Elle doit être étudiée pour mieux connaître son potentiel génétique : étape nécessaire à la construction de nouveaux hybrides mais aussi pour la préservation des populations naturelles existantes le long des fleuves et cours d'eau.

Notre travail s'inscrit dans le projet "Génome à l'école". L'analyse des SNPs présents dans la partie du gène Phy B2 analysée nous permet de conclure à une certaine homogénéité entre les deux populations de peupliers noirs que nous avons étudiée.

La traduction de la partie génétique étudiée permet de constater que les SNPs notés n'ont pas d'incidence sur la structure primaire de la protéine codée par le gène Phy B2.

L'absence de données phénologiques ne nous a pas permis de mettre en relation les différents génotypes trouvés avec les phénotypes correspondants.

Abstract

For a long time confused by the hybrid poplar, the autochthonous black poplar (*Populus nigra*) is gaining more and more interest both at the ecological and genetic point of view. This tree must be studied for a better knowledge of its genetic potential, necessary step for construction of new interspecific hybrid but also for conservation of natural populations along riversides. Our work is included in the project "Génome à l'école". The analysis of the SNPs present in the part of PhyB2 gene analysed permit to conclude that the two poplar populations studied have a certain homogeneity. The genetic translation of the part of PhyB2 studied gene permit to note that the repered SNPs have no repercution in the primary structure of the protein which is coded by the Phy B2 gene. The absence of phenological datas do not permit to search for relationships between the different studied genotypes and the corresponding phenotypes.

Introduction

Les peupliers appartiennent à la famille des Salicacées, genre *Populus* ; ils sont originaires de l'hémisphère nord. En France, trois espèces autochtones sont présentes : le peuplier noir (*Populus nigra* L.) et le peuplier blanc (*Populus alba* L.) rencontrés le long des principaux cours d'eau et le peuplier tremble (*Populus tremula* L.) qui vit en forêt. Dans notre paysage existe également un peuplier noir issu d'Italie, de forme fastigiée, utilisé comme arbre d'ornement mais surtout des plantations de cultivar de peupliers pour l'industrie. Ces arbres, à croissance rapide, sont cultivés pour leur bois qui peut être déroulé ou transformé en papier. La France est le premier producteur de bois de peuplier en Europe. Le peuplier fait donc l'objet d'une exploitation intensive (environ 250 000 ha).

Et le peuplier noir dans tout cela ...

Le peuplier noir : *Populus nigra* (annexe 1) appartient au genre *Populus*, famille des Salicacées.

Le peuplier noir n'a pas de valeur économique, ses principaux intérêts sont les suivants :

- génétique. La populiculture utilise des cultivars¹ provenant de deux croisements :
 - « interaméricains » issus de l'hybridation entre *Populus deltoides* x *Populus trichocarpa* ;
 - « euraméricains » issus de l'hybridation entre *Populus nigra* x *Populus deltoides* . Ces

hybrides cumulent une partie (la moitié) des caractéristiques génétiques de chacun des parents : croissance rapide, peu de branches, rectitude, qualité du bois, résistance aux pathogènes,... . Dans les hybrides « euraméricains » *Populus nigra* apporte sa rusticité vis-à-vis de l'environnement, sa résistance au chancre bactérien et au virus de la mosaïque, sa moindre sensibilité à la rouille brune (*Marsonina brunnea*) et sa grande aptitude au bouturage.

Les peupleraies sont des plantations clonales de cultivars. Les clones plantés sont parfois très proches génétiquement les uns des autres. Cet excès de quasihomogénéité génétique rend la peupleraie très fragile aux attaques parasitaires. En effet, les parasites s'adaptent et contournent les résistances des cultivars employés : la dernière race de rouille a réussi à se répandre en Europe en seulement 5 ans. Il est donc nécessaire de diversifier la base génétique exploitée pour la réalisation des hybrides.

- écologique. *P. nigra* est l'espèce dominante (en taille, en âge) d'un habitat naturel particulier : la ripisylve (forêt alluviale à bois tendre). Elle est le support d'une biodiversité remarquable. En effet, après plusieurs années de croissance, le peuplier noir présente de nombreuses caches et cavités naturelles qui offrent un hébergement à de nombreux insectes, chauves souris et oiseaux. De plus, son système racinaire très développé lui permet de piéger les sédiments, fixer les berges et épurer le milieu par la fixation de polluants issus de l'agriculture ou des stations d'épurations situées en amont (nitrates, phosphates, ...).

Mais le peuplier noir est une espèce végétale menacée :

- Problèmes d'habitat : le peuplier noir est présent naturellement dans les forêts riveraines ou ripisylves. L'activité de l'homme (nettoyage des berges, canalisation des cours d'eau,

¹ Un *cultivar* est une variété de plante (arbres compris) obtenue en culture, généralement par sélection, pour ses caractéristiques « réputées uniques ». Il peut s'agir de qualités esthétiques, techniques, de vitesse de croissance (pour les arbres par exemple), d'adaptation à un biotope ou de résistance à certaines maladies. Chez le peuplier, obtenus par hybridation, ces cultivars ne peuvent pas se multiplier notamment par semis puisque la seconde génération sera dégénérée. Ils sont multipliés par voie végétative (bouturage, ...).

perturbations des régimes hydraulique) entraîne une diminution de cette ripisylve et donc la disparition du peuplier noir.

- Problème de pollution génétique : le peuplier noir peut s'hybrider avec de nombreuses autres espèces de peuplier entraînant de ce fait un appauvrissement de son génome.

Quelles actions pour protéger le peuplier noir ?

L'Etat français a pris conscience de l'intérêt de cette espèce et a toujours soutenu son programme de conservation au sein de la Commission des Ressources génétiques forestières (CRGF). De plus, ce programme de conservation s'inscrit dans une composante européenne : le réseau EUFORGEN, (<http://www.euforgen.org>) permettant une cohérence des études à l'échelle de l'aire naturelle de l'espèce.

Les actions suivantes ont été mise en place :

- Répertoire et conserver les peupliers noirs remarquables présents en France.

Suite aux menaces exercées sur le peuplier noir et son habitat, un programme national de conservation des ressources génétiques a été défini et mis en place par la commission des Ressources génétiques forestières (CRGF) en 1992 sous l'égide du ministère de l'Agriculture et de l'Alimentation. L'objectif principal est de conserver les gènes fondateurs de la variabilité actuelle et de préserver au mieux les adaptations locales comme les mécanismes naturels qui la sous-tendent. Deux programmes de conservation ont été mis en place :

- Programme de conservation *ex situ* : une collection de copies végétatives de Peuplier noir représentant des origines géographiques diverses est actuellement conservée au Conservatoire national de la biodiversité forestière (CNBF) de Guéméné Penfao. Ces individus sont des copies végétatives soit d'arbres isolés remarquables, soit d'individus issus de populations naturelles.
- Programme de conservation *in situ* : il permettra de sauvegarder et de maintenir sur le long terme l'espèce et sa diversité. Les sites naturels retenus doivent permettre de conserver le potentiel d'adaptation de cette espèce sur le long terme (par le jeu de la régénération par voie sexuée) tout en le laissant évoluer dans son milieu en phase avec l'évolution de l'environnement.

- Préservation et revégétalisation des berges,
- Actions de communication autour de l'espèce *P.nigra* (<http://peupliernoir.oreans.inra.fr>),
- Formation des personnels de terrain : techniciens de rivière, ...
- Actions pédagogiques : mise en place de *Populetum* montrant la diversité génétique intraspécifique du peuplier noir, ...

M. Marc Villar : UMR INRA-ONF BioForA : Biologie Intégrée pour la valorisation de la diversité des Arbres et de la Forêt (Centre Inra Val de Loire, site d'Orléans) est l'animateur de ce programme de conservation.

La problématique

Notre activité s'inscrit dans le cadre de "Science à l'école" projet "Génome à l'école" auquel nous participons depuis sa création en 2011.

Les questions que nous nous sommes posées sont les suivantes :

- Est-ce que la population française de peuplier noir est génétiquement homogène?

1. Nous analyserons, à notre échelle, une partie de la séquence du gène qui code pour le phytochrome B2. Les phytochromes sont des protéines qui interviennent entre autre dans la floraison des plantes donc dans le démarrage de la végétation après la période de repos hivernale (voire annexe 2). Le gène du phytochrome B2 comporte 4 exons et 3 introns. La séquence nucléotidique analysée est contenu dans l'exon 2 du gène du phytochrome B2 comme cela est montré dans l'annexe 3. La séquence étudiée a été nommée : Phy B2, en rapport avec le gène qui la contient. Cette séquence Phy B2 a une taille de 663 nucléotides pour une séquence codante de 3639 nucléotides (exon 1 + ex 2 + ex 3 + ex 4). Le gène fait 7448 nucléotides de long.

2. Nous utiliserons le polymorphisme génétique pour différencier les individus. Plusieurs marqueurs du polymorphisme génétiques existent (annexe 4). Dans le cadre du peuplier noir nous utiliseront les SNPs (single nucleotid polymorphism).

Ce travail sera réalisé sur deux populations sauvages de peupliers noirs

- Quel est l'impact de la génétique sur le phénotype ?

Une étude de la phénologie de floraison, réalisée en 2006, a montré que le pic de floraison des peupliers noirs d'une même population n'est pas forcément synchrone : des différences de 3 à 4 semaines entre arbres voisins ont été notées.

Nous espérons donc mettre en évidence une relation entre les génotypes trouvés et les phénotypes (date de floraison) notés sur le terrain pour les arbres étudiés.

Place de notre travail dans cette expérimentation

Notre travail est la suite logique du travail effectué par les élèves des années précédentes. Compte tenu de la lourdeur et des difficultés de la manipulation, il est difficile de réaliser cette étude sur une seule année scolaire. Par conséquent, les séquences d'ADN analysées ont été obtenues par les élèves des promotions 2014/15 et 2015/16. Ces derniers ont réalisé :

- l'extraction de l'ADN des feuilles de peuplier (14 arbres);
- l'amplification de la séquence Phy B2 par PCR ;
- l'électrophorèse qui permet de vérifier la qualité des amplifications (présence d'un seul amplicon de la bonne longueur) ;
- envoi des échantillons au Génomètre (Evry) pour le séquençage (annexe 5).

Notre activité correspond à la dernière étape de l'analyse :

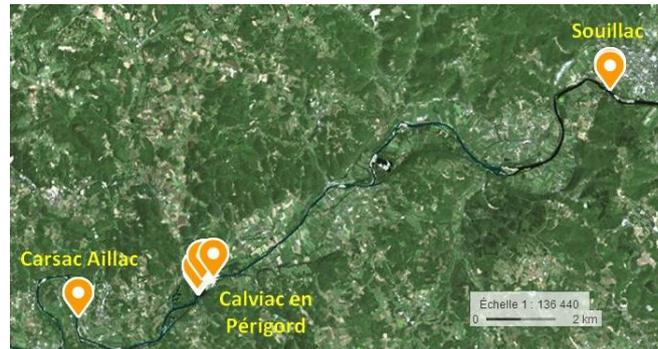
- à partir des électrophorégrammes réalisés et restitués par le Génomètre, vérifier et corriger les séquences nucléotidiques obtenues en utilisant le logiciel "GENALYS" (annexe 6) ;
- comparer les différentes séquences obtenues pour tenter de répondre à notre problématique.

Les problèmes rencontrés

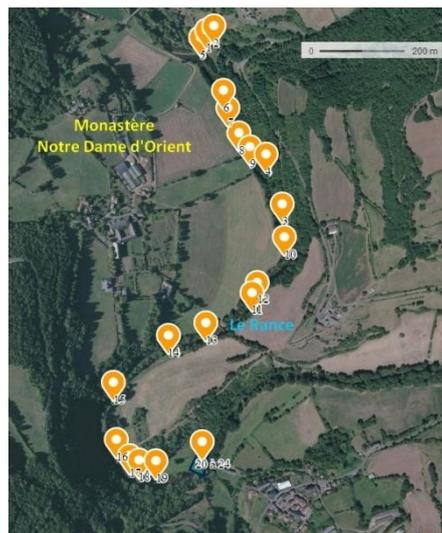
Nos problèmes ont été les suivants :

- un manque de nombreuses connaissances et compétences. Donc, dès le mois de Septembre, notre enseignant:
 - nous a présenté le projet génome dans sa globalité ;
 - nous a expliqué tout le travail et toutes les techniques utilisées qui ont permis d'obtenir ces séquences nucléotidiques par nos prédécesseurs. Nous n'avons pas fait les manipulations à ce niveau de l'année. Ces techniques seront abordées entre le mois de Février et le mois d'Avril dans nos séquences de travaux pratiques de STL "biotechnologie" ;
 - nous a donné un cours sur la structure et les propriétés de l'ADN nécessaires pour comprendre ce que nous allons faire ;
 - nous a appris à utiliser le logiciel "Génalys" : logiciel public développé pour l'analyse du polymorphisme nucléotidique chez l'espèce humaine ;
 - nous a appris à lire les électrophorégrammes obtenus après séquençage (annexe 5) et la conduite à tenir devant une erreur d'interprétation du tracé par ce logiciel (annexe 6) et le fonctionnement du code IUPAC (annexe 7) ;
 - nous avons aussi utilisé un site de cartographie en ligne créé par l'IGN et le BRGM : "Géoportail" qui est un portail Web public permettant l'accès à des services de recherche et de visualisation de données géographiques ou géolocalisées.
- avoir deux populations sauvages de peupliers noirs suffisamment éloignées géographiquement. Ces deux populations sont les suivantes :
 - 14 peupliers noirs provenant de la pépinière expérimentale de Guéméné Penfao. Ces peupliers proviennent de la vallée de la Dordogne entre Souillac (Lot) et Carsac Aillac (Dordogne). Ces peupliers sont référencés de la façon suivante : **DDG Pn xxx**. Ces peupliers nous ont été fournis

sous la forme de boutures que nous avons mis en jauge et fait pousser au lycée (en 2015 et 2016). Après les essais, ces boutures ont été plantées chez M. GALLET (notre professeur de biotechnologie). Nous disposons donc encore de notre matériel végétal pour les études ultérieures.



- 19 peupliers noirs étudiés par les élèves de STL option "biotechnologie" du lycée Raoul Dautry de Limoges qui participent eux aussi au projet "Génome à l'école". M. L. SUAU "professeur de biotechnologie" nous a transmis les séquences Phy B2 non nettoyées de ses arbres. Ces peupliers sont référencés de la façon suivante : **RD ND Pn xxx**
Ces arbres proviennent de la vallée du Rance, à Notre Dame d'Orient, commune de Pousthomy,



Ces deux populations de peupliers noirs sont distantes de 160 km.

- avoir des électrophorègrammes correspondant à des essences végétales autres que *P. nigra* pour s'assurer de la spécificité de notre travail. Ces données étaient disponibles, en effet, dans les études antérieures ont été étudiées :

- 1 peuplier hybride euraméricain (cultivar I 45/51). Il est issu du croisement entre *Populus deltoides* (♀) x *Populus nigra* (♂). Il provient du jardin de M. GALLET. Il est noté **Phy**. Il contient normalement la moitié des gènes de *P.nigra* ;
- 1 peuplier blanc *Populus alba* provenant du bord du Caudeau, commune de Saint Sauveur (Dordogne) (coordonnées : E 0,56748 ; N 44,884848). Il est noté **Pa**. Il appartient au genre *Populus* ;
- 1 saule (*Salix*) de provenance inconnue. Il est noté **Sa**. Il appartient à la famille des Salicacées.

- disposer des données phénologiques qui correspondent à la date de floraison des peupliers noirs analysés.

RESULTATS

Nous avons analysé la séquence Phy B2 de 36 arbres.

A) Qualité et longueur des séquences analysées

Pour chaque arbre analysé, nous avons obtenu une séquence "f" et une séquence "r" correcte.

La séquence de référence a une taille de 663 bases. Sur les 36 séquences analysées, 21 (58.3%) nous ont permis d'avoir la séquence complète Phy B 2.

Pour 2 Pn (RD_ND_Pn_009 et 021) les électrophorégrammes obtenus sont de mauvaises qualités en 5' et en 3'. Il manque 40 bases en 5' pour ces deux séquences et respectivement 40 et 30 bases en 3'. Les résultats obtenus pour ces arbres sont donc incomplets.

Pour les séquences où il manque moins de 10 nucléotides en 5' ; soient les arbres Sa, PHy, DDG Pn 016/030/034, RD_ND_Pn 005/008/013/020 et pour les séquences où il manque moins de 43 nucléotides en 3' soient les arbres : Sa, DDG Pn 030/033/034/038, RD_ND_Pn 004/005/006/009/013/021, nous verrons dans le paragraphe suivant que ces séquences plus courtes peuvent être intégrées à notre analyse.

Tableau 1		Séquence F		Séquence R		Longueur de la séquence analysée	
		Début (5')	Fin (3')	Début(5')	Fin (3')	Début (5')	Fin (3')
Sa		38	659	8	618	8	659
Pa		34	663	1	624	1	663
PHy		41	663	8	637	8	663
DDG Pn	001	38	663	1	629	1	663
	003	58	663	1	631	1	663
	016	35	663	3	629	3	663
	025	42	663	1	622	1	663
	026	43	663	1	621	1	663
	027	47	663	1	621	1	663
	028	41	663	1	621	1	663
	029	43	663	1	628	1	663
	030	59	661	10	620	10	661
	031	40	663	1	631	1	663
	032	41	663	1	629	1	663
	033	38	658	1	621	1	658
	034	40	567	3	621	3	621
038	42	647	1	621	1	647	
RD_ND_PN	001	31	663	1	630	1	663
	002	41	663	1	631	1	663
	003	34	663	1	631	1	663
	004	41	659	1	630	1	659
	005	41	662	2	637	2	662
	006	37	658	1	631	1	658
	007	34	663	1	636	1	663
	008	41	633	10	621	10	633
	009	41	573	145	620	41	620
	011	41	663	1	631	1	663
	013	41	654	2	621	2	654
	014	43	663	1	631	1	663
	015	40	663	1	631	1	663
	016	41	663	1	631	1	663
	018	31	663	1	629	1	663
	020	40	663	10	636	10	663
021	41	570	151	634	41	634	
023	40	663	1	633	1	663	
024	41	663	1	633	1	663	

B) Etude des SNPs

Les SNPs repérés sont catalogués dans le tableau 2.

Hormis les arbres RD_ND_Pn 009 et 021 pour lesquels la séquence analysée commence au nucléotide 41, tous les autres présentent un SNP à la position 14. Parmi ceux-ci, un seul individu (DDG 038) présente un T à la place d'un G, tous les autres (33) présentent un A. Ce SNP est aussi observé pour les autres essences (Sa, Pa et PHy).

Cette position de SNP ne peut donc pas être utilisée pour caractériser nos individus dans les analyses ultérieures.

Cas des peupliers DDG (en mauve dans le tableau 2)

Fournis par la pépinière expérimentale de Guémené Penfao, nous sommes assurés de leur identification. Ils vont nous servir de référence pour notre analyse.

Synthèse des résultats : position des SNPs	46	79	106	Nb d'individus concernés	Individus concernés : DDG
Nucléotides de référence	C	G	G	3	028, 031, 032
SNPs observés	T	A	A	5	001, 003, 025, 033, 034
	Y	R	R	5	016, 026, 027, 029, 030
	Y	A	A	1	038

Sur 14 peupliers noirs analysés :

- trois peupliers noirs ne présentent pas de SNPs. Les nucléotides rencontrés sont identiques à ceux du gène de référence. De plus, ils sont homozygotes.

- dix peupliers noirs présentent 3 SNPs. Ceux-ci sont toujours situés en position 46, 79 et 106. Trois cas sont présents :

- 5 peupliers présentent la combinaison T A A : ils sont homozygotes ;
- 5 peupliers présentent la combinaison Y R R : ils sont hétérozygotes. Selon le code IUPAC(annexe 7), Y correspond à C ou T et R correspond à A ou G) ;

- 1 atypique qui présente 5 SNPs en position 9, 14, 46, 79 et 106. Le SNP en position 9 n'a pas été retrouvé chez les autres individus de notre analyse quant au SNP en position 14 nous avons trouvé un T au lieu de A comme chez tous les autres individus. S'agit-il d'erreurs d'interprétation de notre part sur la lecture des électrophorogramme ou d'un peuplier qui serait issu d'une hybridation avec un autre individu d'une autre espèce de peupliers ou de cultivars voisins ?

Il semblerait donc qu'un peuplier noir se caractérise soit par l'absence totale de SNP soit par la présence de 3 SNPs toujours positionnés au même endroit : 46, 79, 106.

Cas des peupliers RD ND Pn (en vert dans le tableau 2)

Les séquences Phy B2 fournies par les élèves du lycée R. Dautry de Limoges présentent les mêmes caractéristiques que les séquences Phy B2 des peupliers DDG étudiés précédemment : soit une absence de SNPs, soit la présence de trois SNPs situés aux positions 46, 79 et 106.

L'analyse des génotypes montre la présence de trois combinaisons différentes :

Synthèse des résultats : position des SNPs	46	79	106	Nb d'individus concernés	Individus concernés : RD_ND_Pn
Nucléotides de référence	C	G	G	12	001, 002, 003, 004, 005, 007, 009, 011, 013, 014, 016, 018
SNPs observés	T	A	A	2	006, 008
	Y	R	R	5	015, 020, 021, 023, 024

Nous n'avons pas observé de sujets atypiques dans cette dernière population.

Tableau 2	: Catalogue des SNPs repérés																																															
Position sur le gène de référence	9	14	28	46	47	51	67	71	79	93	96	106	124	130	191	203	208	254	275	311	335	340	344	386	391	429	431	446	447	453	462	478	479	480	532	535	546	556	559	574	583	586						
Base sur le gène de référence	A	G	A	C	G	T	T	G	G	C	C	G	G	A	G	T	T	C	C	T	A	A	A	G	A	T	C	A	C	A	A	A	A	A	T	G	G	C	A	A	T	G						
sa		A	G		R	K	C	R	R	T	T		A	M	A	K	Y	T		Y	R	C	T		R	G	T	G		M	G	W	T	G	C	A	A	M	G	R	A	A						
Pa		A						A											T						A		G			G						G	A	A	G				A					
PHy		A		Y			Y		A			R															K									R	R	R										
ddg Pn 001		A		T					A			A																																				
ddg Pn 003		A		T					A			A																																				
ddg Pn 016		A		Y					R			R																																				
ddg Pn 025		A		T					A			A																																				
ddg Pn 026		A		Y					R			R																																				
ddg Pn 027		A		Y					R			R																																				
ddg Pn 028		A																																														
ddg Pn 029		A		Y					R			R																																				
ddg Pn 030		A		Y					R			R																																				
ddg Pn 031		A																																														
ddg Pn 032		A																																														
ddg Pn 033		A		T					A			A																																				
ddg Pn 034		A		T					A			A																																				
ddg Pn 038	C	T		Y					A			A																																				
RD_ND_Pn_001		A																																														
RD_ND_Pn_002		A																																														
RD_ND_Pn_003		A																																														
RD_ND_Pn_004		A																																														
RD_ND_Pn_005		A																																														
RD_ND_Pn_006		A		T					A			A																																				
RD_ND_Pn_007		A																																														
RD_ND_Pn_008		A		T					A			A																																				
RD_ND_Pn_009		⊗																																														
RD_ND_Pn_011		A																																														
RD_ND_Pn_013		A																																														
RD_ND_Pn_014		A																																														
RD_ND_Pn_015		A		Y					R			R																																				
RD_ND_Pn_016		A																																														
RD_ND_Pn_018		A																																														
RD_ND_Pn_020		A		Y					R			R																																				
RD_ND_Pn_021		⊗		Y					R			R																																				
RD_ND_Pn_023		A		Y					R			R																																				
RD_ND_Pn_024		A		Y					R			R																																				

A partir de notre gène de référence et à partir de ces deux populations de peupliers noirs (DDG et RD_ND_Pn) nous pouvons dire que génétiquement un peuplier noir se caractérise soit par l'absence de SNPs, soit par la présence de 3 SNPs positionnés en 46, 79 et 106.

Lorsque les SNPs sont notés alors trois combinaisons sont observées T,A,A ou Y,R,R ou Y,A,A aux positions respectives 46 , 79, 106. La combinaison Y,A,A a été noté chez un seul individu sur les 33 analysés.

C) Comparaison des deux populations

La comparaison de ces deux populations permet de constater que :

- dans la population DDG seulement 21.4% ne présentent pas de SNPs alors que dans la population RD_ND_Pn ce taux est de 63,2%.
- ces deux populations de peupliers noirs présentent les mêmes SNPs malgré leur éloignement (160 km).
- ce qui distingue ces deux populations c'est l'importance relative de chacun des génotypes de SNPs :
 - o 50% de T,A,A et 50% de Y,R,R pour la population DDG ;
 - o 28,5% de T,A,A et 71,5% de Y,R,R pour la population RD_ND_Pn. La présence d'un ilot (5 arbres très proches RN_ND_20 à 24 comme on peut le voir sur la carte page 8) favorisant dans ce cas la combinaison Y,R,R.

Dans ces deux populations, il est possible de distinguer 3 catégories de peupliers noirs :

- ceux ne présentant pas de SNPs sont des homozygotes identiques à la référence C G G ;
- ceux présentant trois SNPs donc en position 46, 79, 106 :
 - les homozygotes T A A ;
 - les hétérozygotes Y R R ;
 - et un atypique.

Les proportions de chacune de ces catégories varient entre les deux populations même si nous avons peu d'individus pour chacune d'elles. Nous rencontrons bien les même génome mais dans des proportions différentes.

D) Spécificité de la technique

Nous avons inclus dans nos essais trois arbres d'essences différentes et plus ou moins proches de *Populus nigra*. Il s'agit d'un saule (*Salix, Sa*), d'un peuplier blanc (*Populus alba, Pa*) et d'un peuplier hybride cultivar I45/51 (Phy) dont un des parents est un peuplier noir. Les résultats montrent :

- le peuplier hybride présente 9 SNPs, parmi ceux-ci, une partie correspond aux marqueurs de *Populus nigra* (position 46, 79 et 106). Les autres SNPs correspondant peut être alors à ceux de *Populus deltoides* le deuxième parent.

- le peuplier blanc présente 10 SNPs. Cette essence appartient au genre *Populus* mais se distingue bien de *P. nigra* : 1 seul SNP en commun en position 79.

- le saule présente 35 SNPs. C'est une Salicacées comme le peuplier mais la lecture des résultats montre une grande divergence dans le nombre de SNPs. Ce qui place cette essence loin de *P.nigra* dans notre étude.

L'utilisation des SNPs permet donc de distinguer d'autres essences végétales : elles se différencient par rapport à la position et par rapport au nombre de SNPs relevés.

E) Conséquence de ces SNPs sur la protéine phytochrome B2

L'utilisation d'un logiciel de traduction à partir de l'ADN sens² permet de situer le cadre de lecture et donc les codons concernés par les SNPs

	Codon initial concerné par SNP	Acide aminé	Codons avec SNP	Acide aminé	Codons avec SNP	Acide aminé
46	TAC ⁴⁶	Tyrosine	TAT	Tyrosine	TAY	Tyrosine
					TAC	
79	GTG ⁷⁹	Valine	GTA	Valine	GTR	Valine
					GTG	
106	CCG ¹⁰⁶	Proline	CCA	Proline	CCR	Proline
					CCG	
En rouge le SNP avec sa position en exposant Dans le code IUPAC (annexe 7) : Y = C ou T ; R= G ou A						

L'analyse des résultats montre que les SNPs observés ne modifient pas la séquence primaire de la protéine et ceci en raison du code génétique dégénéré : un même acide aminé peut être codé par plusieurs codons : par exemple le codon CCX code pour l'acide aminé proline.

Le codon contenant le SNP position 46 est normalement TAC ce qui correspond à l'acide aminé Tyrosine. Le remplacement de C par T ne modifie pas l'acide aminé qui sera utilisé pour réaliser la protéine, TAT codant aussi pour l'acide aminé tyrosine.

Ces mutations n'affectent donc pas la protéine Phy B2 dans son fonctionnement, la séquence primaire n'étant pas modifiée. Ces mutations sont donc silencieuses. Par contre, ces mutations peuvent affecter la quantité de protéine Phy B2 produite. L'expression d'une protéine dépend de deux choses entre autres : la quantité d'ARN messager produit correspondant et la disponibilité des ARN de transfert. Par exemple si l'ARN de transfert correspondant au codon UAU est moins présent que celui correspondant au codon UAC alors la protéine sera fabriquée moins rapidement d'autant plus si l'arbre est homozygote. Dans le cas des hétérozygotes UAY, un chromosome porte le bon codon UAC et l'autre chromosome porte le codon muté UAU, la protéine Phy B2 sera alors exprimée à un taux intermédiaire.

² Brin sens = brin (+) = brin non transcrit = brin complémentaire (même séquence que l'ARNm avec U à la place de T)
 Brin antisens = brin (-) = brin transcrit = brin matrice

Conclusion

Nos analyses de séquences montrent qu'il est possible de caractériser l'espèce *Populus nigra* en utilisant les marqueurs du polymorphisme génétique que sont les SNPs.

Les deux peuplements analysés sont distants de 160 km et cependant nous retrouvons les mêmes marqueurs, certes pas dans les mêmes proportions, mais la population analysée présente un faible effectif.

Cependant, la tendance serait à une certaine homogénéité entre ces deux peuplements.

La traduction de la zone étudiée montre que la protéine Phy B2 n'est pas affectée dans sa structure primaire. Elle sera peut être seulement traduite à des niveaux différents.

L'analyse porte seulement sur deux peuplements (33 individus au total). L'effectif est réduit... Le dispositif "Génome à l'école" fédère plusieurs lycées : il aurait été intéressant d'inclure le travail de chacun d'eux sur ce gène Phy B2. C'est une perspective à envisager.

Faire de la génomique dans le but d'étudier et comparer les séquences présente peu d'intérêts pratiques. Sauf si on lui associe le phénotype. La mise en relation du génotype avec le phénotype permet alors de prédire le phénotype d'un individu à partir de la connaissance de son génotype.

L'utilisation des SNPs permet l'analyse génétique ciblée de certains gènes d'intérêt (qualité du bois, rectitude, résistance à des ravageurs, résistance au froid,...). La connaissance du génotype permettra de prédire le phénotype. Le sélectionneur ne sera plus obligé d'attendre plusieurs années pour "phénotyper" ses plants. Une analyse génétique suffira alors. Cela permettra d'accélérer les programmes de sélection de nouveaux cultivars correspondant à des situations écologiques précises.

Dans notre cas, le gène Phy B2 est impliqué dans la floraison des arbres (annexe 2), les données de date de floraison ne sont pas disponibles pour les arbres étudiés. En pépinière expérimentale, les arbres sont recépés tous les trois ans, la plantation chez M.Gallet a aussi trois ans. La floraison s'obtient à partir de la sixième année chez le peuplier. Pour les arbres provenant de l'Aveyron, les dates de fleurissement n'ont pas été relevées. Il ne nous est pas possible donc de relier le phénotype "floraison" avec les trois génotypes relevés dans notre étude.

Nous avons étudié un seul gène, nous ne disposons pas d'assez de résultats pour les gènes Phy B1, Phy A et Cry2 qui sont eux aussi des gènes intervenant dans la floraison (annexe 2). Le contrôle de la floraison dépend de plusieurs gènes, se contenter d'un seul (Phy B2) est trop restrictif. Notre travail montre ce qu'il est possible de réaliser à notre niveau : élèves de 1ère Sciences et Techniques de Laboratoire (STL) option "biotechnologie".

Pour terminer, ce travail fut une activité de groupe intéressante. Nous avons réussi à surmonter un certain nombre de difficultés ensemble. Les approches innovantes employées nous ont motivées. En plus des savoir-faire, ce travail nous a appris des savoir-être : travailler ensemble en utilisant les mêmes règles rigoureuses, savoir communiquer clairement.

Nous ne regrettons pas le temps passé sur ce projet en plus de notre emploi du temps de la semaine.

Bibliographie

Le peuplier noir : une ressource génétique à l'interface entre habitats naturels d'intérêt communautaire et sylviculture intensive.

François Lefèvre /Dossier de l'environnement de l'INRA n°21

Villar M. & Forestier O. – 2009. Le Peuplier noir en France : Pourquoi conserver ses ressources génétiques et comment les valoriser ? Revue forestière française, n°5, pp. 457-464.

PIÉGAY (H.), PAUTOU (G.), RUFFINONI (C.). — Les forêts riveraines des cours d'eau. Écologie, fonctions et gestion. — Paris : IDF, 2003. — 464 p.

<http://www.sciencesalecole.org/plan-genome-a-lecole-presentation/>

http://www.sciencesalecole.org/wp-content/uploads/2016/06/Cahier_Protocoles_Genome_Version-du-2015_10_15.pdf

Annexe 1 : Description botanique de *Populus nigra*³

Le peuplier noir (*Populus nigra*), également nommé Liard, Piboule, ... est un arbre de la famille des Salicacées.

Appareil végétatif



Le peuplier noir de Brantome (24) : un arbre remarquable depuis 2016

Age
environ 150 ans

Circonférence à 1.5m
10,4 m

Hauteur
Environ 40 m

Envergure du houppier
25 m

Le Peuplier noir peut atteindre une hauteur de 30 à 35 m : il fait partie des grands arbres. Sa longévité est importante certains spécimens ont plus de 400 ans. Les formes varient beaucoup selon le contexte, en particulier selon que l'arbre ait poussé seul et isolé ou dans une haie, un boisement (troncs plus longs et droits) ou une ripisylve (dans ces derniers cas, il est plus élancé, avec des branches plus longues cherchant la lumière).

Les broussins sont assez fréquents.

La branchaison est irrégulière, aux verticilles non marqués, avec pour les grandes branches des sujets âgés une forme d'arche et des rameaux longs de l'année nettement cylindriques à la base, parfois sub-anguleux vers l'apex mais sans côtes.

Taillé en têtard, le tronc peut devenir trapu.

Le tronc possède une écorce brune-noire (à l'origine de son nom botanique) et plutôt sombre à rhytidome rugueux et profondément fissurée longitudinalement, s'épaississant avec l'âge. Sur les vieux sujets (très gros bois) l'écorce forme un réseau losangique caractéristique.

Au niveau des tiges les bourgeons végétatifs sont petits, glabres, visqueux et souvent appliqués.



Les feuilles apparaissent après les fleurs, sont de forme triangulaire à losangique. Elles sont denticulées, vertes et glabres sur les deux faces, visqueuses au débourrement. Leur pétiole, parfois rougeâtre, est aplati au sommet.

Le limbe est plus court sur les rameaux de l'année, et « la feuille y présente une forme losangique et acuminée (feuille se terminant en pointe allongée et effilée). » Le limbe des feuilles des longs rameaux est moins typique, « toujours de surface plus grande et avec une base du limbe largement arrondie ou aplatie. » La surface des feuilles est toujours plus petite qu'en situation équivalente chez les peupliers hybrides.

Appareil reproducteur

L'espèce est dioïque (fleurs mâles et femelles sur individus différents). La floraison a lieu de mars à avril ; la pollinisation ainsi que la dispersion du pollen se fait par le vent (anémophile).

L'arbre atteint sa maturité sexuelle vers six ans, en produisant des chatons pendants ;

 <p>chatons mâles - 8 à 10 cm de long à maturité ; - rouges/pourpres.</p>	 <p>chatons femelles - 6 à 8 cm de long à maturité ; - vert-jaunâtre.</p>	 <p>Les fleurs femelles produisent des capsules ovoïdes ou sphériques à pédicelle court, vert foncé, avec deux valves, conservant la disposition en chaton.</p> <p>Les fruits sont mûrs 6 à 8 semaines après pollinisation. Ils libèrent alors une sorte de bourre de coton hydrophobe contenant les graines.</p>
---	---	---

³ Extrait de : http://peupliernoir.orleans.inra.fr/documents_informatifs.html

Annexe 2 : Les Phytochromes et leurs rôles : quelques données⁴

La lumière est un signal environnemental particulièrement important chez les végétaux. Ce signal comporte au moins deux aspects : l'intensité et la longueur d'onde. Les récepteurs réagissant à la lumière sont appelés photorécepteurs. Ils sont composés d'une fraction protéique et d'un chromophore. Ce dernier est la partie du photorécepteur interagissant directement avec la lumière. Ce type de récepteur est appelé phytochrome, ce photorécepteur réagit à la lumière rouge.

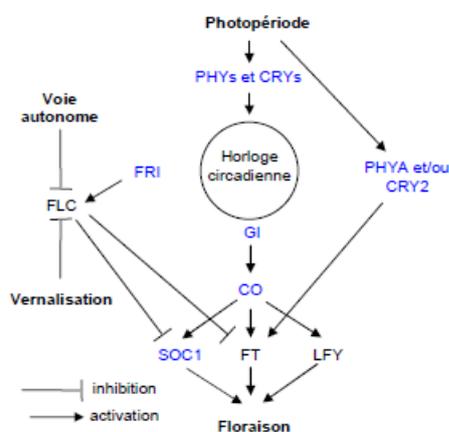
Il est spécialisé dans la reconnaissance de la lumière rouge et de la lumière proche infra-rouge ou rouge lointain. La lumière rouge frappe le phytochrome, il passe de la forme inactive à la forme active et migre du cytoplasme dans le noyau. Là, il se lie sur des gènes et active ou réprime la transcription de ces derniers. Les phénomènes les plus connus contrôlés par le phytochrome sont la germination des graines et la mise à fleurs des plantes à fleurs, mais aussi la croissance des feuilles, la synthèse des pigments, les mouvements des organes et des organites, etc.

Chez les organismes végétaux dont le génome a été entièrement séquencé : arabelle (*Arabidopsis thaliana* : *Brassicaceae*), ..., il a été révélé que les phytochromes sont codés par une famille de 5 gènes, appelés *phyA-phyE*.

En fonction des plantes, des études complémentaires montrent que :

- ces 5 gènes ne sont pas toujours présents (seulement 3 chez le riz)
- le niveau d'expression de ces gènes varie en fonction des conditions de croissance.

Le schéma ci-dessous montre une synthèse de la régulation de la floraison chez la plante modèle : *Arabidopsis thaliana*. Il faut noter la place primordiale des phytochromes en début de la chaîne de régulation de cette floraison.



Modèle simplifié de la voie de contrôle de la floraison chez la plante modèle *Arabidopsis thaliana* (d'après Mouradov et al., 2002 et Izawa et al., 2003). Les gènes en bleu sont ceux dont nous allons étudier les homologues chez *Populus nigra*. CO : constans ; CRY : cryptochromes ; FLC : Flowering Locus C ; FRI : Frigida ; FT : Flowering Locus T ; GI : Gigantea ; LFY : Leafy (sera inclus ultérieurement) ; PHYA : Phytochrome A ; PHYB : Phytochrome B ; SOC : Suppressor of Overexpression of Constans.

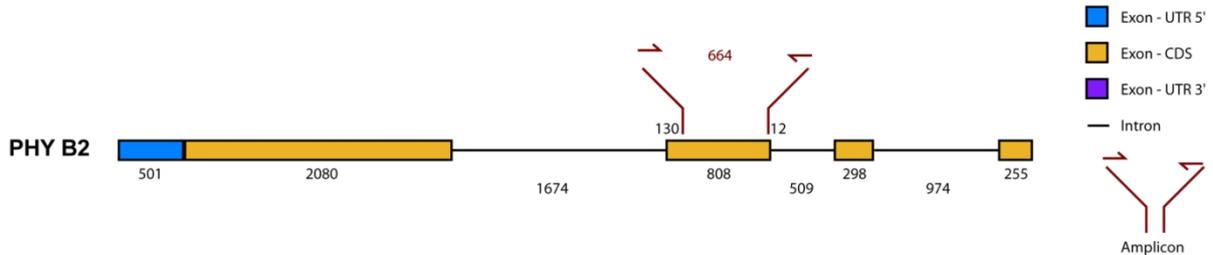
Remarque : les gènes notés en bleu sont ceux qui sont étudiés dans le cadre du projet "Genome à l'école".

⁴ Extrait de : http://ressources.univ-lemans.fr/AccesLibre/UNT/UNISCIEL/biologie1/web/co/module_Biologie_14.html

Annexe 3 : Structure et séquences complètes du gène *Phy B2* chez *P. trichocarpa* étudié dans le projet *Populus* (extrait du document génome à l'école)

1 Structure du gène Phytochrome B2 chez *P. trichocarpa*

Le génome du peuplier fait : 485±10 millions de paires de bases réparties sur 38 chromosomes⁵ (485 Mbs) Situé sur le chromosome X, le gène du phytochrome B2 est formé de 4 exons et de 3 introns. La séquence analysée (663 bases) se situe dans l'exon 2.



2. Séquence du gène Phytochrome B2 chez *P. trichocarpa*

Légendes	<p>En gris : séquences flanquantes (en 5', contient les séquences promotrices)</p> <p>En bleu : séquences non traduites (UTR) en 5' et 3'</p> <p>En noir : intron</p> <p>En rouge : exon</p> <p>Surligné bleu : amorce F utilisée pour la PCR</p> <p>Surligné gris : séquence complémentaire de l'amorce R utilisée pour la PCR</p> <p>Souligné rouge : <u>partie du gène étudié</u></p>
----------	---

```

1   GTCGAGCTGCACACAAGCCGTCGGTATACCTATATTAATAATAAAATAAAAATAAAAAATTCATTACACCATACAATAATATATTAATAAAAAAGAAA
101  AAGAAAAGAAAACAACAAAAGGACGCAATGAATCTTTCTCGTTAATGTACTTTTCGGTTTTATTAATGGTGTCTCCGTATCATATTAGATTAGCTGCT
201  GGTGGACAAA GGGGACTAATAAATTTCTCCTAGCCAAGCCTTAAGCAGGGGCACAGTGTGAATTTTGTGTACTAACGAGATTTCTTCCCTCAAA
301  TTCATTTCTTTCTGGCAAACTCTTTCTCTGCAAATTCACAAAACAAAACCTTCTGTGTTGATTTCTCCCATTTCCCAAAGCGCAGCATAT
401  AAAAGACAAAACCTCAAAAATCTCCCACACTACGCTATACTATACTAACTGCAATCCAACACAAGATCTGAATCTTTTTTCGGTCACTCTCCACCACTAA
501  ATATATTTGGTTAAACAGAGCAACCAAAATCTTAGTTGAAGAAAAGAAATGGCATCACAATCACAAGGCAAAGCAATCAGCCGGTACATAATCAAGC
601  TCAGTCTTCGGGTACCAAGTAAACATGAGGCAACACCATCAGGCCACAGAATCAGTAAGCAAAGCAATTCACAATACACAGTCGATGCACAACCTCCATGCA
701  GTTTTGAACAATCTGGTGGCACTGGTAGGTCATTTGATTACTCAAAGTCTGTAGAACTACTAACCAAGT CAGTACCTGAACAACAATCACAGCGTATT
801  TGTCAAAATTCAAAGGGTGGTCAATTCAGCCATTGG GTGCATGATTGCTGCAGATGAACAGTCTTTTAGGGTCATTGCGTATAGCGAAAACGCCAA
901  AGATATGCTT GGTTAACTCCACAATCAGTCCCTTCTTAGAGAAAACAAGAAATCTCTTTGTTGGGGCTGATGTAAGGA TTCTTTTAGGCCTTCACG
1001 GCTGTTTTGCTTGAAAAGCATTGGTGAAGGAAATAACATGTTGAATCCAATTTGGATTCAATCCAAGAATTCGGGAAGCCCTTTTACGCAATTT
1101 TGCATAGGATTGATGTTGGT ATTGTTATTGATTAGAGCCTGCTAGGACTGAAGACCTGCTTTATCTATAGCAGGGGCTGTCAGTCTCAGAAGTTAGC
1201 AGATTATGGTTTATAAGTTTCATGAGGACGAGCATGGTGAAGTTGTCGCTGAGAATAAAAGGGTTGATTGGAACTTATATCGGTTTGCATTATCCTT
1301 CCACGGATACACCAAGCTTCGAGGTTTTGTTCAAGCAGAATAGGGTTAGGATGATTGGATTGTCA TGCTATACCGGTCGGTGTATTACAGGATGA
1401 GGCCTTATGCAGCCTTATGCTGGTTGGTTCGACTCTT CGTGCTCCTCACGGTTGTCATGCTCAGTATATGGAGAATATGGGGTCGATTGCTCATTG
1501 GCCATGGCTG TTATATTTATGGAATGATGAAGAAGCTATCGTGGGAGGAACTCGATGAGGCTTTGGGGTTGGTGT TTGCCATCACACTTCTGCTA
1601 GGTGATTCATTTCCGCTTCGTTATGCGTGTGAGTTCTTAATGCAAGGCTTCGGGCTTCAATTGAACATGGAATTGCAGTTGGCATCGCAGTTGTTGGA
1701 GAAACATGCTTGAGGACTC AACTCTTATGTGATATGCTTCTCCGCACTCTCCTACTGGCATTGTTACTCAAAGTCCAAGTATCAT GGACCTTGTG
1801 AAGTGTGATGGGGCAGCTCTTATTACCAGGGCAGTATTACCTTTAGGCGTGACCCCA ACTGAAACCCAAATAAAGATATTGGGAGTGGTTGTTGA
1901 CCCTTCATGGAGCCCAACTGGATTAAGCA CTGACAGTTTTGGCCGATGCTGGGTATCCTGGGGCAGCCTTTCTCGGCGATGCAGTTTGTGGGATGGCTG
2001 TGCTTATATTGCTGAGAGAGATTTCTTTTTTGGTTTCGGTCTCACACTGCAAAGGAGTCAAATGGGGT GGTGCGAAGCATCACCCAGAGGACAAGGAT
2101 GATGGGCGAGGATGCATCCGCTTATCATTTCAAGGCATTCTTAGAGTGGTGAAGAGCCGGAGTTTACCGTGGGAGAATGCAGAAATGGATGCCATTC
2201 ATTCTTTGCAGCTTATTTTGCAGACTCATTAGGGATGCTGAAGCAACCAATTTCTAAGGCTGTTGTACATACCCAGCTCAAGGATATGGAATTGCAAGG
2301 TATGGATGAGCTCAGTTTCGGTTGCAAGAGAAATGTTAGACTAATAGAGACTGCAACTGCTCCGATATTTGCTGTAGACGTTGATGGCCGCATAAACGGG
2401 TGGAATGCAAAAGTTGCTGA GTTGACTGGACTCAGTTGAGGAAGCCATGGGGAAGTCTTTGGTTCACGATCTGTTTATAAGGAATATGAAGAAATG
2501 TTGACAAGCTTATTCATCGTGCAGTAAAAGGTACCCACTTTATATCTCACATATCTCAATGAATCTTTTTAAGTGCATCCGAGCTTACAGGTTTGCTCC
2601 TTTACTAGAAAATAACCTTGAAAACATACATGGTTTTGCAAAAAGCTGGAATTATAAAAGCCATGAATTGAAATGTAAGAATTTCAAACCTTCAAAT
2701 CCTGGAATTAATTTATGCAATAAAAGGTTTCAATCAAAATAAGAAAACAGCTGGCTCACTCAGAAGTGTGCGGTACAGAACCTTGAATAGTTTCATTTG
2801 CACAGCATGCCTGAATAGTCTCTGATTAATGATTTCAA TGAAGTGTGATGATAATTAATCTTCTTGAATGCACTCCATTGATAAGGCAAGATGAAA
2901 CGTTGATGCAAAAGCACTGTGGTAACGCACAATTTCTCAAAATCAGAGATAAATAACAAGGAAATTAAGGGTCGTGGGCTTTTCTTAAGGAACCAAGT
3001 AGTTGGGAATTGCAATACTACAATGGAGAACATTTTCTGATAGATCACTCACTAGGATCTTGGGCACCACATCTAGGTCAAAATGGCTCAGGCATTC
3101 CCACTCAATAGCCTTGGCGTGAAGTATTCTTTGCAAAATCTCTTAAAGTGGATAACTGGATAGCTGAGTCTTGATTAAAGTTATCCAGTGAGCTT
3201 AGTTCAATTGATAGTAGCAGCTTGGAGCAAAATACCTAGCTTCTTATTACTGTAGAATTCATAAGGACTGAGATGCATTGCTTTGCGTTCATGAGATC

```

⁵ <https://www.futura-sciences.com/sciences/actualites/recherche-premier-sequencage-genome-arbre-peuplier-9812/>

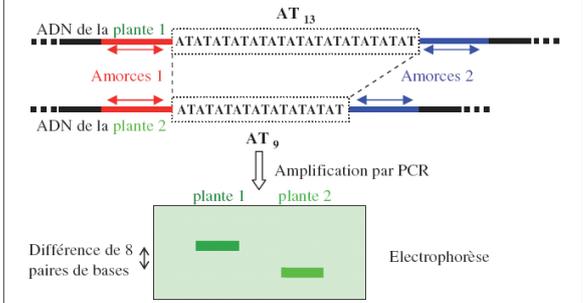
3301 TAGCTTCAAGTAAATCCCTAGGGTCTGAATGTTGAGGCACTCCCAAGAACTGAGTTCATGTCTAAAGAAGCTGCTGGTGAATTACTCTAATTTCC
3401 AGGGAACCTGCACATTACATCTCATCTGATGAAAAATAGTGGTTCATGCATTTTTCTAAATCTAAAAAGT TAGTGAAGAGATCACTTCAAAATGGCTCCT
3501 TCAATTATGCTTCGAGTGATGCTCTCCGTAAAAGAATGT GTTTGTGCTTATATTACCTTTATCTATGCACATTTTTAGGTGTAACAGAGAGATAACC
3601 GACGCAGGACAATGGTAGGAATTGAAGAAAGAGATAGGAAAAGGGGTAGAAATCCTAGAAAATAGGGTCTTAAGATCCACAACAACGAATATATGCA
3701 AATAGCAATTCTCAAATTCATTCACGTTCTCCCAACACCCTAAAGGCTCTAATTAATAAGGAAAGGCTAAAACCTCTAATAGTACTAATTATATCT
3801 AATGTTACTTTACCAATTAACCTCAAGATAATAATAAAGTCTGAACTAATAAAAAATTTAAAAAAGAAAGATAGTAGCACAAAATGCT TGTTCCCTTTG
3901 GCTGCATCAATAACCTAGAATATTATGTGTATTACTACAATAACTCAGTTGTGATTAATAATTAACAATGAAAATGCATGCTATGGTTTTAAATA
4001 GAAATTTGATAGATGCCCTCACACACGGTGCTAATTTGGACAGTGCTTGATCATTCTAATCAGCAGTTTAAACCATAACAGTAGGACTAGCTTTAATATAG
4101 GTTTGAGGAACTGAAAACACATAAAACATGAAAATTTAGTTTGTTCGAAAGTTTTGGATTTTTGATATCCTATTATACCCTATGTTTAACTAAT
4201 GCAGCGGAAGAAGATAAGAACGTGGAGATAAAATTGAGGA CTTTTGCTCTGAACACAAAAGAAGGCTGTTTTGTGGTGGTAAATGCTTGTTCGAGCA
4301 AGGATTACATGGATAATATAGTTGGGGTTTGCTT TGTTGGTCAGGATATTACAGGTCAAAAAGTGTAATGGACAATACGTCCTTATACAAGGTGATTA
4401 TAAGGCTATTGTACACAGTCCCAATCCTTCGATCCCTCCGATTTTTGCTTCAGATGAGAACACATGTTGCTTGGAGTGGAACTGCCATGGAAAAACTC
4501 ACAGGATGGTCCAGGGGCGAAGTTGTTGGGAAGATGTTGGTTGGGGAGGTTTTGGCAGTTGCTGATGGCTCAAGGTCAGATGCACTGACAAAATTC
4601 TGATTGCCCTGCACAATGCAATTGGAGGGATAGATACAGACAAGTTACCTTTTTCTTCTTTGACCGGAATGAAAAAATGTGCAAACTCTCTTGACAGC
4701 TAACAAGAGGGTTAATATGGAGGGAGATATTATGGAGCCTTCTGCTTCTGAGATTGCAAGTCTGAGTTGCAGCAAACCTTTGAAAGTTCAGAAAACAG
4801 CAGGAAAAAAGAGCTTTGCAAGGATGAAAGAGTTGGCTTACATTTGCCAGGAAATAAAAAATCCTTTAA GTGGTATACACTTACCAACTCGCTTTGG
4901 AGAACACAGACTTGACTGAAGATCAACAGCAGTTTCTCGAGACTAGTGCATGTGAAAAACAGATATTGAAGATCATACGAGACATTGATCTTGAGAG
5001 CATTGAAAATGGTGAGCTTTCTAAAGAAATTATCATACTTAAATTTGACAGTGAGGTGAGTATTTCTGCTTTGGTACCATTCTTTGGAAATATATA
5101 TATGACATGAAATCCAATAGATGATGATTATCCTTGAATGTTCAACACTATACTAAAGTTATCGACATATTTGTTTTGGCACTAGGTGATGTCAGTTTTG
5201 ATGCAGGTTCCATTCCCTTGAGTCCATTCCCTTAACAAAATAGCCTTCAAATTTGACTGCGGTGCTCTGTTTTAGTCGTCTAATACTACTGAATTTATC
5301 ATCCATGGATGCCGAGATTGTAACAATATTGATTTCTTTCTGTAATTTATTTTGGGTCTTTTAGTAATCCAGGTGATAACTCGCTTTTGACAAT
5401 CTGTATAATCTTTTTAAGTTAATCATGTGCATTCAAGTGCATTTTCATTAATTTTTCATCTCAATCTCCAGCATACGTACATTCTAATTAGAGAGAGT
5501 CTCCATTTGCTTTGAATTAG TCACTGGAGCTTGAGAAGGCTGAATTTTTACTTGGGAGTGTCAATAA GCTGTTGTAGCCAAGCAATGCTATTGCTG
5301 AGGGAAAAGAAATCTACAATTGCTTCGTGATATTCCAGAAG AAATAAAAAAGCTGGCAGTTTATGGTGATCAGGCAAGAATTCACAAGTATTGGCTGATT
5701 TCTTGTGTAATATGGTGCCTTATGCTCCATCTCAGCAGGTTGGGTGGAGATTGATGTTTGTCCAACACTGAAGCAAATC TCTGATGGACACTCTCGT
5801 GCATACAGAGTTCAAGTATGTTCTCCTCAACTCTCTGCAAGTAAATGAT AACAAATTTTACTGATAGAGGAAAAGCCAAAAATTCACGATGGCTGTT
5901 ACTTTTTTCTTTATTTTCATGTAACCCCTTTTGATATTCTAATAGCTACATAATGTTTTCTTGAGTTAATATAAATGCCACTTCTTATTCAA
6001 GTATTTTTCTTTGTGAGACTCTGATGATTTTTACAGTATCTTTAGTTAAGTACACTCGAATCTTGTTTATTCTTCCCCTCTCATGACCAGCCA
6101 TGCATCCTAATCTGACTTCCAAGCTGAAA TCTCAACCAGAAGGTAGGTTTGTAGTCACTCAACAGGTATAGAATTAGACTTGATCTGCCATTAATCATGA
6201 CAATCACCATTCCACAGAACAATAATTTACAATGTTGAGATCCTCAACTGGCTGCTTCTTGAATCTGTC GACTATTGTTGTATGAACAAGGAAGTTG
6301 ATGATGCTGTAGTGAAGAGCTATCTGTGATTTCTGTTTGT TGTGACATGATAAAGTGGTTTATATGAAGTCAAATTTATGGTCTTCAGTTTGTACGCC
6401 TCTGTTGAAAACTGTATCTCCTCTTTGCTGTCATCATGTTATATACCAAGAAAATTTAGATGATGGTTGAAAACTACATTGTGCTACTTGTTCCT
6501 TTCTTTCTCCTCTTGTAGAGAACGGAAGTCTCGTTAGTGCATTCATTAATAATTGACTAATATATTATTATTATTCTTTTGAAGGATTGTT
6701 TGCCCTGGTGAAGTCTTCC TCCTGAATTAGTTCAGACATGTTCCATAGTAGATGGTTACTCAAGAGGGCTAGGGCTCAGCATG TGCAGGAAGA
6801 TCTTAAAGCTAATGAATGGTGAAGTCCAATATATTAGAGAGTCAGAAAAGATGCTATTTCT TAGTTGCTCGAAGTACCCATGCCTCAAAAAG TGGAAA
6901 GGGTGCTGCAGACTAGGATGCTTGGCTGC AACATCCACTAAGGAGCTTTCTCAATTTCAAACACTGACACTGTTATGCTGCACCGTGAGGGGATGGTTCC
7001 AAAAATCCAGTTTCAAGGTTAAGCGTGTGCCAGGGCCGCTGATTTGGTTAAGTTATCCTGAAAATACA ACTGTTACTCTATCCACTTACAATATGG
7101 TTTTTGTCTTTATCGGCGCCAGTCCAGATCAATAACTCACACATGAGGTAGCTGTTACTTTCTAGCCTGTTAATTCGAAGATCCCATCGATGATAG
7201 AAGCCAAATATAATGCTGTGGTATGCTGTCATTGTTATAATAAAGAAGCATCAAAAAGGATGCCATTAGATTGTTGCAG AGGCTTCCACGATAAGGCC
7301 TCAAAGGTTGATATATTCCCTGCTGCAAGTATATACTGATGTACC (7448)
7401

Annexe 4 : Les marqueurs du polymorphisme génétique⁶

Le polymorphisme des génomes se manifeste de différentes façons :

- Duplications. Des parties plus ou moins grandes du génome sont copiées les unes derrière les autres (en tandem). Le polymorphisme réside alors dans le nombre de copies.
- Réarrangements. Des parties du génome sont inversées par rapport à la situation normale (inversions)
- Insertion / délétion. Des parties du génome sont soit supprimées, soit ajoutées. Elles peuvent être :
 - Grandes (translocations chromosomiques) : un morceau de chromosome est déplacé sur un autre chromosome : délétion puis insertion
 - Petites : une base est ajoutée ou supprimée = indel
- Polymorphisme au niveau d'un seul nucléotide (SNP, *single nucleotide polymorphism*). Il est le résultat d'une erreur lors de la réplication principalement :
 - Transitions entre 2 purines, 2 pyrimidines (A↔G, T↔C)
 - Transversions purine/pyrimidine (A↔T, A↔C, G↔T, G↔C)

Deux exemples de marqueurs de polymorphisme génétique

SNP : <i>single nucleotide polymorphism</i>	Polymorphisme de répétitions : étude des microsatellites
<p>Séquence de référence : ATGGAACTTAAACTGGATGA Séquence à analyser : ATGCAACTTAAACTGTATGA</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; padding: 2px; margin: 5px;">SNP</div> <div style="border: 1px solid black; padding: 2px; margin: 5px;">INDEL</div> </div>	
<ul style="list-style-type: none"> • Ils ne présentent généralement que deux allèles, ils sont donc peu informatifs par rapport aux microsatellites, mais ils sont très abondants. Leur détection est facile à automatiser. • Pour trouver des SNP, il faut avoir des génomes déjà séquencés. • De nombreuses techniques permettent de révéler ce type de polymorphisme. Le séquençage de fragments PCR est celle que nous allons pratiquer dans le cadre de « Génome à l'École » 	<ul style="list-style-type: none"> • Egalement nommés SSR (<i>Simple Sequence Repeats</i>) ou STR (<i>Short Tandem Repeats</i>). Ce sont des répétitions de motifs de 2, 3, 4, 5 ou 6 bases (par exemple, ATATATAT...) • Il y a beaucoup d'allèles différents (chaque nombre de répétitions constitue un allèle, et le nombre de répétitions peut varier entre 3 et 100 !) • Ces marqueurs sont donc très informatifs (la probabilité pour que deux individus possèdent des allèles différents est élevée)

Le polymorphisme des génomes est un outil qui permet :

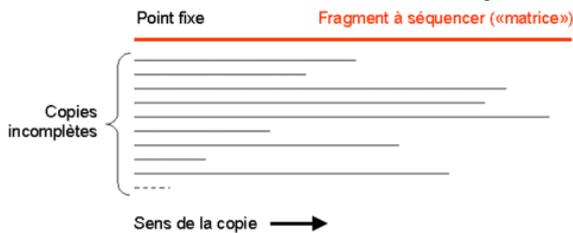
- d'étudier l'évolution, la classification, l'adaptation, la diversité génétique des espèces ;
- d'étudier le fonctionnement du génome (notion de gènes, d'allèles, ...), d'identification de gènes d'intérêt ;
- de sélectionner des individus remarquables pour créer de nouvelles variétés adaptées aux besoins de l'homme et de son environnement.

⁶ Extrait de : **le polymorphisme des génomes** (d'après Patricia Faivre Rampant, Nîmes, juin 2011)

Annexe 5 : Le séquençage

<p>TGCAT TATGCCTGGT CACGTGCAAA</p> <p>CTGTGATTCGAAGACTGCCTAGTAC</p>	<p>Ces deux brins d'ADN possèdent le même nombre de chacune des 4 bases, mais dans un ordre différent : seul le séquençage fera la différence</p>
---	---

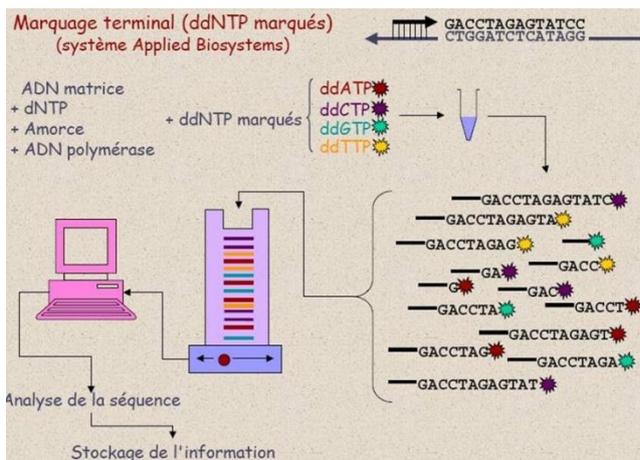
Principe de la méthode de Sanger



Produire à partir d'un point fixe des copies incomplètes de la matrice, qui seront ensuite séparées selon leur taille par électrophorèse par exemple.

En présence des amorces adéquats, et des dNTP, l'élongation de la chaîne est assurée par une ADN polymérase.

L'ajout de didéoxyribonucléotide triphosphates ddNTP dans le milieu réactionnel permet de stopper de façon aléatoire l'élongation de la molécule d'ADN en cours de fabrication. Lorsque l'ADN polymérase incorpore un ddNTP complémentaire de la matrice, l'élongation est stoppée. On obtient de la sorte toute les longueurs de fragments possibles. Le problème est : quel est le ddNTP incorporé au bout des chaînes ?

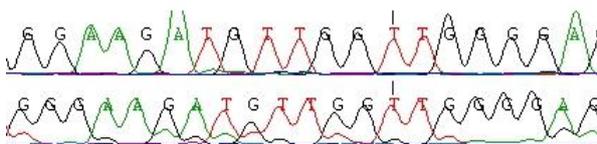


Avec les séquenceurs ABI utilisés au Genoscope, on utilise des ddNTP marqués chacun par un fluorochrome différent. On peut ainsi distinguer les fragments terminés par A, T, G, C.

Dans les séquenceurs capillaires, l'électrophorèse se déroule dans un tube de verre très fin (quelques microns), remplie de polymère : un laser en sortie excite les groupements fluorescents. Le signal de fluorescence est enregistré à mesure que les fragments sortent du gel : chaque base correspond à un pic du signal, la couleur de fluorescence identifie la nature de la base.

Remarque : ces séquenceurs sont encore utiles mais les techniques de séquençage à très haut débit permettent d'aller beaucoup plus vite.

Ces séquenceurs sont multicapillaires, ce qui permet de traiter en même temps autant d'échantillons qu'il y a de capillaires.



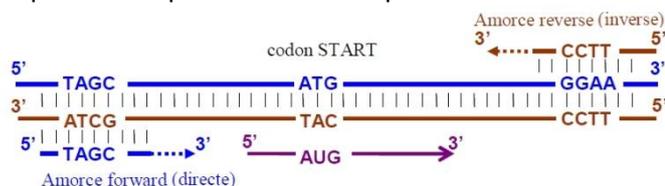
Exemple de tracé obtenu (données personnelles)

Les électrophorègrammes obtenus sont ensuite analysés informatiquement par divers logiciels (GENALYS™, ...). Ils traduisent le tracé obtenu en séquence de base. Des erreurs sont possibles, une correction humaine est nécessaire.

Annexe 6 : Lecture et correction des électrophorogrammes

Pour chaque gène analysé, nous possédons deux électrophorogrammes :

- celui correspondant à la séquence nommée "f" = forward. Elle a la même orientation que le gène de référence, c'est le brin qui porte la séquence codante.
- celui correspondant à la séquence nommée "r" = reverse. Elle a l'orientation inverse par rapport au gène de référence : c'est la séquence complémentaire de la précédente.



Pour éviter les problèmes de lecture, le logiciel calcul la séquence complémentaire de la séquence "r". Nous disposons alors de deux traces quasiment identiques, l'une pouvant servir à corriger les erreurs présentes sur l'autre.

Nous présenterons ici quelques corrections appliquées sur des séquences lors de notre travail. Les séquences sont toutes présentées de la même façon : piste haute = séquence "f", piste basse = séquence "r"

<p>T G G G A A G A T G</p>	<p>L'idéal : le logiciel déduit la même séquence à partir des tracés "f" et "r".</p>
<p>T C A C A G G A T G</p> <p>C A C a G G A T G G</p>	<p>Ici, sur le tracé "r" le logiciel n'a pas pu résoudre le pic fait de A et C il a été résolu en M (code IUPAC). Il y a une discordance avec le pic correspondant sur "f" résolu en A. Que faire ? On constate que C a tendance à trainé est crée un bruit de fond sur "r", donc ce pic est corrigé en a (minuscule car la base a été corrigée)</p>
<p>T A C T T T T T T G</p> <p>T C C T A T T T T T T G</p>	<p>Ici, sur le tracé "f" et "r" on observe un pic formé de deux bases A et G résolu en R (code IUPAC) : erreur ou réalité. Le peuplier est une espèce diploïde donc sur un chromosome, pour ce locus, il y a A et sur l'autre chromosome un G. Ce peuplier est hétérozygote à ce locus.</p>
<p>I A T T T T T G A</p> <p>I N N N N G N N N N N</p>	<p>Sur ces tracés, il est possible de travailler avec la séquence "f" mais impossible de vérifier avec la séquence "r" : cette dernière n'a pas pu être résolue par le logiciel et n'est pas corrigable manuellement. Ces séquences sont retirées de l'analyse. Il y a eu un échec au séquençage ou alors les amplicons que nous avons fournis au génoscope étaient de mauvaise qualité.</p>

Annexe 7 : Analyse des séquences par le logiciel "GENALYS™" : le code IUPAC

IUPAC : International Union of Pure and Applied Chemistry ou **Union internationale de chimie pure et appliquée (UICPA)**, est une organisation non gouvernementale ayant son siège à Zurich, créée en 1919 qui s'intéresse aux progrès en chimie, chimie physique, biochimie, etc. Elle a pour membres des sociétés nationales de chimie. C'est l'autorité reconnue pour le développement de règles à adopter pour la nomenclature, les symboles et la terminologie des éléments chimiques et de leurs dérivés, par le biais de son Comité interdivisionnel de la nomenclature et des symboles (*Interdivisional Committee on Nomenclature and Symbols*) qui fixe la nomenclature de l'UICPA.

Lecture sur l'électrophorégramme par le logiciel "GENALYS™"	Code IUPAC
A ou C	M
A ou G	R
A ou T	W
C ou G	S
C ou T	Y
G ou T	K
A ou C ou T	H
G ou T ou C	B
G ou C ou A	V
G ou A ou T	D
G ou A ou T ou C	N