Production et traitement de données omiques hétérogènes en vue de l'étude de la plasticité de la paroi chez des écotypes de la plante modèle Arabidopsis thaliana provenant d'altitudes contrastées

Harold Duruflé

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Harold DURUFLE

le vendredi 20 octobre 2017

Titre :

Production et traitement de données omiques hétérogènes en vue de l'étude de la plasticité de la paroi chez des écotypes de la plante modèle Arabidopsis thaliana provenant d'altitudes contrastées.

École doctorale et discipline ou spécialité :
ED SEVAB : Développement des plantes

Unité de recherche :
Laboratoire de Recherche en Sciences Végétales (LRSV) UMR UPS / CNRS 5546

Directeur/trice(s) de Thèse :

Pr. Christophe DUNAND et Pr. Philippe BESSE

Jury :

Christine GRANIER, Directeur de recherche, INRA, Montpellier
Hélène FRÉROT, Maître de Conférences, Université de Lille
Fabrice ROUX, Directeur de recherche, CNRS, Toulouse
Jean-Philippe REICHHELD, Directeur de recherche, CNRS, Perpignan

Université de Toulouse

THESE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par :
Université de Toulouse III Paul Sabatier
(UT3 Paul Sabatier)

Présentée et soutenue par :
Harold DURUFLÉ
Le 20/10/2017

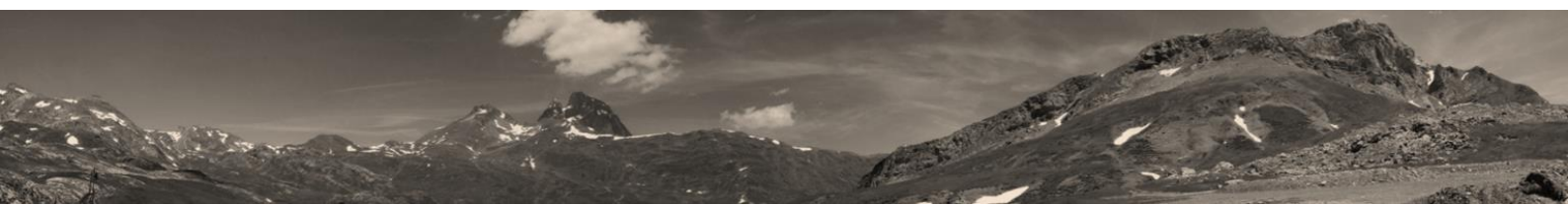
Titre:
Production et traitement de données omiques hétérogènes en vue de l'étude de la plasticité de la paroi chez des écotypes de la plante modèle *Arabidopsis thaliana* provenant d'altitudes contrastées.

Ecole Doctorale SEVAB - Développement des plantes

Unité de recherche:
Laboratoire de Recherche en Sciences Végétales (LRSV)
UMR CNRS-UPS 5546

Directeurs de thèse:
DUNAND Christophe, Professeur, Université Paul Sabatier, Toulouse
BESSE Philippe, Professeur, Université Paul Sabatier, Toulouse

Membres du jury:
GRANIER Christine, Directeur de recherche, INRA, Montpellier
ROUX Fabrice, Directeur de recherche, CNRS, Toulouse
FRÉROT Hélène, Maître de Conférences (HDR), Université de Lille
REICHHELD Jean-Philippe, Directeur de recherche, CNRS, Perpignan



REMERCIEMENTS

"Nobody sees a flower - really - it is so small it takes time - we haven't time - and to see takes time, like to have a friend takes time." Georgia O'Keeffe

Une thèse, ce n'est pas que de la science, c'est aussi une grande aventure humaine. Je tiens donc à remercier toutes les personnes qui m'ont aidé, épaulé et… subit pendant ces 3 années, ainsi que ceux et celles qui m'ont permis d'en arriver là.

Tout d'abord, un grand merci à Christophe DUNAND, mon directeur de thèse « côté Biologie ». Merci de m'avoir fait confiance et de m'avoir permis de réaliser ce beau projet au sein de ton équipe. Ces 3 années ont été riches en émotions et j'espère en être sorti grandi pour me construire un futur dans la science. Cette expérience m'a aussi apporté une vision écoresponsable de la science grâce à toi.

Merci également à Philippe BESSE, mon co-directeur de thèse « côté Statistique » qui a permis d'appuyer mon financement de thèse et de nous aiguiller dans les stratégies d'intégrations.

Ce projet n'aurait pas pris cette dimension sans l'aide ni le temps pris par Sébastien DEJEAN qui fût mon mentor en Bio-statistique durant ces 3 années. Un très grand merci pour ta disponibilité, ton optimisme assuré, ta gentillesse et ta pédagogie sans faille.

Je tenais à remercier les membres de mon comité de thèse : Denis VILE (malgré ton anniversaire) ; Monique BURRUS et Nathalie ESCAVARAGE pour avoir identifié et récolté nos belles Arabettes dans les Pyrénées. La virée en voiture dans les Pyrénées pour installer mes petites plantes au Pourtalet a été un pur plaisir malgré les virages... ; et particulièrement un grand merci à Elisabeth JAMET pour ton soutien, ton expertise et ton efficacité à trouver des coquilles dans mes analyses. Nos discussions m'ont beaucoup apporté tant sur le plan scientifique que sur mes réflexions personnelles. Cette thèse aurait une autre allure sans tes conseils et remarques avisés.

Je souhaite remercier aussi les membres du jury, pour avoir accepté de prendre le temps de lire et juger mon travail.

Un grand merci à Philippe RANOCHA, l'homme de l'ombre de ce projet. Merci pour toute l'aide que tu as pu m'apporter durant ces 3 ans, malgré mes erreurs et mes oublis (encore trop nombreux). Il est très rassurant d'avoir un soutien littéralement infaillible à ses côtés lorsque l'on fait une thèse.

Merci également aux scientifiques qui m'ont épaulé dans ce projet et se sont vu assaillir de multiples doutes et questions : Vincent BURLAT, Cécile ALBENNE, Maxime BONHOMME et Hélène SAN CLEMENTE.

Il est rare de pouvoir réaliser des tests *in natura* durant sa thèse, je tiens donc à remercier Maria-Antonia GUILLEN CASTELLS (Tonia) pour m'avoir permis de venir déterrer un bout de ton jardin et y installer le temps d'un été mes petites plantes.

Cette aventure m'a aussi donné l'opportunité de faire mes premier pas dans l'enseignement, merci à Cécile, Guigui et Julie d'avoir rendu cette expérience mémorable, ainsi que dans l'encadrement de stagiaires Sébastien, Merwann et Duchesse (à toi le tour maintenant !).

Afin de clore la partie sérieuse de ces remerciements, je tenais à dire un grand merci à tous ceux qui m'ont fait grandir cette passion de la recherche tous les étés saison après saison au Lautaret en RDPL : Serge AUBERT (j'aurais vraiment voulu pouvoir discuter de ces recherches avec

toi…), Roland (Mr le Président), Christophe (mi padre cito), Richard, Marielle, Constance, Clodius,…

Un grand remerciement à tous celles et ceux qui rendent le quotidien plaisant au labo : Keuvin, le beau Brubru, Edith, Malo, Aurélie, Catherine, la dream team des pauses de 10h (Emilie, Coco, Chiel et Laurent), Lolo, Ambroise, Morgane et à tous les stagiaires et ceux que j'ai oublié de nommer qui font qu'il est bon de travailler dans ce labo !

Un grand merci également à Hélène, Léa, Manon, Aude (ôde), Ariane, Mathilde, Franck, Rémi et Thomas pour nous avoir permis de fonder l'AJS et de pouvoir se faire quelques bons apéro et barbecues.

Les soutiens infaillibles que sont les amis de longue date : Xavouille, Greg, Pierre, Manu, Shadia, Sylvain, Yo,… ainsi que des nouveaux (la Baronne, Nico, Xav, Maëlys,…).

« *Sans musique la vie serait une erreur* » (Nietzsche), un chaleureux merci à mes amis musiciens de l'OSUM (mimi, Tic et Tac,…), la Garonne, l'OSET et évidement le BigO'rchestra pour me permettre de continuer de vivre cette passion malgré mes faibles disponibilités.

Enfin, et pas des moindres, un grand merci à toute ma famille, qui m'a soutenu et toujours encouragé dans cette passion folle pour le monde végétal, née il y a bien longtemps…

Et pour finir, un remerciement tout particulier à ma plus belle découverte toulousaine, Mathilde, pour tout ce que tu m'apportes tous les jours et qui m'a permis de finir ce doctorat en douceur.

SOMMAIRE

Contexte général

Le réchauffement climatique constitue une problématique d'actualité très préoccupante en raison de ses effets potentiels sur la biodiversité et le secteur agricole. Mieux comprendre l'adaptation des plantes face à ce phénomène récent représente donc un intérêt majeur pour la science et la société.

L'étude de populations naturelles provenant d'un gradient d'altitude permet de corréler l'impact d'un ensemble de conditions climatiques (température, humidité, radiation, *etc*.) à des traits phénotypiques. Ces différentes populations sont dites adaptées à leurs conditions climatiques *in natura*. En cultivant ces plantes dans des conditions standardisées de laboratoire (intensité lumineuse, substrat, température, arrosage, *etc*.), la variabilité observée, le phénotype, est alors dû essentiellement à la variabilité génétique intrinsèque à chaque plante, donc à son génotype. La mise en culture de ces mêmes plantes en changeant une seule variable, par exemple la température, permet de mettre en évidence un phénotype caractéristique. Ce phénotype observé peut être une réponse d'acclimatation d'un génome adapté. Le projet *WallOmics* vise à caractériser l'adaptation des plantes à l'altitude par l'étude de populations naturelles d'*Arabidopsis thaliana* provenant des Pyrénées.

Les acteurs moléculaires de l'adaptation des plantes au climat sont encore mal connus mais il apparaît que la paroi des cellules végétales pourrait avoir un rôle important dans ce processus. En effet, celle-ci représente le squelette des plantes et leur confère une rigidité tout en représentant une barrière externe sensible et dynamique face aux changements environnementaux. Sa structure et sa composition peuvent être modifiées à tout moment. Il est d'ailleurs possible de dire que cette paroi végétale donne la forme générale de la plante (taille, forme, densité, *etc*.), son phénotype observable. Ce projet se consacrera principalement à l'étude des parois des cellules végétales.

Les nouvelles technologies ont permis l'émergence des données dites «omiques», c'est-à-dire de vastes ensembles de données à des niveaux biologiques multiples, comme des données écologiques, de phénotypages, biochimiques, protéomiques, transcriptomiques et génomiques. L'étude et la mise en relation de ces données a favorisé le développement d'approches globales qui visent à établir une réponse à plusieurs échelles. C'est justement par ce type d'approche non mécanistique que le projet *WallOmics* va contribuer à établir les bases moléculaires des modifications des parois face aux changements climatiques.

I. Introduction

I.1 Contraintes des variations d'altitude chez les plantes

I.1.1 Contexte du réchauffement climatique

La réponse des plantes au réchauffement climatique est récemment devenue un enjeu majeur, en raison de ses effets potentiels à la fois sur la biodiversité (Meehl *et al.* 2007), sur le secteur agricole et sur les populations. Les manifestations engendrées par ce réchauffement climatique sont diverses : modification des périodes saisonnières, changements de températures, occurrences plus élevées de stress abiotiques (froid, gelée, tempête, sécheresse, inondation, *etc*.) ; et les conséquences sont multiples : rendements agricoles de plus en plus variables, recrudescence des maladies, redistribution géographique des productions agricoles, *etc*. L'étude des variations phénotypiques liées à ces changements est donc devenue primordial.

En montagne, le déplacement de plusieurs espèces de plantes vers des altitudes plus élevées a déjà été observé parallèlement à l'évolution des conditions climatiques (Beniston 2003).

I.1.2 La montagne, terrain d'étude privilégié

La montagne est un environnement de travail qui permet d'étudier un gradient naturel de stress abiotique. En altitude, les plantes sont exposées à des conditions plus extrêmes qu'en plaine. Elles sont soumises à de plus forts rayonnements UV, accompagnés d'une augmentation de l'intensité lumineuse, de températures extrêmes et d'une réduction importante de la saison végétative.

La distribution de la végétation en milieu montagnard est caractérisée par un étagement graduel. On peut distinguer l'étage collinéen, représenté par une dominance de forêts de divers feuillus, l'étage montagnard, caractérisé par des forêts de sapins, hêtres et pins, l'étage subalpin, occupé majoritairement par l'épicéa, puis l'étage alpin, qui ne possède que des herbacées, et pour finir le nival, où l'on retrouve essentiellement des mousses et des lichens. Le niveau de ces étages n'est pas le même partout dans le monde : ainsi, l'étage alpin est situé aux alentours de 4 000 m à l'équateur, 2 200 m dans les Alpes, et au niveau de la mer aux pôles (Fig. 1).

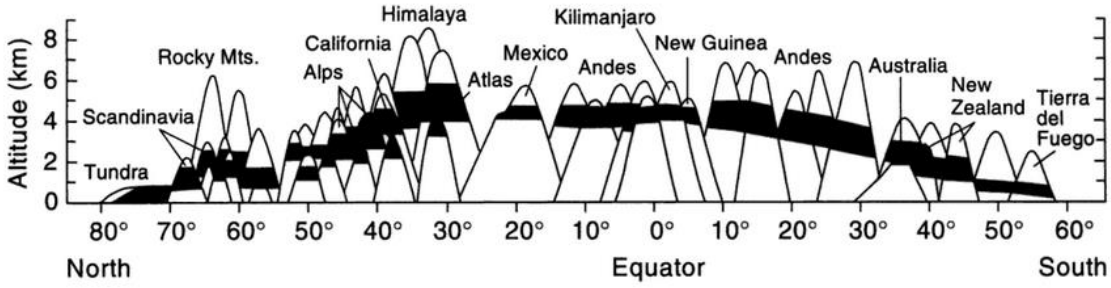


Figure 1 : *Position des zones alpines dans le monde (représentées en noir) (Körner 2003)*

Cette distribution écologique est notamment due au gradient de température qui décroît fortement avec l'altitude (0,5 à 0,8°C tous les 100 m) au cours de la période de végétation (Körner 2003). Par exemple, les précipitations sont fréquentes sous forme de neige durant la période de végétation à partir du niveau alpin, mais elles se font majoritairement sous forme de neige à l'étage nival. De plus, le climat de ces étages est très contrasté en fonction des latitudes où il est présent. L'hygrométrie est la composante d'altitude la plus variable en fonction des régions : il n'existe aucune tendance ou règle sur le gradient de précipitations en fonction de l'altitude (Fig. 2). On peut ainsi retrouver en région équatoriale (courbe E) une décroissance des précipitations en fonction de l'altitude, un optimum de précipitation en moyenne montagne en région subtropicale (courbe S), ou encore un gradient positif de précipitations en fonction de l'altitude en région tempérée (courbe t). De toutes les composantes climatiques, les précipitations et la saisonnalité exercent la plus grande influence sur la variation climatique d'une région (Körner 2007).

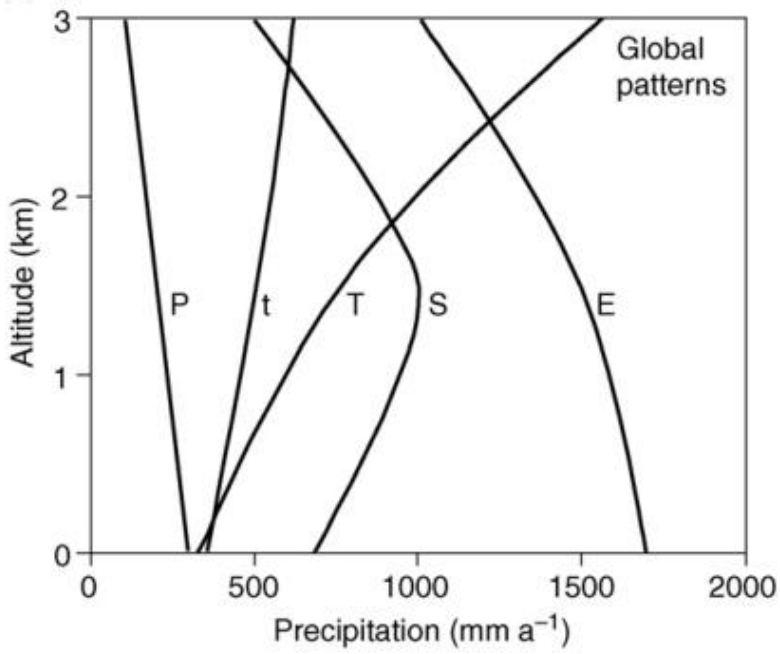


Figure 2 : *Tendance des typologies de précipitations en fonction du climat. E, équatorial (0 - 10° latitude); S, subtropical (10 – 30° latitude); T, transition (30 – 40° latitude); t, tempéré (40 – 60° latitude); P, polaire (Groenland) (Körner 2007).*

I.1.3 Comportement des plantes face à des changements d'altitude

La graduation des composantes du climat affecte fortement les plantes et leurs phénotypes. Cependant, cette plasticité phénotypique permettant l'adaptation à des environnements très changeants varie selon les espèces.

On peut trouver des espèces avec des capacités d'acclimatation très importantes comme *Heliotropium curassivicum*, qui change son optimum thermique en fonction de la température du milieu où elle se trouve (de 20 à 40°C), tout en conservant sa capacité d'assimilation de l'eau et des nutriments, et donc sa capacité de croissance (Berry & Bjorkman 1980). D'autres espèces possèdent des écotypes ou variétés liés aux différentes conditions climatiques associées à leur répartition géographique. Ainsi, les individus d'une même espèce peuvent montrer des différences morphologiques importantes pour une meilleure réponse aux différentes conditions climatiques. Par exemple, le genre *Camelina*, montre une variation intra- et inter- spécifique importante de sa cuticule lui conférant une tolérance différenciée à la sècheresse (Tomasi *et al.* 2017). Une autre étude récente sur différents accessions de la graminée *Setaria viridis* montre une plasticité physiologique et phénotypique lorsqu'ils sont soumis à des stress hydriques et thermiques permettant de distinguer les accessions tolérantes et sensibles à ces stress (Saha *et al.* 2016). La graminée *Festuca eskia* présente une réduction de sa taille et de sa masse, et augmente sa surface foliaire lorsqu'elle est transplantée en altitude (Gonzalo-Turpin & Hazard 2009). Mais les espèces adaptées à des conditions climatiques extrêmes possèdent une diversité de plasticité limitée. Ainsi *Ranunculus glacialis*, adaptée au milieu alpin, ne peut survivre en dessous de 2 400 m dans les Alpes (Streb *et al.* 2003). Ces différents cas ouvrent de nouvelle direction prometteuse en biologie évolutive et en biologie des systèmes grâce à l'étude des variants naturels, jointe à des études moléculaires et climatiques.

Les espèces végétales ayant une large répartition géographique sont souvent constituées de populations qui peuvent présenter ces différences phénotypiques dues aux différentes pressions de sélections des divers habitats qu'elles occupent : *A. thaliana* fait partie de ces espèces possédant une répartition géographique mondiale.

I.2 *Arabidopsis* : la plante, son environnement, son écologie

I.2.1 Généralités

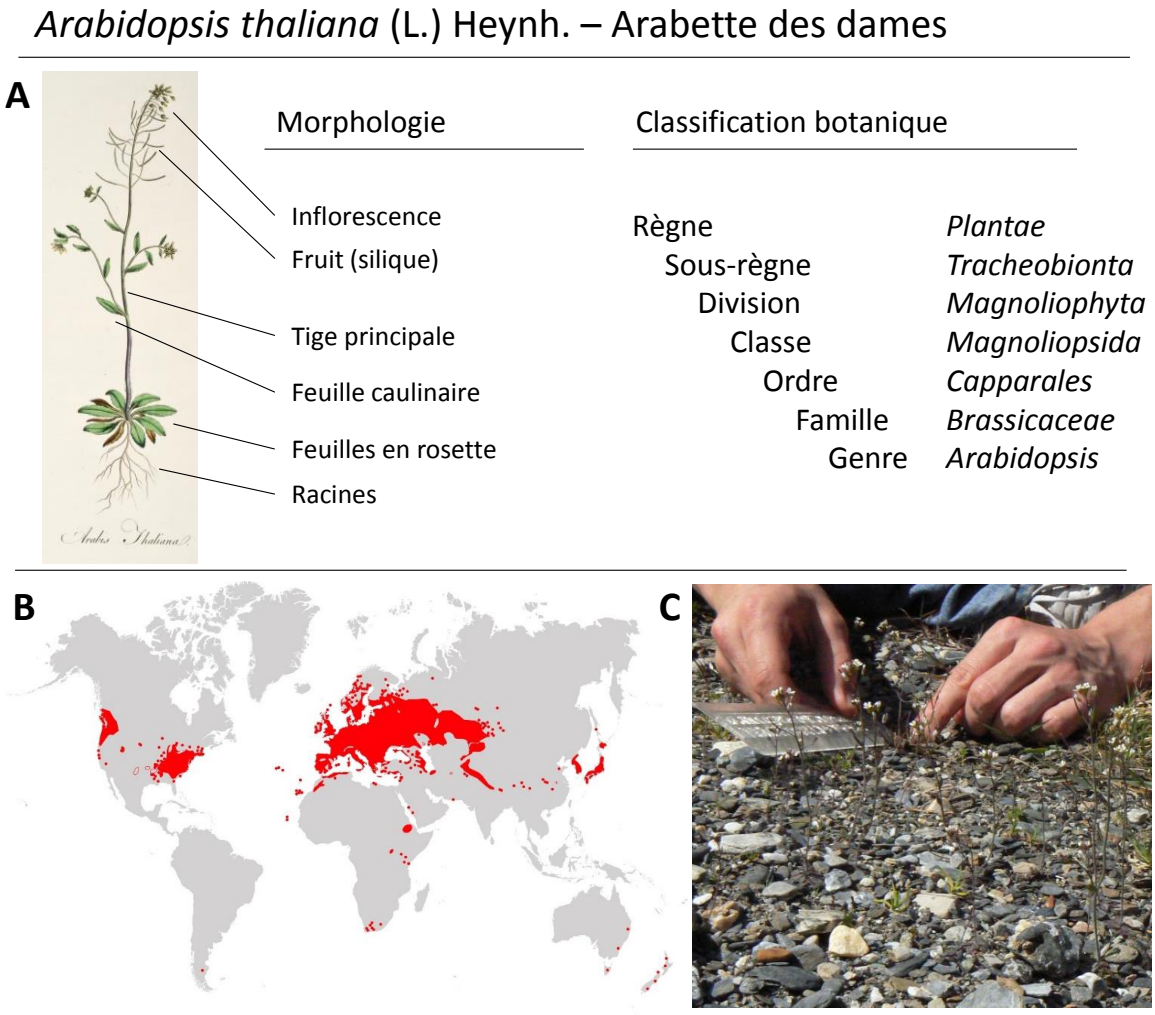


Figure 3 : *A) Information relative à la morphologie (William Curtis, 1835, Flora Londiniensis, John Innes Historical Collections) et à la classification botanique d'*A. thaliana. *B) Distribution mondiale de l'espèce (Krämer 2015) ; C) Exemple d'habitat naturel lors d'une récolte, ici à Gripp en Hautes-Pyrénées.*

A. thaliana est une angiosperme annuelle de la famille des *Brassicaceae*, largement répandue dans l'hémisphère nord. Plante pionnière et opportuniste, elle se trouve dans des habitats perturbés avec peu de compétition interspécifique.

Après germination, les différentes feuilles forment une rosette, au centre de laquelle émerge un bourgeon lors de la phase de transition florale. Celui-ci formera alors une tige qui portera l'inflorescence, puis les fruits, appelés siliques (Fig. 3).

I.2.2 Une plante modèle

A. thaliana est caractérisée par un petit génome (125 millions de paires de bases), une croissance rapide, une petite taille et un fort taux d'autofécondation (> 90%) qui facilite sa culture et son étude en laboratoire. Considérée comme plante « modèle », elle a été la première plante à être séquencée (Huala *et al.* 2001), et de nombreuses études la concernant ont été réalisées en biologie moléculaire, en écologie ou encore en évolution. Grâce à l'expansion des nouvelles technologies, un grand nombre de bases de données lui sont consacrées (TAIR (Swarbreck *et al.* 2008), AHD (Jiang *et al.* 2011), Araport (Cheng *et al.* 2017), AraPheno (Seren *et al.* 2017), *etc.*), permettant notamment des études sur les interactions entre son génotype et son environnement (El-Lithy *et al.* 2004; Mitchell-Olds & Schmitt 2006). Il y a quelques années, un projet de séquençage de 1001 génomes a permis de mieux comprendre l'évolution et la diversité génétique de cette espèce (1001 genome consortium ; (Weigel & Mott 2009).

I.2.3 Son cycle de vie

De par la diversité de ses milieux de vie, *A. thaliana* possède de nombreuses populations identifiées possédant une grande variabilité génétique et phénotypique (Weigel & Mott 2009). Ces variations naturelles suggèrent une forte capacité d'adaptation locale (Mitchell-Olds & Schmitt 2006). L'un des phénotypes les plus marqués est l'existence de deux différents cycles de vie (Fig. 4).

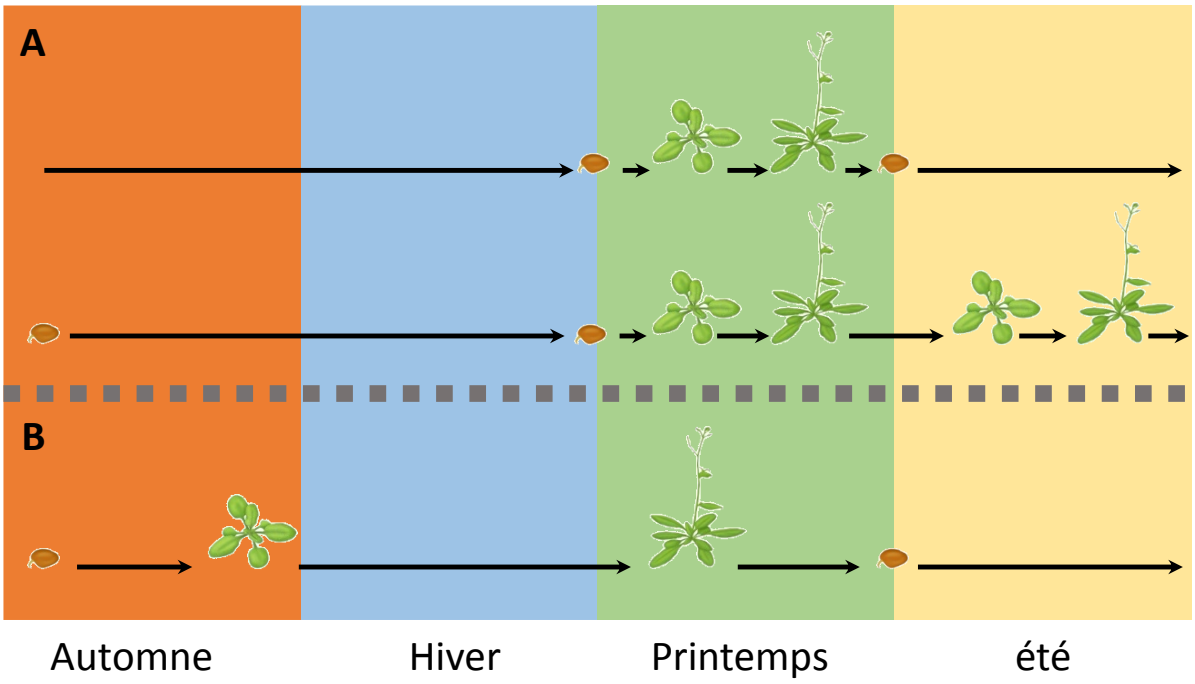


Figure 4 : *Illustration des variations de cycles de vie d'*A. thaliana *(d'après Burghardt et al. 2015) montrant sa capacité à germer et fleurir à différentes saisons. A) les* « summer-annual » *et B) les* « winter-annual ».

Deux cycles de vies distincts contrôlés et régulés par l'environnement ont été identifiés dans les populations de cette espèce. Les deux facteurs environnementaux affectant les cycles de vies d'*A. thaliana* sont la photopériode et la température. Le premier cycle, nommé *summer-annual* (Fig. 4A), est caractérisé par sa rapidité régulée principalement par la photopériode. En jour long, la floraison sera plus rapidement induite qu'en jour court. Ce cycle de vie, majoritaire chez *A. thaliana*, permet de voir plusieurs générations sur une même année. Le second cycle, nommé *winter-annual* (Fig. 4B), est régulé par la température et spécifiquement par la vernalisation. La vernalisation est l'exposition prolongée aux basses températures, induisant la transition florale par l'activation de gènes. Ainsi, pour ces accessions, les rosettes ont besoin de subir les températures d'un hiver pour initier la floraison. Cette combinaison de signaux et de cycles de vie permet à cette espèce de maximiser ses chances de reproduction. Mais si les cycles de vie ne sont qu'un élément spatio-temporel, des traits morphologiques et physiologiques permettent aussi de caractériser les différents écotypes d'*A.thaliana.*

I.2.4 Sa diversité phénotypique

Les différents traits observables chez *A. thaliana* reflètent une adaptation morphologique et physiologique aux différents biotopes qu'occupe l'espèce. Ces variations naturelles offrent une alternative aux études par mutations pour l'observation des fonctions d'un gène. Cette plante modèle offre une large diversité de ces traits au sein de ses populations naturelles réparties au niveau mondial (Fig. 5), lui conférant un large avantage pour des études de la variation de traits complexes *via* l'interaction entre le génotype et l'environnement.

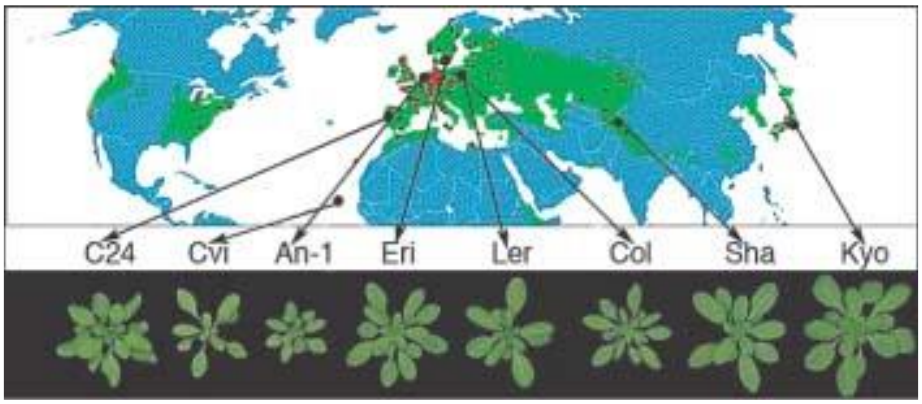


Figure 5 : *Variations morphologiques naturelles d'accessions* A. thaliana *provenant de ses différentes localisations dans le monde (Van Norman & Benfey 2009).*

Les différentes populations naturelles sont des conséquences directes de l'adaptation locale d'une espèce. Depuis quelques années, de nombreux projets visent à identifier les variations génétiques produisant des phénotypes différents à profusion. Ces études ont pour

objectif de relier les variations génétiques intra-espèce aux différents phénotypiques observés. Les études d'association de génomes, ou GWAS (*genome-wide association studies*), constituent un outil privilégié pour ces analyses. Chez *A. thaliana*, 214 000 marqueurs génétiques ont été obtenus à travers diverses techniques d'hybridation de plus de 1 000 génomes (Seren *et al.* 2017). Une étude récente a notamment utilisé cet outil afin de fournir une analyse de l'architecture génétique qui sous-tend à la biosynthèse de la paroi cellulaire végétale (Wood *et al.* 2017). Cependant, les bases moléculaires de ces modifications morphologiques restent largement inconnues. Etant admis que la paroi cellulaire est l'une des composantes principales contribuant à l'architecture des plantes, les modifications de celle-ci doivent être prises en considération lors de l'analyse phénotypique.

I.3 La paroi végétale

I.3.1 Structure et dynamique

La paroi végétale peut être considérée comme le squelette des plantes car elle détermine la forme des cellules qu'elle entoure (Braidwood *et al.* 2014). En effet, cette couche extracellulaire est essentielle au port des plantes, tout en assurant la communication entre les cellules. De plus, située au contact direct de l'environnement, elle contribue à la signalisation, la protection et l'adaptation des plantes face aux stress environnementaux. En raison de ses nombreuses fonctions, elle représente un compartiment dynamique capable de se remodeler durant tout le développement de la plante, aboutissant à une diversité de structure entre les organes tel que la fleur (pétales), la feuille (trichomes et cellules de garde) ou les racines (poils absorbant) (Fig. 6).

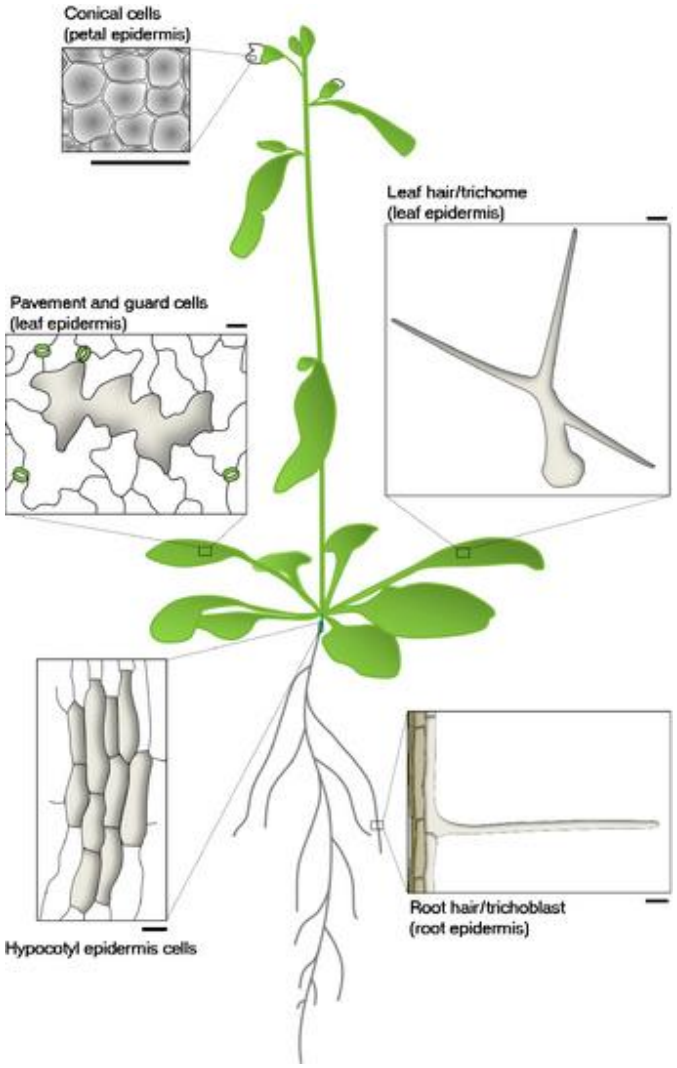


Figure 6 : *Diversité anatomique et morphologique d'*A. thaliana. *Illustration de cinq types différents de cellules : cellules épidermiques de racines, d'hypocotyles, de feuilles, de pétales et de cellules spécialisées (poils absorbants et trichomes) ; Barres d'échelles : 25 µm (Braidwood et al. 2014).*

Formée lors de la division cellulaire et réarrangée en permanence durant la vie de la cellule, cette structure est mise en place sous forme de couches à partir de la lamelle moyenne. On distingue deux types de paroi : la paroi primaire et la paroi secondaire (Fig. 7A), dont les fonctions, les structures et la composition diffèrent.

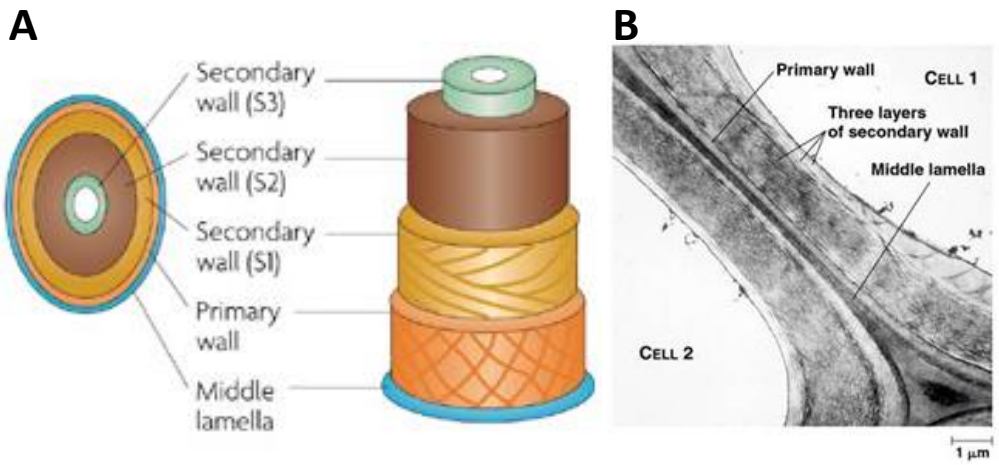


Figure 7 : *La paroi cellulaire secondaire est composée de plusieurs couches. A) Représentation simplifiée des différentes couches de la paroi cellulaire végétale (Sticklen 2008) B) Coupe transversale de paroi cellulaire vue au microscope électronique et localisation des différentes couches de la paroi cellulaire (Benjamin Cummings, Pearson Education).*

- la paroi primaire est composée d'un réseau de polysaccharides (90 – 95 %) et de protéines (5 – 10 %) qui se forme pendant la croissance cellulaire (Voragen *et al.* 2009). Les polysaccharides qui la composent sont constitués de microfibrilles de cellulose, d'hémicelluloses et de pectines. Son épaisseur est relativement minoritaire sur l'épaisseur totale de la paroi. De plus, elle contient des protéines structurales et non-structurales lui conférant une certaine plasticité (Cosgrove 2001).

- la paroi secondaire se forme en fin de croissance de la cellule et sa composition est peu variable : il s'agit majoritairement de polysaccharides peu hydratés, avec notamment une plus forte présence de microfibrilles de cellulose que de pectines (Roland & Vian 1979). Sa structure est plus rigide et plus épaisse que celle de la paroi primaire (Fig. 7B) car elle peut contenir des composants hydrophobes comme la lignine. Les cellules lignifiées, destinées à mourir, forment communément le bois chez les arbres.

- la lamelle moyenne, quant à elle, assure un rôle de cohésion entre les cellules. Elle n'a pas vocation à apporter une fonction mécanique aux plantes. Mise en place lors de la division cellulaire, elle est constituée essentiellement de pectines et ne contient pas de cellulose.

I.3.2 Composition

La paroi végétale est donc un réseau dense de polysaccharides, de protéines et d'autres composés comme la lignine. Les polysaccharides peuvent être subdivisés en trois catégories en fonction des sucres qui les constituent : les pectines, les hémicelluloses et les microfibrilles de cellulose (Fig. 8A).

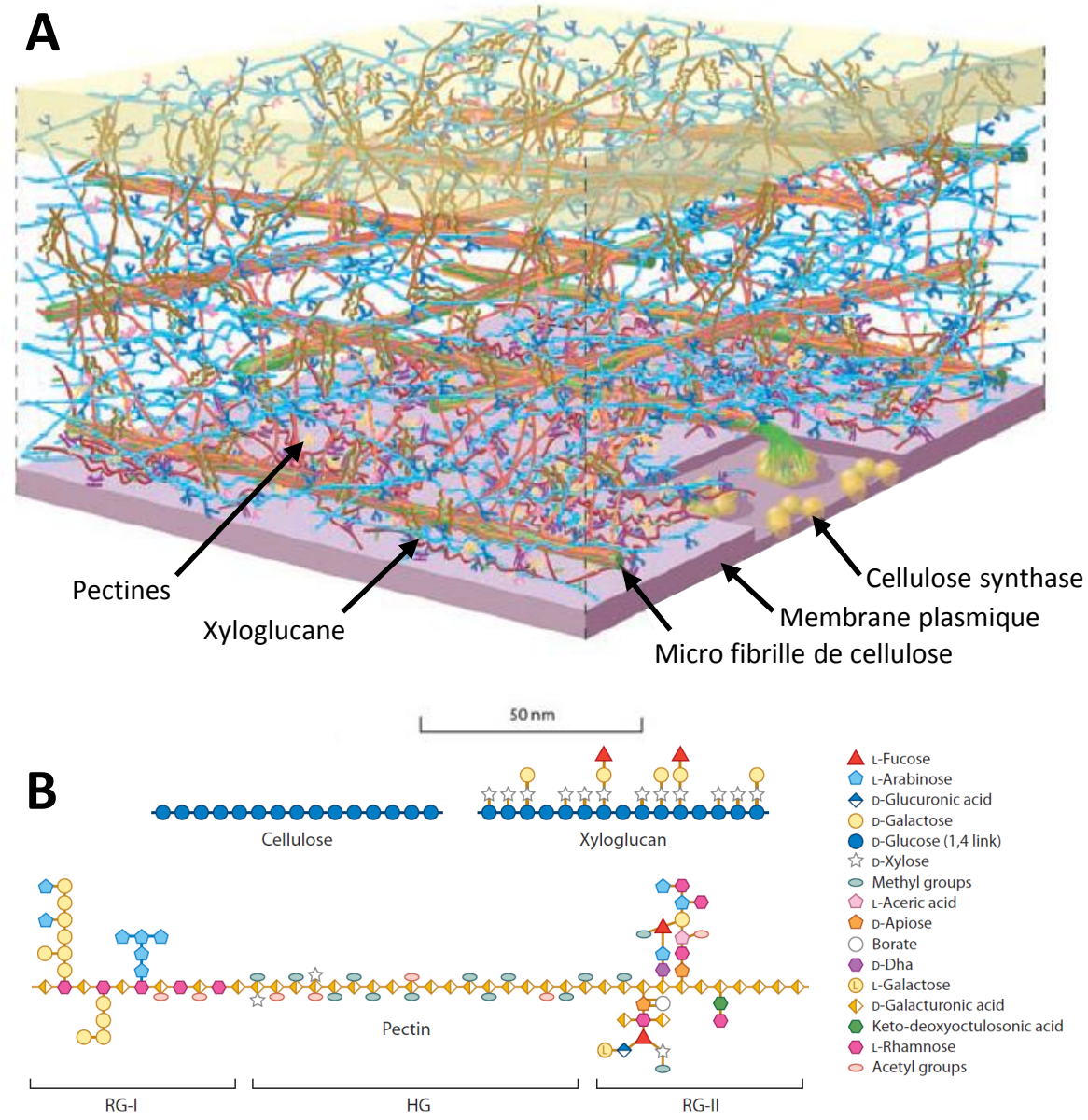


Figure 8 : *Détail du réseau complexe de polysaccharides pariétaux formé de chaînes de sucres. A) Modélisation du réseau de polysaccharides des parois d'*A. thaliana. *Les différents polymères y sont représentés en fonction du contenu en cellulose (qui a été réduit par rapport à une cellule vivante pour plus de clarté ; adapté de Somerville et al. 2004). B) Détail modélisé des structures des composant majoritaire de la paroi : la cellulose, le xyloglucane, ainsi que les pectines (Wolf et al. 2012). XG : xyloglucane ; RGI & RGII : rhamnogalacturonan I & II ; HG : homogalacturonan*

Les pectines, qui assurent de nombreux rôles (adhésion, défense, porosité, *etc*.), sont les composants les plus complexes et divers observés dans la paroi végétale (Cosgrove 2005). Elles forment une matrice hydrophile qui se lie aux divers composants, permettant l'expansion cellulaire et rendant la paroi plastique. Les pectines contiennent 4 domaines caractéristiques (Fig. 8B) :

i) l'homogalacturonane (HG) en large majorité (60 % des parois primaires), qui est une chaîne linéaire d'acides galacturoniques (GalA) lié en α – (1,4) avec des degrés de méthyl estérification variable.
ii) le xylogalacturonane (XGA) qui est une chaîne de HG substituée avec du xylose (Xyl) en position O3 des GalA.
iii) le rhamnogalacturonane I (RG I), comportant des motifs simples de GalA, de rhamnose (Rha), de galactose (Gal) et d'arabinose (Ara) sur une chaîne linaire de GalA et de Rha (α-(1,4)-GalA-α-(1,2)-Rha).
iv) le rhamnogalacturonane II (RG II), qui se compose de motifs de décorations très complexes, pouvant contenir une dizaine de sucres différents reliés par un grand nombre de combinaisons et de liaisons, le tout relié à une chaîne d'HG lié en α – (1,4).

Les hémicelluloses sont des polymères constitués en grande majorité de glucose (Glc), Gal, Xyl, de mannose (Man) et de fucose (Fuc). Leurs liaisons covalentes ou non avec les microfibrilles de cellulose et la lignine leur confèrent un rôle dans le renforcement des parois. Ces polymères sont classés en quatre groupes majeurs :

i) les xyloglucanes (XG), qui sont constitués d'un squelette de Glc (Glc ; β-(1,4)-D-Glc) substitué partiellement de Xyl en α-(1,6).
ii) les xylanes, qui sont des chaînes linaires de Xyl (β-(1,4)-Xyl) pouvant être substitués de GalA et d'Ara.
iii) les mannanes, qui sont des squelettes de Man liés en β-(1,4) pouvant porter des ramifications de Gal lié en α-(1,6).
iv) les glucanes, qui sont des chaînes de Glc non ramifiés (β-(1,4) (1,3)-glucanes).

Les microfibrilles de cellulose sont sûrement les composés pariétaux les plus connus et les plus utilisés en industrie en raison de la stabilité de leur structure simplement constitués de Glc et de leur propriété insoluble. Elles sont en effet exclusivement constituées d'une chaîne d'une dizaine de milliers de cellobioses (dimère de Glc), formant ainsi des fibres rigides. Ces

fibres sont synthétisées par un complexe enzymatique appelé cellulose-synthase (CESA) ou « rosette », ce nom lui venant de la structure que forment ses 6 sous-unités (Kumar *et al.* 2017). Les réseaux de cellulose confèrent la rigidité des parois cellulaires *via* l'interaction avec les hémicelluloses et les pectines.

La paroi végétale peut aussi contenir d'autres polymères comme la lignine. Elle est constituée principalement de trois unités monolignols appelées p-hydroxyphényl (H), guaiacyl (G) et syringyl (S). Il est à noter qu'*A. thaliana,* ne contient que des unités G et S (Raes *et al.* 2003). La composition non polysaccharidique de ce composé lié aux hémicelluloses contribue à la rigidité des parois, et assure des rôles d'hydrophobicité du xylème (transport de l'eau) et de protection contre les pathogènes (Déjardin *et al.* 2010).

Cependant, la paroi n'est pas qu'un complexe de polymères figé, un dixième de sa composition étant fait de protéines de structure ou non. Depuis son séquençage en l'an 2000, environ 2 000 protéines sont prédites comme pouvant être sécrétées (Huala *et al.* 2001). A ce jour, un peu plus de 800 protéines pariétales ont été identifiées dans les protéomes d'*A. thaliana* (Duruflé *et al.* 2017). Les analyses *in silico* de la présence de domaines fonctionnels caractéristiques ont permis de suggérer une classification de ces protéines pariétales en 9 classes fonctionnelles (Fig. 9) (Irshad *et al.* 2008).

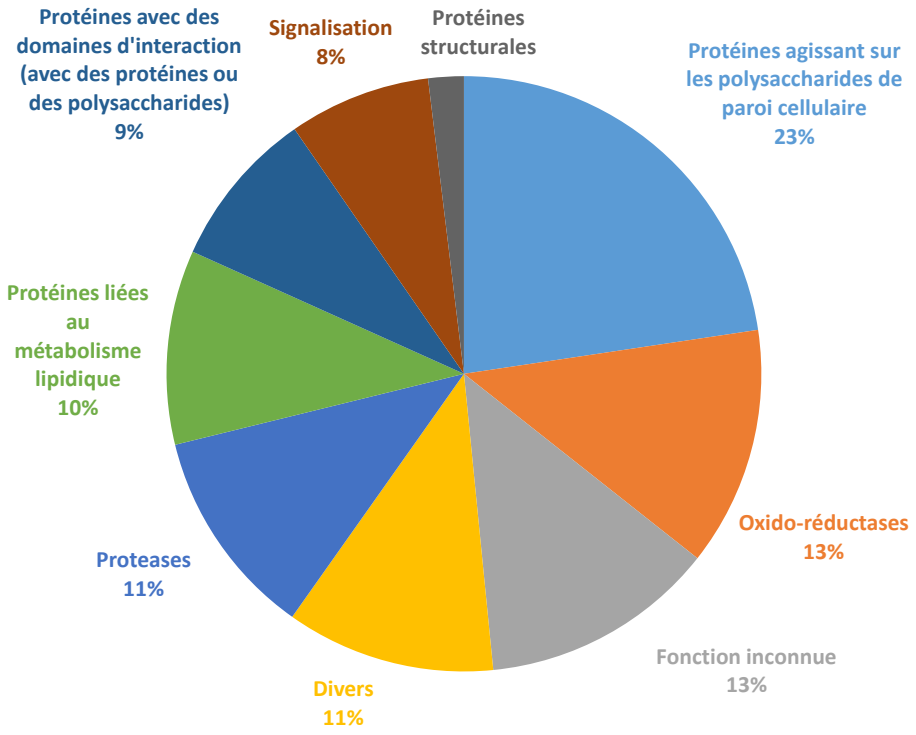


Figure 9 : *Distribution des protéines pariétales identifiées chez* A. thaliana *en fonction de leurs domaines fonctionnels prédits (WallProtDB).*

Les neuf classes fonctionnelles sont les suivantes :

i) les protéines agissant sur les polysaccharides sont les plus représentées (23 %) qui permettent le remodelage dynamique des parois (Franková & Fry 2013). Cette classe contient principalement des glycosides hydrolases (GH), des carbohydrates estérases/lyases ainsi que des expansines ;

ii) les oxydo-réductases (13 %) qui comprend notamment les peroxydases de classe III (CIII Prxs). Ces protéines peuvent réguler l'élongation cellulaire en participant notamment à former des liens covalents entre les protéines structurales. Elles sont aussi impliquées dans les réarrangements des polysaccharides et la réponse immunitaire (Francoz *et al.* 2015) ;

iii) les protéases (12 %), incluant des Asp- Cys- et Ser- protéase et des carboxypeptidases. Elles permettent la dégradation des protéines et pourraient participer dans la maturation des protéines pariétales (Albenne *et al.* 2009) ;

iv) les protéines liées au métabolisme des lipides (10 %), qui interagissent avec les xyloglucanes et la cellulose lors de l'extension pariétale. Elles servent aussi de transporteur de lipides à travers les parois hydrophiles et permettraient des liaisons entre les lipides dans la cuticule (Jacq *et al.* 2017). Cette classe contient des lipases et des acylhydrolases de type GDSL ainsi que des protéines de transfert de lipides (LTP) ;

v) les protéines à domaines d'interaction (8 %), comme les protéines à domaines LRR (*leucine-rich repeat*) qui interagissent avec les polysaccharides et des protéines pariétales;

vi) les protéines de signalisation (8 %), comme les récepteurs kinases et les arabinogalactanes (AGP). Ces dernières pourraient participer aux propriétés biomécaniques de la paroi (Seifert, 2007) ;

vii) les protéines structurales (2 %), qui correspondent essentiellement aux protéines riches en résidus hydroxyproline et glycine. Cette classe est décrite avec plus de détails dans la partie II.6 ;

viii) les protéines de diverses familles (12 %), qui ne sont pas assez importantes pour former une classe fonctionnelle à part entière ;

ix) les protéines de fonctions inconnues (12 %) sont des protéines sans domaine fonctionnel prédit ou ayant une fonction encore inconnue (*domain of unknown function*, DUF).

L’ensemble de ces protéines pariétales sont la clef de la plasticité de la paroi cellulaire et du réarrangement permanent du réseau de polymères. Ce dynamisme est crucial pour la plante car il assure une croissance maîtrisée des cellules tout en conservant une certaine rigidité.

Mais les parois sont aussi en interaction directe avec l’environnement et les facteurs abiotiques. Elles sont donc aussi nécessaires pour la perception du changement d’environnement, et les réponses induites.

I.3.3 Les modifications pariétales en fonction des stress abiotiques

Les plantes sont exposées régulièrement à des conditions défavorables et à de multiples stress abiotiques, lesquels peuvent être récurrents. La paroi végétale étant la première ligne de défense des cellules, ces stress peuvent fortement la perturber. L'arrêt de la croissance, observé fréquemment sous l’effet d’un stress abiotique (Fig. 10B), est notamment dû à la rigidification de celle-ci. Ce phénomène peut être provoqué par une réticulation de composés phénoliques estérifiés avec les glycoprotéines pariétales de structure comme les extensines et/ou des polymères d'hémicelluloses. Provoquée par une variation d’espèces actives de l’oxygène (ROS pour *reactive oxygen species*) (Tenhaken 2014), ce phénomène engendre une densification des réseaux, réduisant l'accès des expansines et des enzymes comme la xyloglucane endo-trans-glucosylases/hydrolases (XTH) au polymère de xyloglucane.

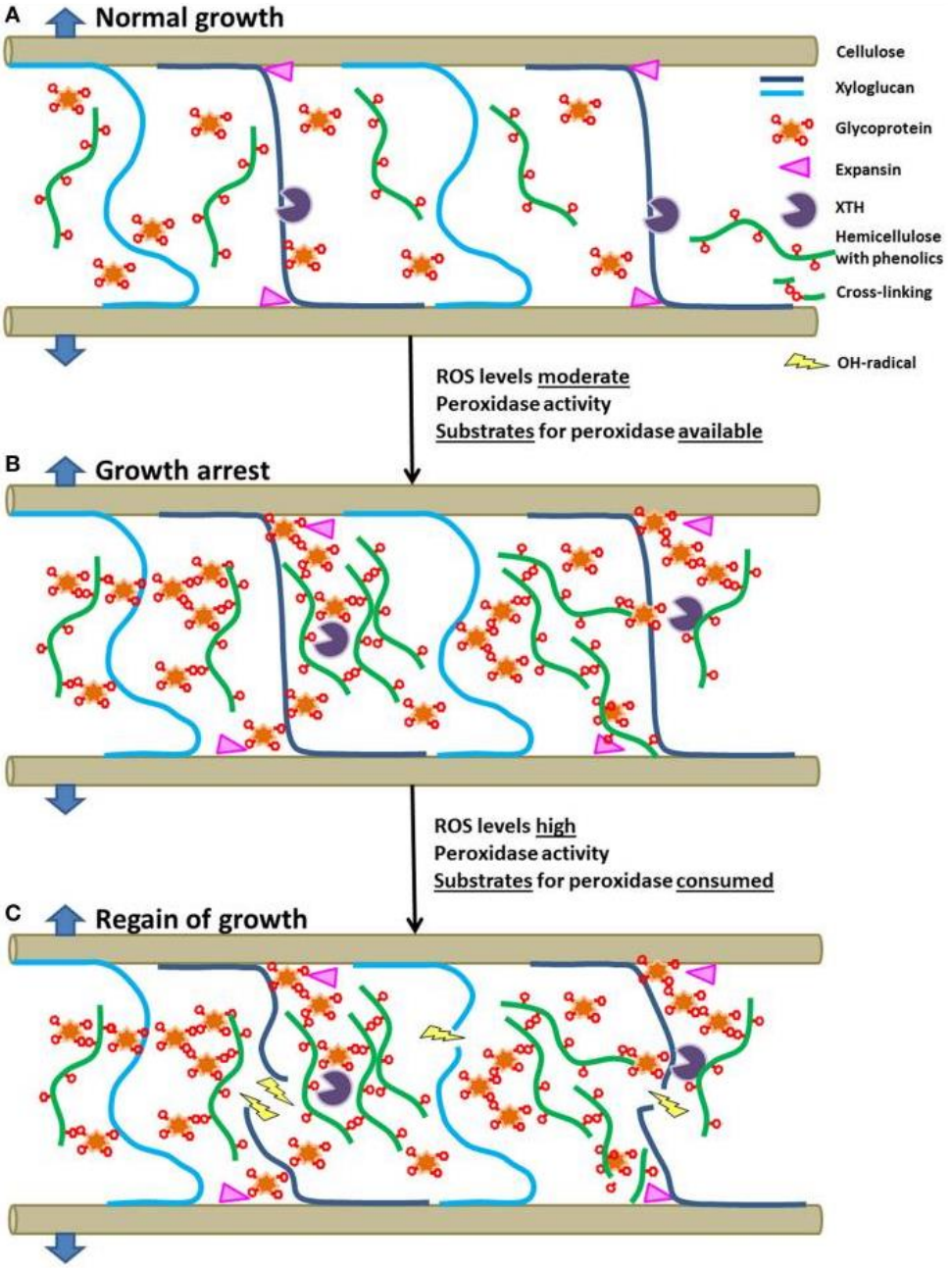


Figure 10 : *Modélisation de croissance de la paroi des plantes : en conditions non stressées (A), sous stress abiotique (arrêt de croissance) (B) tolérantes ou ayant rétabli leur croissance par l'action des ROS sur les polymères (C) (Tenhaken 2014).*

Si la production de ROS se poursuit (Fig. 10C), les niveaux élevés qui en résultent provoquent des clivages non enzymatiques des chaînes polymères, permettant à la cellule de croître à nouveau. Le processus de relâchement de la paroi cellulaire ainsi obtenu est fonctionnellement équivalent à une croissance non stressée. Les expansines et les XTH, responsables du relâchement de la paroi, sont fonctionnelles, et contribuent à la dynamique réversible du stress (Le Gall *et al.* 2015). La balance entre les ROS et les peroxydases, enzymes qui utilisent ces composés comme substrat, est donc essentielle à cette plasticité (Hamann 2012).

Les pectines sont également souvent modifiées lors de période de stress. Par exemple chez le blé, la propension de synthèse des chaînes latérales des RGI et RGII augmente lors d'un stress salin. Les niveaux plus élevés de pectines HG augmentent la proportion de gels hydratés dans les parois, limitant ainsi les dommages des cellules (Leucci *et al.* 2008). Chez *Miscanthus giganteus*, un changement de composition pariétale est induit lors d'une acclimatation au froid (Domon *et al.* 2013) par une augmentation considérable des quantités de N-glucanes ((1 - 3), (1 - 4) – β – d - glucane) et d'alcool cinnamique déshydrogénase (CAD). Lors d'un traitement aux basses températures, l'acclimatation au froid suggère également une implication de la lignification.

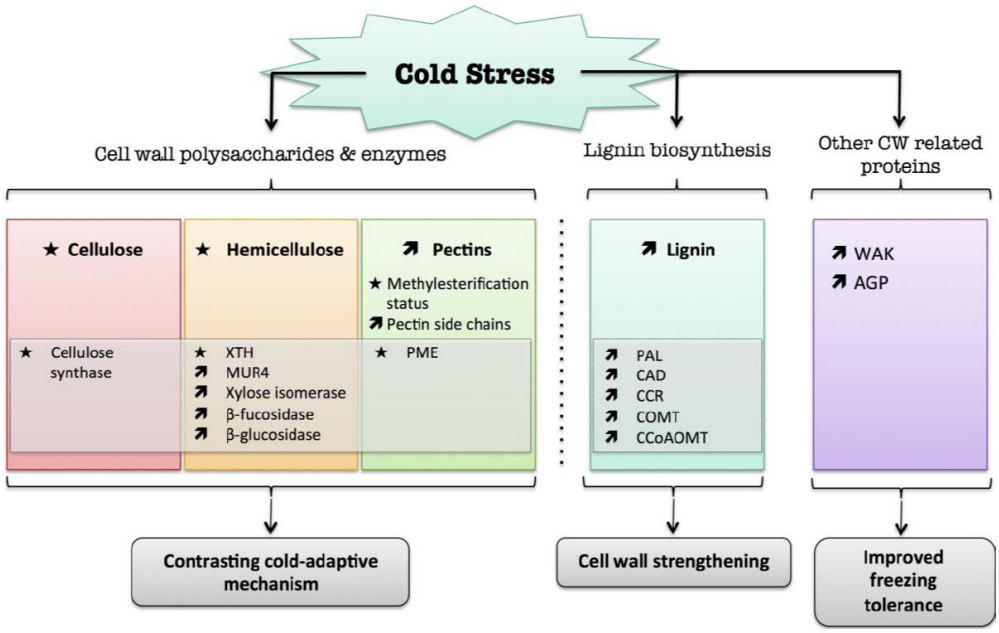


Figure 11 : *Présentation schématique de la réponse des parois végétales à un stress froid. Résultat des différentes études compilées dans l'article de Le Gall et al. 2015. Les flèches indiquent l'augmentation d'abondance des composés ; les étoiles signifient des résultats contrastés entre les données génétiques et protéomiques dans la littérature ; protéine à arabinogalactane (AGP) ; alcool cinnamique déshydrogénase (CAD) ;caffeoyl-CoA 3-O-methyl-transferase (CCoAOMT) ; cinnamoyl-CoA reductase (CCR) ; caffeate O-methyltransferase (COMT) ; paroi végétale (CW) ; phénylalanine ammonia-lyase (PAL) ; pectine methylesterase (PME) ; peroxydase pariétale (PRX) ; UDP- D -xylose 4-epimerase (MUR4) ; kinase associée à la paroi (WAK) ;xyloglucane endo-β-transglucosylases/hydrolases (XET/XTH).*

Par son caractère dynamique, la paroi cellulaire n'est pas facile à analyser. La preuve en est que la majorité des études portant sur les modifications de la paroi végétale de plantes soumises à un stress abiotique se focalise principalement sur les gènes impliqués dans le métabolisme des polysaccharides pariétaux (Fig. 11). L'amélioration des technologies d'analyse protéomique a tout de même permis d'initier des études comparatives entre des protéomes de plantes stressées et des plantes témoins (Ghosh & Xu 2014; Zhang *et al.* 2016).

À ce jour, peu d'études sont réalisées sur l'ensemble de ces jeux de données avec une approche dite intégrative.

I.4 Statistiques et l'intégration de données

L'analyse statistique d'intégration de différents types de données peut être une aide puissante pour identifier de nouveaux candidats (gènes, protéines, métabolites, *etc*.) et ainsi découvrir de nouvelles voies métaboliques qui permettent aux plantes de s'adapter aux contraintes environnementales. Le développement des nouvelles technologies offre la capacité d'étudier la quantité et la diversité des molécules à différentes échelles (du gène à la plante). Le séquençage de génomes d'organismes de référence a permis de prédire avec précision des gènes, des ARN et des protéines facilitant ainsi ces approches. La disponibilité des génomes a permis le développement des technologies d'analyse à haut débit, dites analyses omiques (Fig. 12). Parmi lesquelles se trouvent la génomique (pour l'analyse de l'ensemble des gènes), la transcriptomique (pour l'analyse de l'ensemble des transcrits), la protéomique (pour l'analyse de l'ensemble des protéines), la métabolomique (pour l'analyse de l'ensemble des métabolites), *etc*. Cette terminologie a par la suite été élargie pour couvrir d'autres types de données comme par exemple les données phénomiques (pour l'ensemble des phénotypes).

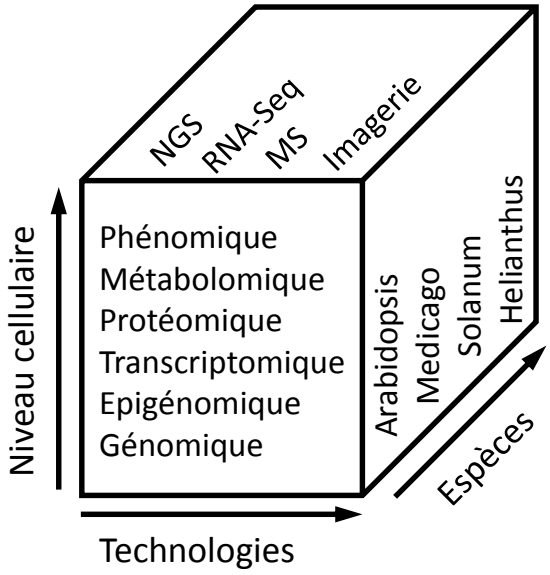


Figure 12 : *Représentation de la complexité multidimensionnelle des données biologiques en lien avec les avancées dans les technologies à haut débit. L'hétérogénéité des données générées peut être attribuée aux différents niveaux d'études, de la cellule à la plante entière en utilisant une large gamme de techniques à travers différentes espèces végétales (d'après Rajasundaram & Selbig 2016).*

Les technologies omiques produisent à débit élevé de grandes quantités de données qu'il n'est plus possible d'analyser manuellement. Une nouvelle approche a donc vu le jour, fournissant de nouveaux éléments pour analyser ces données et mieux comprendre la réponse d'un organisme au stress : cette approche est nommée biologie des systèmes.

La biologie des systèmes, en tant qu'approche holistique, consiste à intégrer les données de diverses disciplines dans des modèles bio-statistiques afin de comprendre la réponse physiologique d'un organisme dans sa globalité (Fig. 13). Ainsi, la biologie des systèmes, analyse d'une part les réseaux impliqués dans la réponse aux stress, et expose d'autre part la dynamique de ces réponses. (Chawla *et al.* 2011).

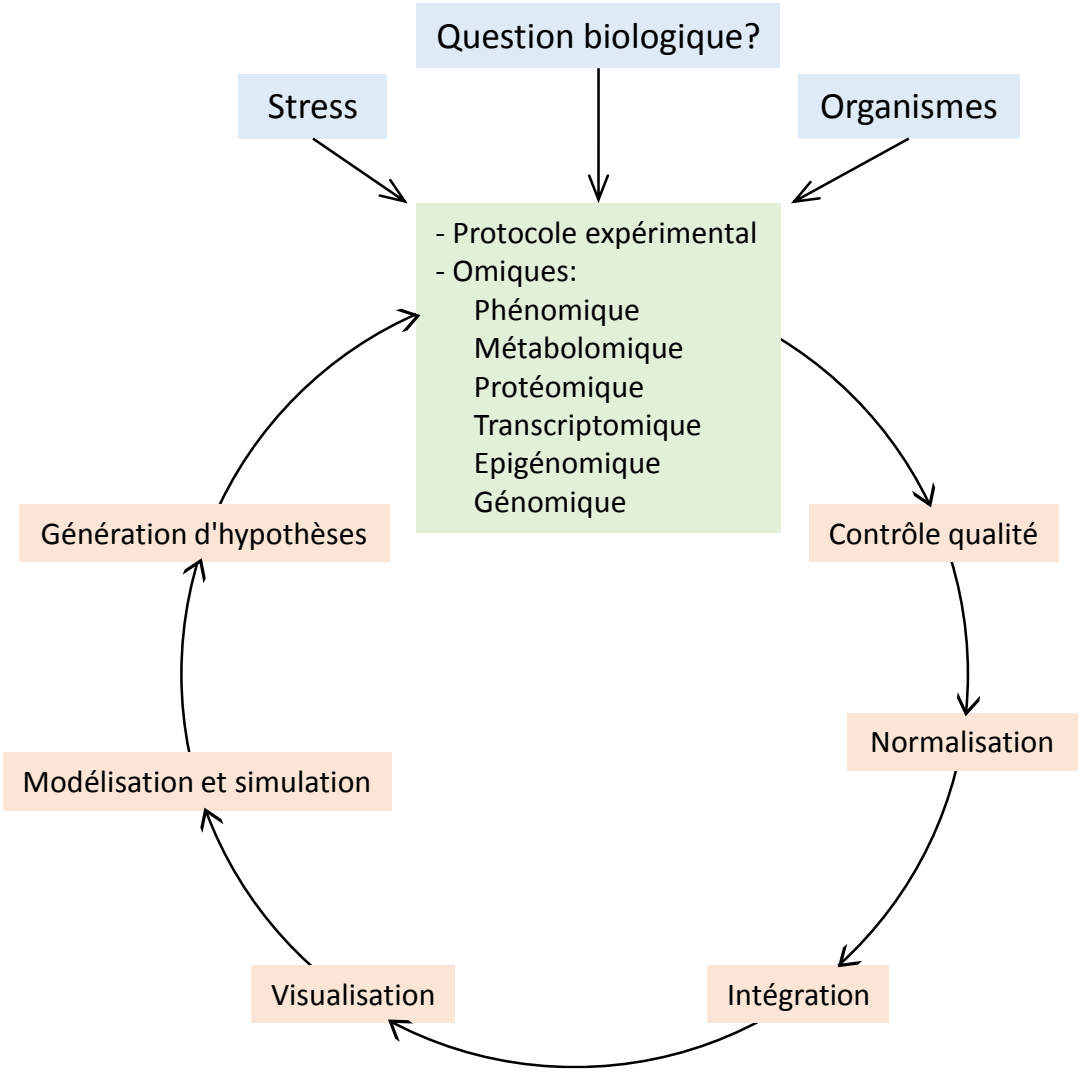


Figure 13 : *Approche cyclique de la biologie des systèmes lors de la réponse d'une plante à différents stress. L'hypothèse biologique est testée et analysée en utilisant un ou plusieurs types de données omiques. L'intégration des données suivie de la modélisation permet de valider ou de proposer une nouvelle hypothèse (d'après Chawla et al. 2011).*

Ces analyses ont largement été développées *via* l'utilisation du logiciel R et des packages associés qui ont vu le jour comme FactoMineR (Lê *et al.* 2008) ou mixOmics (Lê Cao *et al.* 2009). De façon schématique, les analyses statistiques intégratives sont structurées principalement en quatre couches d'analyses, de la plus simple à la plus complexe (Fig. 14) :

i) l'analyse uni-variée explore une seule variable à la fois (l'expression d'un gène, l'abondance d'une protéine, *etc*.) et se traduit par des statistiques élémentaires comme une moyenne, une médiane, des écarts types, *etc*. qui peuvent être représentées par des graphiques simples de type diagramme en barres, boîte à moustaches, *etc*.

ii) l'analyse bi-variée permet d'observer les liens entre deux variables qu'elles soient catégorielles ou quantitatives. C'est l'analyse la plus couramment utilisée en laboratoire car elle permet par exemple d'étudier la corrélation entre le niveau d'expression d'un gène et l'abondance de la protéine associée (coefficients de corrélation de Pearson, de Spearman), ou d'observer l'effet d'un traitement ou d'un génotype sur l'expression de gènes (tests de Student, de Wilcoxon, ANOVA, *etc*.) ou encore de tester l'indépendance de 2 variables catégorielles (test de chi2, test exact de Fisher, *etc*.). Ces analyses sont représentées aussi bien par des diagrammes en boîte que par des matrices de corrélation. Ces aspects sont détaillés dans la publication de Saporta (2006) et vulgarisés dans l'article de Van Eeuwijk *et al.* 2016.

iii) l'analyse multivariée explore un jeu de données dans son ensemble (transcriptomique, protéomique, métabolomique, *etc*.) de façon supervisée ou non :

   a. l'analyse multivariée non supervisée correspond généralement à la mise en œuvre d'une Analyse en Composantes Principales (ACP, ou PCA pour *Principal Component Analysis*) qui explique au mieux les informations de l'ensemble des données (Bro & Smilde 2014) ou encore à des techniques de classification (*clustering*). Une synthèse des approches exploratoires multidimensionnelles est proposée dans Lebart *et al.* 2006.

   b. l'analyse multivariée supervisée permet, par exemple, de classer les échantillons (écotypes, traitements,…) à partir d'informations quantitatives sur ces échantillons. Ainsi la régression des moindres carrés partiels en version discriminante (PLS-DA pour *Partial Least Squares regression Discriminant Analysis*) analyse un jeu de données en discriminant les échantillons en fonction d'une variable qualitative attribuée (écotypes, traitements, *etc*.) (Barker & Rayens 2003; Pérez-Enciso & Tenenhaus 2003). Cette méthode est plus adaptée que la méthode plus classique

d'analyse factorielle discriminante dans les cas où le nombre de variables dépasse largement le nombre d'individus comme c'est le cas pour les études omiques.

iv) l'analyse dite multi-blocs, permet d'intégrer ensemble plusieurs jeux de données, appelés aussi blocs (transcriptomique, protéomique, métabolomique, *etc*.). Cette méthodologie de statistiques intégratives a des origines relativement anciennes (Carroll 1968; Kettenring 1971) mais son usage n'a été généralisé que très récemment en partie grâce à l'essor de la biologie des systèmes mais ne comporte que peu de modèles. Ces analyses multi-bloc peuvent être supervisées ou non :

   a. les analyses multi-blocs non supervisées comme la PLS et la CCA (pour *Canonical Correlation Analysis*) qui mettent en lien deux blocs quantitatifs entre eux (transcrits et protéines, transcrits et climatologies, *etc*.) ont été étendues, d'une part, pour traiter les cas avec de nombreuses variables (González *et al.*) et d'autre part pour traiter plus de 2 blocs simultanément (Tenenhaus 2011 & 2014).

   b. les analyses supervisées de type multi-blocs ont été développées encore plus récemment (Günther *et al.* 2014; Singh *et al.* 2016). Elles permettent d'intégrer ces mêmes blocs mais avec un *a priori* sur une classe de variable qualitative préalablement attribuée.

Enfin, toutes les analyses multivariées et multi-blocs peuvent être effectuées en ajoutant des pénalités de type LASSO (pour *Least Absolute Shrinkage and Selection Operator*) (Tibshirani 1996). Ce mode d'analyse sélectionne les variables les plus informatives des jeux de données en supprimant les moins informatives, permettant ainsi de faire ressortir les candidats ayant les profils les plus intéressants pour le biologiste. Ce mode d'analyse est implémenté dans le package mixOmics sous le nom de *Sparse* comme la s-PCA (*Sparse PCA*) et la s-PLS (*Sparse PLS*). Ces analyses de statistiques sont détaillées plus spécifiquement dans la partie II.7.

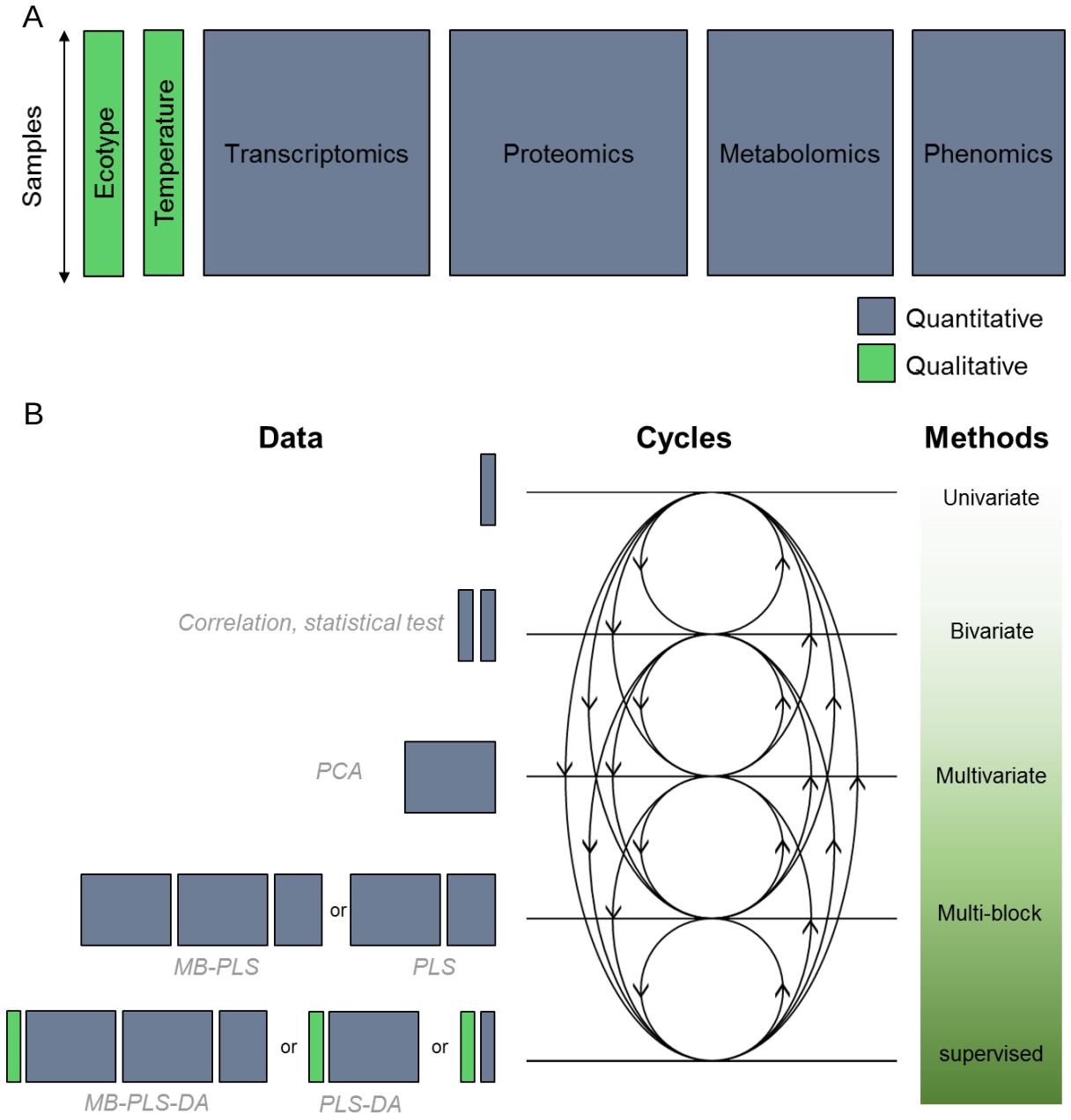


Figure 14 : *Schéma des méthodes statistiques d'analyses de données. A) Structure et nature des données utilisées lors d'analyses intégratives. B) Vue d'ensemble des différentes étapes de la méthodologie globale lors d'une analyse intégrative. A chaque étape supplémentaire de l'analyse, les résultats sont confrontés aux analyses précédentes permettant une décomposition des informations observées.*

L'interprétation des données devient de moins en moins triviale avec l'augmentation du nombre d'analyses cumulées. La discussion entre le statisticien et le biologiste est donc nécessaire afin de fournir les analyses les plus pertinentes pour répondre à une question biologique donnée.

I.5 Les objectifs de la thèse

Les objectifs de ce projet de thèse étaient d'évaluer les réponses d'*A. thaliana* face au réchauffement climatique par une approche intégrative innovante alliant l'écologie, la génétique, les technologies omiques et des données de phénotypages. Pour relier ces résultats à une problématique de réchauffement climatique, ces analyses ont porté sur des populations naturelles provenant d'altitudes contrastées des Pyrénées.

Une approche intégrative n'est possible que lorsque des données aussi différentes que des données d'expression de gènes et de protéines, des données de météo ou des analyses phénotypiques sont produite. La première partie de ce manuscrit portera sur les verrous expérimentaux et analytiques rencontrés et inhérents aux traitements de données hétérogènes.

La seconde partie détaillera les analyses génétique et phénotypique réalisées sur les nouvelles populations identifiées et récoltées dans les Pyrénées. De plus, les liens entre génétique, origine climatique et spécificité phénotypique mettront en évidence la distribution et la variabilité naturelle des populations pyrénéennes.

La dernière partie du manuscrit englobera deux études intégratives de données omiques sur la thématique de la plasticité pariétale soumise à des conditions de températures optimales et sub-optimales. La première étudie des rosettes de deux écotypes connus provenant de conditions de croissances contrastées. La seconde est basée sur l'analyse des rosettes et des tiges de quatre populations naturelles des Pyrénées préalablement sélectionnées.

II. Production de données omiques et méthodologies d'analyses

Au cours de cette étude, deux verrous principaux ont dû être levés : (i) un verrou expérimental lié à l'hétérogénéité des nombreux jeux de données omiques produits et (ii) un verrou analytique lié au traitement de ces données hétérogènes. Afin de répondre à ce premier verrou, une méthodologie et une analyse spécifiques ont étés développées sur l'écotype de référence Columbia (Col).

II.1 Protocole expérimental

Des expériences de phénotypage, d'analyse des monosaccharides pariétaux, de protéomiques pariétales, ainsi que de transcriptomiques ont également étés effectuées en chambre de culture (90 µmol.photons.m^{-2}.s^{-1}, 70 % d'humidité avec une photopériode de 16 h) à deux températures : 22 et 15°C.

Le choix de ces valeurs températures a été fait sur la base des critères suivants. Tout d'abord, la condition de culture à 22°C est considérée comme une condition standard pour cette espèce (Boyes *et al.* 2001) permettant d'obtenir un état physiologique de référence. Au contraire, la condition de culture à 15°C représente une condition dite « sub-optimale » qui est caractérisée généralement par un ralentissement du développement de la plante. Cultivées dès le début à cette température les plantes s'acclimatent à ce stress modéré et vont mettre en avant au travers de leurs plasticités phénotypiques, des capacités d'adaptation.

La première partie de la thèse a consisté à produire de larges jeux de données quantitatives et reproductibles pour être utilisées à des fins d'analyses statistiques. L'analyse de protéines pariétales nécessitait une quantité suffisante de matériel biologique. Afin de répondre à ce prérequis, une vingtaine de plants d'*A. thaliana* ont étés produits dans chaque condition.

II.2 Le macro-phénotypage

Dans ce travail, les données de macro-phénotypage sont essentielles car elles illustrent les variations morphologiques liées à une adaptation que l'on cherche à expliquer.

Le phénotypage des rosettes a été réalisé après 4 et 6 semaines pour les rosettes ayant poussé respectivement à 22 et 15°C, dates auxquelles le bourgeon d'initiation de la tige principale apparaît. Ainsi, le nombre de feuilles, la masse et le diamètre des rosettes ont été

analysés. Une photographie a systématiquement été prise (Fig. 15) avant collecte dans l'azote liquide pour déterminer l'aire de la surface exposée des rosettes *via* le logiciel *ImageJ* v. 1.51.

Les tiges ont été récoltées à 6 semaines à 22 °C et 8 semaines à 15 °C. Dans les deux conditions, lors de la récolte, la hauteur, le diamètre, la masse des tiges et le nombre de feuilles caulinaires ont été analysés. Pour finir, une photographie a été réalisée avant la congélation de chaque plante (Fig. 15).

Figure 15 : *Illustrations de l'écotype de référence Col cultivé à 22°C. A) rosette d'*A. thaliana *au moment de l'apparition du bourgeon d'initiation de la tige principale. B : tige principale lors de la récolte.*

II.3 Le micro-phénotypage

Dans un deuxième temps, l'étude anatomique des rosettes et des tiges (micro-phénotypage) a permis de relier le macro-phénotypage aux analyses moléculaires. Afin d'observer les différences anatomiques entre les rosettes et les tiges d'*A. thaliana*, des coupes transversales de ces tissus ont été réalisées. L'observation en lumière blanche après coloration ou en UV permet de visualiser certains composants de la paroi. Cette approche permet une analyse qualitative, telle que la composition en composés hydrophobes par un ratio de coloration bleu Alcian / rouge Astra (Fig. 16A). Des données quantitatives, comme le dénombrement de vaisseaux conducteurs ou la taille des pétioles ont aussi servi à compléter les informations de phénotypage. (Fig. 16B)

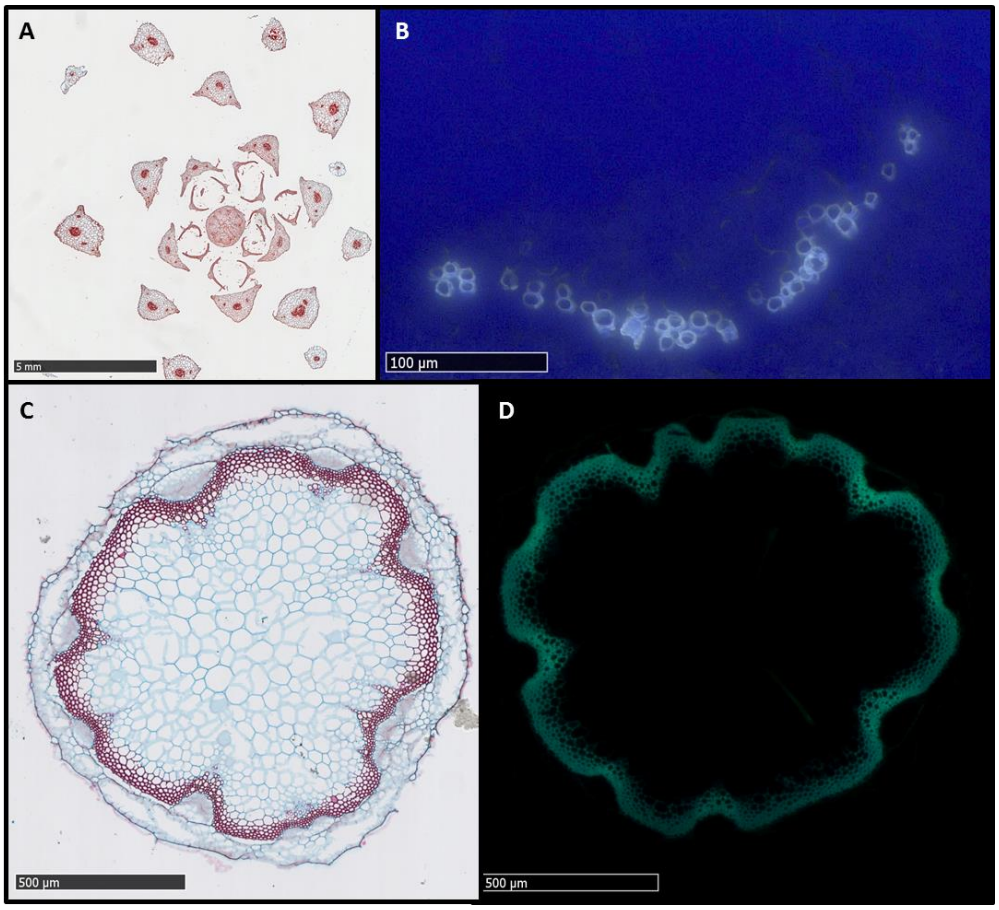


Figure 16 : *A) Coupe transversale de rosette cultivée à 15°C colorée au bleu Alcian et rouge Astra. Barre d'échelle : 5 mm ; B) Analyse par auto-fluorescence d'un pétiole de rosettes de Col poussé à 15°C. Barre d'échelle : 100 µm ; C) Coupe transversale de tige principale cultivée à 15°C colorée au bleu Alcian et rouge Astra et D) analysée par auto-fluorescence. Barre d'échelle : 500 µm*

Les conclusions tirées de ces analyses ont été complétées par des analyses immunologiques de la composition en polysaccharides de la paroi (détaillé en partie IV.1).

II.4 Les analyses des sucres pariétaux

L'extraction des polysaccharides pariétaux nécessite une préparation spécifique des parois. Pour cela, les rosettes et tiges ont tout d'abord été lavées puis mixées dans un tampon d'acétate et saccharose accompagné d'un cocktail inhibiteur de protéases. La purification des parois a été réalisée suivant le protocole décrit par Feiz *et al.* (Feiz *et al.* 2006) ; l'extraction séquentielle des polysaccharides, adaptée d'Irshad *et al.* (Irshad *et al.* 2008), est détaillée dans l'article (Duruflé *et al.* 2017).

Après hydrolyse acide, sept monosaccharides (Rha, Ara, Gal, Fuc, Glc, Xyl et GalA) ont été détectés par HPAEC (*High Performance Anion Exchange Chromatography*) (Fig. 17).

Il est à noter que la dernière fraction (E5) enrichie en cellulose n'a pas été analysée en raison de l'utilisation de cadoxen (cadmium ethylenedianiine) pour son extraction. En effet, ce composé ne permet pas une hydrolyse acide dans nos conditions.

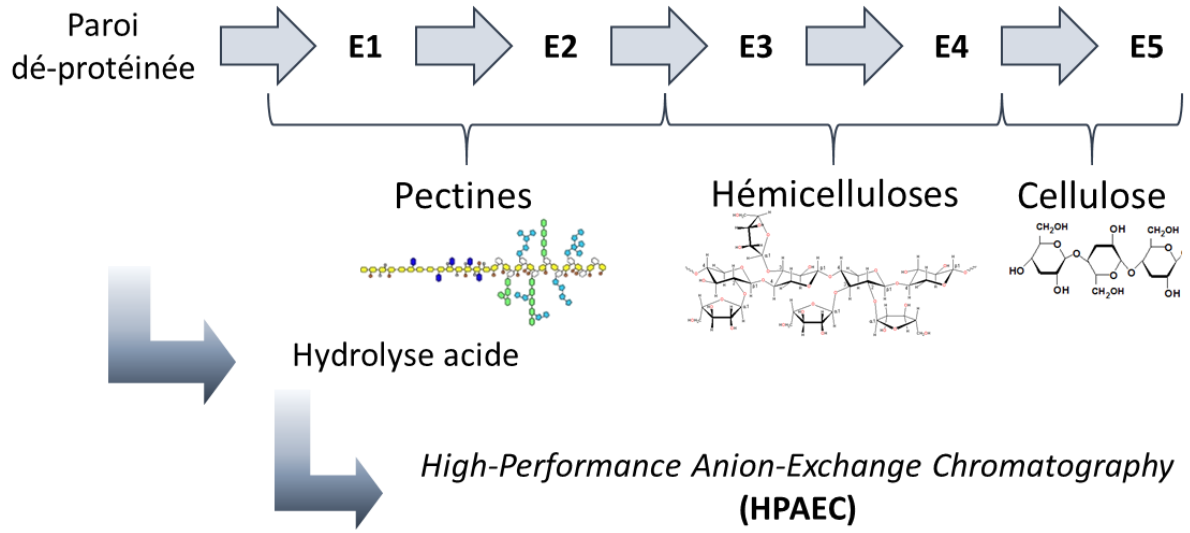


Figure 17 : *Schéma récapitulatif de l'analyse des polysaccharides pariétaux. Après une extraction séquentielle, les 5 extraits obtenus (E1, E2, E3, E4 et E5) sont hydrolysés en condition acide et les monosaccharides ainsi obtenus sont analysés par HPAEC.*

L'analyse des sept monosaccharides provenant des quatre premiers extraits (E1-E4) peut difficilement être associée à une analyse statistique intégrative en raison de la difficulté d'interprétation biologique. Pour permettre l'étude de la variabilité de la composition des parois, les polysaccharides ont été reconstruits à partir des données issues de l'analyse des monosaccharides. Cette reconstruction utilise les formules suivantes adaptées de Houben *et al.* 2011 (Tab. I).

Table. I : *Reconstruction des polysaccharides pariétaux par l'analyse des monosaccharides. Les formules sont adaptées d'Houben et al.2011 ; masse moléculaire du GalA (MGalA), masse moléculaire du Rha (MRha).*

Polysaccharides	Formule
Pectine rhamnogalacturonane 1	(Rha x $(1 + \frac{MGalA}{MRha})$) + Ara + Gal
Pectine homogalacturonane	GalA - (Rha x $(1 + \frac{MGalA}{MRha})$)
Xyloglucane	Fuc + Glc + Xyl
Pectine linéaire	(GalA -Rha) / ((Rha x $(1 + \frac{MGalA}{MRha})$) + Ara + Gal)
Contribution des RG dans les pectines	(Rha x $(1 + \frac{MGalA}{MRha})$) / (GalA - Rha)
RG1 ramifié	(Ara + Gal) / (Rha x $(1 + \frac{MGalA}{MRha})$)

Ces analyses en polysaccharides reconstitués seront utilisées par la suite lors des analyses intégratives. La structuration et le remaniement de ce réseau complexe de polysaccharides est dynamique. Les protéines pariétales sont essentielles à cette plasticité de par leurs interactions qui forment des assemblages supramoléculaires complexes. Il est ainsi impératif de prendre en compte l'identification et l'étude des différentes variations en abondance des protéines pariétales.

II.5 L'analyse du protéome pariétal des rosettes d'*A. thaliana*.

Le protéome pariétal de rosette d'*A. thaliana* est peu connu avec seulement 148 protéines pariétales identifiées sur les 700 prédites comme étant sécrétées (WallProtDB, http://www.polebio.scsv.ups-tlse.fr/WallProtDB/). Les études précédemment réalisées sur cette organe ont été réalisées avec une méthode non destructive d'infiltration sous vide (Boudart *et al.* 2005; Haslam *et al.* 2003; Trentin *et al.* 2015).

A l'occasion de la mise en œuvre de notre projet, deux méthodes d'analyse des protéomes ont été utilisées :

i) une méthodologie destructive d'extraction des protéines pariétales suivie d'une séparation par une electrophorèse-1D courte. Chaque piste a ensuite été découpée en trois bandes pour être analysée en spectrométrie de masse (MS). Cette séparation permettait de diviser par trois le nombre de protéines analysées à chaque passage en spectrométrie de masse, augmentant ainsi la couverture et la finesse de l'analyse ;

ii) une méthodologie reposant sur la même méthode destructive d'extraction de protéines pariétales, mais cette fois suivie d'une stratégie de type *shotgun* pour l'analyse en MS. Cette technique a notamment permis de réduire considérablement le temps d'analyse.

Il était important de comparer les différentes méthodologies et techniques d'analyse avant de poursuivre par une exploitation quantitative. Cette étude comparative est d'ailleurs discutée dans une publication parue dans la revue *Proteomics*, reprise intégralement ci-après.

DATASET BRIEF

An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes

Vincent Hervé[1,2], Harold Duruflé[1*], Hélène San Clemente[1], Cécile Albenne[1], Thierry Balliau[3,4], Michel Zivy[3,4], Christophe Dunand[1] and Elisabeth Jamet[1]*

[1] Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Castanet Tolosan, France
[2] INRS—Institut Armand Frappier, Laval, Canada
[3] CNRS, PAPPSO, UMR 0320/UMR 8120 Génétique Végétale, Gif sur Yvette, France
[4] INRA, PAPPSO, UMR 0320/UMR 8120 Génétique Végétale, Gif sur Yvette, France

Received: July 7, 2016
Revised: September 28, 2016
Accepted: October 21, 2016

Plant cells are surrounded by cell walls playing many roles during development and in response to environmental constraints. Cell walls are mainly composed of polysaccharides (cellulose, hemicelluloses and pectins), but they also contain proteins which are critical players in cell wall remodeling processes. Today, the cell wall proteome of *Arabidopsis thaliana*, a major dicot model plant, comprises more than 700 proteins predicted to be secreted (cell wall proteins—CWPs) identified in different organs or in cell suspension cultures. However, the cell wall proteome of rosettes is poorly represented with only 148 CWPs identified after extraction by vacuum infiltration. This new study allows enlarging its coverage. A destructive method starting with the purification of cell walls has been performed and two experiments have been compared. They differ by the presence/absence of protein separation by a short 1D-electrophoresis run prior to tryptic digestion and different gradient programs for peptide separation before mass spectrometry analysis. Altogether, the rosette cell wall proteome has been significantly enlarged to 361 CWPs, among which 213 newly identified in rosettes and 57 newly described. The identified CWPs fall in four major functional classes: 26.1% proteins acting on polysaccharides, 11.1% oxido-reductases, 14.7% proteases and 11.7% proteins possibly related to lipid metabolism.

Keywords:
Arabidopsis thaliana / Cell wall / Leaf / Plant proteomics / Rosette

Additional supporting information may be found in the online version of this article at the publisher's web-site

Plant cells are surrounded by cell walls playing many roles during development and in response to environmental constraints. Cell walls are mainly composed of complex polysaccharidic networks of cellulose, hemicelluloses and pectins [1]. They contain many proteins which diversity has been revealed by dedicated proteomics studies [2–4]. These proteins also called cell wall proteins (CWPs) are critical players in cell wall remodeling processes and in signaling [5, 6]. Our knowledge of plant cell wall proteomes is quickly enlarging since the beginning of the development of mass spectrometry (MS) technologies applied to protein analysis in the early 2000's. Today, the cell wall proteome of *Arabidopsis thaliana*, a major dicot model plant, is the best described. It comprises more than 700 proteins predicted to be secreted which have been identified in different organs or in cell suspension cultures, which is between one-third and one-half of its expected size [7]. However, the cell wall proteome of rosettes is poorly represented with only 148 CWPs identified after extraction of proteins from the rosette apoplast by vacuum infiltration [8–10], described as a non-destructive method [3]. This new study aims at enlarging the coverage of this proteome using

Correspondence: Dr. E. Jamet, UMR 5546 UPS/CNRS, Laboratoire de Recherche en Sciences Végétales, BP 42617, F-31326 Castanet-Tolosan, France
E-mail: jamet@lrsv.ups-tlse.fr
Fax: +33(0)534 32 38 02

Abbreviation: CWP, cell wall protein

*Co-first authors
Colour Online: See the article online to view Figs. 1 and 2 in colour.

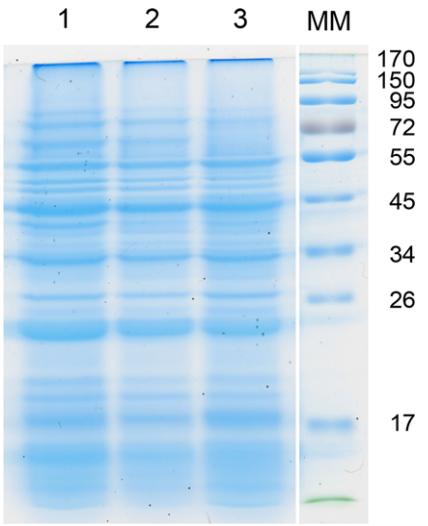


Figure 1. 1DE patterns of *A. thaliana* proteins extracted from purified rosette cell walls by $CaCl_2$ 0.2 M and LiCl 2 M solutions. Forty μg of proteins of each sample (**1**, **2**, **3**: biological repeats 1, 2 and 3 of experiment 1, respectively) have been separated by regular 1DE and stained with Coomassie blue. Molecular mass (MM) markers are indicated in kDa.

a destructive method for protein extraction starting with the purification of cell walls [11, 12]. In addition, two different techniques of protein preparation prior to tryptic digestion and MS analysis have been compared. They differ at two levels: (i) one protein extract has been separated by a short 1D-electrophoresis (1DE) run (experiment 1), whereas the second has been subjected to shotgun analysis (experiment 2) and (ii) the separation of tryptic peptides by liquid chromatography (LC) has been performed with a shorter column (150 mm vs. 300 mm) and a gradient program steeper in experiment 1 than in experiment 2 (Supporting Information Fig. S1).

A. thaliana plants have been cultivated in growth chambers at 22°C with a photoperiod of 16 h light/8 h dark. Rosettes have been collected after 4 weeks. Twenty plants have been used per biological replicate and three biological replicates have been performed for each experiment. Briefly, cell walls have been purified as described [12]. Proteins have been extracted from lyophilized cell walls in four steps using a 5 mM acetate buffer pH 4.6 complemented with 0.2 M $CaCl_2$ (two successive extractions) or 2 M LiCl (two successive extractions) [11]. The four salt extracts were pooled to get the protein samples to be analyzed. Typically, 35 mg dry cell walls/g fresh material and 0.57 mg proteins/g fresh material were obtained. The quality of the protein extracts has been checked by regular 1DE. The presence of thin well-separated bands after Coomassie blue staining showed the absence of protein degradation. In addition, all three biological repeats for each experiment looked similar (see Fig. 1 showing the samples of experiment 1).

Experiment 1 included a short separation (5 mm run) of proteins by 1DE, followed by a cutting of the gel in three pieces prior to tryptic digestion of proteins in gelo [13]. Experiment 2 was a shotgun protein analysis directly after their extraction from purified cell walls. Proteins were digested in solution by trypsin as described [14]. For the two experiments, LC-MS/MS analyses were performed with a Q-exactive mass spectrometer (Thermo Fisher Scientific, Villebon-sur-Yvette) as described in Supporting Information Fig. S1. The X!Tandem software was used for protein identification [15] and the X!Tandem pipeline for MS data processing http://www.thegpm.org). The detailed procedure for LC-MS/MS analysis and the parameters used for peptide and protein identification are detailed in Supporting Information Tables S1–S2 (sheets "parameters"). The false discovery rates (FDRs) used for peptides and proteins are the following ones: 0.079% and 0%, respectively, for experiment 1; 0.076% and 0%, respectively, for experiment 2. All the MS/MS data have been made publicly available in the PROTICdb database (/http://moulon.inra.fr/protic/cell_wall_athaliana_rosettes) [16]. Those regarding the CWPs can be found in *WallProtDB* as well as their functional annotations (http://www.polebio.lrsv.ups-tlse.fr/WallProtDB) [17]. The identification of a given protein has been validated in one experiment when at least two proteotypic peptides of this protein were present in at least two biological replicates. Proteins have been annotated using the *ProtAnnDB* pipeline to predict sub-cellular localization and functional domains (http://www.polebio.lrsv.ups-tlse.fr/ProtAnnDB) [18]. Only the proteins having a predicted signal peptide, no endoplasmic reticulum retention signal and no more than one trans-membrane domain have been retained as CWPs. These predictions are done with several bioinformatics program and literature data are used when available to refine this analysis.

Altogether, 1093 proteins have been identified among which 328 have been predicted as CWPs. Experiments 1 (1D-E) and 2 (shotgun) have allowed the identification of 286 and 250 CWPs, respectively (Fig. 2A and Supporting Information Tables S1–S3). Even if a great proportion of the identified CWPs are in common between the two experiments, 78 have been specifically identified in experiment 1 and 42 in experiment 2. These features may be due to (i) differences in tryptic digestion efficiency, (ii) the splitting of the protein extract in three parts in experiment 1 and/or (iii) biological variability between the two experiments. It seems that CWPs specific to Experiment 1 or 2 were identified with less proteotypic peptides than those common to the two experiments, suggesting that they are less abundant proteins (see Supporting Information Fig. S2). Finally, the shotgun experiment was more convenient to handle, but it has led to the identification of less proteins.

This study has led to a significant enlargement of the rosette cell wall proteome to 361 CWPs, with 57 CWPs newly identified, 213 newly identified in rosettes, and 93 specific to this proteome (Table 1). One-hundred-and-fourteen CWPs

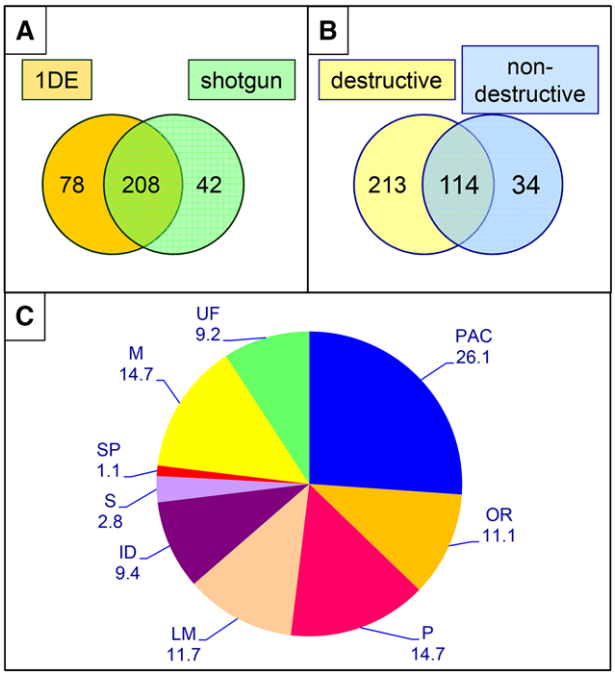


Figure 2. Characteristics of the new rosette cell wall proteome of *A. thaliana*. (A) Number of CWPs identified in experiment 1 (1DE) vs. number of CWPs identified in experiment 2 (shotgun). (B) Comparison of the number of CWPs identified in this study (destructive method) and in previous studies (non-destructive method) [8–10]. (C) Distribution (%) of CWPs identified in the overall rosette cell wall proteome (non-destructive and destructive methods) in functional classes according to bioinformatic prediction of functional domains: PAC (proteins acting on carbohydrates), OR (oxido-reductases), P (proteases), LM (proteins possibly related to lipid metabolism), ID (proteins with interactions domains with carbohydrates or proteins), S (proteins possibly involved in signaling), SP (structural proteins), M (miscellaneous proteins), UF (proteins of yet unknown function).

Table 1. Number of proteins specific to the enlarged rosette cell wall proteome and common to other cell wall proteomes: roots [21–23], stems [24], etiolated hypocotyls [11,12,25], cell suspensions culture and culture medium [26–31]

CWPs	Number
Newly identified	57
Newly identified in rosettes	213
Specific to rosettes	93
Common to roots	219
Common to stems	64
Common to etiolated hypocotyls	135
Common to cell suspension cultures/liquid medium	109

Data are from *WallProtDB* (http://www.polebio.lrsv.ups-tlse.fr/WallProtDB/index.php).

were found to be common between the new rosette cell wall proteome (destructive method) and the previously described ones (non-destructive methods) and 34 specific to the previous ones (Fig. 2B). These differences could be related to the fact that CWPs interact in various ways with cell wall components. The non-destructive method usually allows recovering proteins with weak interactions with cell wall components. Such proteins may be lost during the cell wall purification procedure used in the destructive method [2]. These two methods are complementary and have allowed increasing the coverage of the rosette cell wall proteome.

The distribution of CWPs in functional classes has revealed four major groups: 26.1% proteins acting on polysaccharides, 11.1% oxido-reductases, 14.7% proteases and 11.7% proteins possibly related to lipid metabolism (Fig. 2C). This distribution differs slightly from that of *Solanum tuberosum* mature leaves with less proteins acting on polysaccharides (33.6% in *S. tuberosum*), less proteases (19.8%) and more proteins possibly related to lipid metabolism (5.3%) [19]. It is also different from that of *Brachypodium distachyon* leaves with less oxido-reductases (13.9% in *B. distachyon*), less proteases (18.3%) and more proteins with interactions domains with polysaccharides or proteins (4.8%) [20].

The comparison of the rosette cell wall proteome to those of other *A. thaliana* organs has shown that 219 CWPs were common with that of roots [21–23], 64 to that of stems [24] and 135 to that of etiolated hypocotyls [11,12,25]. Besides, 109 CWPs were common to the cell wall proteomes of rosettes and cell suspension cultures [26–31] (Table 1). Altogether, the cell wall proteome of *A. thaliana* remains the best described with 766 CWPs indexed in *WallProtDB*. Each new set of data allows completing the CWP atlas and identifying CWPs specific to a given organ or present as housekeepers. This rosette cell wall proteome has been obtained from 4-week-old plants. It would be interesting to look for changes in its composition during plant development or in response to environmental cues. Such an approach has been undertaken in *B. distachyon* and has revealed changes between the protein profile of young and mature leaves [20]. Finally, the shotgun strategy appears as an interesting tool to get in a few steps a cell wall proteome from small amounts of material. It paves the way for the development of micro-proteomics in order to identify tissue-specific CWPs and associate these results to micro-transcriptomics or -metabolomics data for example.

The authors are thankful to Université Paul Sabatier (Toulouse, France) and CNRS for supporting their research. They also wish to thank Olivier Langella for introducing the LC-MS/MS data in the PROTICdb public repository. HD is supported by a grant of the Midi-Pyrénées region and of the Federal University of Toulouse. LC-MS/MS analyses were performed at the Plateforme d'Analyse Protéomique de Paris Sud-Ouest (PAPPSO).

The authors have declared no conflict of interest.

References

[1] Carpita, N. C., Gibeaut, D. M., Structural models of primary cell walls in flowering plants, consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* 1993, *3*, 1–30.

[2] Albenne, C., Canut, H., Hoffmann, L., Jamet, E., Plant cell wall proteins: a large body of data, but what about runaways? *Proteomes* 2014, *2*, 224–242.

[3] Albenne, C., Canut, H., Jamet, E., Plant cell wall proteomics: the leadership of *Arabidopsis thaliana*. *Front. Plant Sci.* 2013, *4*, 111.

[4] Rose, J., Bashir, S., Giovannoni, J., Jahn, M., Saravanan, R., Tackling the plant proteome: practical approaches, hurdles and experimental tools. *Plant J.* 2004, *39*, 715–733.

[5] Franková, L., Fry, S. C., Biochemistry and physiological roles of enzymes that "cut and paste" plant cell-wall polysaccharides. *J. Exp. Bot.* 2013, *64*, 3519–3550.

[6] Van der Hoorn, R., Plant proteases: from phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* 2008, *59*, 191–223.

[7] Jamet, E., Canut, H., Boudart, G., Pont-Lezica, R., Cell wall proteins: a new insight through proteomics. *Trends Plant Sci.* 2006, *11*, 33–39.

[8] Boudart, G., Jamet, E., Rossignol, M., Lafitte, C. et al., Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: identification by mass spectrometry and bioinformatics. *Proteomics* 2005, *5*, 212–221.

[9] Haslam, R. P., Downie, A. L., Raventon, M., Gallardo, K. et al., The assessment of enriched apoplastic extracts using proteomic approaches. *Ann. Appl. Biol.* 2003, *143*, 81–91.

[10] Trentin, A., Pivato, M., Mehdi, S., Barnabas, L. et al., Proteome readjustments in the apoplastic space of *Arabidopsis thaliana ggt1* mutant leaves exposed to UV-B radiation. *Front. Plant Sci.* 2015, *6*, 128.

[11] Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R., Jamet, E., A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers. *BMC Plant Biol.* 2008, *8*, 94.

[12] Feiz, L., Irshad, M., Pont-Lezica, R. F., Canut, H., Jamet, E., Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from Arabidopsis hypocotyls. *Plant Methods* 2006, *2*, 10.

[13] Ligat, L., Lauber, E., Albenne, C., San Clemente, H. et al., Analysis of the xylem sap proteome of *Brassica oleracea* reveals a high content in secreted proteins. *Proteomics* 2011, *11*, 1798–1813.

[14] Blein-Nicolas, M., Albertin, W., da Silva, T., Valot, B. et al., A systems approach to elucidate heterosis of protein abundances in yeast. *Mol. Cell. Proteomics* 2015, *14*, 2056–2071.

[15] Craig, R., Beavis, R., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, *20*, 1466–1467.

[16] Langella, O., Valot, B., Jacob, D., Balliau, T. et al., Management and dissemination of MS proteomic data with PROTICdb: example of a quantitative comparison between methods of protein extraction. *Proteomics* 2013, *13*, 1457–1466.

[17] San Clemente, H., Jamet, E., *WallProtDB*, a database resource for plant cell wall proteomics. *Plant Methods* 2015, *11*, 2.

[18] San Clemente, H., Pont-Lezica, R., Jamet, E., Bioinformatics as a tool for assessing the quality of sub-cellular proteomic strategies and inferring functions of proteins: plant cell wall proteomics as a test case. *Bioinform. Biol. Insights* 2009, *3*, 15–28.

[19] Lim, S., Chisholm, K., Coffin, R. H., Peters, R. D. et al., Protein profiling in potato (*Solanum tuberosum* L.) leaf tissues by differential centrifugation. *J. Proteome Res.* 2012, *11*, 2594–2601.

[20] Douché, T., San Clemente, H., Burlat, V., Roujol, D. et al., *Brachypodium distachyon* as a model plant toward improved biofuel crops: search for secreted proteins involved in biogenesis and disassembly of cell wall polymers. *Proteomics* 2013, *13*, 2438–2454.

[21] Basu, U., Francis, J. L., Whittal, R. W., Stephens, J. L. et al., Extracellular proteomes of *Arabidopsis thaliana* and *Brassica napus* roots: analysis and comparison by MUdPIT and LC-MS/MS. *Plant Soil* 2006, *286*, 357–376.

[22] Chen, Y., Ye, D., Held, M. A., Cannon, M. C. et al., Identification of the abundant hydroxyproline-rich glycoproteins in the root walls of wild-type Arabidopsis, an *ext3* mutant line, and its phenotypic revertant. *Plants* 2015, *4*, 85–111.

[23] Nguyen-Kim, H., San Clemente, H., Balliau, T., Zivy, M. et al., *Arabidopsis thaliana* root cell wall proteomics: increasing the proteome coverage using a combinatorial peptide ligand library and description of unexpected Hyp in peroxidase amino acid sequences. *Proteomics* 2016, *16*, 491–503.

[24] Minic, Z., Jamet, E., Negroni, L., der Garabedian, P. A. et al., A sub-proteome of *Arabidopsis thaliana* trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. *J. Exp. Bot.* 2007, *58*, 2503–2512.

[25] Zhang, Y., Giboulot, A., Zivy, M., Valot, B. et al., Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry* 2011, *72*, 1109–1123.

[26] Bayer, E. M., Bottrill, A. R., Walshaw, J., Vigouroux, M. et al., Arabidopsis cell wall proteome defined using multidimensional protein identification technology. *Proteomics* 2006, *6*, 301–311.

[27] Borderies, G., Jamet, E., Lafitte, C., Rossignol, M. et al., Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: a critical analysis. *Electrophoresis* 2003, *24*, 3421–3432.

[28] Chivasa, S., Ndimba, B. K., Simon, W. J., Robertson, D. et al., Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis* 2002, *23*, 1754–1765.

[29] Kwon, H.-K., Yokoyama, R., Nishitani, K., A proteomic approach to apoplastic proteins involved in cell wall regeneration in protoplasts of Arabidopsis suspension-cultured cells. *Plant Cell Physiol.* 2005, *46*, 843–857.

[30] Borner, G. H., Lilley, K. S., Stevens, T. J., Dupree, P., Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. *Plant Physiol.* 2003, *132*, 568–577.

[31] Oh, I. S., Park, A. R., Bae, M. S., Kwon, S. J. et al., Secretome analysis reveals an Arabidopsis lipase involved in defense against Alternaria brassicicola. *Plant Cell* 2005, *17*, 2832–2847.

Cette étude nous a permis de comparer notre méthodologie à celles suivies par les études précédentes. En tenant compte de la faible quantité de protéines produites et du nombre d'échantillons traités, la stratégie *shotgun* est apparue très pertinente pour la suite de notre étude. C'est donc avec cette méthodologie que l'analyse des protéomes de tige a été traitée.

II.6 Analyse du protéome pariétal des tiges d'*A. thaliana.*

Le projet *WallOmics* tient aussi à étudier la plasticité pariétale au niveau des tiges d'*A. thaliana*. Cet organe, servant essentiellement de support aux fleurs et à la propagation des graines, doit résister aux contraintes climatiques et mécaniques. Il se trouve donc que la paroi des cellules des tiges est très différente de celle des autres organes de la plante.

Cependant, malgré son rôle important, une unique analyse du protéome pariétal des tiges d'*A. thaliana* était disponible, permettant l'identification de seulement 86 protéines pariétales (Minic *et al.* 2007). Notre méthodologie, utilisant une méthode destructive de purification des parois, devrait permettre d'accroître la couverture et la connaissance du protéome pariétal des tiges d'*A. thaliana*.

Dans le contexte de l'étude qui suit, une comparaison entre les protéomes de tiges existant chez d'autres espèces végétales a été discutée. Cette étude comparative est détaillée dans la publication ci-après parue dans la revue *Proteomics*.

DATASET BRIEF

Cell wall proteome analysis of *Arabidopsis thaliana* mature stems

Harold Duruflé[1], Hélène San Clemente[1], Thierry Balliau[2], Michel Zivy[2], Christophe Dunand[1] and Elisabeth Jamet[1]

[1] Laboratoire de Recherche en Sciences Végétales, CNRS, UPS, Université de Toulouse, Auzeville, Castanet Tolosan, France
[2] PAPPSO, GQE - Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France

Received: November 17, 2016
Revised: January 18, 2017
Accepted: January 31, 2017

Plant stems carry flowers necessary for species propagation and need to be adapted to mechanical disturbance and environmental factors. The stem cell walls are different from other organs and can modify their rigidity or viscoelastic properties for the integrity and the robustness required to withstand mechanical impacts and environmental stresses. Plant cell wall is composed of complex polysaccharide networks also containing cell wall proteins (CWPs) crucial to perceive and limit the environmental effects. The CWPs are fundamental players in cell wall remodeling processes, and today, only 86 have been identified from the mature stems of the model plant *Arabidopsis thaliana*. With a destructive method, this study has enlarged its coverage to 302 CWPs. This new proteome is mainly composed of 27.5% proteins acting on polysaccharides, 16% proteases, 11.6% oxido-reductases, 11% possibly related to lipid metabolism and 11% of proteins with interacting domains with proteins or polysaccharides. Compared to stem cell wall proteomes already available (*Brachypodium distachyon*, *Sacharum officinarum, Linum usitatissimum, Medicago sativa*), that of *A. thaliana* stems has a higher proportion of proteins acting on polysaccharides and of proteases, but a lower proportion of oxido-reductases.

Keywords:
Arabidopis thaliana / Cell wall / Proteome / Stem

Additional supporting information may be found in the online version of this article at the publisher's web-site

Plants, being sessile, are adapted to withstand environmental stresses. In natura, the major impact consists in mechanical disturbance by other organisms and also by environmental factors, such as rain, temperature or wind. The plant cell wall represents an external physical barrier crucial to perceive and limit the effects of environmental changes on plant physiology [1]. As major contributors to the plant shape and to the function of the different organs, cell wall rigidity or viscoelastic properties can be modified [2]. They define specific individual cell surfaces and manage the remodeling. Cell walls also participate in cell-to-cell adhesion required for organ integrity and for the robustness required to withstand mechanical impacts and environmental stresses [3].

Plant cell walls are mostly composed of complex networks of polysaccharides (cellulose, hemicelluloses, pectins) [4] and proteins [5]. These proteins also called cell wall proteins (CWPs) are estimated to represent about 10% of the cell wall mass [6]. CWPs contribute to most of the modifications of the cell wall composition and architecture as well as to the phenotype by perceiving fluctuations of the environment and remodeling the cell wall [7, 8].

Arabidopsis thaliana stems play many roles in development, transport of waters and nutrients and takes on the role of erecting flowers and siliques as high as possible to increase

Correspondence: Dr. E. Jamet, UMR 5546 UPS/CNRS, Laboratoire de Recherche en Sciences Végétales; BP 42617, F-31326 Castanet-Tolosan, France
E-mail: jamet@scsv.ups-tlse.fr

Abbreviations: CE, carbohydrate esterase; **CWP**, cell wall protein; **GH**, glycoside hydrolase; **PME**, pectin methyl esterase; **Prx**, class III peroxidase

Colour Online: See the article online to view Figs. 1 and 2 in colour.

the success species propagation. As a consequence, their cell walls exhibit a dynamics, a structure and a composition different from those of other organs [9]. A previous study of the cell wall proteome of *A. thaliana* stems has led to the identification of 86 CWPs after extraction of total proteins with a buffer containing 0.2 M $CaCl_2$ and affinity chromatography on Concanavalin A to trap *N*-glycoproteins [10]. This new study aims at increasing the coverage of the *A. thaliana* stem cell wall proteome by using a destructive method relying on the purification of cell walls followed by an extraction of proteins with buffers containing 0.2 M $CaCl_2$ and 2 M LiCl.

A. thaliana plants have been cultivated in growth chambers at 22°C with a photoperiod of 16 h light/8 h dark. Mature stems have been collected after 6 weeks, i.e. 2 weeks after the bolting stage. Twenty plants have been used per biological replicate and three biological replicates have been performed for each experiment. Cell walls have been purified as described [11]. Proteins have been extracted by four successive washings of cell walls with 5 mM acetate buffer pH 4.6 complemented with 0.2 M $CaCl_2$ (twice) or 2 M LiCl (twice). The four protein extracts were combined prior to further analysis. Typically, 100 mg dry cell walls/g fresh material and 2 mg proteins/g dry material were obtained. The same amount of each protein extract from the three biological replicates have been separated by 1DE and stained with Coomassie blue to check their quality.

The procedure used to identify proteins was the same as the one described as experiment 2 in a previous work, which is a shotgun MS analysis [12] (Supporting Information Fig. S1). Briefly, the protein extracts have been digested in solution as described [13]. The samples were then analyzed by LC-MS/MS with a Q-Exactive mass spectrometer (Thermo Fisher Scientific, Villebon-sur-Yvette, France) (see Supporting Information Fig. S1). The parameters used for peptide and protein identification are indicated in Supporting Information Table S1 (sheet "parameters"). The false discovery rates (FDRs) for proteins and peptides were 0.0% and 0.0615%, respectively. The X!Tandem software (www.thegpm.org) and the X!Tandem pipeline have been used for MS data processing and protein identification (pappso.inra.fr/bioinfo/xtandempipeline/) [14]. All the MS/MS data have been made publicly available in PROTICdb (http://proteus.moulon.inra.fr/w2dpage/proticdb/angular/) [15] and those regarding CWPs in *WallProtDB* (www.polebio.lrsv.ups-tlse.fr/WallProtDB) [16]. The identification of a given protein has been validated when at least two specific peptides were found in at least two biological replicates. The annotation of the proteins has been performed with the *ProtAnnDB* pipeline which includes several bioinformatics programs to predict the sub-cellular localization of the proteins as well as their functional domains (www.polebio.lrsv.ups-tlse.fr/ProtAnnDB) [17]. Only the proteins with a predicted signal peptide, no ER retention signal and eventually one trans-membrane domain have been considered as CWPs [18]. This latter case corresponds to that of plasma membrane receptors.

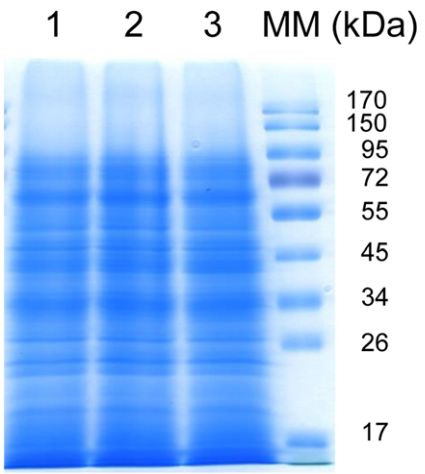


Figure 1. 1DE patterns of *A. thaliana* proteins extracted from purified stem cell walls with 0.2 M $CaCl_2$ and 2 M LiCl buffered solutions. 40 μg of each sample (1, 2, 3: biological replicates 1, 2, 3, respectively) have been separated by 1DE and stained with Coomassie blue. Molecular mass markers (MM) are in kDa.

On the three 1DE patterns corresponding to the three biological repeats (Fig. 1), protein bands appeared well resolved as expected in the absence of significant protein degradation. In addition, the three profiles looked very similar, thus showing the reproducibility of the experiment. This was confirmed after the identification of the proteins by LC-MS/MS since 85% of the proteins identified have been found in the three replicates (Supporting Information Table S1). Altogether, 951 proteins have been identified in at least two biological replicates, among which 808 (85.0%) in the three biological replicates and 143 (15.0%) in two out of three of them: 42 proteins (4.4%) were identified in biological replicates 1–2, 50 (5.3%) in biological replicates 2–3, and 51 (5.4%) in biological replicates 1–3. This elevated level of reproducibility is very promising for comparative analysis using large scale proteomics approaches, skipping the step of protein separation prior to MS/MS analyses.

Among the 951 identified proteins, 302 (31.8%) have been predicted to be secreted and named CWPs (Supporting Information Tables S1 and S2). The other 649 proteins (68.2%) have been considered as intracellular proteins contaminating the cell wall preparation. This percentage of intracellular proteins was higher than that observed in other plant stem proteomes obtained using similar methods to extract proteins from purified cell walls, such as *Brachypodium distachyon* (39.8% and 24.5% of intracellular contamination in apical and basal internodes, respectively) [19], *Saccharum officinarum* (44%) [20], *Medicago sativa* (27–37%) [21, 22] and *Linum usitatissimum* (31%) [23]. This percentage of intracellular proteins was also much higher than that observed (12.7%) in a previous *N*-glycoproteome study performed on *A. thaliana* stems [10]. In this latter case, the study focused on proteins targeted to the secretion pathway where the glycosylation of proteins occurs and which is the classical route to the cell wall

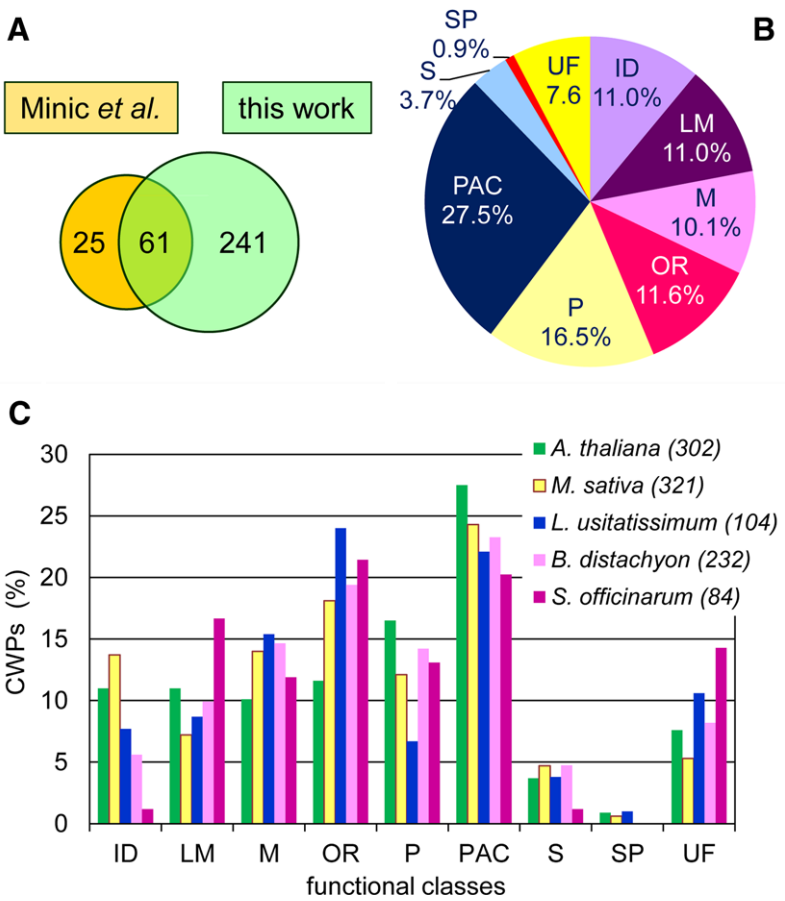


Figure 2. Features of the *A. thaliana* stem cell wall proteome. (A) Comparison of this work to that of Minic et al. (2007) [10] based on ConA affinity chromatography to isolate secreted proteins. (B) Distribution of CWPs into functional classes: proteins with interacting domains with proteins or polysaccharides (ID), proteins possibly related to lipid metabolism (LM), miscellaneous proteins (M), oxido-reductases (OR), proteases (P), proteins acting on cell wall polysaccharides (PAC), proteins possibly involved in signaling (S), structural proteins (SP), and proteins of yet unknown function (UF). (C) Comparison between the stem cell wall proteomes of *A. thaliana*, *M. sativa*, *L. usitatissimum*, *B. distachyon* and *S. officinarum*. The size of each cell wall proteome is indicated between brackets.

[24]. This enrichment in glycoproteins probably explains the low level of proteins predicted to be intracellular in this experiment. Altogether, the coverage of the stem cell wall proteome (302 CWPs) has been much enlarged with the identification of 241 CWPs not identified in the previously published *A. thaliana* stem cell wall proteome (Fig. 2A) [10]. In addition, 39 new CWPs not yet identified in *A. thaliana* cell wall proteomes (www.polebio.lrsv.ups-tlse.fr/WallProtDB) were found.

The CWPs have distributed into functional classes according to the prediction of functional domains [5] (Fig. 2B). As in other cell wall proteomes, the class of proteins acting on cell wall polysaccharides like glycosyl hydrolases (GHs), pectin methylesterases (PMEs), carbohydrate esterases (CEs) and expansins is the most populated (27.5%). The class of proteases is the second most important one (16.5%) with Asp proteases, Cys proteases, Ser carboxypeptidases and Ser proteases of the subtilisin type as the main families. Of similar importance are the classes of oxido-reductases (11.6%), proteins possibly related to lipid metabolism (11.0%), and proteins with interaction domains with proteins or polysaccharides (11%). Among the oxido-reductase, there are several class III peroxidases (Prxs) playing roles in the cross-linking of structural proteins and of monolignols and/or in the non-enzymatic breakage of polysaccharides [25], laccases possibly involved in lignification or in oxidative polymerization of flavonoids [26, 27], and multicopper oxidases of the SKU (SKEWED ROOT) family [28]. Miscellaneous proteins and proteins of yet unknown function represent 10.1% and 7.6% of the proteome, respectively. The signaling and structural proteins functional classes are minor ones representing 3.7% and 0.9% of the proteome respectively. As in previous cell wall proteomics studies, it has not been possible to extract covalently cross-linked structural proteins for lack of using a dedicated protocol [29]. Regarding signaling proteins, several fasciclin-like arabinogalactan proteins (FLAs) have been identified. Such proteins have been found to be associated to secondary wood formation in [30] or to tension wood in poplar [31].

The enlarged stem cell wall proteome can be compared to that of rosettes which has been obtained using a similar protocol [12] (see Supporting Information Table S2 and *WallProtDB*, http://www.polebio.lrsv.ups-tlse.fr/WallProtDB/). Two hundreds and 41 CWPs were common to both proteomes whereas 86 were specific to the stem cell wall proteome. The distribution into the nine functional classes was similar in both organs. The organ-specificity was mainly observed at the level of members of multigene families, as in many plant cell wall proteomes [18]. As an example, five Prxs (AtPrx03, 17, 22, 28, 64) were only identified in stems whereas four Prxs (AtPrx32, 37, 39, 51) were only found in rosettes. In the same way, two expansins (AtEXPA23, AtEXPB3) were

only identified in stems, while five (AtEXPA1, 3, 4, 6, 11) were only found in rosettes. Besides, a few protein families were only found in stems, e.g. laccases (AtLAC2, 4, 11).

As mentioned above, several stem cell wall proteomes have been already studied in *A. thaliana* [10], *M. sativa* [21, 22], *L. usitatissimum* [23], *B. distachyon* [19] and *S. officinarum* [20]. These proteomes have been obtained using similar strategies. Briefly, cell walls have been purified prior to extraction with salt solutions and proteins have been identified by LC-MS/MS after tryptic digestion. The proteomics data were retrieved from *WallProtDB* in which all the proteins have been annotated and classified in the same way using the *ProtAnnDB* pipeline [16, 17] and they have been compared (Fig. 2C). An overall picture of these proteomes can be obtained by comparing the distribution of the CWPs identified into functional classes. Differences in these distributions can be observed. However, it was not possible to distinguish dicots (*A. thaliana*, *L. usitatissimum* and *M. sativa*) and monocots (*B. distachyon* and *S. officinarum*) as two different groups of plants, although they have different types of cell walls [4]. The stem cell wall proteome of *A. thaliana* has three main characteristics when compared to the four other proteomes (Fig. 2C and Supporting Information Table S3): (i) a higher proportion of proteins acting on polysaccharides with the presence of more GHs and CEs; (ii) a lower proportion of oxido-reductases due a low number of Prxs; (iii) a higher proportion of proteases which are mainly Ser proteases of the subtilisin type and Ser carboxypeptidases. On the contrary, (i) the stem cell wall proteome of *L. usitatissimum* has a low proportion of CEs and proteases with the exception of Asp proteases, but a high proportion of expansins, Prxs, multicopper oxidases, protease inhibitors, fasciclin AGPs, germins and purple acid phosphatases [23] whereas (ii) that of *S. officinarum* has a particularly high proportion of Prxs, Asp proteases, LTPs and thaumatins [20]. These differences could be due to different stages of development and/or differences in cell wall composition, but also to the higher copy number of some protein families such as Prxs (e.g. 73 and 141 in *A. thaliana* and *L. usitatissimum*, respectively, peroxibase.toulouse.inra.fr). Only the *B. distachyon* cell wall proteomics study has considered stems at two developmental stages, basal and apical internodes [20]. It has been shown that the proportion of proteins acting on carbohydrates was slightly higher in apical internodes than in basal internodes (23.5% vs. 19.4%), whereas it was the contrary for oxido-reductases (18.0% vs. 21.2%). The stems of *L. usitatissimum* are very different from those of *A. thaliana* and *M. sativa* since they contain fibers which have thick secondary walls mainly composed of cellulose microfibrils and non-cellulosic polysaccharides such as hemicelluloses and pectins, but very low amount of lignin [32]. Only the full understanding of the roles of all CWPs will allow correlating CWP profiling to cell wall composition and structure.

With this new cell wall proteomics study, the total number of *A. thaliana* CWPs identified is of 805, that is between two fifths and one half of the expected total number of CWPs [33]. The cell wall proteomics atlas of *A. thaliana* thus remains the best described, including large proteomes for all vegetative organs (roots, rosettes and stems) as well as etiolated hypocotyls and cell suspension cultures [12, 18, 34]. This description should now be completed by the analysis of different stages of development of each organ and of post-translational modifications. Recent improvements in cell wall proteomics strategies such as reduction of the required amount of starting material [35] and development of lectin-affinity chromatography techniques to characterize glycosylations [36, 37] have paved the way toward the enlargement of our present vision of the plant cell wall proteomes.

All the MS/MS data have been made publicly available in PROTICdb (http://proteus.moulon.inra.fr/w2dpage/proticdb/angular/) [15] and those regarding CWPs in WallProtDB (www.polebio.lrsv.ups-tlse.fr/WallProtDB) [16].

The authors are thankful to Université Paul Sabatier (Toulouse, France) and CNRS for supporting their research. HD is supported by a grant of the Midi-Pyrénées region and of the Federal University of Toulouse. LC-MS/MS analyses were performed at the Plateforme d'Analyse Protéomique de Paris Sud-Ouest (PAPPSO). The authors also wish to thank Dr Cécile Albenne for her advices for protein extraction.

The authors have declared no conflict of interest.

References

[1] Le Gall, H., Philippe, F., Domon, J., Gillet, F. et al., Cell wall metabolism in response to abiotic stress. *Plants* 2015, *4*, 112–166.

[2] Cosgrove, D., Plant cell wall extensibility: connecting plant cell growth with cell wall structure, mechanics, and the action of wall-modifying enzymes. *J. Exp. Bot.* 2015, *67*, 463–476.

[3] Wolf, S., Hematy, K., Höfte, H., Growth control and cell wall signaling in plants. *Ann. Rev. Plant Biol.* 2012, *63*, 381–407.

[4] Carpita, N. C., Gibeaut, D. M., Structural models of primary cell walls in flowering plants, consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* 1993, *3*, 1–30.

[5] Jamet, E., Albenne, C., Boudart, G., Irshad, M. et al., Recent advances in plant cell wall proteomics. *Proteomics* 2008, *8*, 893–908.

[6] Cassab, G. I., Varner, J. E., Cell wall proteins. *Annu. Rev. Plant. Physiol. Plant Mol. Biol.* 1988, *39*, 321–353.

[7] Van der Hoorn, R., Plant proteases: from phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* 2008, *59*, 191–223.

[8] Franková, L., Fry, J., Biochemistry and physiological roles of enzymes that 'cut and paste' plant cell-wall polysaccharides. *J. Exp. Bot.* 2013, *64*, 3519–3550.

[9] Strabala, T., Macmillan, C., The Arabidopsis wood model-the case for the inflorescence stem. *Plant Sci.* 2013, *210*, 193–205.

[10] Minic, Z., Jamet, E., Negroni, L., der Garabedian, P. A. et al., A sub-proteome of *Arabidopsis thaliana* trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. *J. Exp. Bot.* 2007, *58*, 2503–2512.

[11] Feiz, L., Irshad, M., Pont-Lezica, R. F., Canut, H., Jamet, E., Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from Arabidopsis hypocotyls. *Plant Methods* 2006, *2*, 10.

[12] Hervé, V., Duruflé, H., San Clemente, H., Albenne, C. et al., An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes. *Proteomics* 2016, *16*, 3183–3187.

[13] Blein-Nicolas, M., Albertin, W., da Silva, T., Valot, B. et al., A systems approach to elucidate heterosis of protein abundances in yeast. *Mol. Cell. Proteomics* 2015, *14*, 2056–2071.

[14] Craig, R., Beavis, R., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, *20*, 1466–1467.

[15] Langella, O., Valot, B., Jacob, D., Balliau, T. et al., Management and dissemination of MS proteomic data with PROTICdb: example of a quantitative comparison between methods of protein extraction. *Proteomics* 2013, *13*, 1457–1466.

[16] San Clemente, H., Jamet, E., *WallProtDB*, a database resource for plant cell wall proteomics. *Plant Methods* 2015, *11*, 2.

[17] San Clemente, H., Pont-Lezica, R., Jamet, E.,Bioinformatics as a tool for assessing the quality of sub-cellular proteomic strategies and inferring functions of proteins: plant cell wall proteomics as a test case. *Bioinform Biol. Insights* 2009, *3*, 15–28.

[18] Albenne, C., Canut, H., Jamet, E., Plant cell wall proteomics: the leadership of *Arabidopsis thaliana*. *Front Plant Sci.* 2013, *4*, 111.

[19] Douché, T., San Clemente, H., Burlat, V., Roujol, D. et al., *Brachypodium distachyon* as a model plant toward improved biofuel crops: search for secreted proteins involved in biogenesis and disassembly of cell wall polymers. *Proteomics* 2013, *13*, 2438–2454.

[20] Calderan-Rodrigues, M., Jamet, E., Douché, T., Rodrigues Bonassi, M. et al., Cell wall proteome of sugarcane stems: comparison of a destructive and a non-destructive extraction method showed differences in glycoside hydrolases and peroxidases. *BMC Plant Biol.* 2016, *16*, 14.

[21] Printz, B., Dos Santos Morais, R., Wienkoop, S., Sergeant, K. et al., An improved protocol to study the plant cell wall proteome. *Front Plant Sci.* 2015, *6*, 237.

[22] Verdonk, J., Hatfield, R., Sullivan, M., Proteomic analysis of cell walls of two developmental stages of alfalfa stems. *Front Plant Sci.* 2012, *3*, 279.

[23] Day, A., Fénart, S., Neutelings, G., Hawkins, S. et al., Identification of cell wall proteins in the flax (*Linum usitatissimum*) stem. *Proteomics* 2013, *13*, 812-825.

[24] van de Meene, A., Doblin, M., Bacic, A., The plant secretory pathway seen through the lens of the cell wall. *Protoplasma* 2017, *254*, 75–94.

[25] Francoz, E., Ranocha, P., Nguyen-Kim, H., Jamet, E. et al., Roles of cell wall peroxidases in plant development. *Phytochemistry* 2015, *112*, 15–21.

[26] Pourcel, L., Routaboul, J. M., Kerhoas, L., Caboche, M. et al., TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. *Plant Cell* 2005, *17*, 2966–2980.

[27] Wang, Y., Bouchabke-Coussa, O., Lebris, P., Antelme, S. et al., LACCASE5 is required for lignification of the *Brachypodium distachyon* culm. *Plant Physiol.* 2015, *168*, 192–204.

[28] Jacobs, J., Roe, J. L., *SKS6*, a multicopper oxidase-like gene, participates in cotyledon vascular patterning during *Arabidopsis thaliana* development. *Planta* 2005, *222*, 652–666.

[29] Chen, Y., Ye, D., Held, M. A., Cannon, M. C. et al., Identification of the abundant hydroxyproline-rich glycoproteins in the root walls of wild-type Arabidopsis, an *ext3* mutant line, and its phenotypic revertant. *Plants* 2015, *4*, 85–111.

[30] MacMillan, C., Taylor, L., Bi, Y., Southerton, S. et al., The fasciclin-like arabinogalactan protein family of *Eucalyptus grandis* contains members that impact wood biology and biomechanics. *New Phytol.* 2015, *206*, 1314–1327.

[31] Lafarguette, F., Leple, J. C., Dejardin, A., Laurans, F. et al., Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytol.* 2004, *164*, 107–121.

[32] Morvan, C., Andème-Onzighi, C., Girault, R., Himmelsbach, D. et al., Biolding flax fibres: more than one brick in the walls. *Plant Physiol. Biochem.* 2003, *41*, 935–944.

[33] Jamet, E., Canut, H., Boudart, G., Pont-Lezica, R., Cell wall proteins: a new insight through proteomics. *Trends Plant Sci.* 2006, *11*, 33–39.

[34] Nguyen-Kim, H., San Clemente, H., Balliau, T., Zivy, M. et al., *Arabidopsis thaliana* root cell wall proteomics: increasing the proteome coverage using a combinatorial peptide ligand library and description of unexpected Hyp in peroxidase amino acid sequences. *Proteomics* 2016, *16*, 491–503.

[35] Francin-Allami, M., Merah, K., Albenne, C., Rogniaux, H. et al., Cell wall proteomics of *Brachypodium distachyon* grains: a focus on cell wall remodeling proteins. *Proteomics* 2015, 2296–2306.

[36] Ruiz-May, E., Rose, J., Progress toward the tomato fruit cell wall proteome. *Front Plant Sci.* 2013, *4*, 159.

[37] Zhang, Y., Giboulot, A., Zivy, M., Valot, B. et al., Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry* 2011, *72*, 1109–1123.

Notre étude a permis d'étendre considérablement la connaissance des protéines pariétales présentes dans les tiges d'*A. thaliana*. En effet, l'identification de 241 nouvelles protéines offre une vision plus complète de la paroi de cet organe chez *A. thaliana*. La caractérisation de ces protéines permet de mieux appréhender les analyses quantitatives dont a besoin l'analyse intégrative.

Les données générées par ces analyses en MS, de protéomes de rosettes et de tiges d'*A. thaliana* ont permis d'évaluer des règles déjà connues de modifications post-traductionnelles. Ces modifications ont un impact important sur les protéines en modifient leur conformation, leur activité biologique et / ou leur capacité à interagir avec des composants de la paroi. L'une de ces principales modifications est la glycosylation. À partir de l'étude des modèles d'hydroxylation des prolines de cinq protéines pariétales obtenues, des règles sont proposées et discutées et que des nouveaux modèles sont décrits.

Ce travail, qui ouvre la voie à une meilleure connaissance des protéines pariétales et de leurs modifications post-traductionnelles est détaillé dans l'article « *Hydroxylation of Pro residues (Hydroxyproline) in cell wall proteins: is it possible to define a rule? »* soumis pour publication dans *Frontiers in Plant Science* (Annexe)

Cette première partie de notre travail a permis de produire en grandes quantités des données expérimentales diverses pouvant être utilisées par la suite dans des analyses statistiques intégratives. Ces méthodes ont pu être appliquées lors des analyses du projet *WallOmics* portant sur des populations naturelles d'*A. thaliana* cultivés à 15 et 22°C.

Le verrou expérimental limitant la production de données étant levé, le verrou analytique, précédemment évoqué, peut être abordé.

II.7 Un cadre pour l'analyse des données omiques: des statistiques univariées à l'analyse intégrative multi-blocs

L'analyse d'un simple jeu de données omiques ne permet pas de comprendre un système biologique complexe dans son ensemble ; pour cela, il est nécessaire de combiner et d'intégrer plusieurs jeux de données à des échelles différentes. Les analyses multi-échelles permettent au biologiste d'appréhender l'impact que génère un stress biotique ou abiotique sur le développement d'une plante. Par comparaison avec un état ou un écotype témoin, il est alors possible d'identifier et de mettre en lien de nouveaux gènes/protéines candidats possiblement impliqués dans la réponse à ce stress. Ce type d'analyse est bien adapté pour l'étude de grands jeux de données qui ne peuvent pas être analysés manuellement en détail. Il devient même fondamental lors d'études comportant plusieurs jeux de données omiques, nommément « blocs ».

Dans notre étude, quatre populations naturelles d'*A. thaliana* provenant des Pyrénées (Roch, Grip, Hern et Hosp) sont analysées (détaillées dans la partie III) en présence de l'écotype de référence Col. Après une mise en culture de *triplica* biologiques à deux températures (22 et 15°C), représentant chacun une vingtaine de plantes, les rosettes et les tiges ont alors été étudiées *via* une analyse multi-blocs. De par la grande quantité de données obtenues, il nous a paru pertinent de considérer l'intégration de données omiques hétérogènes pour lever le verrou analytique de notre projet.

Des discussions entre biologistes et statisticiens, initiée à l'occasion de ce projet, ont soulevés de véritables interrogations comme la mise en place d'un protocole expérimental adapté, l'analyse des données et le lien existant entre la nature des analyses mises en œuvre, et les conclusions qu'il est possible d'en tirer.

L'ensemble de ces réflexions, ainsi que la méthodologie statistique, sont détaillées dans la publication présentée ci-après, laquelle sera soumise en *back to back* avec l'article *WallOmics* (Partie III).

A framework for omics data analysis: from univariate statistics to multi-block integrative analysis

Harold Duruflé[1], Merwann Selmani[1], Philippe Ranocha[1], Christophe Dunand[1*], Sébastien Déjean[2*]

[1] Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, BP 42617, 31326 Castanet-Tolosan, France.

[2] Institut de Mathématique de Toulouse, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

[*] Corresponding authors:

Sébastien Déjean (sebastien.dejean@math.univ-toulouse.fr; +33 (0)5 61 55 69 16)

Christophe Dunand (dunand@lrsv.ups-tlse.fr; +33 (0)5 34 32 38 58)

ABSTRACT

The high-throughput measurements of new techniques available used in biological studies require specific and adapted statistical treatments. We propose a framework to manage and analyse multi-omics heterogeneous data to carry an integrative analyse out. Natural populations of the model plant *Arabidopsis thaliana* are employed in this casework. In this project, four omics data sets (phenomics, metabolomics, cell wall proteomics and transcriptomics) have been collected, analysed and integrated in order to study the cell wall plasticity of plants exposed to different sub-optimal temperature growth conditions. Because each biological questions are associated to one method, the methodology presented starts from account basic univariate statistics to a multi-block integration analysis. The question about the utility to conduct an integrative analysis brought the question of the data processing, the missing data, the numerical and graphical outputs up. We do not comment any biological interpretation of the results discussed in the article (Duruflé, *in preparation*). We highlight instead each method of the statistical analysis limits, and show how a biologist has to interpret and take over the conclusions of a statistical study. This article provides to biologists a better understanding of the statistical integration and how to include it into a global reflexion.

Keywords: Statistical framework, Systems biology, Arabidopsis, Abiotic stress, Cell wall

1. INTRODUCTION

1.1. Context:

Biological processes require ever more complex measurements. Today, biologists have plethora of new techniques available to address their questions. The high-throughput measurements have revolutionized the way to predict and evaluate the behaviour of organisms towards their environmental changing. Nowadays, one biological sample can deliver many types of “big” data, such as genome sequences (genomics), genes and proteins expressions (transcriptomics and proteomics), metabolites profiles (metabolomics) and phenotypes observations (phenomics). The revolution of high throughput has also greatly reduced the cost of producing the omics data, opening the door to the development of tools for data treatment and analysis.

The heterogeneous data used at different cellular levels are associated to a wide variety of techniques sometimes specific to species. These data require a specific experimental design and a certain methodology for their monitoring to be valued as best as possible. An experimental design not adapted to integrative analysis could complicate the final interpretation of the collected data. Moreover, a suitable methodology of analysis can optimize and have point of better visibility on the data. To use multi-omics data make it possible to a better understanding of a biological system.

Furthermore, quantification technologies improve accuracy and create great potential for elucidating new questions in biology. But this revolution must be carefully used because the goal of a quantification analysis is not to validate another. For example, it is known that is somewhere difficult to correlate transcriptome and proteome profiles. In a common experiments, collecting different types of data helps to understand the effects of one or more various experimental conditions. A cohort of hypotheses could be done with multi-omics analysis. We can identify biological candidates as biomarkers (e.g. genes, proteins) under

complex environmental conditions, or find new complex biochemical regulation at different biological level.

This article focuses on the framework to manage and analyse multi-omics heterogeneous data in order to studying cell wall plasticity of the model plant *Arabidopsis thaliana* in two different temperature of growth. It will consider from basic univariate statistics to multi-block integration analysis. The graphical visualizations are simple to accomplish with all the existing tutorials, but interpret them is not easy. That is why this article will provide to biologists a better understanding of the statistical integration and how to include it in a global reflexion.

1.2. Why considering integrative analysis?

Henceforth, it is generally admitted that studying a single omics data is not enough to understanding the effects of a treatment in a biological system. To obtain a holistic view, it could be interesting and necessary to combine multiple omics analysis.

To highlight the interest of integrative approaches, let consider a toy example with two variables (Vx and Vy) measured on 12 individuals (6 from one group called WT, and 6 from another one, called Mut). The values are presented in the table I.

Statistical tests (Wilcoxon rank sum test and Student t test) have not revealed any significate differences between the two groups for none of the two variables Vx and Vy analysed separately (p-values greater than 0.3 in Table 1). But, a simple scatterplot (Figure 1) highlights the interest of combining the two variables. Indeed, it clearly appears that the two groups are linearly separated if we consider both variables Vx and Vy.

Table I: *Toy data set containing 12 individuals and 3 variables (1 qualitative Group, and 2 quantitatives Vx and Vy). P-values indicated on the last two rows are related to two samples tests (Wilcoxon rank sum test and Student t test) comparing separately for each variable the observations between WT to Mut.*

Individual	Group	Variable Vx	Variable Vy
1	WT	2	2
2	WT	4.5	3.5
3	WT	6	4.3
4	WT	8	5.1
5	WT	10	6
6	WT	11	6.5
7	Mut	3	2.5
8	Mut	5	3.25
9	Mut	5.5	3.3
10	Mut	7	4.2
11	Mut	8.5	4.8
12	Mut	9	5
Test difference WT vs Mut	Wilcoxon rank sum test (p-value)	0.82	0.31
	Student t test (p-value)	0.74	0.39

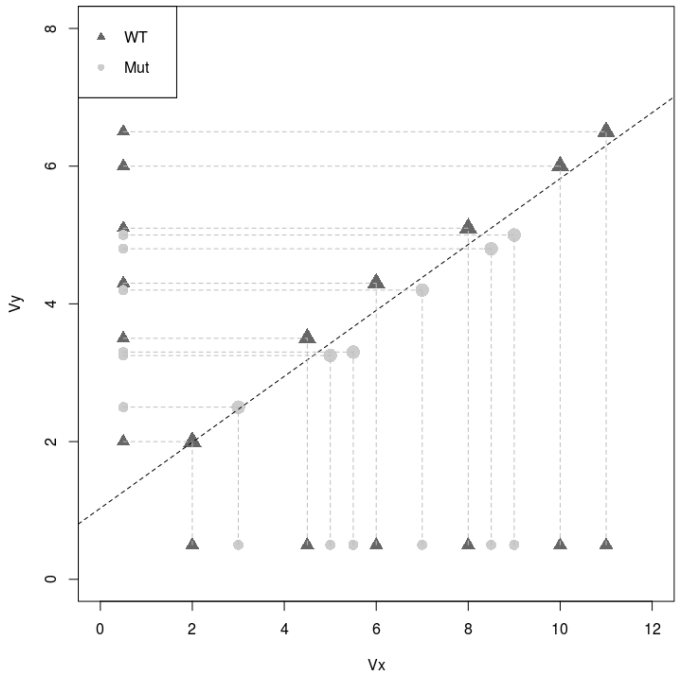


Figure 1: *Scatterplot representing Vy values (vertical axis) according to Vx values (horizontal axis). WT and Mut observations are represented as dark grey triangle and light grey circle respectively.*

In the same vein, we claim that an integrative analysis of several data sets acquired on the same individuals will reveal information that single data set analysis would keep hidden. Furthermore, this toy example also highlight the interest of relevant graphical representation: what the table 1 hides, the Figure 1 shows it clearly. To be strongly convinced about data visualisation, the recent work proposed in (Matejka and Fitzmaurice, 2017) is certainly a good way.

In this article, we describe a framework for the analysis of several data sets related to the same samples. We illustrate the gaps bridged by each method from the computation of univariate statistics to a thorough implementation of multi-block exploratory analysis. The first section presents the background of our study and the data sets we have dealt with. More details can be found in (Duruflé, *in preparation*). Then, we presents the several statistical methods used to address specific biological questions; numerical and graphical outputs of these methods are also detailed. Afterwards, we presents the results, give clues to interpret them and provide directions for biological researches.

2. BACKGROUND

2.1. Biological context

In the global warming context, seasons are altered with a modification of the temperature. The elevation of the temperature is the most studied because it is already observed (Savo et al., 2016). Moreover, the occurrence of freezing stress can also appear without any previous chilling period and it could become critical to maintain agricultural productivity in the future (Gray and Brady, 2016). The model plant *Arabidopsis thaliana* (L.) Heynh of the *Brassicaceae* family has a worldwide geographical distribution and by the way, is adapted to multiple and contrasted environmental conditions. The huge accumulation of molecular data

concerning this plant is very helpful for studying complex responses at multiple levels. All the scientists advance that this model plant could be applied to other plant species which have an economic interest for translational pipeline (Sibout, 2017).

2.2. Experimental design

First a compromise is necessary to determine the ideal number of biological replicates. It is hard to find an agreement between the reality of the biological experimentation (e.g. limitation in material, space, time, work force and cost) and the necessity to get robust information for the statistical analyses. The method used about the randomization also needs to be considered. The experimental protocol must minimize potential external impacts within and between the replicates and avoid confounded effects.

To strengthen the results, biological replicates can be an average of technical replicates, if the type of analysis allows it. For the biologist, it is important to know the number of experimental repetitions to appreciate the variability between the different conditions. But, for a statistician, the information resides into the intrinsic variability of the different samples or repetitions. For all these reasons, one sample considered as out of norms by the biologist could be interesting in a multi-omics analysis.

Our experimental design is built with two crossed factors: i) ecotypes with 5 levels (4 Pyrenees Mountain ecotypes Roch, Grip, Hern, Hosp living at different altitudes and Col, 1 reference ecotype from Poland living at low altitude), ii) temperature with 2 levels (22°C and 15°C). For each ecotypes, rosettes and floral stems were collected and analysed. At 22°C, rosettes were collected at four weeks, corresponding to the emergence of the floral stem. Floral stems were collected at 6, 7 and 8 weeks respectively for Col, Roch and Grip and, Hern and Hosp at 22°C. At 15°C, rosettes and stems have been collected two weeks later. More details

about the plant culture conditions can be found in (Duruflé, *in preparation*). At the end, three independents biological replicates are analysed. To compensate consequent mass of fresh material request for the cell wall purification, 20 plants per sample was used in each condition. To minimize the experimental effect, each plant has grown in a randomly place according to the experimental design, represented in Figure 2.

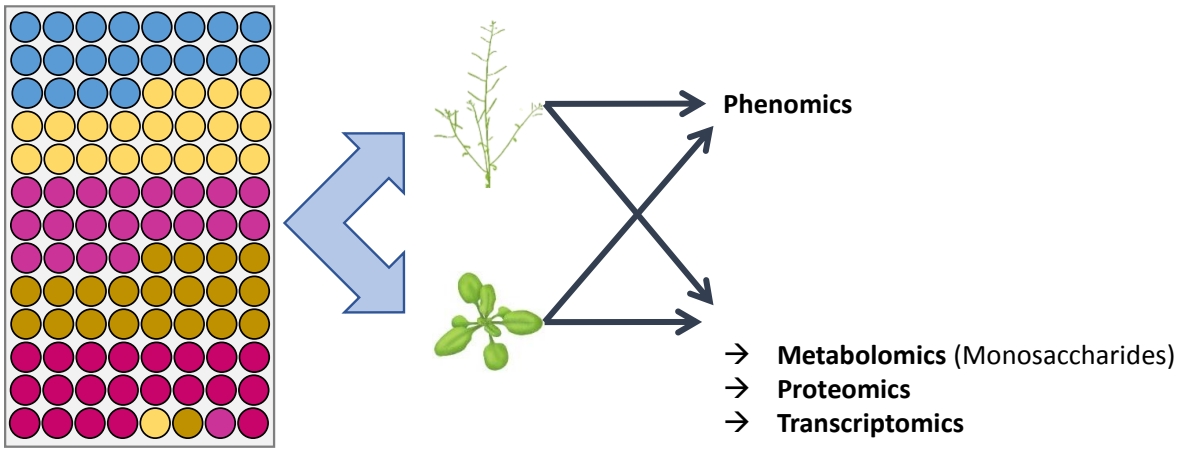


Figure 2: *Schematic overview of the strategy and experimental protocol used in this study. Each circle represents one plant and each colour stands for one ecotype of* A. thaliana*. This experimental design was made in triplica, taking care to change randomly the position of the ecotype at each repetition to avoid positioning effects.*

2.3. Omics data sets and curation

In this project, the four following omics data sets (called blocks thereafter) have been collected (Figure 3).

(i) Phenomics, which corresponds to a macro phenotyping analysis was performed on two organs. At the end of the experiment and before to be frozen, 9 parameters were measured: 5 on rosettes (weight, diameter, number of leaves, density, projected rosettes area) and 4 related to the floral stems (height, diameter, number of cauline leaves, length).

(ii) Metabolomics, which corresponds the identification and the quantification of seven cell wall monosaccharides (fucose, rhamnose, arabinose, galactose, glucose, xylose and the galacturonic acid) was performed by HPAEC-PAD. Theoretical cell wall polysaccharides composition has been reconstructed based on the monosaccharide analyses following the formula in (Duruflé, *in preparation*). Finally, 6 calculations of polysaccharides were performed.

(iii) Proteomics (LC-MS/MS analyses) were performed at the PAPPSO proteomics platform (pappso.inra.fr/) Proteins quantification was only performed for the cell wall proteins previously identified (with ProtAnnDB tool) using the MassChroQ software (Valot et al., 2011). At the end, 364 cell wall proteins were identified and quantified in the rosettes and 414 in the floral stems.

Missing data have been replaced with a background noise, in order to facilitate statistical analysis. It has been used differently to distinguish validated proteins (identification with at least two specific peptides in at least two of the three biological replicates of the ecotype/temperature combination), non-validated ones (identification of only one specific peptide and/or in only one biological replicate) and undetectable proteins (no peptide identified in this combination). If the protein was validated, a background noise, corresponding to the minimum, and the first statistical quartile of the biological replicate, was applied. If the protein was undetectable, a background noise of 6 (value lower than the minimum value found in the whole experiment) was applied. This treatment allows to combine the quantification process with a qualitative study and provides more confidence on the final result.

(iv) Transcriptomics (RNA-seq) have been performed on an Illumina HiSeq 3000 (get.genotoul.fr, Auzeville, France) according to the standard Illumina

protocols. Short pair-end sequencing reads generated were analysed using CLC Genomic Workbench 8.0 software (CLC bio, Aarhus, Denmark). A threshold was applied with just one simple rule: each gene had to be expressed more than one RPKM in at least one sample. Finally, 19,763 transcripts were analysed in the rosettes and 22,570 in the floral stems.

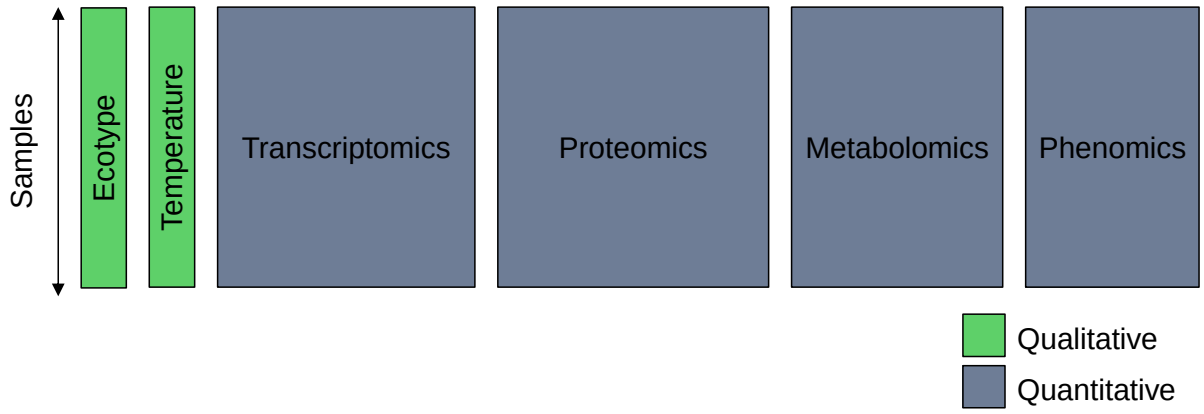


Figure 3: *Schematic representation of all the different blocks (or data set) used in this study. The samples are represented in rows and variables in columns. Qualitative and quantitative blocks are represented in green and grey respectively.*

3. TIDY DATA

Statistical data analysis requires efficient data pre-processing. As expressed in (Wickham, 2014), "*It is often said that 80% of data analysis is spent on the process of cleaning and preparing the data"*. So in an integrative analysis framework, each data set need to be structured in the same way. He also mention the following statements: *1/ Each variable forms a column. 2/ Each observation forms a row* (Wickham, 2014). So, in our context, each data set is presented with biological samples in rows and variables (transcripts, proteins...) in columns.

Handling missing data is always a big deal. As expressed by Gertrude Mary Cox (American statistician of the 20th century), "*the best thing to do with missing values is not to have any"*, however, and fortunately, many methods exists to deal with missing values. For

instance, methodologies implemented in the missMDA package (Husson and Josse, 2013) are dedicated to the handling of missing values in the context of multivariate data analysis. More recently, (Voillet et al., 2016) focused on missing rows in datasets in an integrative framework. Inside an integrative study, we can easily be in this latter case if, for instance, the number of biological samples are not the same when acquiring transcriptomic and proteomic analyses. The main idea to remember would be to deal with missing values with an ad-hoc method taking into account the specificity of the data.

4. METHODS

4.1. Software

As mentioned in the Comprehensive R Archive Network (CRAN, cran.r-project.org), *R "is a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc*."

R is not user-friendly but it gives access to very recent methodological developments and the R community is very active (R-bloggers, R-help, UseR conference...). Furthermore, efficient tools such as RStudio (www.rstudio.org) were developed in order to make the initiation to R easier. And many resources are available on CRAN to get started with R. Then, read, use and adapt existing scripts seems highly reasonable after few hours of practice. Regarding specifically the community of biologists, the repository Bioconductor (Gentleman et al., 2004) provides tools (including softwares, annotation data and experiment data) for the analysis of high-throughput genomic data.

The energy around R appears in the packages developed by and for the community. So, several packages exists to address statistical integrative study. We focus on the mixOmics (Lê

Cao et al., 2009) package as it has been partly developed in our team but other packages such as FactoMineR (Lê et al., 2008) can also be used for nearly similar purposes. Methodologies presented in (Bécue-Bertaut and Pagès, 2008) and (Sabatier et al., 2013) are also alternatives in this context. Regarding commercial softwares, we mention that, for instance, SIMCA-P (Umetrics) propose several methods to perform integrative analysis and toolboxes for Matlab are also available. As mentioned above, we will focus on the use of mixOmics and the comparison of softwares is beyond the scope of the present article.

4.2. One purpose one method

In this section, partly inspired from the tutorial of mixOmics, we want to highlight the link between a statistical method and the biological question it intends to address.

- *Purpose: explore one single quantitative variable (e.g. the expression of one gene or the values get for one phenotypic variable)*: univariate elementary statistics such as mean, median for main trends, and standard deviation or variance for dispersion can be completed with graphical representation such as boxplot.
- *Purpose: assess the influence of one single categorical variable (e.g. ecotype) on a quantitative variable (same examples as above)* : statistical significance test such as Student t test or Wilcoxon rank sum test for two groups and ANOVA or Kruskall-Wallis for more groups will address this question. In this context, a special attention must be payed to the structure of the data: independent samples (e.g. independent groups observed in various conditions) or paired samples (e.g. same samples observed twice or more).
- *Purpose: evaluate the relationships between two quantitative variables*: correlation coefficients (Pearson for linear relationships and Spearman for monotonous ones).

Graphical representations of correlation matrices may provide a global overview of pairwise indicators.

- *Purpose: explore one single data set (e.g. microarray data) and identify the trends or patterns in the data, experimental bias or, identify if the samples 'naturally' cluster according to the biological conditions*: unsupervised factorial analysis such as Principal Component Analysis (PCA) provides such information about one data set without any prior (Mardia et al., 1980). Scaling the data before performing PCA is frequently useful when dealing with omics data to make the PCA results meaningful.
- *Purpose: classifying samples into known classes (e.g. ecotype) based on one single data set (e.g. transcriptomic data)*: Supervised classification methods such as PLS-Discriminant Analysis (PLS-DA) assess how informative the data are to rightly classify samples, as well as predict the class of new samples.
- *Purpose: unravel the information contained in two data sets, where two types of variables are measured on the same samples (e.g. metabolomics and transcriptomics data)*: Using PLS related methods enable to know if common information can be extracted from the two data sets (or highlight the relations between the two data sets).
- *Purpose: the same as above but considering more than two data sets (e.g. metabolomics, transcriptomics and phenotypic data)*: multi-block PLS related methods was recently developed (Lê Cao et al., 2008) to address this issue.
- *Purpose: the same as above but in a supervised context (e.g. metabolomics, transcriptomics phenotypic data and ecotype categories)*: multi-block PLS-DA (referred as DIABLO for Data Integration Analysis for Biomarker discovery using Latent variable approaches for Omics studies) was recently developed (Singh et al., 2016, Günther et al., 2014) to address this issue.

For the last two methods (multi-block related), a design matrix must be defined to express the relationships between each block. A default version used a full design matrix: each block is related to all the others. But other definitions of the design matrix can be justified.

A schematic view of the data sets are presented in Figure 3. Each rectangle stand for a data set (or a block): samples (plants) are represented in rows and variables in columns. Each method used is then presented with the piece(s) of the whole data sets it deals with.

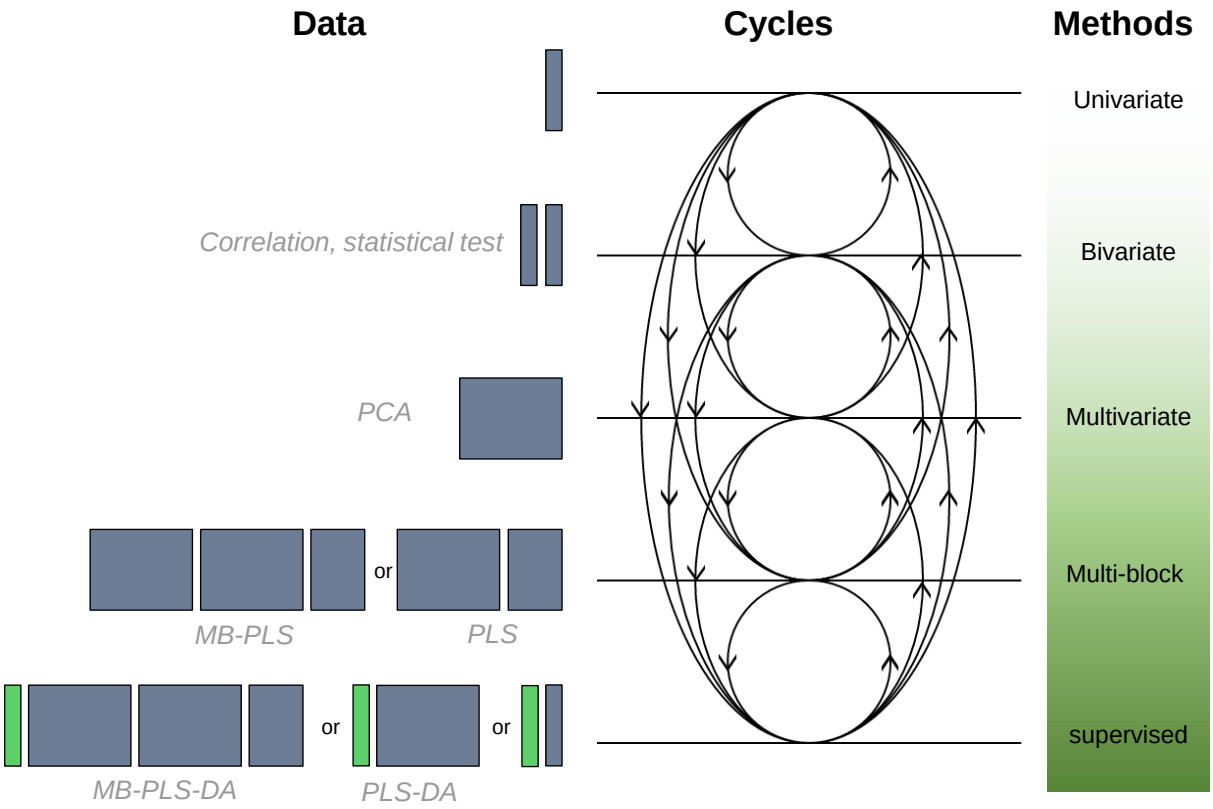


Figure 4: *Schematic overview of the methods implemented of the associated data sets represented by cycles of all the retro analysis advice within an integrative study. PCA: Principal Component Analysis; MB: Multi-Blocs; PLS: Partial Least Squares regression; DA: Discriminant Analyse. Qualitative and quantitative blocks are represented in green and grey respectively.*

The cycles presented in Figure 4 highlights the better way to perform an integrative statistical study. We prefer this view rather than a straightforward pipeline beginning with univariate analysis and ending with multi-block approaches. Each method contributes to the

global comprehension of the data and can challenge others methods. For instance, univariate statistics may highlight outliers or essential variables, but are they always unavoidable in a multivariate approach? On the other hand, multi-block approaches may focus on new samples and/or variables showing specific behaviour that should be studied through a univariate method. We claim that, facing integrative studies, relevant statistical analysis has to pass through these cycles, with progress and returns.

4.3. Sparse extensions

Every methods developed in mixOmics are proposed with a sparse extension (sparse PCA, sparse PLS...). Sparse methods are useful to remove some of the non-informative variables regarding the purpose of the method. Regarding PCA for instance, the sparse version select only the variables that highly contribute to the definition of each PC, removing the others. Sparsity is mathematically achieved *via* Least Absolute Shrinkage and Selection Operator penalizations LASSO (Tibshirani, 1996).

In practice, using sparse methods in the context of omics data is very useful as it reduced the number of potentially relevant variables displayed on the graphical outputs, and thus it makes easier the biological interpretation of the results and reduces the list of potential candidates for further investigations.

4.4. Numerical and graphical outputs

As previously mentioned, statistical analysis could be associated with graphical representations (Figure 5). A famous sentence assigned to Francis John Anscombe (english statistician of the 20th century) emphasized this point of view: "*...make both calculations and graphs. Both sorts of output should be studied; each will contribute to understanding."* Based

on this principle, a recent work by Matejka and Fitzmaurice (Matejka and Fitzmaurice, 2017) illustrates in a quite funny way how same numerical outputs can provide very different graphical representations.

The results of univariate and bivariate approaches are mainly reported as p-values for statistical testing. Boxplots and barplots, as produced by the ggplot2 package (Wickham, 2016), may complete and reinforce the interpretation of the results. Regarding barplots, one core question relies on the error bar that are frequently added: should they be based on standard deviation or standard error of the mean? A thorough explanation about the difference is provided in (Cumming et al., 2007). The authors also mention this statement that may seem obvious but that is sometimes forgotten: "*However, if n is very small (for example n = 3), rather than showing error bars and statistics, it is better to simply plot the individual data points."*

We also used graphical representations of correlation matrices such as ones produced by the corrplot package (Wei and Simko, 2016) for the R software. This is essential when dealing with (not so) many variables: with 50 variables, 1225 (50 x 49 / 2) pairwise correlation coefficients are computed and have to be analysed and interpreted.

Regarding multivariate analysis (from PCA to multi-block analyses), we used the graphical outputs provided by the mixOmics R package (Lê Cao et al., 2009). They are based on the representation of individuals and variables projected on specific sub-spaces. We recall the main trends in interpreting such plots. A thorough discussion about the complementarity of several graphical display is given in (González et al., 2012).

First, the individuals (biological samples) of the study are represented as points located in a specific sub-space defined by the first PLS-components. Interpretation is based on the relative proximities of the samples and on the equivalent representation for variables. The standard representation for the variables plots is frequently referred as correlation circle plot. It was primarily used for PCA to visualise relationships between variables but it has been

extended to deal with multi-block analysis. In such a plot, the nature of the correlation between two variables can be visualised through the cosine of the angle between two vectors. The representation of variables can also be done through a relevance network. Bipartite networks are inferred using a pairwise similarity matrix directly obtained from the outputs of the integrative approaches. Circos plot can be viewed as a generalization of relevance network where the nodes are located on a circle. Then, based on the same pairwise similarity matrix used for relevance network, a clustered image map can be displayed. This type of representation is based on a hierarchical clustering simultaneously operating on the rows and columns of a real-valued similarity matrix. This is graphically represented as a 2-dimensional coloured image, where each entry of the matrix is coloured on the basis of its value, and where the rows and columns are reordered according to the hierarchical clustering.

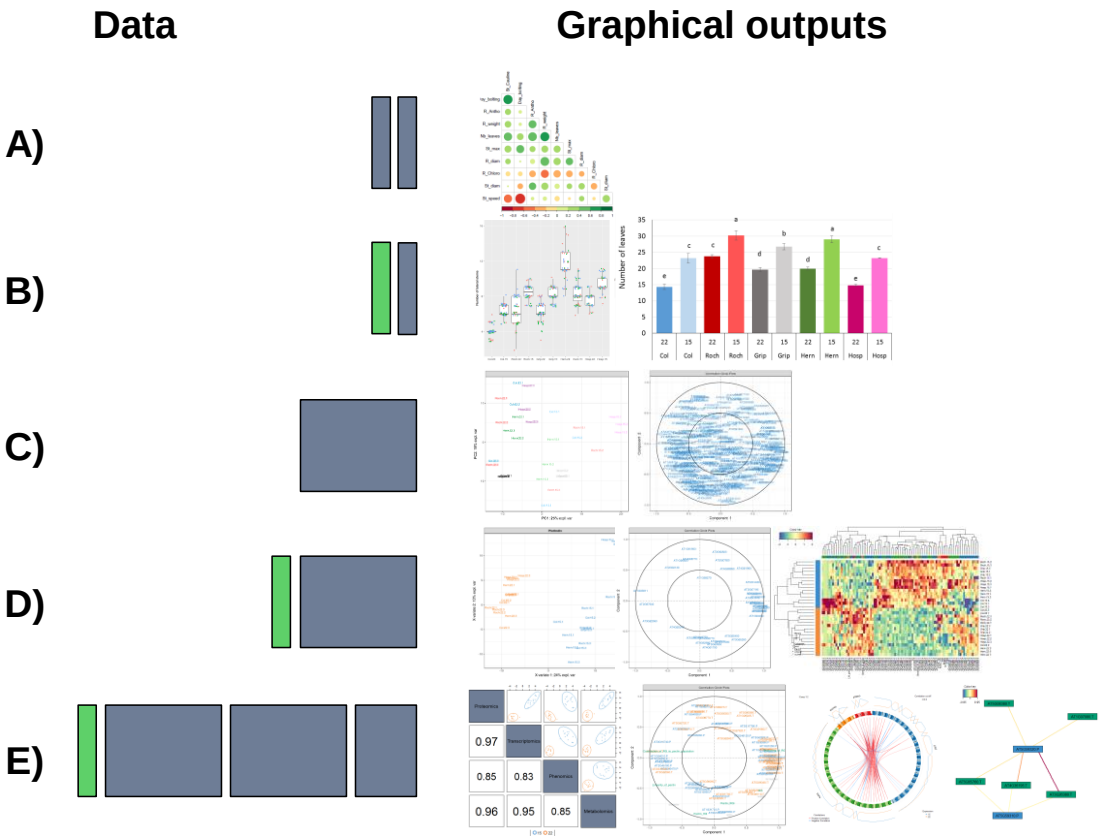


Figure 5: *Example of graphical outputs obtained with different variates. A) Bivariate analyse, B) Bivariate supervised analyse, C) Multivariate analyse, D) Multivariate supervised analyse, E) Multi-block analyse. Qualitative and quantitative blocks are represented in green and blue respectively.*

5. RESULTS

In this section, we do not provide neither a thorough biological interpretation of the results, the reader will refer to (Duruflé, *in preparation*) for this purpose, nor a comprehensive view of every statistical analyses performed. Instead, we highlight the limits of each method leading to the next step of the statistical analysis and show how a biologist has to interpret and to take over the conclusions of a statistical study.

5.1. Bivariate analysis

We illustrate bivariate analysis through some graphical representations of phenotypic data. Figure 6A displays parallel boxplots as well as individual observations of the number of leaves for the 5 ecotypes in the 2 conditions of temperature. Figure 6B only displays the averaged values over plants for each ecotype and temperature. These values are used in further analyses.

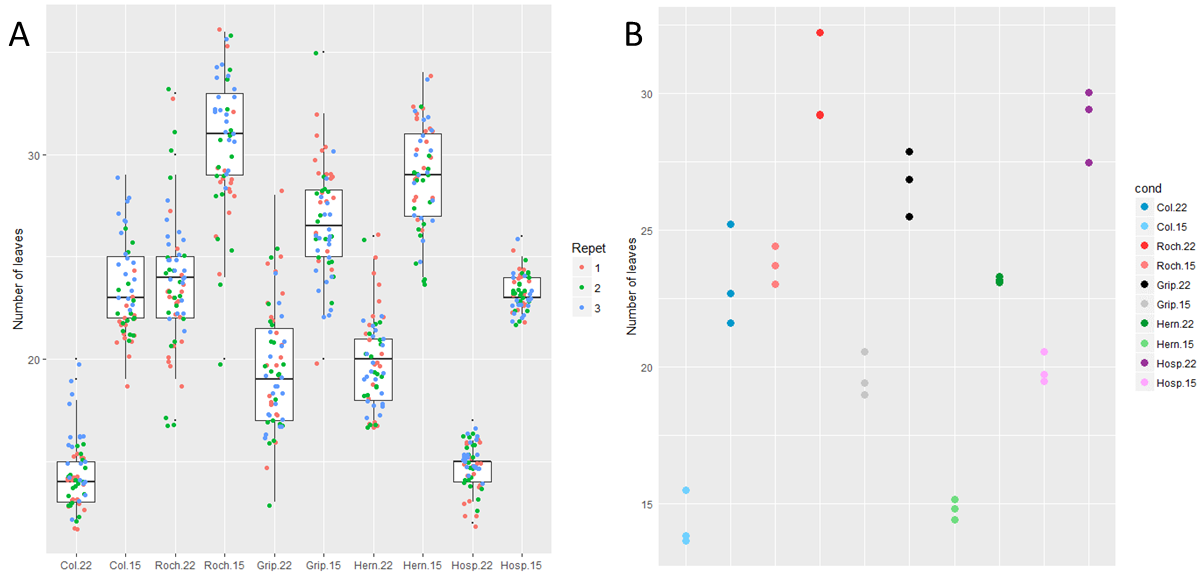


Figure 6: *Example of graphical outputs of bivariate analyse supervised with A) a box plot (each colour corresponds to the different values obtained for one triplicate) and B) an individual plot of the number of leaves for the 5 ecotypes in the 2 conditions of temperature (each colour corresponds to the average obtained for one triplicate).*

The main information extracted from these graphics are related to a quality control of the data. The scattering of points indicates a rather good reproducibility between all the samples and between each repetition. So, we could average the values from several plants of one biological repetition, to go on the analyses. The visual impression provided by Figure 6B regarding temperature and ecotype effects can be confirmed via statistical testing such as two-way ANOVA (results not shown).

This kind of analysis does not provide any information about the potential relationships between several variables. This will justify the next analysis step which deals globally with a whole data set.

5.2. The multivariate analysis

The multivariate approach is illustrated on the cell wall transcriptomic data set acquired on the rosettes. It is composed of 364 variables. The first way to question the whole data set can be through the computation of pairwise correlation coefficients. For instance, Figure 7 displays the correlation matrix between samples. It indicates that the expression of genes for each sample are positively correlated (only green colour and identically oriented ellipses) with the others.

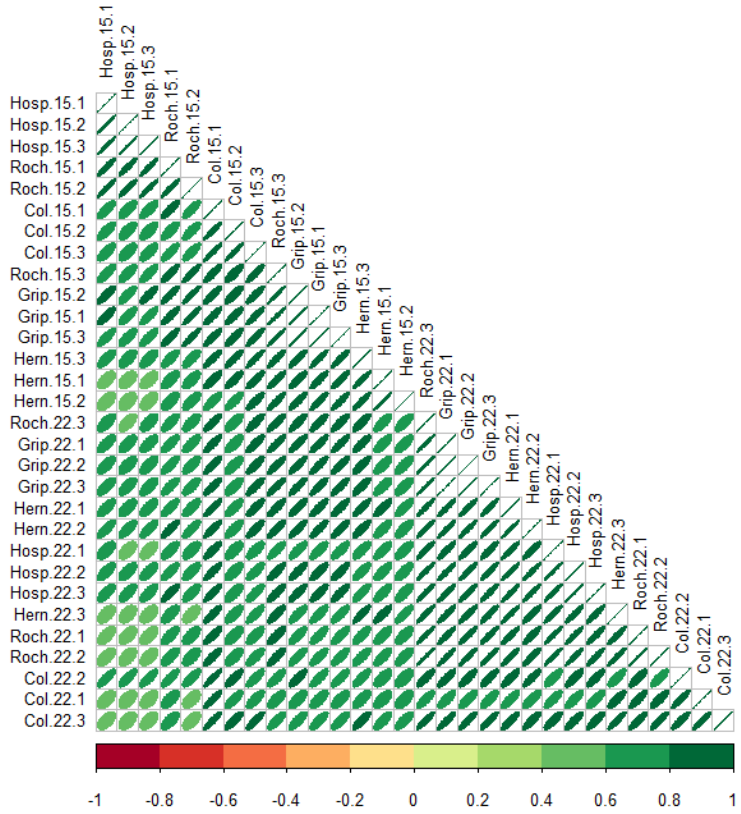


Figure 7: *Example of graphical representation of the multivariate analysis, a pairwise correlation coefficients of the cell wall transcriptomic in rosettes. Colour code and ellipses represent the correlation of genes for each sample.*

Then, a PCA can be performed as an extension of the quality control. For instance, Figure 8A highlights the distance between the three replicates corresponding to one condition. And, we can observe that the ecotype Grip is well gathered whereas Col is more scattered. This information must be moderate because of the rather low proportion of variance explained by the first two principal components displayed here. Having a look at the other components could be meaningful to consolidate and complete this information.

Furthermore, the interpretation of the PCA also brings a first result. Indeed, the samples are clearly separated along the first (horizontal) axis according to the temperature: samples at 22°C are all located on the left (negative coordinates on PC1) whereas samples at 15°C are on the right. This indicates that the effect of temperature is stronger than the ecotype's because

PC1 capture the most important source of variability in the data. The representation of variables (Figure 8B) is not of great interest at this step; it mainly highlights the need for selection methods to make easier the interpretation of the results in terms of gene expression. However, the interpretation of such a plot jointly with the individual plot, enable, for instance, to identify over-expressed genes in samples at 15°C: they are located on the right of the variables plot (in the same area as samples at 15°C in the individual plot).

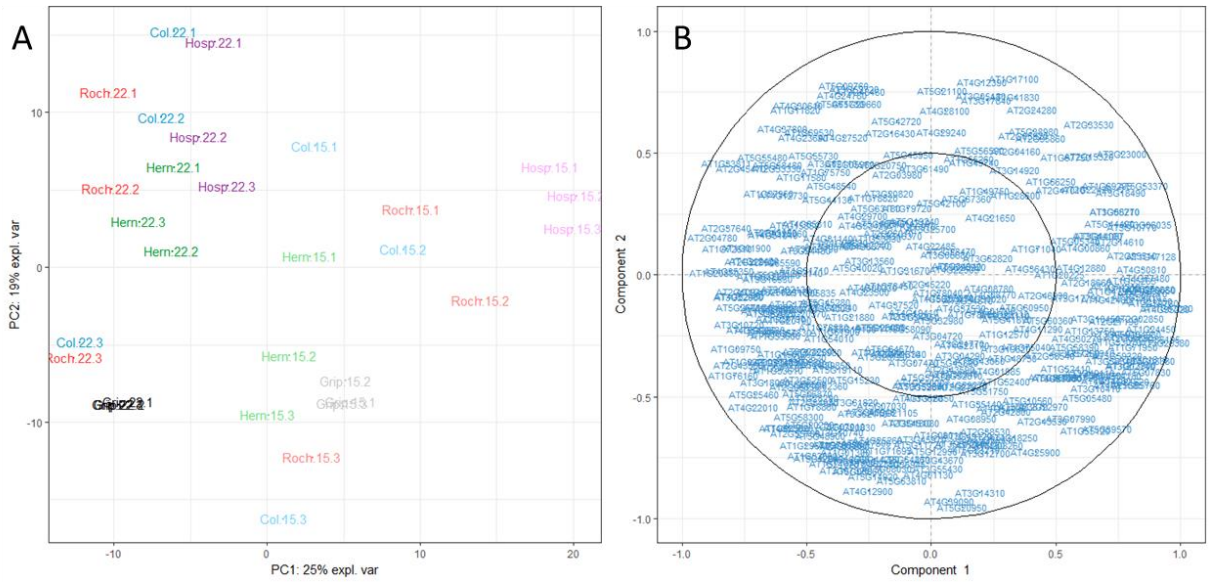


Figure 8: *Example of graphical representation of a PCA. A) In the individual plot are represented, according to their relation to two dimensions among those chosen in the analysis, all the samples are link to one colour specific to one condition (points tends to aggregate together when they are share similarities), and B) the variables plot that represent all the genes in a correlation circle. Strongly associated (or correlated) genes are projected in the same direction from the origin such as the greater the distance from the origin the stronger the association. Two circumferences of radius 1 and 0.5 are plotted to reveal the correlation structure of the variables.*

5.3. Supervised analysis and variable selection

To illustrate supervised analysis, we deal with the same data set as before (cell wall transcriptomic for the quantitative block) to discriminate the samples according to the temperature (qualitative block) performing a PLS-DA analysis. We could have made it with the ecotype, but interpretation would have been more complicated with 5 categories instead of 2

for temperature. And we have already seen that temperature effect is the strongest in the data set. Furthermore, to address the problem of interpretability of the results, we also consider the sparse version of PLS-DA (Lê Cao et al., 2011) to select the most discriminant genes for the temperature effect. The number of variables to select has to be determined by the user. It depends on the way the potential candidates will be validated. For instance, if validation has to be done through new focused biological experiments, the number of selected variables must not be too large.

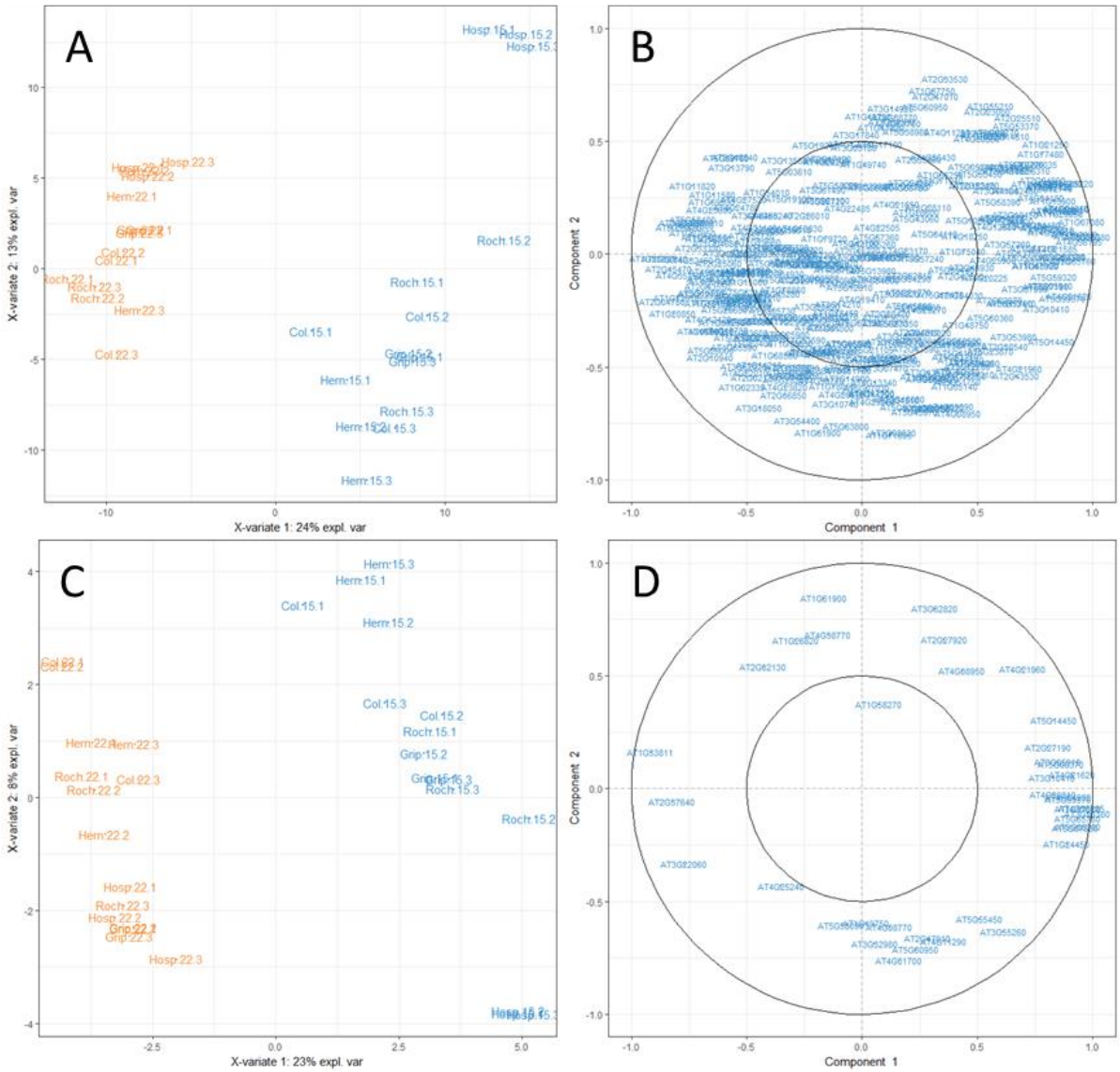


Figure 9: *Example of graphical representation of the supervised analysis. A) Individual plot of a PLS-DA associated to the B) variables plot and C) Individual plot of an S-PLS-DA associated to the D) variables plot.*

Figure 9 displays the results of PLS-DA (A, B) and S-PLS-DA (C, D). Individual plots (A, C) and variable plots (B, D) are interpreted in the same way as PCA's. Individual plots only use two colours corresponding to the two temperatures. Whether for PLS-DA or S-PLS-DA, the discrimination of the samples is clean-cut. This confirm the overriding effect of the temperature. In other words, the variability due to the five ecotypes does not impede to detect the temperature effect. The results of S-PLS-DA indicates that the discrimination can be observed with quite few genes. Indeed, the difference between PLS-DA and S-PLS-DA relies on the number of genes involved in the discrimination process. The list of the most relevant variables, displayed in 9D, are presented and studied in details in (Duruflé, *in preparation*).

To support this idea, we also performed a discriminant analysis considering the ecotype effect. The results (Figure 9) indicates that despite the strong effect of temperature, the method succeeded in discriminating the 5 ecotypes.

The two examples presented here highlights the specificity of a supervised analysis: the biologist can focus on a specific point of its experimental design.

5.4. Multi-block analyses

Multi-block analyses can address the main purpose of an integrative study by analysing all together every blocks acquired on the samples. As an illustration, we expose the results of a five-block supervised analysis focused on the rosettes, considering phenotypic, cell wall transcriptomic, proteomic and metabolomic as quantitative variables and temperature as the qualitative (or categorical) block.

Relationships between blocks must be defined by the user through a design matrix. This matrix, of size (number of blocks) x (number of blocks), contains values between 0 and 1. A value close to 1 (resp. 0) indicates a strong relationship (resp. weak or no relationship) between

the blocks to be modelled. To fix the values in the design matrix is crucial and very complicated because it requires to express biological relationships into numerical values (e.g. can we consider that the link between proteomic and transcriptomic data is stronger than the link between proteomic and metabolomic?). For the sake of simplicity, 0 and 1 values can be used in a binary point of view. In a supervised context, the values also enables to balance the optimisation between, on the one hand the relationships between quantitative blocks and, on the other hand, the discrimination process. In our example, we considered a design matrix composed of zeros to favour the discrimination task rather than the links between the quantitative blocks. A full design matrix (composed of ones) highlights more clearly relationships between blocks but leads to misclassified samples.

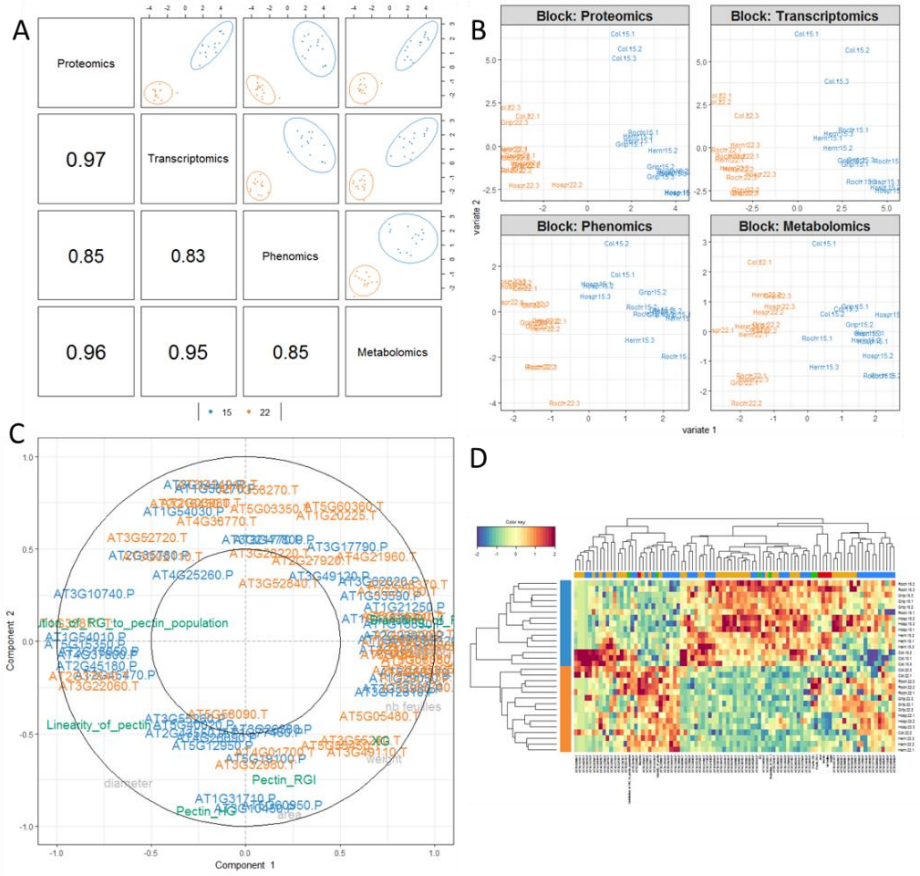


Figure 10: *Example of graphical representation a multi-block analyses. A) plotDIABLO show the correlation between components from each data set maximized as specified in the design matrix. B) Induvial plots projects each sample into the space spanned by the components of each block associated to the C) variables plot that highlights the contribution of each selected variable to each component, D) clustered image map of the variables to represent the multi-omics profiles for each sample.*

The interpretation of a multi-block supervised analysis requires several graphical outputs. Some of them are presented in Figure 10. First, Figure 10A checks whether the correlation between components from each data set has been maximized as specified in the design matrix (Tenenhaus et al., 2014). Globally, correlation values are close to 1 and mainly due to the separation of the two categories (because of our design matrix favouring discrimination). With a full designed matrix, we get higher values but with less separated groups. Regarding the individual plots (Figure 10B), it appears that the discrimination is better for transcriptomic and proteomic blocks. The samples plot has also to be interpreted regarding the variables plot (Figure 10C). To make the interpretation easier, we present here the results of the sparse version of multi-block analysis. So, we can identify variables from each block mainly involved in the discrimination according to the temperature. For instance, they are located on the left of Figure 10C for samples at 22°C (located on the left of Figure 10B). Another way to display the results is presented in Figure 10D. The clustered image map highlights the profiles of selected variables among the samples. It also includes the results of hierarchical clustering performed jointly on variables and samples. Regarding the samples, the two groups based on temperature are visualized through the dendrogram on the left. However, let's note that the cluster gathering the samples at 15°C can be split into two sub-clusters with col ecotype isolated. Regarding the variables, it mainly points out global trends of the behaviour of selected variables. The interpretation can then lead to retro analyses to validate potential candidates. This can be done through new statistical analyses as well as new biological experiments (Chawla et al., 2011).

5.5. Relevance networks

Another way to interpret the results of a multi-block approach consist in producing relevance networks between variables. On the Figure 11A, each selected variables is a node

located on a circle. Variables are sorted first according to their block then depending on their importance in discrimination. An edge links two nodes if their correlation is greater than a given threshold (0.9 in Figure 11A).

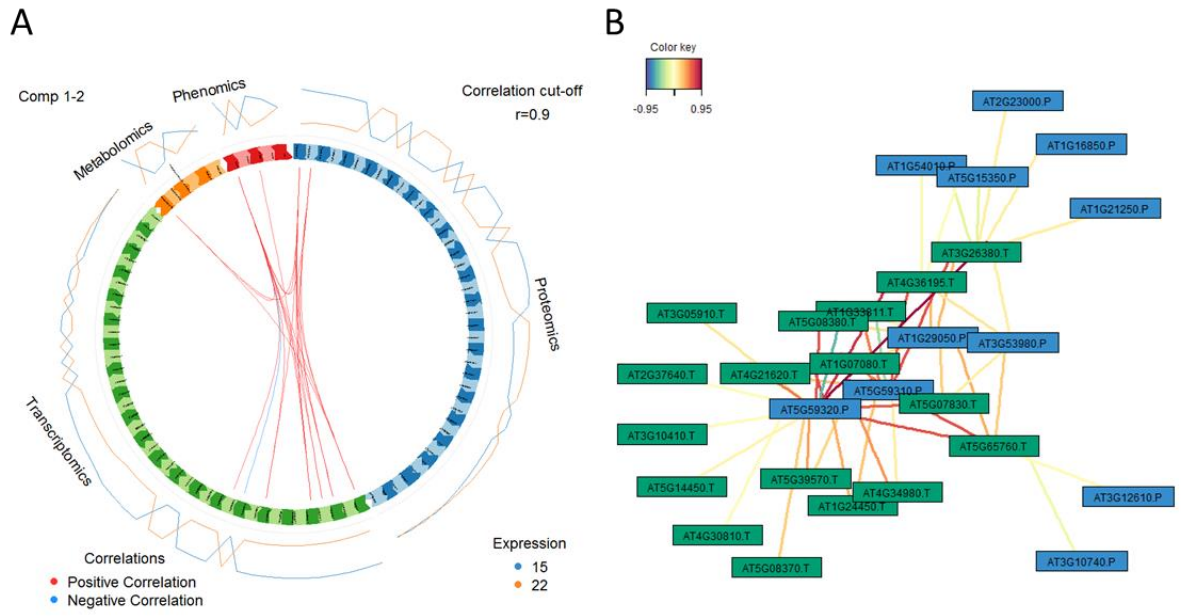


Figure 11: *Example of networks representation by A) a circos Plot represents the correlations between variables of each blocks and show expression levels of each variable according to each class within and between each blocks. B) a network that allows to display the correlation between the transcriptomics data (green) and then proteomics data (blue)*

The correlation are mainly positive and concerns few variables from each block. To complete the interpretation, we focus on another network generated with only two blocks (transcriptomic and proteomic). It accentuates the relationships between pairs of proteins and transcripts. At this stage, the biologist has to appropriate the results to validate them and express conclusions in biological terms (Duruflé, *in preparation*).

6. CONCLUSION

In an integrative biology context, the huge quantity of data produced requires adapted and specific statistical methods. Even if the multi-block approaches can be viewed as "The tool"

to address the purpose, other standard statistical methods more basic (e.g. univariate) must not be omitted. Good comprehension of the biological phenomenon goes through a sequence of various approaches to analyse the data. Finally, we consider that each method contributes to a better interpretation of the others like we intended to express it with the schematic view of the protocol as intertwined cycles. With the vast omics data sets, the statistical analysis of data sets can be a never ending story for biologists because each step of the framework provides information's.

Acknowledgment

The authors are thankful to the Paul Sabatier Toulouse 3 University and to the *Centre National de la Recherche Scientifique* (CNRS) for granting their work. This work was also supported by the French Laboratory of Excellence project "TULIP" (ANR-10-LABX-41; ANR-11-IDEX-0002-02). HD is supported by a Midi Pyrenees Region and the federal university of Toulouse.

References

BÉCUE-BERTAUT, M. & PAGÈS, J. 2008. Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. *Computational Statistics & Data Analysis,* 52**,** 3255-3268.

CHAWLA, K., BARAH, P., KUIPER, M. & BONES, A. M. 2011. Systems biology: a promising tool to study abiotic stress responses. *Omics and Plant Abiotic Stress Tolerance***,** 163-172.

CUMMING, G., FIDLER, F. & VAUX, D. L. 2007. Error bars in experimental biology. *J Cell Biol,* 177**,** 7-11.

DURUFLE, H., RANOCHA, P., DEJEAN, S., BURRUS, M., BALLIAU, T., ZIVY, M., ALBENNE, C., BURLAT, V., JAMET, E. & DUNAND., C. WallOmics: an integrative study of cell wall adaptation to sub-optimal growth conditions of natural population of Arabidopsis. *In preparation*

GENTLEMAN, R. C., CAREY, V. J., BATES, D. M., BOLSTAD, B., DETTLING, M., DUDOIT, S., ELLIS, B., GAUTIER, L., GE, Y., GENTRY, J., HORNIK, K., HOTHORN, T., HUBER, W., IACUS, S., IRIZARRY, R., LEISCH, F., LI, C., MAECHLER, M., ROSSINI, A. J., SAWITZKI, G., SMITH, C., SMYTH, G., TIERNEY, L., YANG, J. Y. & ZHANG, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol,* 5**,** R80.

GONZÁLEZ, I., CAO, K. A., DAVIS, M. J. & DÉJEAN, S. 2012. Visualising associations between paired 'omics' data sets. *BioData Min,* 5**,** 19.

GRAY, S. B. & BRADY, S. M. 2016. Plant developmental responses to climate change. *Dev Biol,* 419**,** 64-77.

GÜNTHER, O. P., SHIN, H., NG, R. T., MCMASTER, W. R., MCMANUS, B. M., KEOWN, P. A., TEBBUTT, S. J. & LÊ CAO, K. A. 2014. Novel multivariate methods for integration of genomics and proteomics data: applications in a kidney transplant rejection study. *Omics: a journal of integrative biology,* 18**,** 682-695.

HUSSON, F. & JOSSE, J. 2013. Handling missing values with/in multivariate data analysis (principal component methods). *Agrocampus Ouest-Laboratoire de mathématique appliquée, Rennes*.

LÊ CAO, K. A., BOITARD, S. & BESSE, P. 2011. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics,* 12**,** 253.

LÊ CAO, K. A., GONZÁLEZ, I. & DÉJEAN, S. 2009. integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics,* 25**,** 2855-6.

LÊ CAO, K. A., ROSSOUW, D., ROBERT-GRANIÉ, C. & BESSE, P. 2008. A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol,* 7**,** Article 35.

LÊ, S., JOSSE, J. & HUSSON, F. 2008. FactoMineR: an R package for multivariate analysis. *Journal of statistical software,* 25**,** 1-18.

MARDIA, K. V., KENT, J. T. & BIBBY, J. M. 1980. Multivariate analysis (probability and mathematical statistics). Academic Press London.

MATEJKA, J. & FITZMAURICE, G. Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017. ACM, 1290-1294.

SABATIER, R., VIVIEN, M. & REYNÈS, C. 2013. Une nouvelle proposition, l'analyse discriminante multitableaux: Statis-lda. *Journal de la Société Française de Statistique,* 154**,** 31-43.

SAVO, V., LEPOFSKY, D., BENNER, J. P., KOHFELD, K. E., BAILEY, J. & LERTZMAN, K. 2016. Observations of climate change among subsistence-oriented communities around the world. *Nature Climate Change,* 6**,** 462-473.

SIBOUT, R. 2017. Crop breeding: Turning a lawn into a field. *Nat Plants,* 3**,** 17060.

SINGH, A., GAUTIER, B., SHANNON, C. P., VACHER, M., ROHART, F., TEBUTT, S. J. & LE CAO, K. A. 2016. DIABLO-an integrative, multi-omics, multivariate method for multi-group classification. *bioRxiv***,** 067611.

TENENHAUS, A., PHILIPPE, C., GUILLEMOT, V., LE CAO, K. A., GRILL, J. & FROUIN, V. 2014. Variable selection for generalized canonical correlation analysis. *Biostatistics,* 15**,** 569-83.

TIBSHIRANI, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)***,** 267-288.

VALOT, B., LANGELLA, O., NANO, E. & ZIVY, M. 2011. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics,* 11**,** 3572-7.

VOILLET, V., BESSE, P., LIAUBET, L., SAN CRISTOBAL, M. & GONZÁLEZ, I. 2016. Handling missing rows in multi-omics data integration: multiple imputation in multiple factor analysis framework. *BMC Bioinformatics,* 17**,** 402.

WEI, T. & SIMKO, V. 2016. corrplot: Visualization of a Correlation Matrix. R package version 0.77. *CRAN, Vienna, Austria*.

WICKHAM, H. 2014. Tidy data. *Journal of Statistical Software,* 59**,** 1-23.

WICKHAM, H. 2016. *ggplot2: elegant graphics for data analysis,* Springer.

La précédente publication vise à mettre en avant l'ensemble des analyses statistiques qui seront réalisées dans ce projet. Il y est notamment justifié et détaillé l'utilisation des diverses analyses statistiques utilisées en lien avec les questions biologiques posées. En raison de la grande quantité d'informations rassemblées, cette réflexion d'analyse a été poursuivie lors de la caractérisation des bases moléculaires des modifications des parois d'*A thaliana* cultivées à différentes températures.

La mise en place d'un protocole expérimental adapté, la standardisation des prétraitements des données ainsi que la méthodologie d'analyse des différents blocs présentés précédemment, nous permettent d'appréhender plus sereinement et plus efficacement le nombre conséquent des données générés durant la thèse.

Dans l'objectif d'étudier l'adaptation d'*A. thaliana* aux températures sub-optimales, par l'étude de sa variabilité naturelle, une partie du projet *WallOmics* s'est concentré sur l'étude de cette espèce dans la région de la chaîne de montagnes des Pyrénées.

III. Caractérisation de populations pyrénéennes d'*A. thaliana*

Le projet *WallOmics* a permis, entre autre, d'appréhender l'adaptation d'*A. thaliana* par l'étude de sa variabilité naturelle. La chaîne de montagnes des Pyrénées représente une barrière physique majeure séparant la péninsule ibérique du reste de l'Europe, ce qui en fait une zone d'étude privilégiée. Dans le cadre de ce projet, une campagne de terrain effectuée au printemps 2013 a permis d'identifier et de récolter 241 individus répartis en 30 nouvelles populations provenant d'altitudes contrastées. La localisation des zones de collecte a également permis d'établir le climat local spécifique (température, précipitation, altitudes, *etc*.) pour chaque population.

La caractérisation de ces nouvelles populations a été réalisée en trois temps :

- Premièrement, ces populations ont fait l'objet d'une analyse moléculaire afin d'évaluer la variation génétique des populations d'*A. thaliana* dans cette région. Deux approches complémentaires ont été réalisées : la statistique bayésienne et la phylogénie. Il en est ressorti une séparation génétique claire et spécifique de certaines populations des Pyrénées.

- Deuxièmement, une étude phénomique sur différents traits phénotypiques de ces nouvelles populations a été réalisée en laboratoire. Cette caractérisation porte sur deux organes : les rosettes ainsi que les tiges, observées à deux températures : 22°C et 15°C. L'ensemble de ces données phénomiques ont mis en évidence une plasticité phénotypique contrastée entre les populations en fonction des températures de culture.

- Enfin, une étude intégrative a pu confronter les données climatiques d'origine à un cluster génétique et à la plasticité phénotypique de ces populations.

Cette étude, qui sera soumise pour publication, est présentée ci-après.

Phenotypic trait variation in genetically distinct *Arabidopsis thaliana* populations from the Pyrenees Mountains highlight acclimation to environmental constraints

Running title: Distribution of *A. thaliana* in the Pyrenees

Harold Duruflé[†], Philippe Ranocha[†] *et al.*

[†] co-first authors

* Corresponding author:

Christophe Dunand (Professor): Email: dunand@lrsv.ups-tlse.fr // Tel: +33 (0)5 34 32 38 58

Abstract

The natural diversity is a free reservoir of variation for studying morphological and developmental traits. Studying natural variation will help to identify genetic mechanisms of complex traits. Mountain habitats provide genuine environmental temperature gradients where plants need to adapt to multiple environments. In this study, we report the identification and genetic study of 341 individuals of *Arabidopsis thaliana* (representing 30 new natural populations), collected between 200 and 1800 m in the Pyrenees Mountains. Class III peroxidases and ribosomal RNA sequences were used as genetic marker to determine the genetic relationships between populations along an altitudinal gradient in the Pyrenees. Using Bayesian-based statistics and molecular phylogenetic analyses the Pyrenees Mountains appears with clear demarcation from any other known ecotypes. Then the individuals are genetically separated in two major clusters and some homogeneous populations are found. These populations exhibited also great phenotypic variability (*e.g.* germination speed, chlorophyll content, height…) when grown at sub-optimal temperature (22°C vs 15°C) to mimic natural environment. These genetic variations are also expressed with natural intraspecific morphological variations. Therefore this study sheds a new light regarding the west European population structure and phenotypic plasticity of *A. thaliana* at contrasted temperatures, and illustrates the fact that integrative analysis combining genetic, phenotypic and climatology variation is a powerful tool for predicting acclimation of population in a global warming context. These new Pyrenean populations are made available to the scientific community and represent a new tool of choice for studying cold tolerance and development.

Keywords: *Arabidopsis*, Population genetics, Natural variation, Genetic diversity, Genetic structure, Peroxidase

Introduction

Plant diversity represents a huge reservoir of variation, due to spontaneous mutations and maintained by different evolutionary processes, like natural or artificial selection. The intraspecific natural variation, *i.e.* the within-species phenotypic variation, reflects species adaptations to different natural environments (Clauss & Mitchell-Olds 2006; Fournier-Level *et al.* 2011; Hancock *et al.* 2011). Analysing the natural variation in wild species will help to elucidate the molecular bases of phenotypic differentiation and thus understand species adaptation to specific natural environments from an ecological and evolutionary point of view (Mitchell-Olds & Schmitt 2006).

Mountainous areas provide unique natural gradients with multiple environments acting as open air laboratories. For instance, through altitudinal gradients, plants must compose with multiple environmental variations including the decrease of temperatures and air humidity, the increase of UV radiations and a diminution of atmospheric pressure with rising altitudes (Körner 2007). In response to these different climatic variations, plants must tightly regulate their physiological processes and modify their phenotypic traits. Additionally, there is evidence that global warming due to climate change is ongoing (Meehl *et al.* 2007). In this context of global climatic changes (seasons often altered, temperature changes, occurrence of freezing stress), studying phenotypic variations along altitudinal gradients will help to understand plant ability to cope, acclimate and adapt (Warren 1998). The temperature increase, and even a few more degrees may lead to ecological and physiological constraints with consequences on plant development. Genetic mechanisms of complex traits can be identified by molecular approaches by biologists who exploit natural variation to identify these mechanisms (Alonso-Blanco *et al.* 2005). Thus, it is imperative to better understand the impact of temperature variation and to integrate these elements into plant breeding programs.

Arabidopsis thaliana (L.) Heyhn. (Brassicaceae) is an annual plant species with worldwide distribution. It is a well-known model for plant molecular sciences, as well as for ecological and evolutionary studies, and numerous genetic resources are available. In addition, it can be maintained as pure lines due to its highly self-fertilizing nature. Several studies about natural genetic variation of *A. thaliana* have been published at local scale in its native European distribution range (Bakker *et al.* 2006; Brennan *et al.* 2014; Horton *et al.* 2012; Le Corre 2005; Nordborg *et al.* 2005; Picó *et al.* 2008; Stenøien *et al.* 2005; Suter *et al.* 2014; Wilczek *et al.* 2014). For example, a large genetic diversity of *A. thaliana* has been found in the Iberian Peninsula in which a strong geographic structure raises the hypothesis of multiple Iberian

glacial refuges that contributed to the post-glacial recolonization of Europe (Beck *et al.* 2008; Consortium 2016). The Pyrenees Mountains chains is the physical barrier that separates the Iberian Peninsula from France and the north of Europe. Studying the genetic and phenotypic diversity in this region is interesting to analyse the genetic mixture and the plant colonization of Europe from the Iberian Peninsula.

The class III peroxidases (CIII Prx) have been detected in all green plants and they constitute a large multigenic family in land plants (Passardi *et al.* 2004). They oxidize various compounds in the presence of peroxide (H_2O_2) or oxygen (O_2) and play a major role in response to many biotic and abiotic stresses, during auxin degradation, cell wall lignification and plant senescence (Francoz *et al.* 2015). Although, peroxidase families are subjected to high rate of duplication, numerous residues belonging to the catalytic domain are conserved between species and ecotypes. The genetic variability between the peroxidase family genes justifies the use of CIII Prx sequences as genetic marker to establish the structure of populations originating from contrasted biotope (Gulsen *et al.* 2010; Nemli *et al.* 2014; Pinar *et al.* 2016). Furthermore, they have already been used to study evolutionary relationships and genotypic diversity on an intra and inter-specific basis (Gulsen *et al.* 2010 2016; Uzun *et al.* 2014).

Another valuable genetic marker are ribosomal RNA (rRNA) genes or rDNA. In plants, two major classes of nuclear rDNA can be distinguished: 45S rDNA and 5S rDNA (Layat *et al.* 2012). The genome of *A. thaliana* contain hundreds of 45S rDNA units, each of them encoding the 18S, 5.8S and the 25S rRNA genes separated by external and internal transcribed spacer (5'ETS and 3'ETS). Significant repeat number polymorphisms for 45S rDNA has been reported among different *A. thaliana* accessions (Davison *et al.* 2007) and lines from Sweden (Long *et al.* 2013); provoking substantial variations in genome sizes. On other hand, the characterization of the 45S rDNA 3'ETS sequence from various *A. thaliana* ecotypes originated from different geographic regions , allowed to reveals the existence of one to four 45S rDNA variants (Abou-Ellail *et al.* 2011; Chandrasekhara *et al.* 2016; Pontvianne *et al.* 2010).

In the present study, our goals were (i) to estimate the CIII Prx nucleotidic polymorphism and the genetic diversity of *A. thaliana* populations from the Pyrenees Mountains, which had never been analysed before, (ii) to analyse the phenotypic variability in these populations at the micro- and macro- levels and (iii) to perform integrative analyses including genetic, phenotypic and environmental data and to estimate the capacity of a given population to acclimate or adapt to different growth temperatures. Studying the genetic diversity of *A. thaliana* in the Pyrenees Mountains help to understand west European population

structure and to evaluate the phenotypic plasticity in response to climate change. This study highlights the acclimation to environmental constraints by putting in relation genetic variations and phenotypic data within and among local populations of *A. thaliana* in the Pyrenees Mountains.

Materials and Methods

Population sampling, growth conditions and seed availability

341 individual divided into 30 new natural populations of *Arabidopsis thaliana* were collected (seeds and whole plants, GPS coordinates are listed in Tab. 1) at different altitudinal levels all along the Pyrenees Mountains (Tab. 1). Their taxonomic belonging to *A. thaliana* species was confirmed through DNA sequencing (cf. genetic analysis). Population names correspond to the first four letters of the closest village or location where the plants were found (Tab. 1). Three populations were used as out-groups: one population located in Lanta, Lant, (10 accessions) close to Toulouse (France) (Bartolli et al. 2017), and two well-known ecotypes corresponding to contrasted altitudes (publiclines.versailles.inra.fr), *i.e.* Columbia, Col (200 m; Poland; 52.745416, 15.235557) and Shahdara, Sha (3,400 m; Tajikistan, 39.250103, 68.249919), and already studied in the lab under the same growth conditions (Duruflé *et al.* 2017).

Field-collected seeds from about 10 individuals per population (Tab. 1) were amplified at least one time in a growth chamber to obtain homogeneous batches of seeds with less parent-of-origin effects, before being used in all experiments of this study.

Seeds were sown in Jiffy-7® peat pellets (Jiffy International, Kristiansand, Norway). After 48 h of stratification at 4°C in darkness, plants were grown at a light intensity of 90 µmol.photons.m-2.s-1, a 70 % humidity and under a 16 h light/8 h dark photoperiod at two different temperatures: 22°C (optimal growth condition for Col) or 15°C (sub-optimal temperature for Col). For rosette phenotyping and biochemical analyses, four- and six week-old rosettes were collected respectively at 22°C and 15°C, corresponding to the bolting stage (emergence of the first flower buds; stage 5.10) for Col and Sha (Boyes *et al.* 2001 2017). Each experiment was done in triplicate with 2 to 5 plants.

Table 1. Characteristics of the populations collected in the Pyrenees Mountains.

N: Number of individuals per population; Localization (GPS coordinates) and bio-climatic variables of the populations of A. thaliana analysed in this study. Geolocalisation and altitude were determined during the sampling. We used the average of the monthly data to calculate the annual minimum, mean and maximum temperatures and annual precipitation obtained from the WorldClim project (Hijmans et al. 2005; www.worldclim.org). Annual UV radiation was obtained from the Photovoltaic Geographical Information System of the European Community (Huld et al. 2012). Climate PC1 values are the position of populations on the first principal component of the PCA of bioclimatic variables that allows to classify populations in function of their environmental conditions (Fig. S4).

Accession name	*N*	*Location*	*Latitude*	*Longitude*	*Altitude (m)*	*Minimum monthly T (°C)*	*Mean annual T (°C)*	*Maximum monthly T (°C)*	*Accumulated annual rain (mm)*	*Accumulated annual radiations (kWh/m²)*	*Climate PC1*
Arag	10	Aragnouet	42.7806	0.1950	1316	-4.3	6.3	19.6	1095	1553.2	-2.0
Argu	10	Argut-Dessous	42.8889	0.7176	723	-1.3	10.2	23.9	931	1394.6	1.0
Bedo	11	Bedous	42.9951	-0.6001	403	0.0	11.5	24.7	885	1350.3	1.9
Belc	11	Belcaire	42.8217	1.9681	967	-0.9	9.6	23.2	925	1485.9	0.3
Biel	10	Bielle	43.0523	-0.4339	450	-0.3	11.1	24.3	890	1395.4	1.6
Bier	8	Biert	42.8993	1.3140	590	-0.4	11.0	24.9	840	1456.5	1.5
Bran	11	Pas de barane	42.9708	0.2312	920	-2.6	8.5	21.9	1001	1482.6	-0.4
Camu	13	Camurac	42.8001	1.9139	1198	-1.9	8.3	21.8	999	1514.1	-0.6
Cast	10	Castet	43.0696	-0.4184	431	-0.1	11.3	24.4	882	1371.4	1.8
Chau	9	Chaum	42.9395	0.6504	492	-0.4	11.4	25.3	821	1430.3	1.8
Col	1	Columbia	52.7454	15.2356	200	-5.2	9.0	24.5	546	1017.6	3.0
Eaux	12	Eaux Chaudes	42.9521	-0.4399	659	-1.9	9.2	22.5	953	1410.5	0.5
Eget	10	Eget Cité	42.7907	0.2611	1115	-3.6	7.2	20.6	1049	1523.1	-1.3
Fos	12	Fos	42.8738	0.7302	531	-0.9	10.8	24.7	880	1431.4	1.4
Gava	13	Gavarnie	42.7361	0.0101	1359	-8.1	1.8	14.3	1363	1500.6	-4.5
Gedr	13	Gedre	42.7920	0.0180	992	-3.6	7.2	20.6	1025	1405.5	-0.7
Grip	13	Gripp	42.9281	0.2041	1190	-3.5	7.3	20.5	1068	1526.9	-1.4
Guch	12	Guchen	42.8639	0.3428	755	-1.6	9.8	23.5	908	1511.8	0.4
Hern	10	Herran	42.9730	0.9150	780	-1.4	9.9	23.7	946	1426.8	0.7
Herr	11	Herrère	43.1689	-0.5401	323	0.5	12.1	25.0	910	1349.9	2.1
Hosp	12	Hospitalet-pres-l'Andorre	42.5862	1.7968	1424	-2.9	7.0	20.3	1072	1536.3	-1.6
Jaco	13	Jacoy	42.9061	1.4073	989	-1.5	9.4	23.2	946	1456.9	0.3
Lant	10	Lanta	43.5649	1.6524	246	0.8	12.5	26.7	749	1372.2	2.9
Lave	12	Lavelanet	42.9308	1.8414	539	0.4	11.7	25.5	811	1444.5	1.9
Mari	12	Sainte Marie de Campan	42.9843	0.2255	840	-2.0	9.3	22.7	952	1447.9	0.2
Mere	13	Merens-les-Vals	42.6580	1.8383	1069	-1.2	9.2	22.7	937	1473.0	0.1
Mong	13	Mongie	42.9098	0.1795	1800	-5.8	4.5	17.3	1231	1559.3	-3.5
Pont	10	Pont d'espagne	42.8512	-0.1407	1456	-7.9	2.1	14.7	1365	1473.2	-4.3
Prad	12	Prades	42.7874	1.8812	1214	-2.1	8.2	21.6	1006	1527.6	-0.8
Roch	12	Chapelle Saint Roch	43.0040	0.1909	696	-1.7	9.7	23.0	922	1411.0	0.7
Savi	11	Savignac-les-Ormeaux	42.7299	1.8153	690	0.2	11.3	25.0	822	1496.6	1.4
Sha	1	Pamiro-Alay	39.2501	68.2499	3400	-16.9	0.5	19.7	893	1787.4	-3.0
Urdo	12	Urdos	42.8725	-0.5545	775	-1.7	9.4	22.9	920	1415.5	0.6

Climatic data

Climatic variables were obtained from WorldClim dataset (Hijmans *et al.* 2005; www.worldclim.org). The values used are the mean of 30 years (1960 to 1990) with a resolution of *ca.* 1 km² per grid cell. Solar radiation reading were obtained from the Photovoltaic Geographical Information System of the European Communities (Huld *et al.* 2012) for the 2001-2012 period (http://re.jrc.ec.europa.eu/pvgis/). The specific solar radiation of Sha was estimated by linear regression between altitude level and solar radiation of the Pyrenees Mountains.

DNA extraction

Genomic DNA was extracted using a standard CTAB protocol from leaves of 351 samples from the 30 populations of *A. Arabidopsis* as well as from the Col, Sha and Lant accessions. DNA concentration was measured by spectrophotometer ND-1000 (NanoDrop, Wilmington, Delaware, USA). Five couple of primers (sequences given in Fig. S1A) were designed to amplify 5 genomic areas of about 1,000 bp each. The 5 regions are distributed on the 5 chromosomes of *A. thaliana*; they encompass *ca.* 500 bp upstream of the ATG and *ca.* 500 bp both coding and non-coding region (including exon and intron). They correspond to the loci of the 5 following CIII Prx (*At1g44970* (*AtPrx09*), *At2g41480* (*AtPrx25*), *At3g50990* (*AtPrx36*), *At4g33870* (*AtPrx48*), *At5g39580* (*AtPrx62*)).

PCR were performed using high fidelity recombinant *Pfu* DNA polymerase (Promega, Madison, WI, USA) according to the manufacturer's instructions. PCR products were sequenced using the same primers and on both strands to ensure reading quality.

Paired end sequencing and manual editing using BioEdit (Hall 2011) allowed a good sequencing quality. The 5 sequences obtained for each of the 353 individual were concatenated and used for genetic and phylogenetic analysis.

PCR amplification of 3'ETS 45S rRNA genes was done using p3/p4 primers as described in (Pontvianne *et al.* 2010). The QGIS software 2.18 was used to georeference the *A. thaliana* populations on a map and to display associated camembert diagrams illustrating the proportion of R haplotypes identified for each population (Quantum GIS Development Team, 2009. QGIS Geographic Information System. Open Source Geospatial Foundation. URL http://qgis.osgeo.org).

Genetic analysis

The sequences of 30 Pyrenean populations, the 3 out-group (Lant, Col and Sha) and 22 ecotypes available from the 1001 Genomes consortium (11 in South France and 11 in North Spain) were analysed (Tab S1). Tajima's D neutrality test was performed using DnaSP v5 software (Librado & Rozas 2009). For Col, Sha and the French/Spanish ecotypes, one individual only was used to minimise the impact of the frequencies of these new haplotypes but also because no information regarding the homogeneity (genetic diversity) of these populations was available. The haplotype class of each sequence has been determined using the Col sequence as reference (home-made Perl script). Population structure was analysed using a Bayesian clustering method implemented in the STRUCTURE software version 2.3.4. (Pritchard *et al.* 2000), Assuming K differentiated genetic clusters in the sample, this method allows estimating the proportion of membership of any individual to any of the K clusters. For each population, 10^6 iterations and 10^5 burn-in period options were used. For each value of K (between 1 to 10), 10 independent calculations were performed, and likelihood values obtained from these 10 calculations were averaged. To identify the appropriate K value, we used Evanno's method based on data likelihood variation over successive K (Evanno *et al.* 2005). The optimal K value was estimated with the largest delta K value with higher likelihood than that of K-1 runs (Fig. S2A-B).

Molecular phylogenetic analysis by maximum likelihood method.

The evolutionary history was inferred by using the maximum likelihood method based on the Tamura-Nei model (Tamura & Nei 1993) performed with a concatenated sequence containing the 5 CIII Prx sequences. The tree with the highest probability is shown with 50% of cut-off value to condense the tree done with 100 bootstraps. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree for the heuristic search was obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value (Tamura *et al.* 2004). The analysis involved 375 nucleotide sequences (from 341 natural individuals from Pyrenean populations, 10 individuals from Lant, the 2 references Col and Sha and 22 ecotypes from the 1001 Genomes consortium). All positions containing gaps and missing data were eliminated. The 5 sequences were concatenated for a total of 4074 positions in the final dataset. Evolutionary analyses were conducted in MEGA v6 (Tamura *et al.* 2013). Genetic distances

were measured as pairwise differences among haplotypes and 100 bootstraps were generated to determine the phylogenetic structure.

Germination

Arabidopsis thaliana seeds were sterilized (30 % bleach (v/v), 0.1 % triton 100 X (v/v)) and sown in Petri dishes containing half-concentrated MS medium (Murashige & Skoog 1962) in 1 % agar. After 48 h of stratification at 4°C in darkness, seeds were incubated in a germination cabinet at 24°C, 60 % of humidity and 110 µmol.photons.m-2.s-1 continuous white light. The percentage of ruptures of two protective envelopes, testa (the outer envelope) and endosperm (the inner envelope) ruptures were obtained by counting two times 50 to 100 seeds from three independent biological replicates. Data were expressed as the mean percentage of the total seed number with standard error of the mean. Testa and endosperm ruptures were quantified, using the testa integrity and the radicle tip protrusion as markers, respectively (Fig. S3A, (Lariguet *et al.* 2013)).

Macrophenotyping

Rosettes grown at 22°C and 15°C were analysed when harvested. Rosette diameter and fresh weight were determined and the number of leaves was counted. The bolting stage was checked for each ecotype. From this point, stems length was measured up to the final height. When stem growth was over, stem diameter and length was measured and the lateral stems/cauline leaves were counted.

Chlorophyll content

Whole rosettes were ground in liquid nitrogen and used for anthocyanin and chlorophyll extraction. Ground material (0.1 g) was vigorously vortexed with 1 mL of 80% acetone solution for 5 min. The samples were centrifuged for 10 min at 1,000 g. The absorbance of chlorophyll α and β was measured in the supernatant, respectively at 663 and 647 nm. Chlorophyll concentration was calculated using the following formula: (A_{663} x 7.15 + A_{647} x 18.71) / mg of fresh material = µg chlorophyll/mg fresh material (Cosio & Dunand 2010).

Anthocyanin content

Ground rosette material (0.1 g) was mixed with 1 mL of 95% ethanol /1% HCl and stored at 5°C for 24 h in darkness. The samples were centrifuged for 10 min at 1,000 g. The absorbance of the supernatant was measured at 530 and 657 nm. The anthocyanin content was

calculated using the formula ((A_{530} – 0.25 x A_{657}) / mg of fresh material)) as described in (Mancinelli *et al.* 1988).

Stem maximum growth speed

To identify the maximum growth speed of the floral stems, a logistics function was modeled (Krislov Morris & Kuhn Silk 1992; Lièvre *et al.* 2016). The data-fitting curve was created for each kinetic. The logistic function used to model stem growth is given by [Y= Asym / (1 + exp ((xmid - log(t)) / scal))] and is parameterized by three parameters corresponding to the final stem length (Asym), the time corresponding to the inflexion point (xmid) and the characteristic growth speed (Scal). The estimated parameter was used to characterize individual stem growth speed.

Data Analysis

Most data analyses were performed with R software (version 3.2.3). ANOVA statistical tests were carried out in order to determine the temperature effects on the populations. In addition, Duncan's multiple range tests were performed to allow a better visualization of the different sets when they correspond to averages. In addition, Kruskal-Wallis and Wilcoxon nonparametric tests were carried to compare groups. To investigate the underlying variation between population data (phenotypic), multilevel principal component analysis (PCA) was used with the mixOmics package (Lê Cao *et al.* 2011) available in CRAN (cran.r-project.org/package=mixOmics). This combination of multilevel and multivariate methods allowed distinguishing variations between populations and between temperatures. Multilevel PCA was done only with contrasted data collected at 15 and 22°C. In addition, distance between the same populations according to the temperature in the scaled PCA multilevel was measured. Graphical representation of the STRUCTURE results was done with the "leaflet" package (cran.r-project.org/web/packages/leaflet).

Results and discussion

Local natural environmental conditions are highly contrasted in the Pyrenees Mountains

The 30 natural populations of *Arabidopsis thaliana* studied in this project were all collected from the Pyrenees Mountains (Tab. 1). Lant was used as a control population since it grows on the plain near the Pyrenees Mountains. Because they are well-known and living in contrasted environments of origin, Sha growing at 3,400 m and Col growing at low altitude

were also included in the study as controls. Each population was described using the altitude (m) and the following 5 climatic variables: annual minimum, annual maximum and annual mean temperatures (°C), total annual precipitation (mm), total annual UV radiations (KW h / m^2).

The variables describing temperature variability (minimum, mean and maximum) are highly correlated and also anti-correlated with the altitude, the annual rain accumulation and radiation. Climatic trends are not generally related to altitude and depend on the topology of temperate mountains. Precipitation and seasonality exert the largest influence on the regional climate variation (Körner 2007). Pyrenees Mountains appear with a temperate typology where the gradient of precipitation is altitude dependent. To analyse the environmental datasets (climates and altitudes), PCA were used to produce principal components (PCs) that explain the climatic variance (Fig. S4). The PCs score approach is routinely used for climate quantification (Wolfe & Tonsor 2014) and represent an index that describes the climate gradients of the populations (Tab. 1). In the principal component analysis (Fig. S4A), the climate PC1 explained 74 % of the multivariate variance across the 6 variables, while climate PC2 only explained 14 % (see Fig. S4A). Climate PC1 is most strongly associated with all the variables, while climate PC2 is only associated with the atypical climatic profile of Sha (Fig. S4B). Then, the environmental data characteristic of each population can be illustrated with climate PC1 value (Tab. 1).

It is difficult to compare the characteristics of all populations but several information could nevertheless be obtained from environmental data. Climate PC1 value allows to classify populations as a function of their environmental conditions considering altitude levels: plain (PC1 value > 0), medium (0 > PC1 value > 1) and high altitudes (PC1 value < 1). The climate PC1 values are highly correlated with the population altitude but some differences could be noted. Eaux has a lower climate PC1 value than other populations living at the same altitude like Savi or Bier. That could be explained by the lowest temperatures at this location as compared to Hern and Roch. On the other hand, Bran and Gedr have a negative climate PC1 value, as do the Pyrenean populations living above 1,100 m of altitude which could be explained by the lowest temperatures and the high precipitations. Climate PC1 value will be useful to compare environmental conditions and phenotypes data in the integrative study.

Both the genetic diversity and the geographic structure reflect a specific genetic lineage in the Pyrenees Mountains

Sequence polymorphism is currently used to analyse genetic variability and structure of natural populations (Gulsen *et al.* 2010; Nemli *et al.* 2014) and the study evolutionary relationships and genotypic diversity on an intra and inter-specific basis (Gulsen *et al.* 2010 2016; Uzun *et al.* 2014). In this study, the polymorphism of five *CIII Prx* loci (*AtPrx09*, *AtPrx25*, *AtPrx36*, *AtPrx48* and *AtPrx62*) was analysed. These five amplified sequences include both coding and non-coding regions (Fig. S1B). They have been selected for their high degree of polymorphism (Single Nucleotide Polymorphism (SNP) and Insertion/deletion (Indel)) detected between ecotype sequences available from 1,001 Genomes public data (http://1001genomes.org/; Weigel, 2009) and with no *a priori* concerning their putative implication in altitude adaptation.

In addition to the sequences obtained from the 341 Pyrenean individuals, the populations used as out-groups (10 individuals of Lant, and one Col and Sha) and 22 ecotypes sequences available from the 1,001 Genomes Project (Consortium 2016) and localized in the south of France and the north of Spain were added. These ecotypes originating from both sides of the Pyrenees Mountains improved the sampling in the regional context. Based on the Col sequence, 338 positions of SNP and Indel were identified and used to define different haplotypes (combination of alleles of different markers in the sequences). According to our analyses, several populations used in this study share the same haplotype (Tab. S2). This could mean that these populations have the same origin, either natural or related to anthropogenic introduction/displacement (e.g. transportation, road construction…). Tajima's D was calculated to compare the total number of segregating sites (polymorphisms) to the average pairwise differences between sequences (Tajima 1989). This test allows to distinguish DNA sequences evolving under selective pressures from neutrality-evolving ones. Only neutrality-evolving sequences can be used for genetic diversity analyses aiming at inferring population structure. The result was not significantly different from zero (D = 0.420), meaning that we could not reject the null hypothesis of neutral mutation–drift equilibrium for these sequences. As a consequence, these markers could then be considered as neutral. On the 55 populations (30 Pyrenean, 22 from the 1001 genomes project and 3 out-group accessions) 58, 39, 54, 93 and 45 different haplotypes were found for *AtPrx09, AtPrx25, AtPrx36, AtPrx48 and AtPrx62*, respectively (Tab. S2). The polymorphism within populations was evaluated using different parameters: the number of haplotypes and the haplotypic frequency distribution, the number of

SNPs and the percentage of identity of haplotypes contents in the populations (Tab. 2). By this systematic analysis and even though *AtPrx48* gene marker presents an important rate of polymorphism, it is possible to say that 22% of the 30 Pyrenean populations (i.e. Camu, Gava, Grip, Hosp, Jaco, Pont and Roch) share the same haplotypes among their individuals. However, based on the identity between the haplotypes and the percentage of SNPs, two other populations (Hern and Prad) appear to be homogeneous (Tab. 2). A population is considered as homogeneous if all the individuals of this population share the same haplotype.

Table 2. Polymorphism observed between the individuals of the populations based on haplotypes characteristics. The following parameters are shown for the populations: % Poly-morphic loci: number of haplotypes detected among the individuals of one population expressed in percentage; SNP (%): percentage of SNP found on the concatenated sequences (including the 5 CIII Prx loci) per population; haplotypic frequency distribution: number of individuals belonging to the same haplotype within a population; identity (%): minimum and maximum percentage of identity within population of the concatenated sequences graphically represented in Fig. 2.

Accession name	% Poly-morphic loci	SNP (%)	Haplotypic frequency distribution	Identity (%)
Arag	100	2.70	1/1/1/1/1/1/1/1/1/1	0.974 - 0.999
Argu	70	1.72	1/1/1/1/1/1/2/3	0.983 - 1
Bedo	45	1.00	1/1/1/1/7	0.989 - 1
Belc	100	3.08	1/1/1/1/1/1/1/1/1/1/1	0.978 - 0.999
Biel	90	3.92	1/1/1/1/1/1/1/1/2	0.962 - 1
Bier	75	2.20	1/1/1/1/1/3	0.982 - 1
Bran	100	2.08	1/1/1/1/1/1/1/1/1/1/1	0.98 - 0.999
Camu	15	0.07	1/12	0.998 - 1
Cast	90	2.94	1/1/1/1/1/1/1/1/2	0.976 - 1
Chau	89	3.25	1/1/1/1/1/1/1/2	0.974 - 1
Eaux	50	0.53	1/1/2/2/3/3	0.994 - 1
Eget	40	0.72	1/1/3/5	0.992 - 1
Fos	67	1.79	1/1/1/1/1/1/3/3	0.982 - 1
Gava	15	0.02	1/12	0.999 - 1
Gedr	62	2.29	1/1/1/1/1/2/2/4	0.977 - 1
Grip	15	0.57	1/12	0.994 - 1
Guch	67	0.19	1/1/1/1/1/1/3/3	0.998 - 1
Hern	70	0.12	1/1/1/1/1/1/4	0.998 - 1
Herr	73	0.79	1/1/1/1/1/2/2/2	0.993 - 1
Hosp	17	0.05	1/11	0.999 - 1
Jaco	15	0.02	4/9	0.999 - 1
Lant	91	4.28	1/1/1/1/1/1/1/1/1/1	0.966 - 0.999
Lave	55	2.82	1/1/1/2/3/4	0.972 - 1
Mari	83	1.91	1/1/1/1/1/1/1/1/3	0.984 - 1
Mere	62	0.33	1/1/1/1/1/1/2/5	0.997 - 1
Mong	100	4.06	1/1/1/1/1/1/1/1/1/1/1/1	0.966 - 0.998
Pont	20	0.02	1/9	0.999 - 1
Prad	33	0.05	1/2/2/7	0.999 - 1
Roch	17	0.02	1/11	0.999 - 1
Savi	100	2.15	1/1/1/1/1/1/1/1/1/1/1	0.981 - 0.999
Urdo	100	3.01	1/1/1/1/1/1/1/1/1/1/1	0.975 - 0.999

To determine and analyse the genetic structure and the polymorphism of the *A. thaliana* populations collected in the Pyrenees Mountains two different approaches were used: (i) a Bayesian clustering analysis and (ii) a molecular phylogenetic analysis. These two analyses are complementary to the systematic approach previously exposed. In order to infer population structure, they use haplotype frequencies to partition within and between populations diversity.

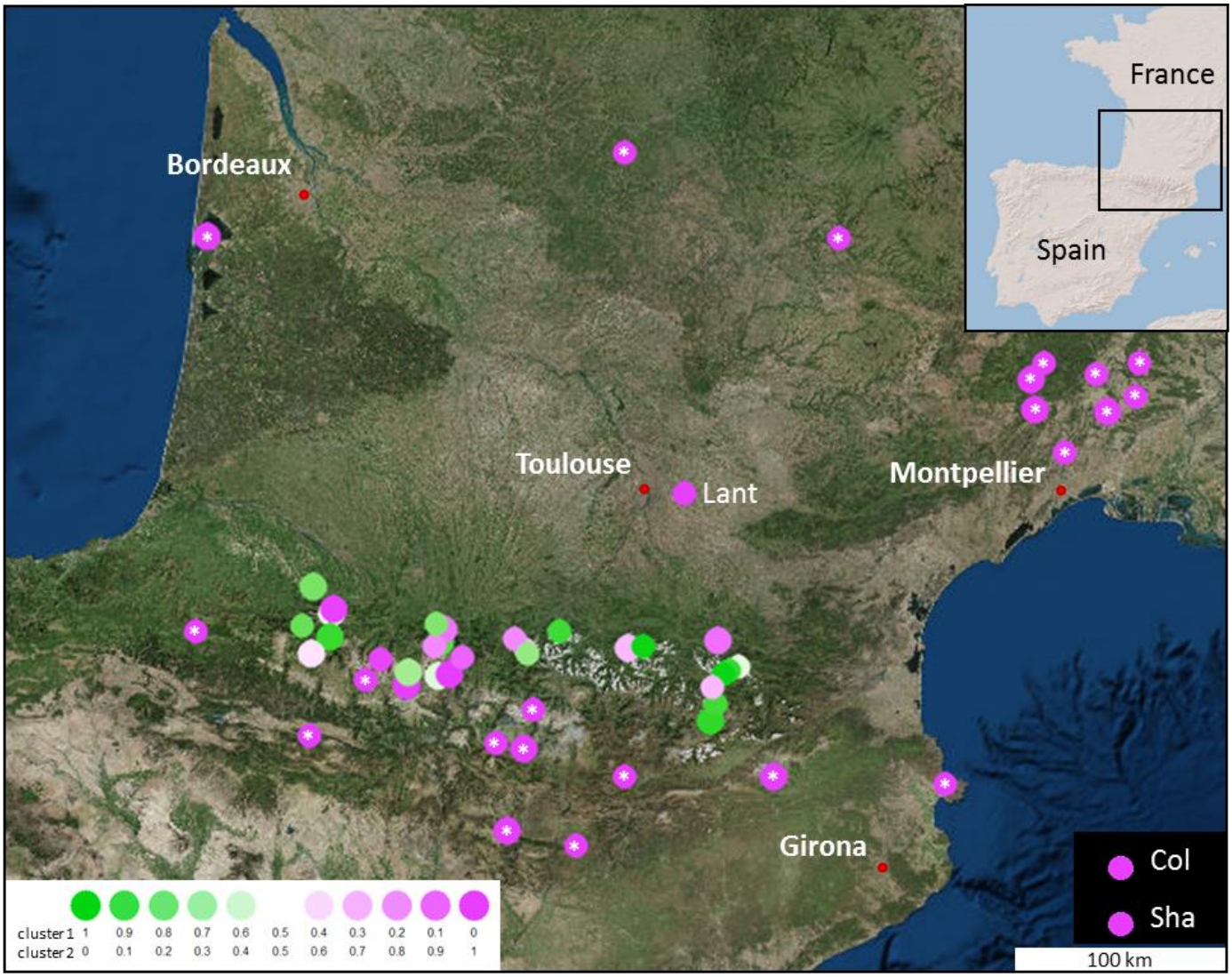


Figure 1. Geographic location and genetic group assignment of 30 populations of A. thaliana in Pyrenees Mountains, 22 accessions of 1001 genomes and 3 out group populations, Lant close to Toulouse and the 2 well known ecotypes, Col and Sha. Relationships inferred with STRUCTURE are illustrated by colored circle. Each individual circle represents the populations allocation into their estimate membership proportions in each genetics cluster determined by STRUCTURE results (K=2). White stars stand for 1001 genomes populations.

In the clustering analysis implemented in STRUCTURE, various genetic clusters (K value) are possible to test, each of which is characterized by a set of haplotype frequencies. The

STRUCTURE analyses allowed to unambiguously infer two genetic clusters (Fig. S2B). Populations with a membership proportion of > 0.8 to a cluster are considered as homogenous (Fukunaga *et al.* 2005; Uzun *et al.* 2014) and populations with < 0.8 proportion are considered as admixed (heterogeneous). The clustering analysis identified 36 populations homogenous among the 55 analysed and considered them as homogeneous and non-admixed. All the ecotypes from 1001 genomes, the populations used as outgroup (Lant, Col and Sha) and 67% of the Pyrenees Mountains populations were found homogeneous and belonged to the same genetic cluster (Fig. S2C). In all the Pyrenees populations, 11 (37%) shared the same genetic cluster differing from the known French and Spanish populations (Fig. 1). Three genetic clusters were defined with the structuring Bayesian analysis as following: i) populations assigned to the clusters 1 (green) and 2 (pink) are homogeneous populations with membership proportion > 0.8 to one of the cluster, ii) populations assigned to the clusters 3 are considered as "hybrid cluster" because they do not belong at one of the two previous clusters.

To have a better view of the natural genetic diversity among the *A. thaliana* populations collected in the Pyrenees, a series of phylogenetic analyses was carried out. Genetic relationships among the different haplotypes were determined by the Neighbor-Joining (NJ) method using the concatenated sequences including the 5 CIII Prx markers. The dendrogram separates the individuals of the Pyrenees populations into three major branches well supported (Fig. 2). Moreover, several populations are found grouped on the same subtree (e.g. Hosp, Hern, Prad) or closed to each other on the same branch (e.g. Roch, Grip, Jaco) as expected with the systematic approach.

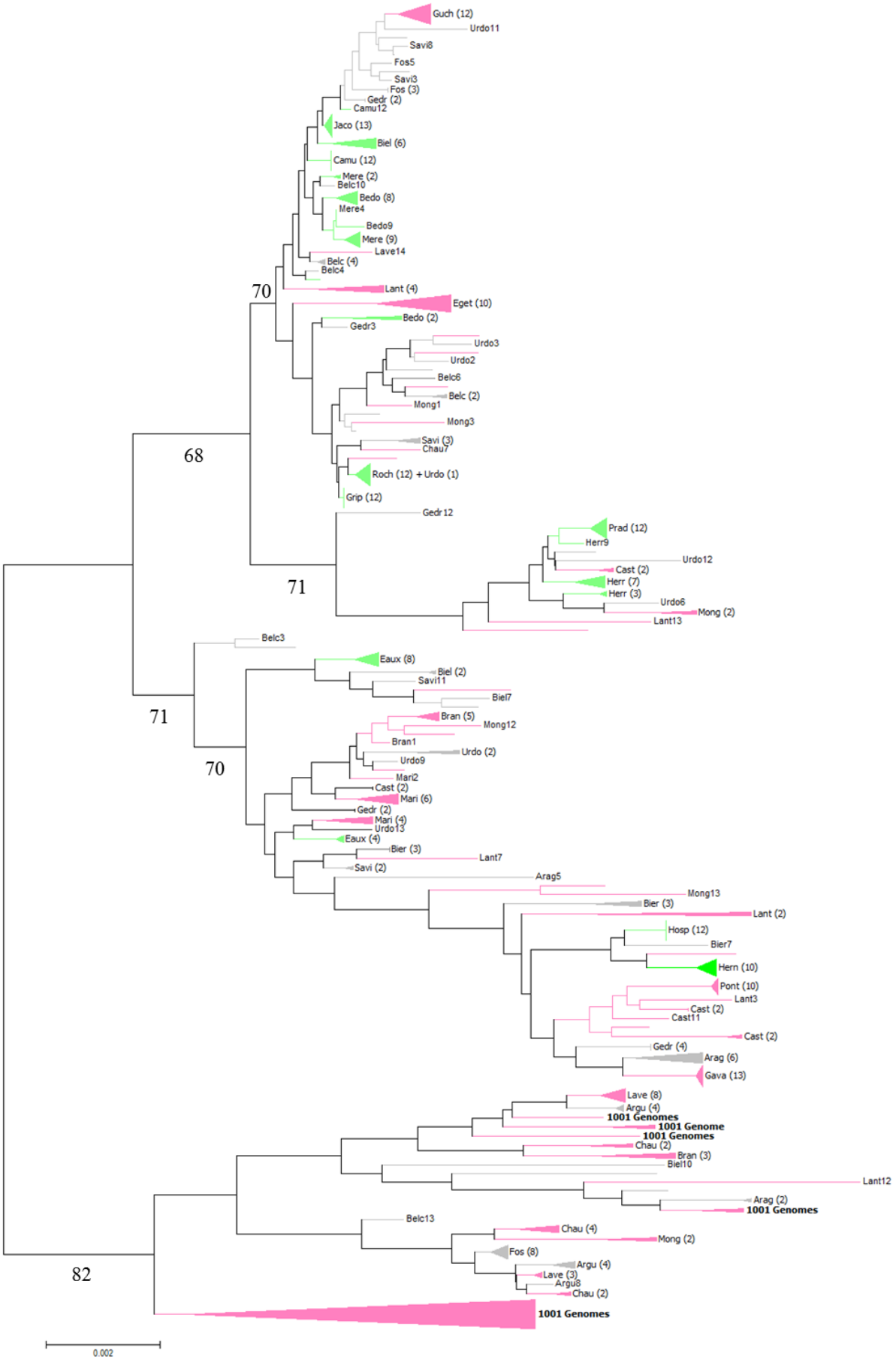


Figure 2. Molecular phylogenetic analysis based on the peroxidase gene markers (maximum likelihood method) of 341 individuals of 30 *A. thaliana* populations from the Pyrenean mountains, 22 ecotypes from the 1001 genome consortium (11 South France and 11 North Spain) and 3 out-group individuals (10 individuals of Lant and the 2 well known ecotypes, Col and Sha). Color represents the population allocations into their estimate membership proportions in each genetics cluster determined by STRUCTURE and illustrated in the Fig. 1. Numbers in brackets stand for the number of individual per cluster and numbers without brackets represent the individual within the population. The Branches and clusters are colored with the same colours as STRUCTURE analysis (Fig.1) with the admixed population in grey.

Most of the *A. thaliana* populations collected in the Pyrenees Mountains are genetically separated from the 22 ecotypes from the 1001 genomes project. The specific structure of these populations in this geographic area makes them very interesting to study in order to understand regional population structure in *A. thaliana*. However, the genetic diversity in the Pyrenees Mountains is not just defined by three genetic clusters. Indeed the diversity measured by molecular phylogenetic analysis describes a fine-scale variation. These inter- and intra-population variations may be associated with the variability of the natural environment and may attest of a specific adaptation of *A. thaliana* to this physical barrier. They could also have resulted from a prehistoric introduction or from a glacial refuge. It is interesting to note the complementarity of both methods with the sequence diversity approach. Indeed, populations that are found non-admixed according to the STRUCTURE approach are also those with a high percentage of identity of sequences and are found in one branch of the tree (such as Grip, Hosp, Hern, Prad and Roch). In contrary, the populations considered as admixed were found distributed into two branches of the tree (such as Belc, Fos, Gedr).

In order to confirm the homogeneity of these 9 populations (Camu, Gava, Grip, Hern, Hosp, Jaco, Pont, Prad and Roch), we performed PCR analyses of the 3'ETS from 45S rRNA gene sequences as an additional genetic marker. This marker is multicopy, as there are about 800 copies of 45S rRNA genes in the Col. These copies can be subdivided into 4 classes or variants in Col. Three of these variants (VAR1, 2, 3) represent almost all the rDNA copies, VAR1 accounting for ~50 % of the copies. The VAR4 is relatively rare (VAR4) (Abou-Ellail *et al.* 2011; Pontvianne *et al.* 2010). Col is one of the most complex ecotype worldwide as the majority of the other ecotypes that were analysed to date display only one (e.g. VAR3 in Sha) or two variants (VAR1 and 3) (Abou-Ellail *et al.* 2011; Chandrasekhara *et al.* 2016). Figure 3 shows the different R haplotypes that we encountered in the 9 Pyrenean populations analysed. The VAR4 was not used to define the R haplotype classes but rather that the relative proportion might vary between variants. For example, the distinction was made between R5.1, R5.2 and R5.3 regarding the relative abundance of VAR1 and VAR3. The abundance is relatively similar in R5.1 or not in R5.2 or R5.3 with VAR1 or VAR3 were the most abundant respectively. Camembert diagrams illustrate the proportion of these R haplotypes for each population. Consistently to the data obtained with CIII Prxs markers, the 9 populations (Camu, Gava, Grip, Hern, Hosp, Jaco, Pont, Prad and Roch) were found 90-100% homogenous compared to the out-group Lant population which is more heterogeneous (Fig 3 and Table S3).

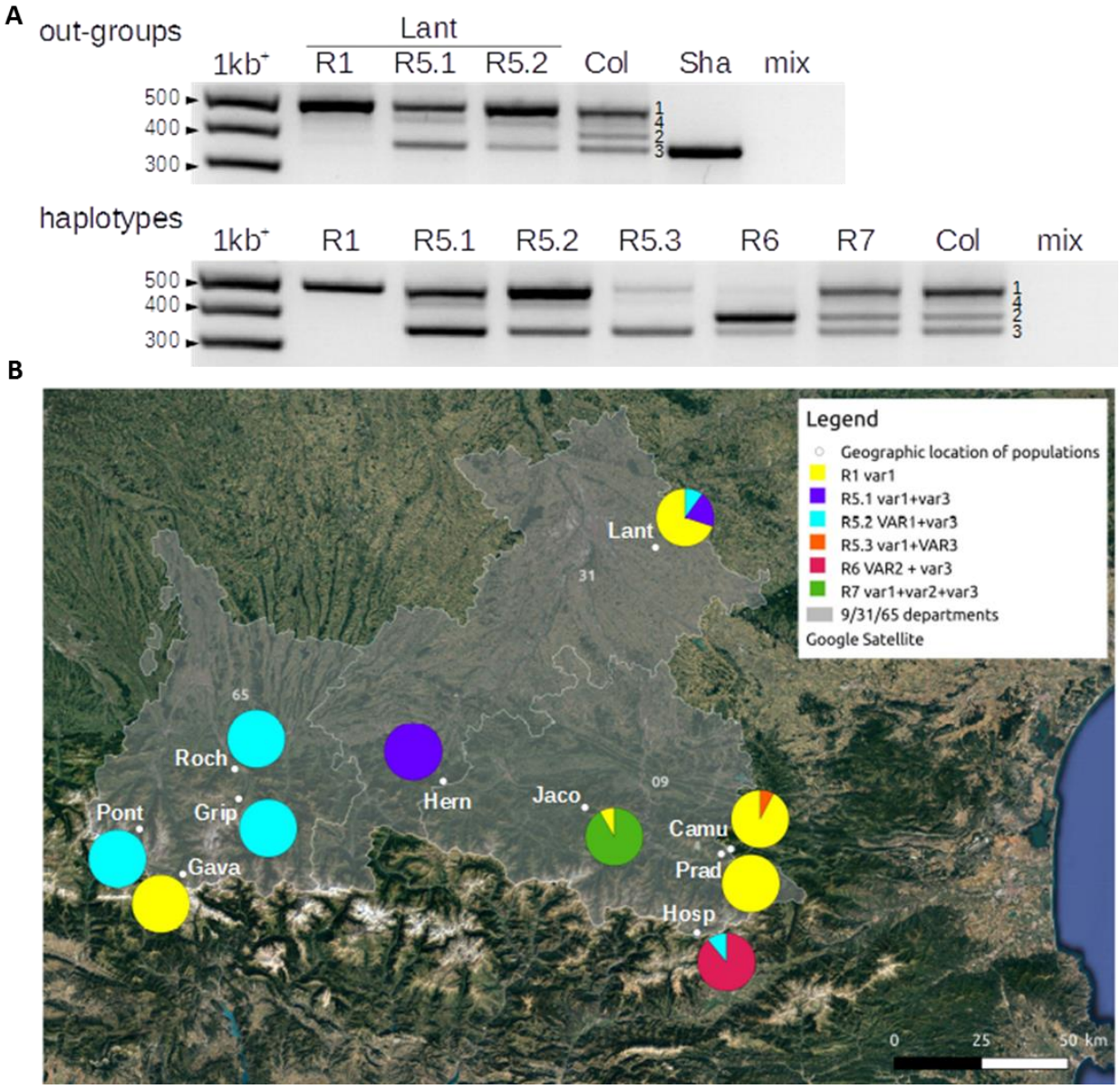


Figure 3. 3'ETS 45S rDNA haplotypes in the 9 natural populations (Camu, Gava, Grip, Hern, Hosp, Jaco, Pont, Prad and Roch). (A) PCR amplification of 3'ETS 45S rRNA gene sequences resuming the different haplotypes encountered in 9 Pyrenean populations (R1, R5, R6 and R7), the out-group Lant population close to Toulouse, France and the two control ecotypes, Col and Sha. Numbers show 3'ETS 45S rRNA gene variants (VAR1 to 4) based on expected sizes known in the control Col ecotype (Pontvianne et al., 2010 ; Abou-Ellail et al., 2011). mix indicates the control reaction without a genomic DNA template. (B) Camembert diagrams representing the proportion of these haplotypes in the 9 Pyrenean populations and the Lant out-group population. Capital and lowercase letters (VAR or var) indicate which variant is the most abundant in the haplotype when appropriate. The populations were georeferenced using Quantum GIS software 2.18 (http://qgis.osgeo.org).

Systematic (Tab 2), structuring Bayesian (Fig. 1) or Neighbor-Joining analyses (Fig. 2) highlight an unexpected natural and local genetic diversity within and between populations. Surprisingly indeed, these three approaches identified the same populations as homogeneous (Jaco, Grip, Hosp, Hern, Roch and Prad). Our study shows that the natural variability can be important even in highly self-fertilizing plants and must not be neglected in omics projects. Populations with genetic diversity and originated from contrasted environmental could present different characteristic phenotypes.

The germination speed varies according to natural variability.

The germination step is crucial in the adaptation of plant populations to different climatic conditions. To test the viability of these accessions, germination was tested in *Arabidopsis* standard growth conditions (22°C). To analyse the germination, only one individual per population was randomly chosen regardless the results of the genetic analyses. It should be noted that the Mong population was absent from the phenotyping in general, and from the germination tests in particular because only fresh material without seeds was initially collected in the Pyrenees Mountains. Unfortunately, its natural location was destroyed by human activity, preventing us from collect seeds.

The germination rate was constant among the 29 Pyrenean populations tested (higher than 90%, data not shown) but considerable variations in the germination speed were observed between them.

In *Arabidopsis thaliana,* it is known that the early steps of germination comprise the successive rupture of two protective seed envelopes: first the testa (the seed coat), then the endosperm (the inner nutritive tissue) (Fig. S3). The percentages of testa rupture (TR) and endosperm rupture (ER) were evaluated at 22°C, 24 and 30 h after imbibition, respectively (Fig. 4, Fig. S3 B-E), because the differences observed are maximised at these two time points. We observed that the germination speed varies considerably between populations: some like Gedr, Sha and Bedo having high percentage of TR/ER and others like Pont, Lant and Mari with low germination.

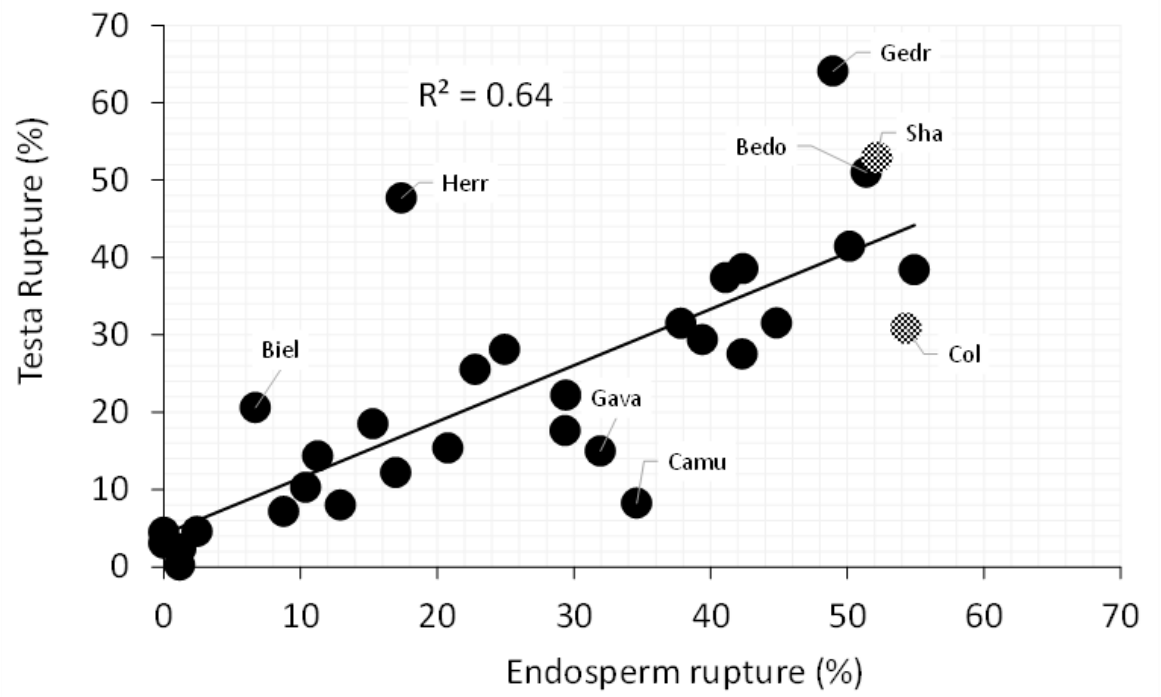


Figure 4. Correlation between testa and endosperm rupture in natural populations of A. thaliana seeds observed 24 and 30 h after imbibition. To compare with known ecotypes, Col and Sha are represented by checkered circles. Data are available in supporting Information S3.

We also observed that the percentage of testa rupture was proportional to the endosperm rupture for the populations ($R^2 = 0.64$), albeit not all of them (Fig. 4). Indeed, the TR/ER ratio was significantly higher for Herr, Biel, Gedr, Bedo and for the non-Pyrenean ecotype, Sha as compared to the other Pyrenean populations, meaning that in the former populations the testa layer could be weaker or that the embryo growth may be slower than the majority of the other ecotypes. Conversely, Gava and Camu showed a similar profile than the ecotype Col, with a percentage of ER higher than the TR. That could be explained by the resistance of the testa or the different forces (*e.g.* pressure, turgescence) applied by the embryo (Müller *et al.* 2006).

This early step in seed life is crucial for plant survival, facing natural climatic variations. The cell wall composition of the testa could be analysed to confront these results and to link germination rate to the natural growth environments. The control of germination can be a predominant phenotypic trait for plants to be more adapted to environmental constraints *in natura* with two different strategies. Rapid maturation allows the plant to maturate faster under unfavourable climatic conditions. Alternatively, slow germination rate allows the plant to control growth timing by waiting for better conditions. The differential growth speed observed

may explain the contrasted time to reach at the bolting stage and, in this way, makes possible to increase the chance to reproduce faster within diverse climatic constraints.

Cold growth conditions induced large range of phenotype response.

The same individuals, used for germination assays, were also chosen for phenotyping analysis at later steps. To better understand phenotypic variations in response to the abiotic gradient and specifically to temperature fluctuation, the phenotypes of *A. thaliana* growing at 22°C (optimal growth condition for Col) and 15°C (to reproduce altitudinal growth conditions and sub-optimal growth for Col) were compared. The rosette of 4- and 6-week old plants grown at 22°C and 15°C, respectively, corresponded to the bolting developmental stage for the well-known populations Col and Sha (Duruflé *et al.* 2017). Contrasted rosettes phenotypes growing at 15°C and at 22°C were observable (Fig. 5A and B). The rosette phenotyping was performed by counting the number of leaves (Fig. S5), weighting the rosettes (Fig. S6), measuring their diameter (Fig. S7), estimating the time to reach the bolting stage (Fig. S8), and assessing their anthocyanin (Fig. S9) and chlorophyll contents (Fig. S10). The floral stem phenotyping involved counting the number of cauline leaves on the main stem (Fig. S11), measuring the length (Fig. S12) and the diameter of the stem (Fig. S13), and also estimating the stem growth speed (Fig. S14). It should be noted that Pont appeared to be a "winter-annual" *Arabidopsis* ecotype (Chouard 1960; Gazzani *et al.* 2003 2005; Michaels & Amasino 1999) and only the rosette data were analysed. However, floral stems data of this population were too different to be compared with those of the other Pyrenean populations.

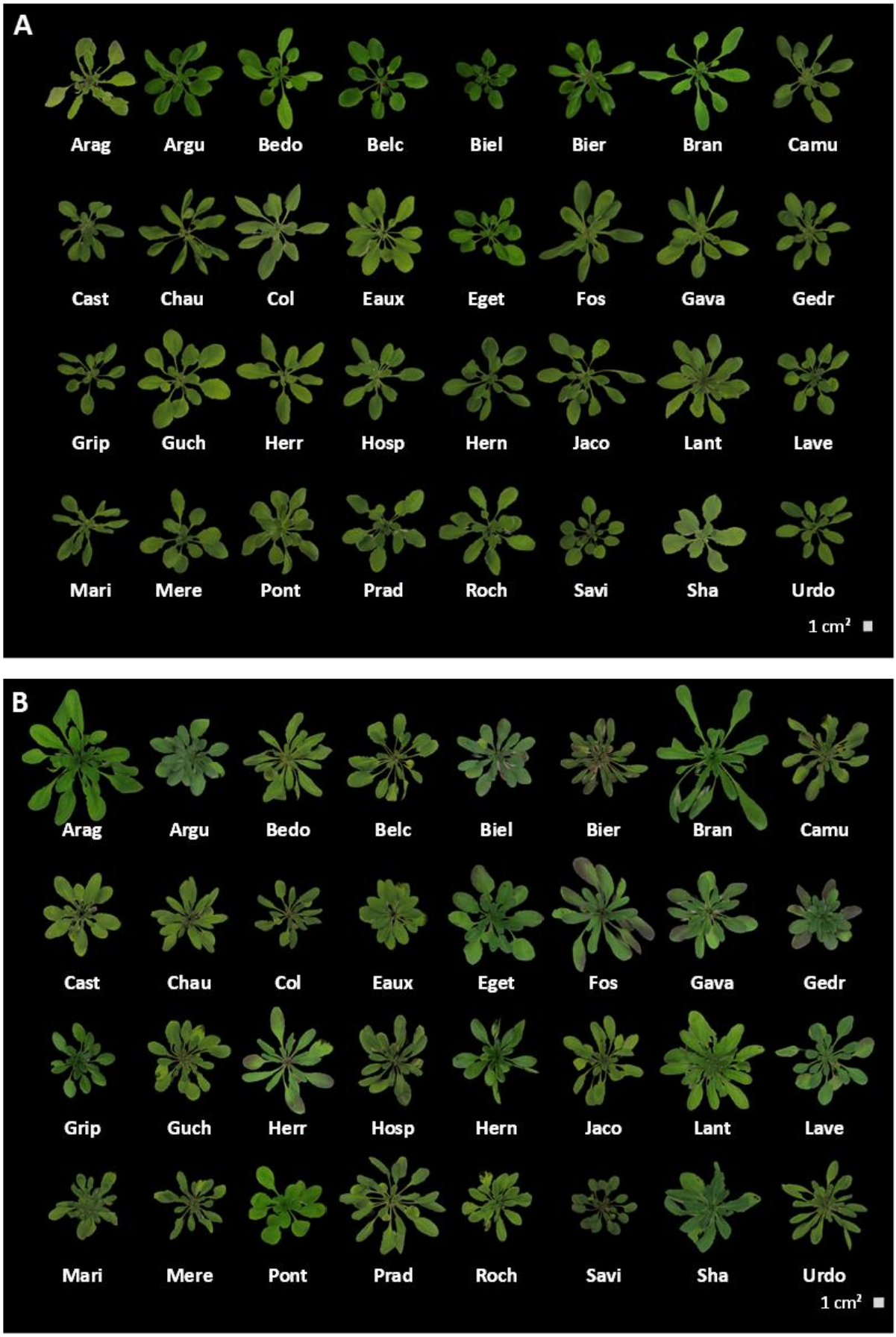


Figure 5. Intraspecific morphology variation in Arabidopsis thaliana. Photographic images of representative phenotypic differences of rosettes observed at (A) 22°C and (B) 15°C of the populations of A. thaliana in Pyrenees Mountains and the 3 out group populations Lant, Col and Sha. Photographs taken just before harvest.

In general, the number of leaves, the weight and the diameter of the rosettes were higher at 15°C than at 22°C, with the exception of Bran and the well-known Col and Sha ecotypes whose rosettes had a smaller diameter at 15°C as compared to 22°C (Duruflé *et al.* 2017) or Pont which had a reduced rosette weight at 15°C than at 22°C (Fig. S5A, 6A). The absence of clear correlation for all the populations between the weight of rosette and the number of leaves could be explained by different densities of the leaves. Indeed, leaves and petioles could be thicker due to a higher quantity of cell wall.

In parallel, increased stem diameter, longer stems, more cauline leaves were observed in plants grown at low temperature, with the notable exceptions of Hern and Prad (less cauline leaves), and Biel (shorter stems at 22°C than at 15°C). The stems growth speed was also significantly positively affected by low temperature. For example, Eaux, Fos and Chau stems grew slower at 22°C than at 15°C (Fig. S14).

Anthocyanin accumulation in leaves is considered as a general stress marker in plants (Mori *et al.* 2009) but in given situations such as for highland species, this could be a protection against UV radiations (Close & Beadle 2003). To evaluate the individual level of resistance to stress of *A. thaliana* Pyrenean plants grown at different temperatures, we quantified anthocyanin in the leaves of individuals cultivated either at 15°C or at 22°C. Most of the populations (e.g. Argu, Jaco, Eaux and Bier) showed an increase of anthocyanin content of rosettes grown at 15°C as compared to 22°C (Fig. S9), reflecting difficulties for these populations to withstand the more stressful cooler growth conditions. However, as expected from its natural localization (1.456 m), the Pont population (and the non-Pyrenean altitudinal control Sha) is well adapted to grow at low temperature as demonstrated by the fact that it does not accumulate high level of anthocyanin, even when grown at 15°C.

Chlorophyll content is an indicator of photosynthesis activity as well as of the whole metabolism. Based on their relative chlorophyll content when grown at 15°C or 22°C, the Pyrenean populations could be divided into two groups: those with significantly more chlorophyll at 15°C than at 22°C (e.g. Mere, Chau, Eget, Bran, Biel) and those with more chlorophyll at 22°C than at 15°C (e.g Cast, Hosp, Prad, Roch, Hern). It should be noted that all the populations have about the same level at 15°C (Fig. S10D). This phenomenon can be due to a compensatory effect to face the reduction of the metabolic activity at 15°C (Strand *et al.* 1997) and the variability of Rubisco catalysis present *in natura* (Stitt & Schulze 1994; Walker *et al.* 2013). This natural variability observed at 22°C might prove useful for acclimation to the global warming context.

In this study, some Pyrenean populations show reduction of the initiation of flowering time characterized by early bolting stage at 22°C compared to Col and Sha. Indeed with references to Col and Sha, a group of precocious populations can be viewable such as Eaux, Cast, Fos, Chau or also Gava (Fig. S8). It appeared that populations with a later bolting stage at 22°C showed a flowering time equal or advanced at 15°C like Prad, Argu and Belc. Activity of the regulators of the floral repressor FLC (flowering locus C), FCA and FVE (flowering locus CA and VE) pathways or the SVP (short vegetative phase) gene could explain this natural variation of flowering time (Blázquez *et al.* 2003; Lee *et al.* 2007; Simpson & Dean 2002) with different ambient temperatures.

Multilevel analyses reveal different phenotype diversity of plasticity

All the phenotypes can be analysed independently to explain the acclimation to the different growth temperatures. But as the 10 traits measured were variable between the two growth conditions and between the populations/ecotypes, PCA was performed to obtain more conclusive results. Multilevel methods of PCA can be used if repeated measurements from different treatments are applied on the same individual. This multilevel multivariate approach was developed to highlight treatment effects within subject separately from the biological variation between subjects (Liquet *et al.* 2012). Under these conditions, we propose an overview with a multilevel PCA of the phenotypic data (Fig. 6A, B).

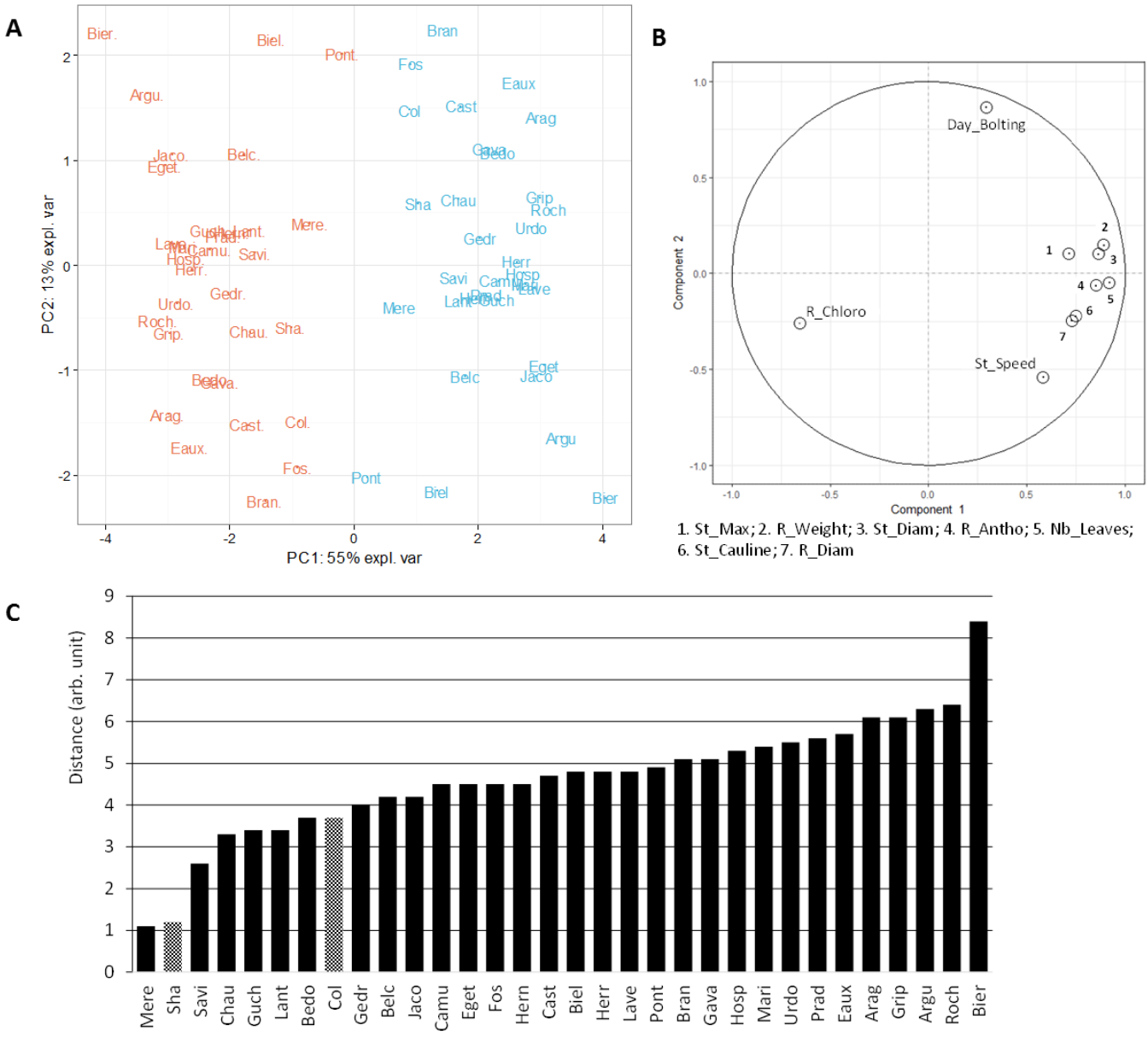


Figure 6. Overview of the phenotype observed at 22°C and 15°C. (A) Multilevel scaled PCA of the phenotypes observed at 22°C (in orange) and at 15°C (in blue) on the day of the harvest. Values for x and y axes are those of PC1 and PC2, respectively. (B) Two dimensional canonical graph for a multilevel normed PCA (correlation circle), the positions of the variables show the quality of the correlation between them. To visualize correctly the variables highly correlated, some of them are combined in a same box. (C) Distance separating the populations according to the temperature in the scaled PCA multilevel. To compare with known ecotypes, Col and Sha are represented with pattern. R_Chloro: chlorophyll content in rosette; R_antho: Anthocyanin content in rosette; Day_bolting: time to growth at the stage 5.1; R_weight: fresh weight of the rosette; Nb_leaves: number of leaves of the rosette; R_diam: rosette diameter; St_diam: diameter of the floral stems; St_speed: calculated stem speed; St_cauline: number of lateral stems/cauline leaves. St_max: length of the stems

The objective of this multivariate approach was to characterize the phenotypic diversity between populations. In this analysis, populations are clearly separated by the temperature effect in the first axis (55%). This separation is mainly due to the chlorophyll overall content which is more abundant in rosette from plant cultivated at 22°C as compared to 15°C, and to the other phenotypic parameters generally higher at 15°C than 22°C in most of the populations. Chlorophyll content appears to be a good proxy to characterize the acclimation of populations to different temperature growth conditions. Indeed, highland population like Prad and Hosp appears with more chlorophyll content at 22°C than 15°C in contrario to the lowland population Chau and Biel. The second axis differentiates Pyrenean populations by the time to reach the bolting stage, a characteristic that is negatively correlated with the floral stem speed. For example, Bran, Fos and Eaux reach the bolting stage earlier when grown at 22°C vs. 15°C and have a faster stem growth speed than the other Pyrenean populations when cultivated at 22°C. These parameters (time to reach bolting stage and stem growth speed) appear to be independent from the developmental stages and could be used to compare different populations.

Contrasted temperature growth conditions highlights variable phenotypic plasticity in more or less genetically heterogeneous populations. The global acclimation difference between the populations can be visualized by the distance separating the populations according to the temperature in multilevel PCA (Fig. 6C). A longer distance between the two temperature growth conditions could mean that the phenotype is more plastic and then more variable. The Pyrenean population Roch, Grip, Prad and Eaux showed an elevated phenotypic plasticity and also a low genetic variability (homogeneous populations). In contrast, the populations Mere, Savi, Chau, Guch, Lant and Bedo showed less phenotypic plasticity and a large genetic variability. Since, only one individual per population has been phenotyped, it can be hazardous to conclude about phenotypic plasticity for a whole population. But, we hope this research will help to predict the evolution of plastic responses to environmental change (Auge *et al.* 2017; van Kleunen & Fischer 2005; Vitasse *et al.* 2010).

This phenotypic analysis showed high intraspecific morphology variation that could be explained by the natural diversity of acclimation to contrasted temperature variations (Hamilton *et al.* 2015). Cell wall is crucial in growth control, for the structural integrity of the global plant shape (Sasidharan *et al.* 2011) and contributes to plant architecture variation in response to environmental changes (Houston *et al.* 2016; Le Gall *et al.* 2015). These morphological variations may be due, in part, to cell wall modifications (Duruflé *et al.* 2017).

Relationship between phenotypic genetic and environmental data.

Kruskal-Wallis nonparametric tests were used to evaluate the effect of genetic clusters or environmental data, on phenotypic traits variation (Fig. 7A, B) and discover potential correlations between them.

In the populations assigned to 3 distinct genetic clusters by the STRUCTURE analysis, rosettes and floral stems phenotypes are impacted. Indeed, of the Pyrenean populations considered as homogeneous (cluster 1 and 2) have systematically wider stems than the hybrid populations (cluster 3) at both growth temperature conditions (Fig. 7C). We hypothesize that this trait might allow homogeneous populations to resist to strong winds in order to sustain seeds dissemination. Moreover, rosettes of populations from the cluster 3 are smaller than those of populations from cluster 1 at the same temperature. At 22°C, the same tendencies are visible, albeit not significant despite the low variability of the data in this genetic cluster (Fig. 7D). Because populations of the cluster 1 are specific of the Pyrenean Mountain, one phenotypic characteristics of *A. thaliana* in the Pyrenees is to have bigger rosette at low temperature growth conditions.

A

Accession name	Genetics Cluster	Environmental (Climate PC1)
Bedo	1	Plain
Biel	3	Plain
Cast	2	Plain
Chau	2	Plain
Col	2	Plain
Herr	1	Plain
Lant	2	Plain
Argu	3	Medium
Belc	3	Medium
Bier	3	Medium
Bran	2	Medium
Eaux	1	Medium
Fos	3	Medium
Gedr	3	Medium
Guch	2	Medium
Hern	1	Medium
Jaco	1	Medium
Lave	2	Medium
Mari	2	Medium
Roch	1	Medium
Savi	3	Medium
Urdo	3	Medium
Arag	3	High
Camu	1	High
Eget	2	High
Gava	2	High
Grip	1	High
Hosp	1	High
Mere	1	High
Pont	2	High
Prad	1	High
Sha	2	High

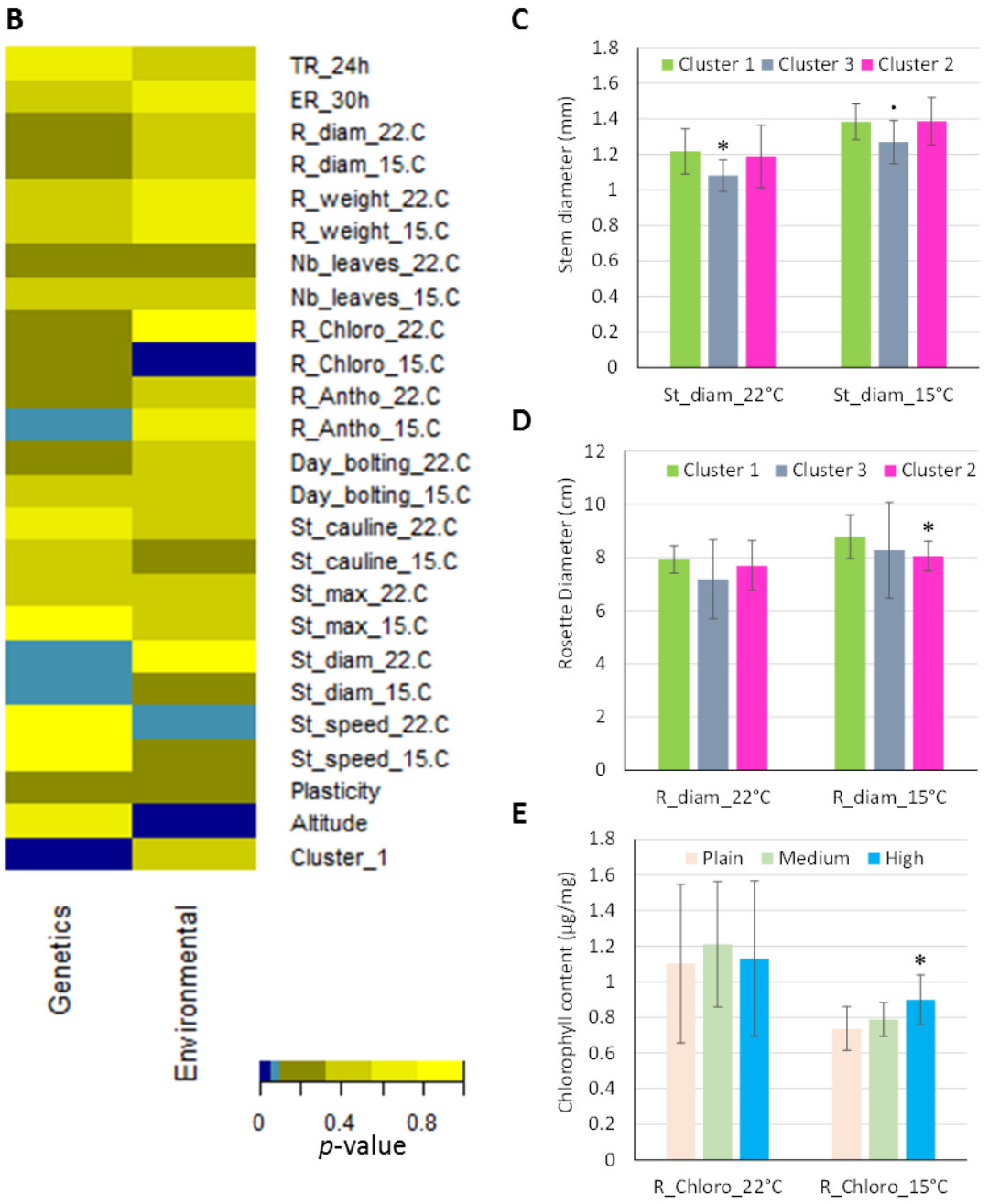


Figure 7. Relationships between phenotypic data observed at 22°C and 15°C, genetic and environmental classification of the populations. (A) Classification of the populations in function of the genetics allocation into their estimate membership proportions in each genetics cluster determined by STRUCTURE (see Fig.S2C) and environmental classification of the populations according to their environmental conditions considering as plain, medium and high altitudes (Climate PC1 rank). (B) Results of the Kruskal-Wallis nonparametric test between variables and classification in function of the genetics and the environmental groups. Colour code represents the p-value of the test (Blue light: $P \leq 0.1$; Dark blue: $P \leq 0.5$). (C) Measurements of the diameter of stems depending upon the genetics clusters. (D) Measurements of the diameter of rosettes depending to the genetics clusters. (E) Measurements of the chlorophyll content in rosettes depending to the environmental classification. Mean (± SEM) and associate to the significant differences between data are shown with Wilcoxon test: '*', $P \leq 0.05$; '.', $P \leq 0;1$).

TR_24h: Testa rupture at 24 h; ER_30h: Endosperm rupture after 30 h; R_diam: rosette diameter; R_weight: fresh weight of the rosette; Nb_leaves: number of leaves of the rosette; R_Chloro: chlorophyll content in rosette; R_antho: Anthocyanin content in rosette; Day_bolting: time to growth at the bolting stage; St_cauline: number of lateral stems/cauline leaves; St_max: length of the stems; St_diam: diameter of the floral stems; St_speed: calculated stem speed; Plasticity: Distance separating the populations according to the scaled PCA multilevel; Cluster_1: Assignment at the genetics Cluster 1 determined by STRUCTURE.

Finally, populations were also distributed in 3 distinct environmental conditions (namely “plain”, “medium” and “high” altitudes) in agreement with the climate PC1 value (Fig. 7A). Indeed, clear correlations exists between phenotypic traits and the original environmental conditions characterized by the climate PC1 value. For example, populations living in high altitude show significantly higher chlorophyll content at 15°C compared to the other populations. However, this was not observed at 22°C (Fig. 7E). Acclimation of the highland populations to low temperature environmental conditions could be due to these phenotypic characteristics. Indeed, chlorophyll content serves as an indicator for the overall availability of photoreaction centres (Ballottari *et al.* 2007), at high altitude this accumulation may play a role in cold acclimation.

No clear correlation can be established between genetic cluster assignment and environmental data (Fig. 7B). However, Fig. 7A highlights the case that hybrid populations are mostly located at medium altitude

In summary, the present characterization of 30 populations of *A. thaliana* using 5 CIII Prx markers has allowed to estimate the population structure and the genetic relationship between these populations within the Pyrenees Mountains. Genetic structure as well as inter- and intra-population variation emphasized the unexpected variability found in this region. Regarding the various genome sequencing projects (1001 Genomes Consortium): 9 Pyrenean populations appear genetically close and homogeneous to other ecotypes known in the region; 11 Pyrenean populations appear genetically close, homogeneous and specific of these mountains and 9 Pyrenean populations appear as hybrids of these two genetics clusters.

This study also revealed phenotypic variations in acclimation of *A. thaliana* across abiotic gradient characterized here by the temperature. Some of them are correlated with identified genetic clusters or with environmental data. These analyses contribute to enrich knowledge on abiotic stress acclimation in natural plant populations. Preliminary analyses suggest that the phenotypic acclimation may be due to the cell wall modification affected by low temperature (Duruflé *et al.* 2017). Transcriptomic, proteomics and cell wall polysaccharides analysis could prove helpful in study cell wall plasticity.

Plant adaptation to the environment is a complex process in the context of global warming. Is not possible to conclude that the most adapted phenotype to the temperature

elevation is not always the most plastic phenotype. Because plasticity have a cost, research from interdisciplinary work could prove helpful to select the best crop genotype in future.

Acknowledgments

The authors are thankful to the Paul Sabatier-Toulouse 3 University and to the Centre National de la Recherche Scientifique (CNRS) for granting their work. This work was also supported by the French Laboratory of Excellence project "TULIP" (ANR-10-LABX-41) and IdEx (ANR-11-IDEX-0002-02). HD is supported by a Midi-Pyrénées Region and the Federal University of Toulouse. We also grateful to the National Botanical Conservatory of Pyrenees and the Midi-Pyrénées Region (CBNPMP) and to the National Park of the Pyrenees (PNP) to provide us with population coordinates; thanks to the PNP for sampling authorisation within the park limitations.. Thanks to Etienne Florence for their help during the sampling of populations, to Nicolas Pouilly (LIPM, Toulouse) for advice and assistance with the large scale extraction of *Arabidopsis* DNA, and to Sarah Pignoly (IRD, Marseille) for initial sequences management. The authors also thanks Sophie Brando and Marie Morlans for their technical assistance.

Data accessibility

Sampling locations are available in the Table 1 and seeds will be available after request.

Sequence data used in this study for the polymorphism analysis are available in the Table S2.

Supporting Information:

Table S1. Details of the populations used for the STRUCTURE analyses from the 1001 Genomes consortium (Weigel, 2009).

Table S2. Concatenated sequences (fasta format) of the 375 individual for the 5 CIII Prx loci (*AtPrx09*, *AtPrx25*, *AtPrx36*, *AtPrx48*, *AtPrx62*) belonging to the 351 Pyrenean populations, the 22 populations of the 1,001 genomes and the ecotype Col and Sha. Each CIII Prx sequence is separated to each other by 5 "N". The name of the population is followed by the specific number of the individual and to the haplotype class to each sequence.

Table S3. Haplotype class information for each individual belonging to the different Pyrenean populations

Author Contributions:

Harold Duruflé: performed the experiments, the macro-phenotyping, the genetics and the statistical analyses, analysed the data and wrote the article.

Philippe Ranocha: performed some of the genetics analyses, retrieved the climatic data and helped with the writing.

Duchesse Lacour Mbadinga Mbadinga: performed most of the macro-phenotyping analyses.

Sébastien Déjean: designed the statistical analyses and analysed the data.

Maxime Bonhomme: supervised the genetics analyses and analysed the data.

Hélène San Clement: performed genetic analyses.

Valérie Hinoux: performed rRNA genetic analyses.

Julio Saez-Vasquez: performed rRNA genetic analyses.

Jean-Philippe Reichheld: supervised the performed rRNA genetic analyses.

Nathalie Escaravage: collected of natural population and the supervised the genetics analyses.

Monique Burrus: collected of natural population and the supervised the genetics analyses.

Christophe Dunand: conceived the project, supervised and designed the experiments, and wrote the article.

LITERATURE CITED

Abou-Ellail M, Cooke R, Sáez-Vásquez J (2011) Variations in a team: major and minor variants of *Arabidopsis thaliana* rDNA genes. *Nucleus* **2**, 294-299.

Alonso-Blanco C, Mendez-Vigo B, Koornneef M (2005) From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *Int J Dev Biol* **49**, 717-732.

Auge GA, Leverett LD, Edwards BR, Donohue K (2017) Adjusting phenotypes via within- and across-generational plasticity. *New Phytol*.

Bakker EG, Stahl EA, Toomajian C, *et al.* (2006) Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Mol Ecol* **15**, 1405-1418.

Ballottari M, Dall'Osto L, Morosinotto T, Bassi R (2007) Contrasting behavior of higher plant photosystem I and II antenna systems during acclimation. *J Biol Chem* **282**, 8947-8958.

Bartoli C, Frachon L, Barret B, Rigal M, Zanchetta C, Bouchez O, Carrere S, Roux F (2017) *In situ* relationships between microbiota and potential pathobiota in *Arabidopsis thaliana* ISME J *submited.*

Beck JB, Schmuths H, Schaal BA (2008) Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol Ecol* **17**, 902-915.

Blázquez MA, Ahn JH, Weigel D (2003) A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. *Nat Genet* **33**, 168-171.

Boyes DC, Zayed AM, Ascenzi R, *et al.* (2001) Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* **13**, 1499-1510.

Brennan AC, Méndez-Vigo B, Haddioui A, *et al.* (2014) The genetic structure of *Arabidopsis thaliana* in the south-western Mediterranean range reveals a shared history between North Africa and southern Europe. *BMC Plant Biol* **14**, 17.

Chandrasekhara C, Mohannath G, Blevins T, Pontvianne F, Pikaard CS (2016) Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in Arabidopsis. *Genes Dev* **30**, 177-190.

Chouard P (1960) Vernalization and its relations to dormancy. *Annual Review of Plant Physiology* **11**, 191-238.

Clauss MJ, Mitchell-Olds T (2006) Population genetic structure of *Arabidopsis lyrata* in Europe. *Mol Ecol* **15**, 2753-2766.

Close DC, Beadle CL (2003) The ecophysiology of foliar anthocyanin. *The Botanical Review* **69**, 149-161.

Consortium G (2016) 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481-491.

Cosio C, Dunand C (2010) Transcriptome analysis of various flower and silique development stages indicates a set of class III peroxidase genes potentially involved in pod shattering in *Arabidopsis thaliana*. *BMC Genomics* **11**, 528.

Davison J, Tyagi A, Comai L (2007) Large-scale polymorphism of heterochromatic repeats in the DNA of *Arabidopsis thaliana*. *BMC Plant Biol* **7**, 44.

Duruflé H, Hervé V, Ranocha P, *et al.* (2017) Cell wall modifications of two *Arabidopsis thaliana* ecotypes, Col and Sha, in response to sub-optimal growth conditions: an integrative study. *Plant Science*.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Fournier-Level A, Korte A, Cooper MD, *et al.* (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**, 86-89.

Francoz E, Ranocha P, Nguyen-Kim H, *et al.* (2015) Roles of cell wall peroxidases in plant development. *Phytochemistry* **112**, 15-21.

Fukunaga K, Hill J, Vigouroux Y, *et al.* (2005) Genetic diversity and population structure of teosinte. *Genetics* **169**, 2241-2254.

Gazzani S, Gendall AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. *Plant Physiol* **132**, 1107-1114.

Gulsen O, Kaymak S, Ozongun S, Uzun A (2010) Genetic analysis of Turkish apple germplasm using peroxidase gene-based markers. *Scientia Horticulturae* **125**, 368-373.

Hall T (2011) BioEdit: an important software for molecular biology. *GERF Bull Biosci* **2**, 60-61.

Hamilton JA, Okada M, Korves T, Schmitt J (2015) The role of climate adaptation in colonization success in *Arabidopsis thaliana. Mol Ecol* **24**, 2253-2263.

Hancock AM, Brachi B, Faure N, *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83-86.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* **25**, 1965--1978.

Horton MW, Hancock AM, Huang YS, *et al.* (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* **44**, 212-216.

Houston K, Tucker MR, Chowdhury J, Shirley N, Little A (2016) The plant cell wall: a complex and dynamic structure as revealed by the responses of genes under stress conditions. *Front Plant Sci* **7**, 984.

Huld T, Muller R, Gambardella A (2012) A new solar radiation database for estimating PV performance in Europe and Africa. *Solar Energy* **86**, 1803-1815.

Krislov Morris A, Kuhn Silk W (1992) Use of a flexible logistic function to describe axial growth of plants. *Bulletin of Mathematical Biology* **54**, 1069 - 1081.

Körner C (2007) The use of 'altitude' in ecological research. *Trends Ecol Evol* **22**, 569-574.

Lariguet P, Ranocha P, De Meyer M, *et al.* (2013) Identification of a hydrogen peroxide signalling pathway in the control of light-dependent germination in Arabidopsis. *Planta* **238**, 381-395.

Layat E, Sáez-Vásquez J, Tourmente S (2012) Regulation of Pol I-transcribed 45S rDNA and Pol III-transcribed 5S rDNA in Arabidopsis. *Plant Cell Physiol* **53**, 267-276.

Le Corre V (2005) Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Mol Ecol* **14**, 4181-4192.

Le Gall H, Philippe F, Domon JM, *et al.* (2015) Cell wall metabolism in response to abiotic stress. *Plants* **4**, 112-166.

Lee JH, Yoo SJ, Park SH, *et al.* (2007) Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. *Genes Dev* **21**, 397-402.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Liquet B, Lê Cao KA, Hocini H, Thiébaut R (2012) A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. *BMC Bioinformatics* **13**, 325.

Lièvre M, Granier C, Guédon Y (2016) Identifying developmental phases in the *Arabidopsis thaliana* rosette using integrative segmentation models. *New Phytol* **210**, 1466-1478.

Long Q, Rabanal FA, Meng D, *et al.* (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat Genet* **45**, 884-890.

Lê Cao KA, Boitard S, Besse P (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**, 253.

Mancinelli AL, Hoff AM, Cottrell M (1988) Anthocyanin production in chl-rich and chl-poor seedlings. *Plant Physiol* **86**, 652-654.

Meehl GA, Stocker TF, Collins WD, *et al.* (2007) Global climate projections.

Michaels SD, Amasino RM (1999) The gibberellic acid biosynthesis mutant ga1-3 of *Arabidopsis thaliana* is responsive to vernalization. *Dev Genet* **25**, 194-198.

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. *Nature* **441**, 947-952.

Mori IC, Utsugi S, Tanakamaru S, *et al.* (2009) Biomarkers of green roof vegetation: anthocyanin and chlorophyll as stress marker pigments for plant stresses of roof environments. *Journal of Environmental Engineering and Management* **19**, 21-27.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio-assay with tobacco tissue culture. *Physiol Plant* **15**, 473-497.

Müller K, Tintelnot S, Leubner-Metzger G (2006) Endosperm-limited Brassicaceae seed germination: abscisic acid inhibits embryo-induced endosperm weakening of *Lepidium sativum* (cress) and endosperm rupture of cress and *Arabidopsis thaliana*. *Plant Cell Physiol* **47**, 864-877.

Nemli S, Kaya HB, Tanyolac B (2014) Genetic assessment of common bean (*Phaseolus vulgaris* L.) accessions by peroxidase gene-based markers. *J Sci Food Agric* **94**, 1672-1680.

Nordborg M, Hu TT, Ishino Y, *et al.* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* **3**, e196.

Passardi F, Longet D, Penel C, Dunand C (2004) The class III peroxidase multigenic family in rice and its evolution in land plants. *Phytochemistry* **65**, 1879-1893.

Picó FX, Méndez-Vigo B, Martínez-Zapater JM, Alonso-Blanco C (2008) Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian peninsula. *Genetics* **180**, 1009-1021.

Pinar H, Unlu M, Ercisli S, Uzun A, Bircan M (2016) Genetic analysis of selected almond genotypes and cultivars grown in Turkey using peroxidase-gene-based markers. *Journal of Forestry Research* **27**, 747-754.

Pontvianne F, Abou-Ellail M, Douet J, *et al.* (2010) Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in Arabidopsis thaliana. *PLoS Genet* **6**, e1001225.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Sasidharan R, Voesenek LACJ, Pierik R (2011) Cell wall modifying proteins mediate plant acclimatization to biotic and abiotic stresses. *Critical reviews in plant sciences* **30**, 548-562.

Simpson GG, Dean C (2002) Arabidopsis, the Rosetta stone of flowering time? *Science* **296**, 285-289.

Stenøien HK, Fenster CB, Tonteri A, Savolainen O (2005) Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe. *Mol Ecol* **14**, 137-148.

Stitt M, Schulze D (1994) Does Rubisco control the rate of photosynthesis and plant growth? An exercise in molecular ecophysiology. *Plant, Cell & Environment* **17**, 465-487.

Strand A, Hurry V, Gustafsson P, Gardeström P (1997) Development of *Arabidopsis thaliana* leaves at low temperatures releases the suppression of photosynthesis and photosynthetic gene expression despite the accumulation of soluble carbohydrates. *Plant J* **12**, 605-614.

Suter L, Rüegg M, Zemp N, Hennig L, Widmer A (2014) Gene regulatory variation mediates flowering responses to vernalization along an altitudinal gradient in Arabidopsis. *Plant Physiol* **166**, 1928-1942.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512-526.

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* **101**, 11030-11035.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729.

Uzun A, Gulsen O, Seday U, Yesiloglu T, Kacar YA (2014) Peroxidase gene-based estimation of genetic relationships and population structure among *Citrus spp.* and their relatives. *Genetic Resources and Crop Evolution* **61**, 1307-1318.

van Kleunen M, Fischer M (2005) Constraints on the evolution of adaptive phenotypic plasticity in plants. *New Phytol* **166**, 49-60.

Vitasse Y, Bresson CC, Kremer A, Michalet R, Delzon S (2010) Quantifying phenological plasticity to temperature in two temperate tree species. *Functional Ecology* **24**, 1211-1218.

Walker B, Ariza LS, Kaines S, Badger MR, Cousins AB (2013) Temperature response of in vivo Rubisco kinetics and mesophyll conductance in *Arabidopsis thaliana*: comparisons to *Nicotiana tabacum*. *Plant, Cell & Environment* **36**, 2108-2119.

Warren GJ (1998) Cold stress: manipulating freezing tolerance in plants. *Curr Biol* **8**, R514-516.

Wilczek AM, Cooper MD, Korves TM, Schmitt J (2014) Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **111**, 7906-7913.

Wolfe MD, Tonsor SJ (2014) Adaptation to spring heat and drought in northeastern Spanish *Arabidopsis thaliana*. *New Phytol* **201**, 323-334.

Table S1. Details of the populations used for the STRUCTURE analyses from the 1001 Genomes consortium (Weigel, 2009).

name	*Accession name*	*Country*	*Latitude*	*Longitude*
Ved	VED-10	France	43.74	3.89
Lec	LEC-25	France	43.91	4.14
Iss	ISS-20	France	43.92	3.71
Mou	MOU2-25	France	43.98	4.31
Arr	ARR-17	France	44.05	3.69
Qui	QUI-8	France	44.07	4.08
Bez	BEZ-9	France	44.12	3.77
Noz	NOZ-6	France	44.12	4.33
Et	Et-0	France	44.6447	2.56481
Pyl	PYL-6	France	44.65	-1.16667
Ag	Ag-0	France	45	1.3
Vie	Vie-0	Spain	42.63	0.76
Ber	Ber-0	Spain	42.52	-0.56
Bis	Bis-0	Spain	42.49	0.54
Coc	Coc-1	Spain	42.31	3.19
Moa	Moa-0	Spain	42.46	0.7
Orb	Orb-10	Spain	42.97	-1.23
Pal	Pal-0	Spain	42.34	1.3
Pan	Pan-0	Spain	42.76	-0.23
Ria	Ria-0	Spain	42.34	2.17
Tol	Tol-7	Spain	42.11	0.6
Vdm	Vdm-0	Spain	42.04	1.01

Figure S1. (A) Primers set used in this study and (B) positions of the amplified sequence on the genes. Grey boxes represent exons, empty boxes introns, and bold lines 5' UTR.

A

Gene ID	Gene Name	Forward primer sequence	Reverse primer sequence	Amplicon length (bp)
At1g44970	AtPrx09	GCATTGAAACAAGCTATATTTGC	AGGAGACAGTTTGAGGACAAGC	871
At2g41480	AtPrx25	CTACAAGTGAGGCTGATGTGAAC	AACAGAGTCACGAGCAGCAA	871
At3g50990	AtPrx36	GTAATGTTAGCACACGGGTGG	ATCGGTGAGATCAAGTCCTTG	730
At4g33870	AtPrx48	GTGATGTAGGAGACAAGTACACC	AAAAATCTACTGACCAAGCATAATTAC	825
At5g39580	AtPrx62	GTAGATGAAAATATCATGCACTAATACG	GCCTTTGTTAATCCTCTTTACTTACTC	869

B

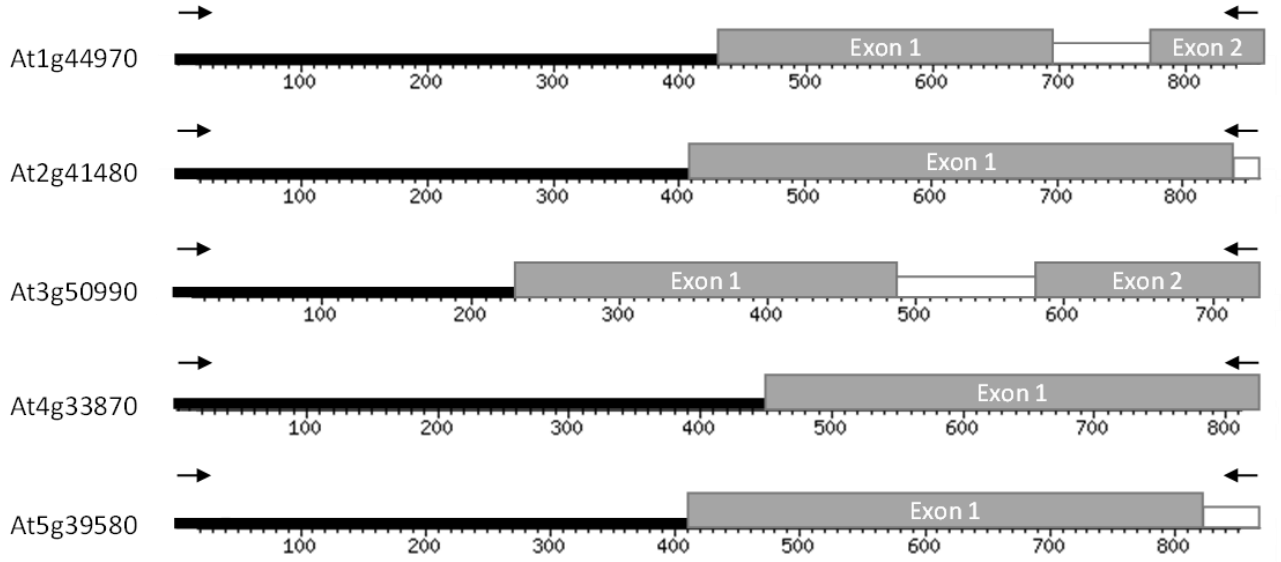


Figure S2. Detail of the steps for the graphical method allowing detection of the number of clusters K estimated with STRUCTURE software. (A) Mean L (K) (± SD) over 10 runs for each K value and (B) delta K calculated as described (Evanno, 2005). The modal value of this distribution is the optimal K or the uppermost level of structure, here K = 2. (C) Cluster assignment of Pyrenees populations based on majority membership proportion inferred from K = 2. Each population is represented by a single vertical bar divided into coloured segments representing the estimated membership proportions of genetic clusters (K) fitted in the model.

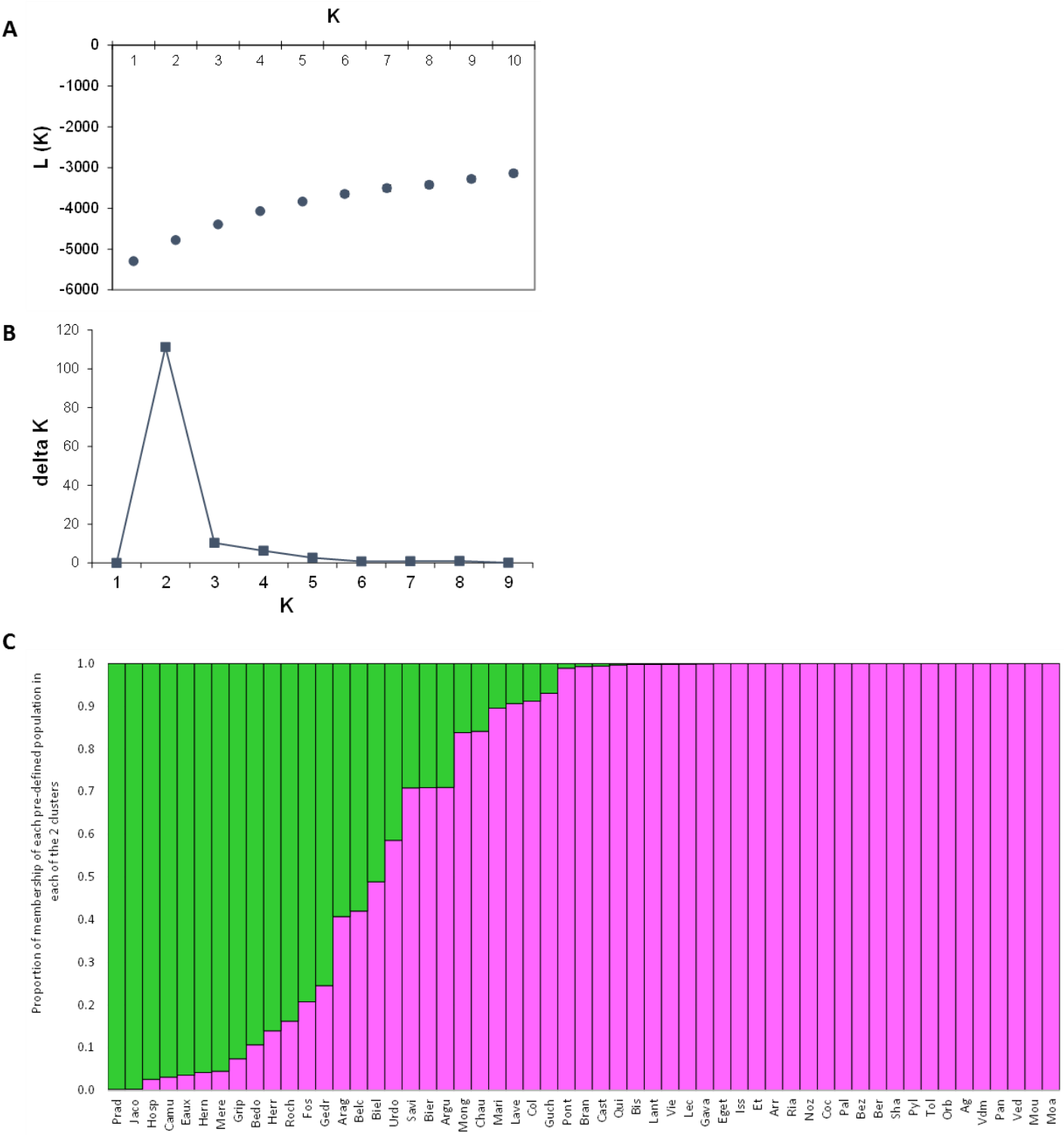
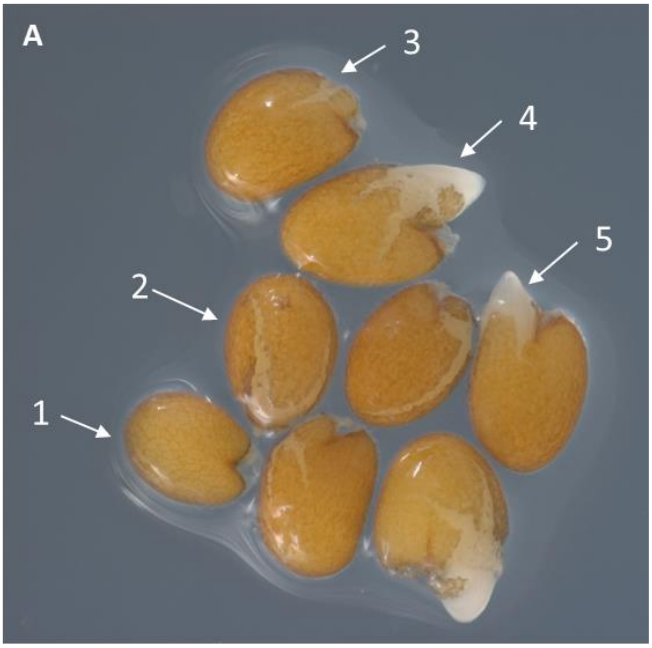
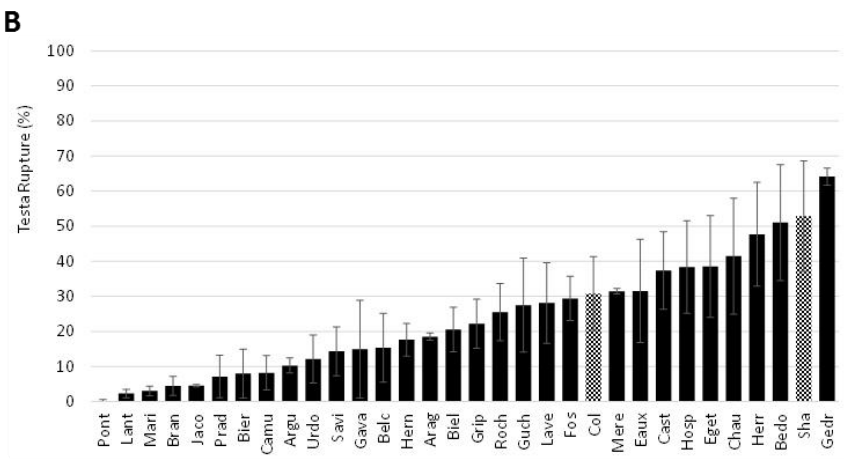
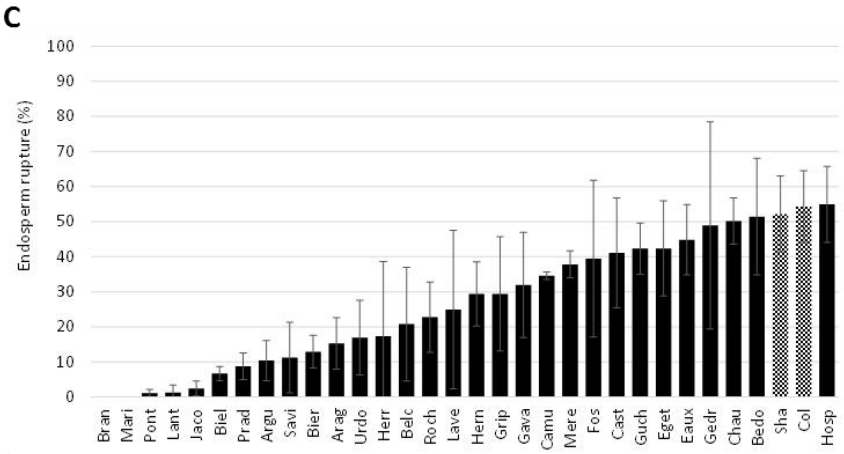


Figure S3. Natural variability of the germination rate .(A) Visualization of the different steps of *Arabidopsis* Col (well-known reference ecotype) seeds germination observed after 30 h post imbibition. (1) Not yet germinated seed, (2, 3) rupture of the testa and the (4, 5) endosperm. Rupture of (B) the testa and (C) the endosperm of *A. thaliana* seeds assessed respectively at 24 and 30 h post imbibition. To compare with known ecotypes, Col and Sha has represented with pattern. Populations were classified in ascending order of percentage of the mean of three independent experiments of testa rupture (B) and the endosperm rupture (C). Duncan's test were performed for each germination step: (B) testa rupture and (C) endosperm rupture. Different letters indicate significant differences between populations with $P \leq 0.01$.







D

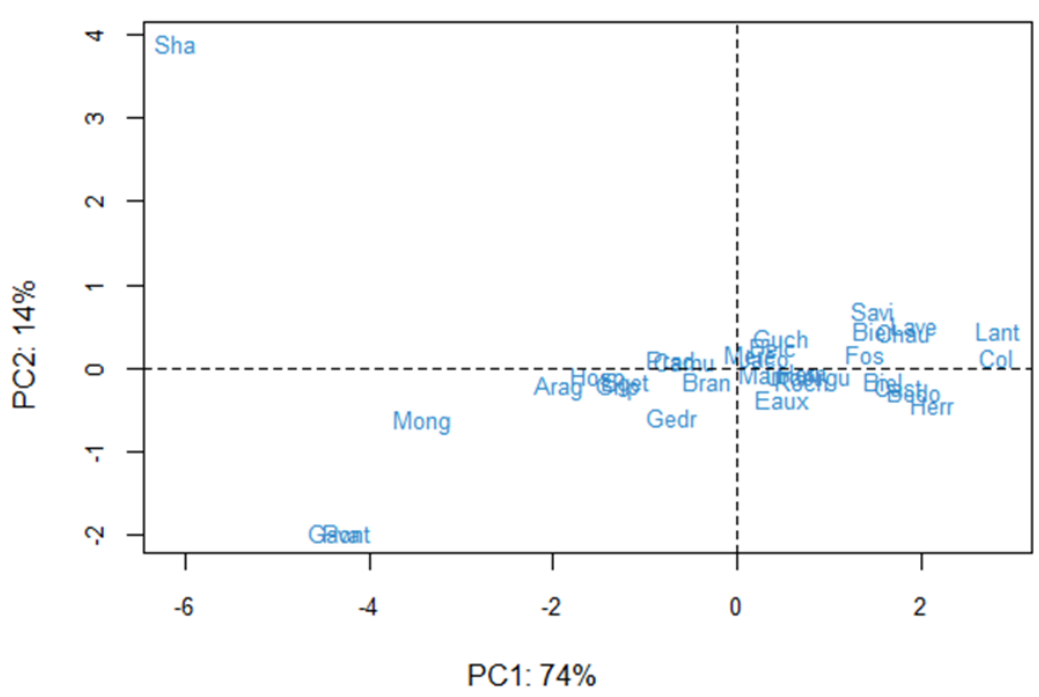
Population	Duncan
Gedr	a
Sha	ab
Bedo	abc
Herr	abcd
Chau	abcde
Eget	bcdef
Hosp	bcdef
Cast	bcdef
Eaux	bcdefg
Mere	bcdefg
Col	bcdefgh
Fos	bcdefghi
Lave	bcdefghij
Guch	bcdefghij
Roch	cdefghijk
Grip	defghijk
Biel	efghijk
Arag	efghijk
Hern	efghijk
Belc	efghijk
Gava	efghijk
Savi	fghijk
Urdo	fghijk
Argu	ghijk
Camu	ghijk
Bier	ghijk
Prad	ghijk
Jaco	ghijk
Bran	hijk
Mari	ijk
Lant	jk
Pont	k

E

Population	Duncan
Hosp	a
Col	a
Sha	ab
Bedo	ab
Chau	abc
Gedr	abcd
Eaux	abcde
Eget	abcdef
Guch	abcdef
Cast	abcdefg
Fos	abcdefgh
Mere	abcdefgh
Camu	abcdefghi
Gava	abcdefghij
Grip	abcdefghij
Hern	abcdefghij
Lave	abcdefghij
Roch	abcdefghij
Belc	bcdefghij
Herr	cdefghij
Urdo	defghij
Arag	efghij
Bier	efghij
Savi	fghij
Argu	fghij
Prad	ghij
Biel	hij
Jaco	ij
Lant	j
Pont	j
Bran	j
Mari	j

Figure S4. Overview of the bio-climatic variables of the populations of *A. thaliana*. (A) Normed PCA of bio-climatic variables. Values for x and y axes are those of PC1 and PC2, respectively. (B) Two dimensional canonical graph for a normed PCA (correlation circle), the positions of the variables show the quality of the correlation between them.

A



B

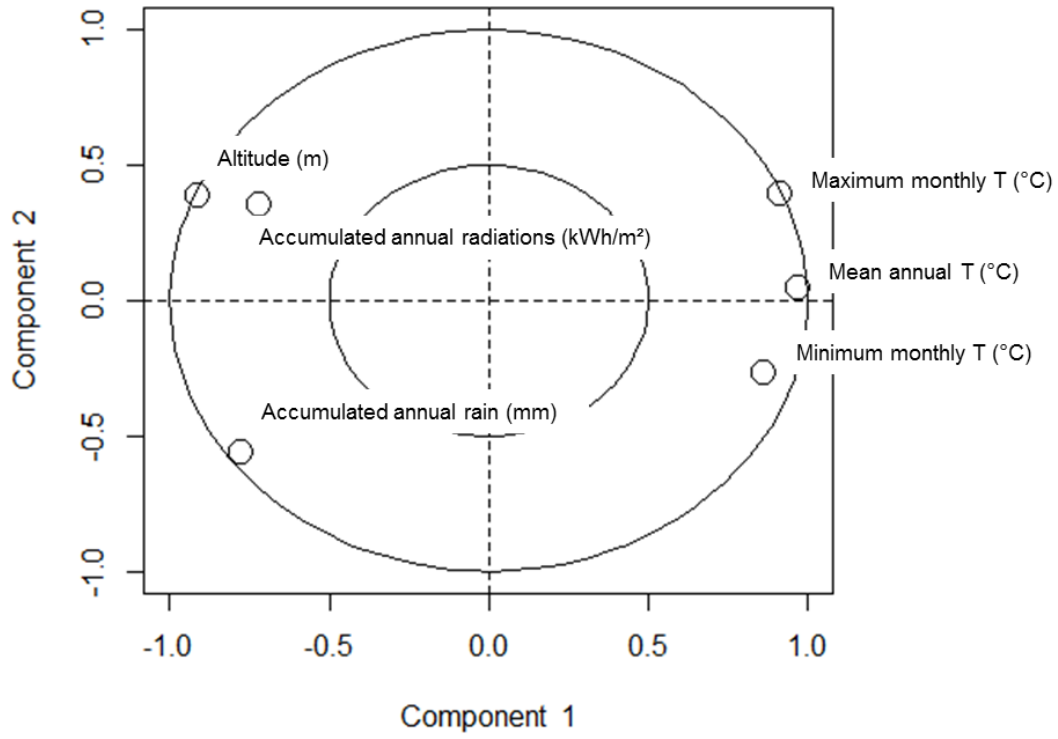
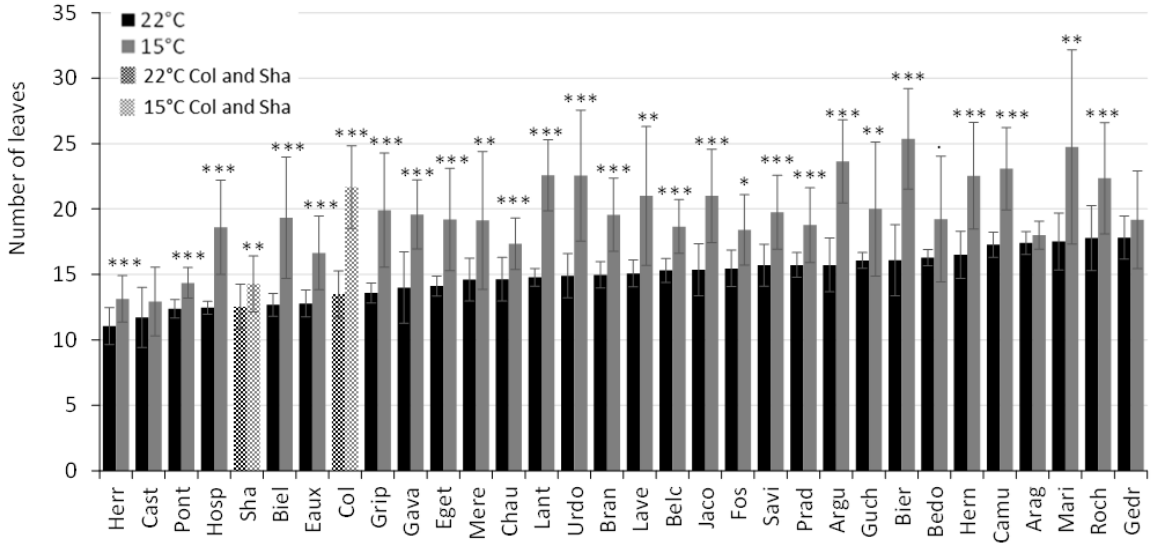


Figure S5. The number of leaves per rosettes increases at 15°C as compared to 22°C. (A) Measurements of the number of leaves per rosettes were done on the day of the harvest. Populations were classified according to the data gathered for plants grown at 22°C. 3 independent batches have been analyzed (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Duncan's test were performed for each growth condition: (B) 22°C and (C) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01.

A



B

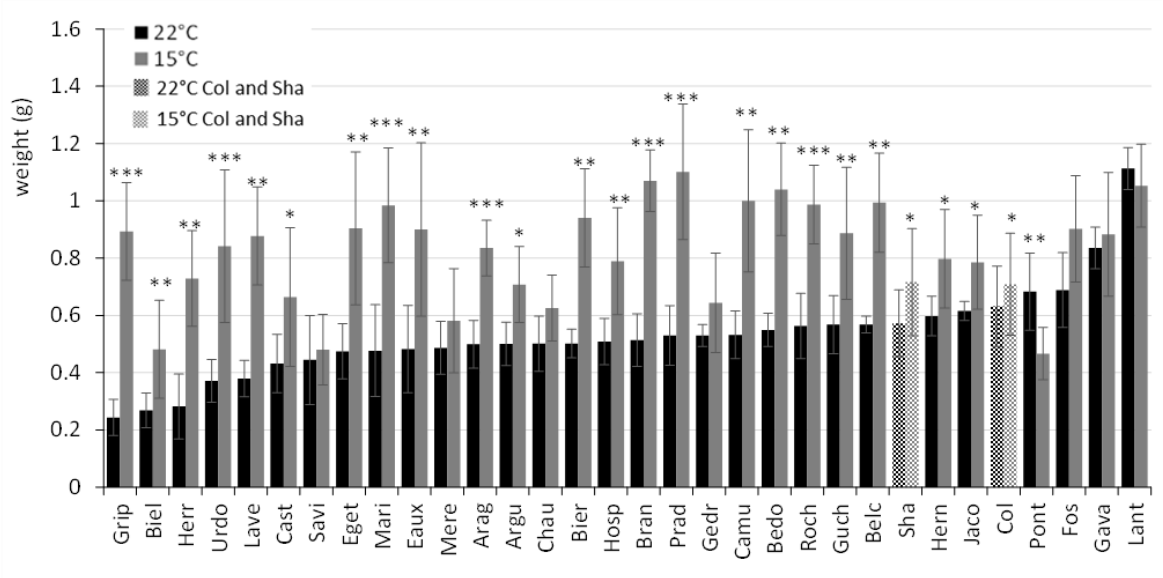
Population	Duncan
Gedr	a
Roch	a
Mari	ab
Arag	ab
Camu	abc
Hern	abcd
Bedo	abcde
Bier	bcde
Guch	bcde
Argu	cdef
Prad	cdef
Savi	cdef
Fos	defg
Jaco	defg
Belc	defg
Lave	defgh
Bran	defgh
Urdo	defgh
Lant	defgh
Chau	efgh
Mere	efgh
Eget	fghi
Gava	ghij
Grip	hij
Col	hij
Eaux	ijk
Biel	ijk
Sha	ijkl
Hosp	jkl
Pont	jkl
Cast	kl
Herr	l

C

Population	Duncan
Bier	a
Mari	ab
Argu	abc
Camu	abcd
Lant	abcde
Urdo	abcde
Hern	abcde
Roch	abcde
Col	abcdef
Jaco	bcdefg
Lave	bcdefg
Guch	cdefgh
Grip	cdefgh
Savi	cdefgh
Gava	cdefgh
Bran	cdefgh
Biel	defgh
Bedo	defgh
Eget	defgh
Gedr	defgh
Mere	defgh
Prad	efgh
Belc	efgh
Hosp	efgh
Fos	efgh
Arag	fghi
Chau	ghi
Eaux	hij
Pont	ij
Sha	ij
Herr	j
Cast	j

Figure S6. The weight of rosettes increases at 15°C as compared to 22°C. (A) Measurements of the weight of rosettes were done on the day of the harvest. Populations were classified according to the data gathered for plants grown at 22°C. 3 independent batches have been analyzed (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Duncan's test were performed for each growth condition: (B) 22°C and (C) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01.

A



B

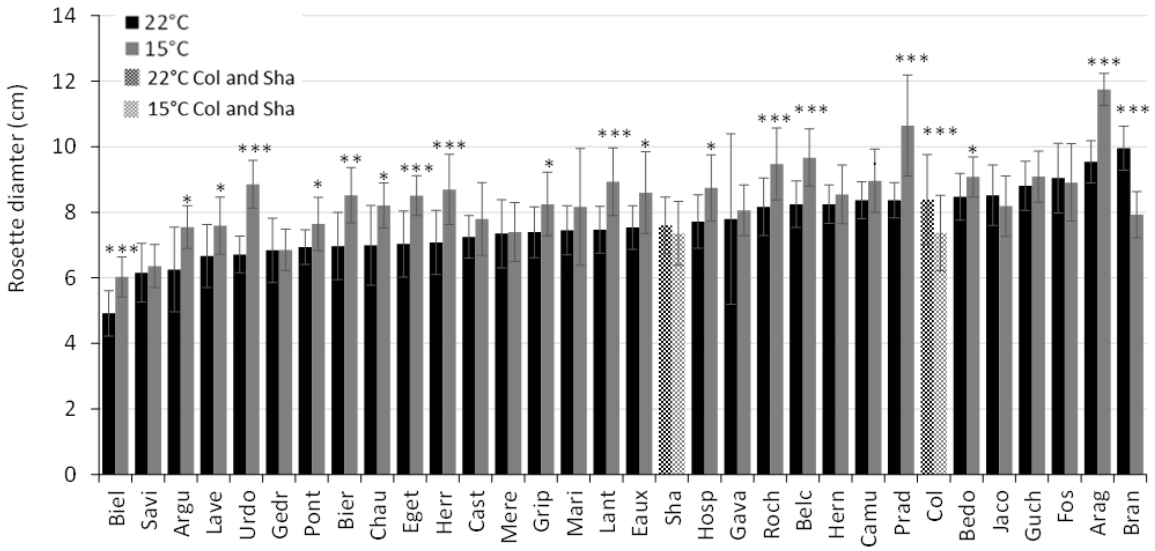
Population	Duncan
Lant	a
Gava	b
Fos	bc
Pont	bc
Col	cd
Jaco	cd
Hern	cde
Sha	cdef
Belc	cdef
Guch	cdef
Roch	cdef
Bedo	cdef
Camu	cdef
Gedr	cdef
Prad	cdef
Bran	cdef
Hosp	cdef
Bier	cdef
Chau	cdef
Argu	cdef
Arag	cdef
Mere	cdefg
Eaux	cdefg
Mari	cdefg
Eget	cdefg
Savi	defgh
Cast	defgh
Lave	efgh
Urdo	fgh
Herr	gh
Biel	h
Grip	h

C

Population	Duncan
Prad	a
Bran	a
Lant	ab
Bedo	abc
Camu	abcd
Belc	abcd
Roch	abcde
Mari	abcde
Bier	abcdef
Eget	abcdefg
Fos	abcdefg
Eaux	abcdefg
Grip	abcdefg
Guch	abcdefg
Gava	abcdefg
Lave	abcdefg
Urdo	abcdefg
Arag	abcdefg
Hern	abcdefgh
Hosp	abcdefgh
Jaco	abcdefgh
Herr	bcdefghi
Sha	cdefghi
Col	defghi
Argu	defghi
Cast	efghi
Gedr	fghi
Chau	fghi
Mere	ghi
Biel	hi
Savi	hi
Pont	i

Figure S7. The diameter of rosettes increases at 15°C as compared to 22°C. (A) Measurements of the rosette diameter were done on the day of the harvest. Populations were classified according to the data gathered for plants grown at 22°C. 3 independent batches were analyzed (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Duncan's test were performed for each growth condition: (B) 22°C and (C) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01.

A



B

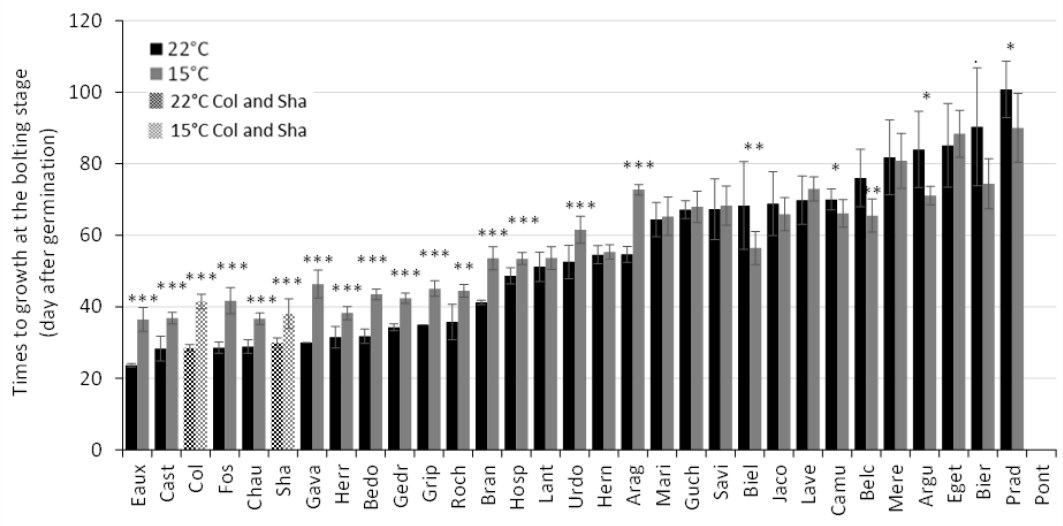
Population	Duncan
Bran	a
Arag	ab
Fos	abc
Guch	bcd
Jaco	cde
Bedo	cdef
Col	cdefg
Prad	cdefg
Camu	cdefg
Hern	cdefgh
Belc	cdefgh
Roch	cdefgh
Gava	defghi
Hosp	efghij
Sha	efghij
Eaux	efghij
Lant	efghij
Mari	fghij
Grip	ghij
Mere	ghij
Cast	hijk
Herr	ijkl
Eget	ijkl
Chau	ijkl
Bier	ijkl
Pont	ijkl
Gedr	ijkl
Urdo	jkl
Lave	jkl
Argu	kl
Savi	l
Biel	m

C

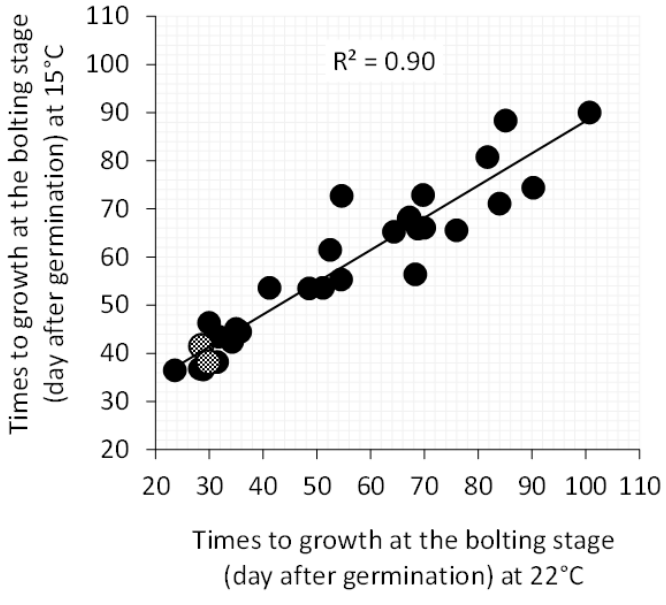
Population	Duncan
Arag	a
Prad	b
Belc	c
Roch	cd
Guch	cde
Bedo	cde
Camu	cdef
Lant	cdefg
Fos	cdefg
Urdo	cdefg
Hosp	cdefgh
Herr	cdefghi
Eaux	cdefghij
Hern	defghij
Bier	defghijk
Eget	defghijk
Grip	efghijkl
Chau	efghijkl
Jaco	efghijkl
Mari	efghijkl
Gava	efghijkl
Bran	fghijklm
Cast	ghijklm
Pont	hijklm
Lave	ijklm
Argu	jklm
Mere	klm
Col	lmn
Sha	lmn
Gedr	mno
Savi	no
Biel	o

Figure S8. The time to growth at the bolting stage increases at 15°C as compared to 22°C. (A) Different timing to reach to the bolting stage of plant grown in contrasted temperature conditions (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Populations were classified according to the data gathered for plants grown at 22°C. (B) Comparison between the time to growth at the bolting stage for population grown at 22 and 15°C. Duncan's test were performed for each growth condition: (C) 22°C and (D) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01. 3 independent batches have been analyzed.

A



B



C

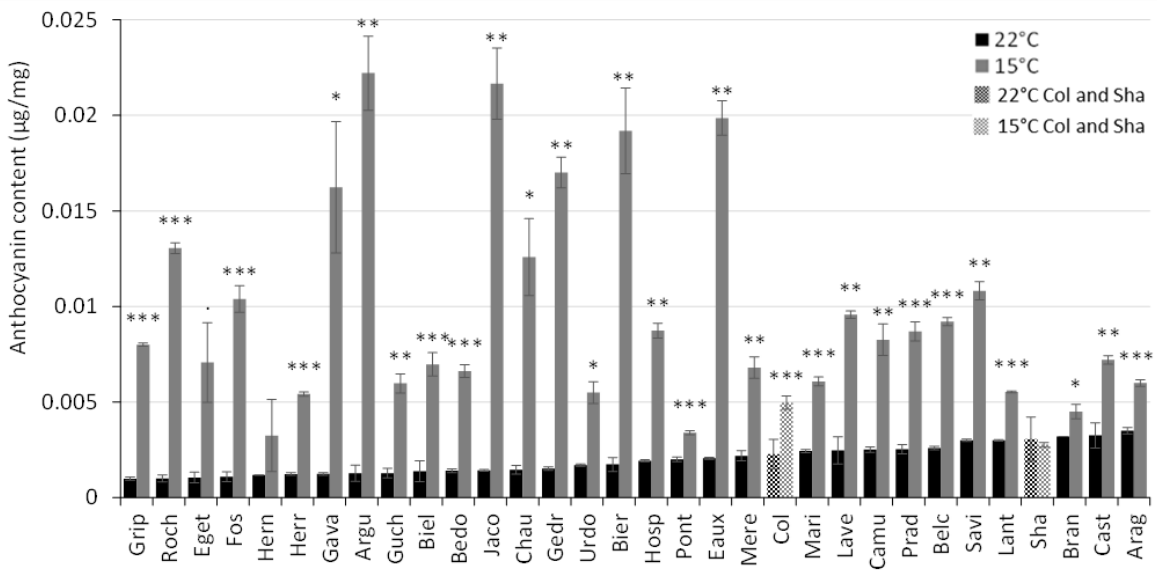
Population	Duncan test
Prad	a
Bier	b
Eget	bc
Argu	bc
Mere	cd
Belc	de
Camu	ef
Lave	ef
Jaco	ef
Biel	ef
Savi	f
Guch	f
Mari	f
Arag	g
Hern	g
Urdo	g
Lant	g
Hosp	gh
Bran	hi
Roch	ij
Grip	ij
Gedr	ij
Bedo	jk
Herr	jk
Gava	jk
Sha	jk
Chau	jk
Fos	jk
Col	jk
Cast	jk
Eaux	k

D

Population	Duncan test
Prad	a
Eget	a
Mere	b
Bier	c
Lave	cd
Arag	cd
Argu	cde
Savi	def
Guch	def
Camu	efg
Jaco	fg
Belc	fg
Mari	fg
Urdo	g
Biel	h
Hern	h
Lant	h
Bran	h
Hosp	h
Gava	i
Grip	i
Roch	i
Bedo	i
Gedr	ij
Fos	ijk
Col	ijk
Herr	jk
Sha	jk
Cast	k
Chau	k
Eaux	k

Figure S9. The anthocyanin content increases in rosettes at 15°C as compared to 22°C. (A) Measurements of the anthocyanin content were done on the day of the harvest. Populations were classified according to the data gathered for plants grown at 22°C. 3 independent batches have been analysed (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Duncan's test were performed for each growth condition: (B) 22°C and (C) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01.

A



B

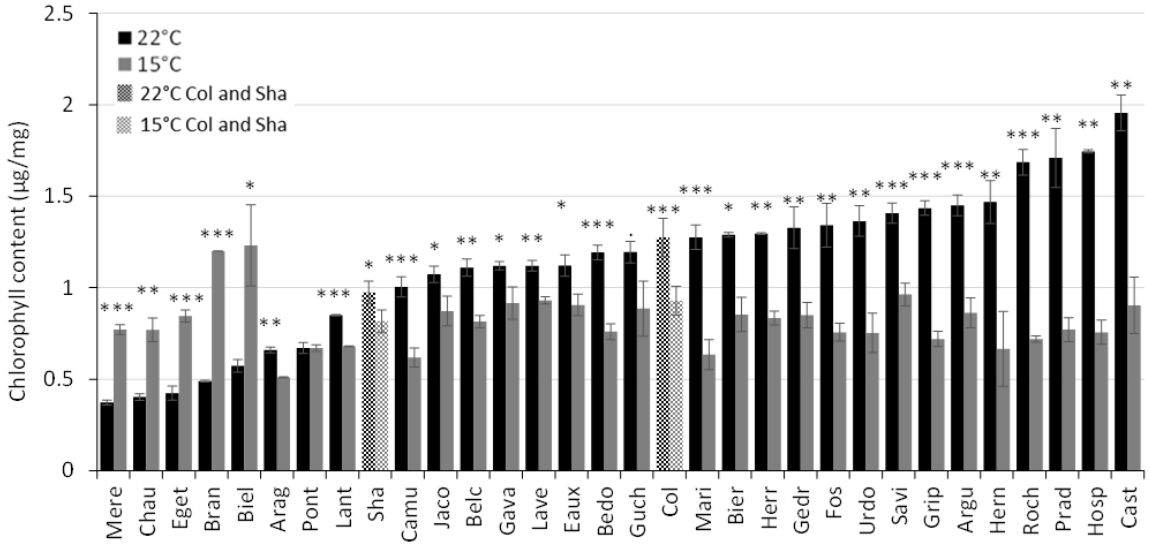
Population	Duncan test
Arag	a
Cast	ab
Bran	ab
Sha	abc
Lant	abc
Savi	abc
Belc	abcd
Prad	abcde
Camu	abcdef
Lave	abcdef
Mari	abcdef
Col	bcdefg
Mere	bcdefgh
Eaux	cdefghi
Pont	cdefghi
Hosp	cdefghi
Bier	defghi
Urdo	defghi
Gedr	defghi
Chau	efghi
Jaco	efghi
Bedo	efghi
Biel	fghi
Guch	ghi
Argu	ghi
Gava	ghi
Herr	ghi
Hern	ghi
Fos	hi
Eget	hi
Roch	i
Grip	i

C

Population	Duncan test
Argu	a
Jaco	a
Eaux	ab
Bier	abc
Gedr	bc
Gava	c
Roch	d
Chau	de
Savi	def
Fos	defg
Lave	efgh
Belc	fghi
Hosp	fghij
Prad	fghij
Camu	fghijk
Grip	fghijk
Cast	ghijkl
Eget	ghijkl
Biel	hijkl
Mere	hijkl
Bedo	hijklm
Mari	ijklmn
Arag	ijklmn
Guch	ijklmn
Lant	jklmn
Urdo	jklmn
Herr	jklmn
Col	klmn
Bran	lmn
Pont	mn
Hern	mn
Sha	n

Figure S10. Chlorophyll content is more stable in rosettes of different population grown at 15°C. (A) Measurements of the chlorophyll content were done on the day of the harvest Populations were classified according to the data gathered for plants grown at 22°C. 3 independent batches have been analysed (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Duncan's test were performed for each growth condition: (B) 22°C and (C) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01. (D) Different chlorophyll content in all the populations at 22 and 15°C growing conditions.

A



B

Population	Duncan test
Cast	a
Hosp	b
Prad	b
Roch	b
Hern	c
Argu	cd
Grip	cde
Savi	cde
Urdo	cde
Fos	cdef
Gedr	cdef
Herr	def
Bier	def
Mari	efg
Col	efg
Guch	fgh
Bedo	fgh
Eaux	ghi
Lave	ghi
Gava	ghi
Belc	hi
Jaco	hi
Camu	i
Sha	ij
Lant	j
Pont	k
Arag	k
Biel	kl
Bran	lm
Eget	lm
Chau	m
Mere	m

C

Population	Duncan test
Biel	a
Bran	a
Savi	b
Lave	bc
Col	bc
Gava	bc
Eaux	bc
Cast	bc
Guch	bcd
Jaco	bcd
Argu	bcd
Bier	bcde
Gedr	bcde
Eget	bcde
Herr	bcdef
Sha	bcdef
Belc	bcdef
Prad	bcdef
Mere	bcdef
Chau	bcdef
Bedo	bcdef
Fos	bcdef
Hosp	bcdef
Urdo	bcdef
Grip	cdef
Roch	cdef
Lant	defg
Pont	defg
Hern	defg
Mari	efg
Camu	fg
Arag	g

D

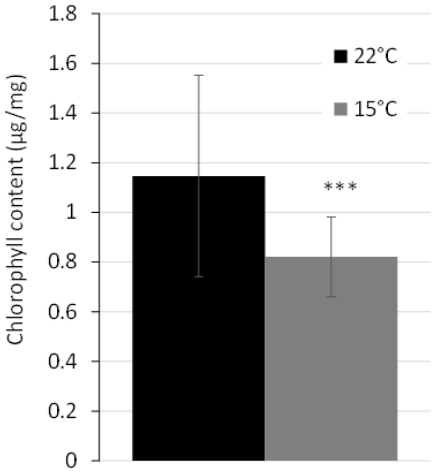
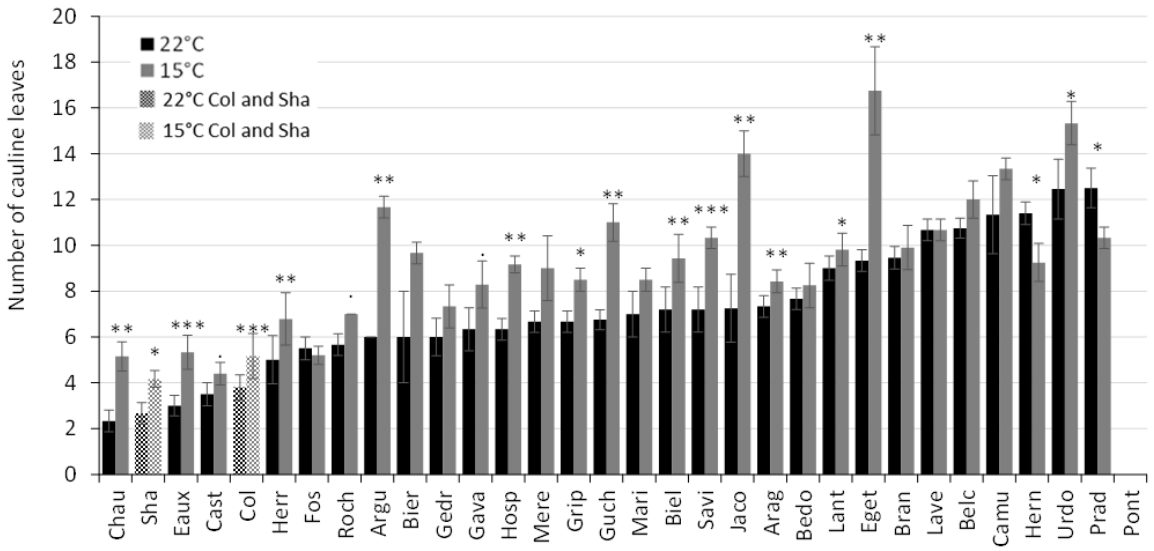


Figure S11. The number of cauline leaves in stems increases at 15°C as compared to 22°C. (A) Measurements of the number of cauline leaves were done on the day of the harvest. Populations were classified according to the data gathered for plants grown at 22°C. 3 independent batches have been analyzed (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Duncan's test were performed for each growth condition: (B) 22°C and (C) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01.

A



B

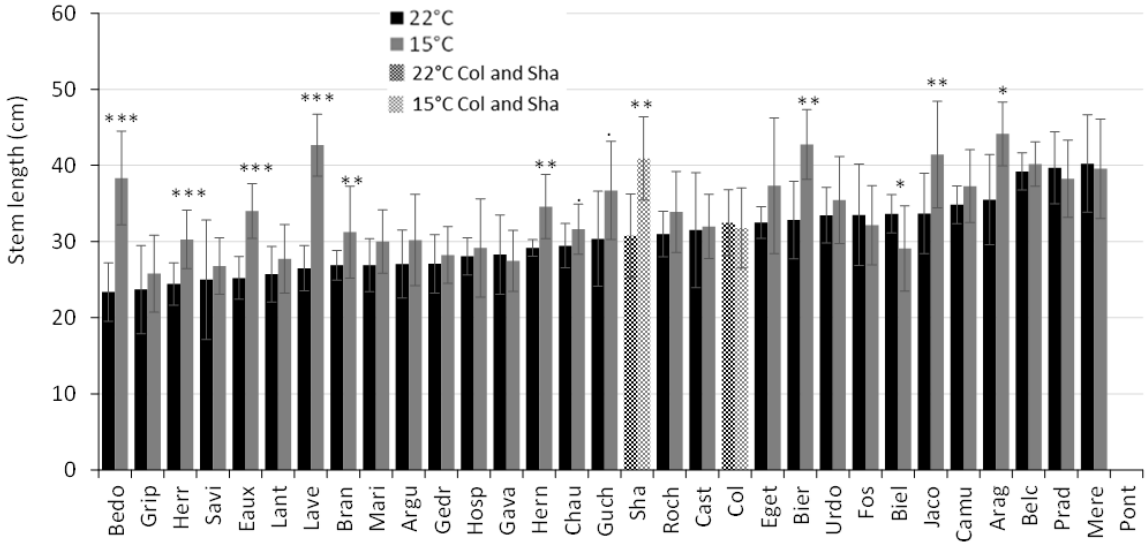
Population	Duncan test
Prad	a
Urdo	a
Hern	a
Camu	a
Belc	ab
Lave	ab
Bran	b
Eget	bc
Lant	bcd
Bedo	cde
Arag	def
Jaco	def
Biel	def
Savi	def
Mari	ef
Guch	efg
Grip	efg
Mere	efg
Gava	efg
Hosp	efg
Argu	efg
Bier	efg
Gedr	efg
Roch	fg
Fos	fgh
Herr	ghi
Col	hij
Cast	ij
Eaux	j
Sha	j
Chau	j

C

Population	Duncan test
Eget	a
Urdo	ab
Jaco	bc
Camu	cd
Belc	de
Argu	def
Guch	efg
Lave	efgh
Prad	efghi
Savi	efghi
Bran	fghij
Lant	ghij
Bier	ghij
Biel	ghij
Hern	ghij
Hosp	ghijk
Mere	hijk
Grip	ijkl
Mari	ijkl
Arag	ijkl
Gava	jkl
Bedo	jkl
Gedr	kl
Roch	lm
Herr	lmn
Eaux	mno
Fos	no
Col	no
Chau	no
Cast	o
Sha	o

Figure S12. The final stems length increases at 15°C as compared to 22°C. (A) Measurements of final stems length were done on the day of the harvest. Populations were classified according to the data gathered for plants grown at 22°C. 3 independent batches have been analyzed (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Duncan's test were performed for each growth condition: (B) 22°C and (C) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01.

A



B

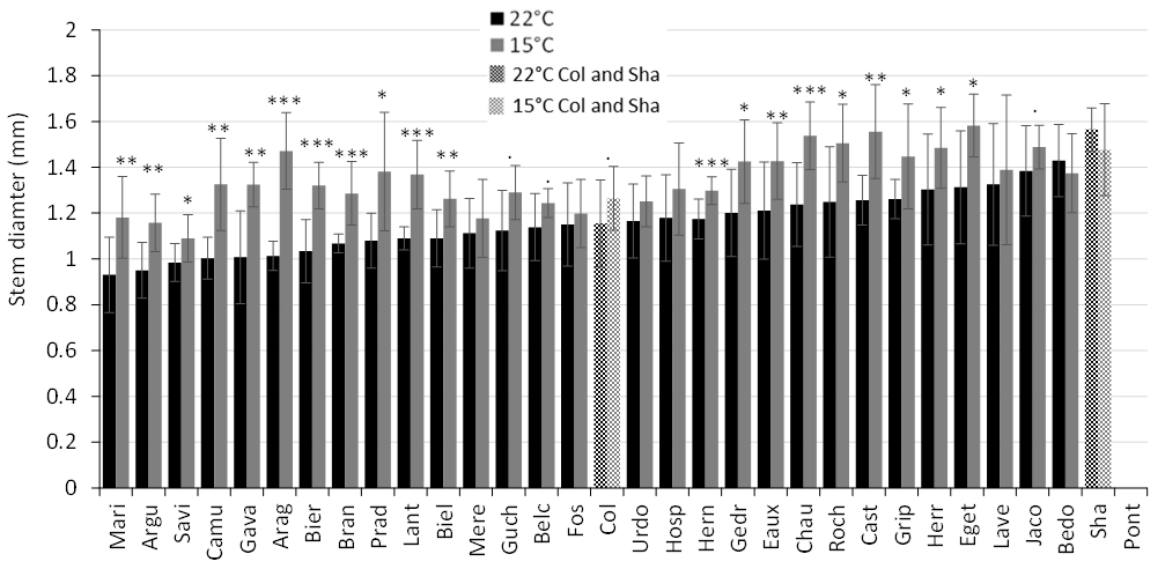
Population	Duncan test
Mere	a
Prad	a
Belc	ab
Arag	abc
Camu	abc
Jaco	bcd
Biel	bcd
Fos	bcd
Urdo	bcd
Bier	cde
Eget	cde
Col	cde
Cast	cdef
Roch	cdefg
Sha	cdefgh
Guch	cdefgh
Chau	cdefghi
Hern	cdefghi
Gava	defghi
Hosp	defghi
Gedr	efghi
Argu	efghi
Mari	efghi
Bran	efghi
Lave	efghi
Lant	fghi
Eaux	fghi
Savi	ghi
Herr	hi
Grip	i
Bedo	i

C

Population	Duncan test
Arag	a
Bier	ab
Lave	ab
Jaco	abc
Sha	abcd
Belc	abcd
Mere	abcde
Bedo	abcdef
Prad	abcdef
Eget	abcdefg
Camu	abcdefg
Guch	abcdefg
Urdo	bcdefgh
Hern	cdefghi
Eaux	cdefghij
Roch	defghij
Fos	efghijk
Cast	fghijk
Col	fghijk
Chau	fghijk
Bran	fghijk
Herr	ghijk
Argu	ghijk
Mari	ghijk
Hosp	hijk
Biel	hijk
Gedr	hijk
Lant	ijk
Gava	ijk
Savi	jk
Grip	k

Figure S13. The diameter of stems increases at 15°C as compared to 22°C. (A) Measurements of the diameter of stems were done on the day of the harvest. Populations were classified according to the data gathered for plants grown at 22°C. 3 independent batches were analyzed (mean ± SD) and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Duncan's test were performed for each growth condition: (B) 22°C and (C) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01.

A



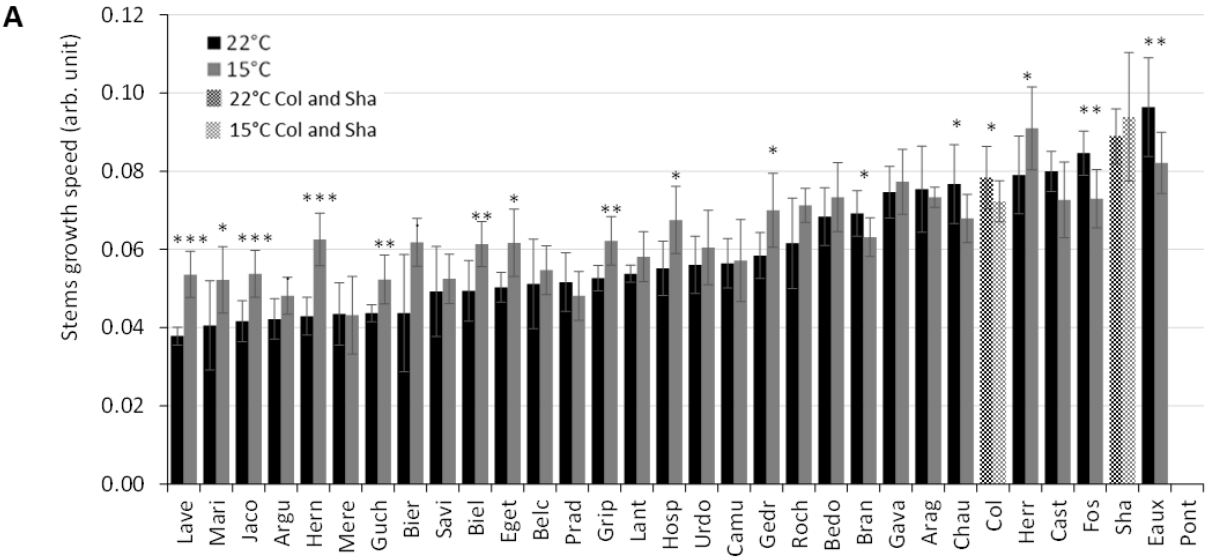
B

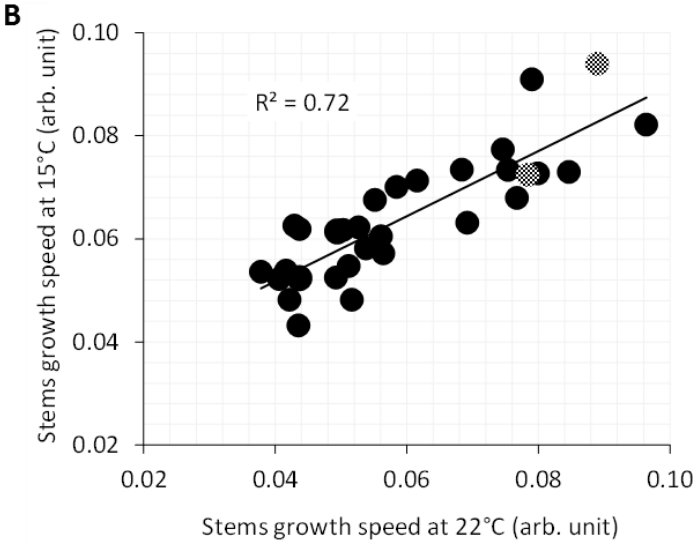
Population	Duncan test
Sha	a
Bedo	ab
Jaco	abc
Lave	bcd
Eget	bcde
Herr	bcdef
Grip	bcdefg
Cast	bcdefg
Roch	bcdefgh
Chau	bcdefghi
Eaux	bcdefghij
Gedr	bcdefghij
Hosp	cdefghijk
Hern	cdefghijk
Urdo	cdefghijkl
Col	cdefghijkl
Fos	cdefghijkl
Belc	defghijkl
Guch	defghijkl
Mere	defghijkl
Biel	defghijkl
Lant	defghijkl
Prad	efghijkl
Bran	fghijkl
Bier	ghijkl
Arag	hijkl
Gava	hijkl
Camu	ijkl
Savi	jkl
Argu	kl
Mari	l

C

Population	Duncan test
Eget	a
Cast	ab
Chau	abc
Roch	abcd
Jaco	abcde
Herr	abcde
Sha	abcdef
Arag	abcdef
Grip	abcdef
Eaux	abcdefg
Gedr	abcdefg
Lave	abcdefgh
Prad	abcdefgh
Bedo	abcdefgh
Lant	abcdefgh
Camu	bcdefgh
Gava	bcdefgh
Bier	cdefghi
Hosp	cdefghi
Hern	defghi
Guch	defghi
Bran	defghi
Col	efghi
Biel	efghi
Urdo	efghi
Belc	fghi
Fos	ghi
Mari	hi
Mere	hi
Argu	hi
Savi	i

Figure S14. The stems growth speed increases at 15°C as compared to 22°C. (A) Different stems growth speed of plants grown in contrasted temperature conditions and significant differences between temperature data are shown with Student test: '***', P ≤ 0.001; '**', P ≤0.01; '*', P ≤ 0.05; '.', P ≤ 0.1). Populations were classified according to the data gathered for plants grown at 22°C. (B) Comparison between the stems growth speed at 22 and 15°C. Duncan's test were performed for each growth condition: (C) 22°C and (D) 15°C. Different letters indicate significant differences between populations with P ≤ 0.01. 3 independent batches have been analyzed (mean ± SD)





C

Population	Duncan test
Eaux	a
Sha	ab
Fos	abc
Cast	bcd
Herr	bcd
Col	bcd
Chau	bcd
Arag	cd
Gava	cd
Bran	de
Bedo	def
Roch	efg
Gedr	efg
Camu	fgh
Urdo	fgh
Hosp	ghi
Lant	ghij
Grip	ghij
Prad	ghij
Belc	ghijk
Eget	ghijk
Biel	ghijk
Savi	ghijk
Bier	hijk
Guch	hijk
Mere	hijk
Hern	hijk
Argu	ijk
Jaco	ijk
Mari	jk
Lave	k

D

Population	Duncan test
Sha	a
Herr	ab
Eaux	bc
Gava	cd
Bedo	cde
Arag	cde
Fos	cdef
Cast	cdef
Col	cdef
Roch	cdef
Gedr	defg
Chau	defgh
Hosp	defgh
Bran	efghi
Hern	efghi
Grip	efghi
Bier	efghi
Eget	efghi
Biel	efghi
Urdo	fghi
Lant	ghij
Camu	hij
Belc	ijk
Jaco	ijk
Lave	ijk
Savi	ijk
Guch	ijk
Mari	ijk
Argu	jk
Prad	jk
Mere	k

L’étude de la structure génétique ainsi que des variations inter et intra-population ont mis en évidence la variabilité spécifique des populations d’*A. thaliana* identifiées dans cette région. Cette étude a également révélé des variations de plasticités phénotypiques lors de l'acclimatation d’*A. thaliana* à un gradient abiotique caractérisé ici par la température. Certains de ces traits phénotypiques se trouvent corrélés à l’appartenance à un cluster génétique ou aux caractéristiques climatiques d’origines de ces populations. Cette étude contribue à enrichir les connaissances sur l'acclimatation et la plasticité phénotypique des plantes face au changement climatique.

L’étude d’*A. thaliana* dans la région de la chaîne de montagnes des Pyrénées nous a également permis d’identifier quatre populations (Roch, Grip, Hern et Hosp) de par leurs homogénéités, leurs différences génétiques ainsi que leurs altitudes contrastées. Ces populations seront utilisées par la suite dans le projet *WallOmics* détaillé en partie IV.2. Cette dernière partie sera consacrée au dernier objectif du projet *WallOmics :* l’analyse, le traitement et l’intégration de données omiques hétérogènes dans le but de mieux comprendre la plasticité pariétale d’*A. thaliana* en étudiant des populations naturelles cultivées à des températures sub-optimales.

IV. Études de cas : plasticité pariétale et intégration de données omiques hétérogènes

IV.1 Cell wall adaptation of two contrasted ecotypes of *Arabidopsis thaliana*, Col and Sha, to sub-optimal growth conditions: an integrative study

Afin de mettre en place un protocole expérimental pour la création du projet *WallOmics*, un projet a été initié en 2014. Ce projet visait à étudier deux écotypes provenant d'altitudes et a *fortiori* de conditions de croissances contrastés. Ainsi, les rosettes de l'écotype Col provenant des plaines de Pologne (200m) et Sha, provenant de la haute vallée de la rivière de Shahdara au Tadjikistan (3400m), ont fait l'objet d'une première étude multi-blocs après avoir été acclimatées à des conditions de température de croissance sous-optimales (22°C et 15°C). Cette étude combine des données phénomiques, de composition de polysaccharides pariétaux, de protéomique pariétale, ainsi que des données de transcriptomique (analyse par RNA-seq).

Les auteurs contribuant à ce travail appartiennent pour la plupart au COST (*European cooperation in science and technology*) FA1306 intitulé : *The quest for tolerant varieties: phenotyping at plant and cellular level*. L'un des principaux objectifs de ce groupe est de discuter des outils aidant à traiter de grands jeux de données hétérogènes dans un but d'analyse intégrative. Cette étude permet, de par une première approche descriptive de chaque bloc et une analyse intégrative, la mise en lien des différents blocs entre eux.

Cette étude est ciblée sur les parois cellulaires en tant que barrières physiques dynamiques variant en structure et en composition, en réponse aux changements environnementaux. Ce projet confirme que les parois végétales peuvent jouer un rôle pendant l'acclimatation de la plante à des températures de croissance sub-optimales. De plus, de putatifs gènes / protéines candidats d'intérêt pour de futures études fonctionnelles ont pu être identifiés. Outre la quantité importantes de résultats, obtenus cette étude a également mis en évidence la complexité des analyses intégratives en ne considérant qu'un seul organe provenant de deux écotypes.

Ce travail pose donc les premières bases pour de futurs travaux réalisés avec une approche intégrative pour la communauté COST FA1306 ainsi que pour le projet *WallOmics*. Enfin, il met en lumière un certain nombre de candidats susceptibles d'être utilisés pour des analyses fonctionnelles

Cette étude, publiée dans la revue *Plant Science* est présentée ci-après.

Plant Science 263 (2017) 183–193

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Cell wall modifications of two *Arabidopsis thaliana* ecotypes, Col and Sha, in response to sub-optimal growth conditions: An integrative study

Harold Duruflé[a,1], Vincent Hervé[a,1], Philippe Ranocha[a], Thierry Balliau[b,c], Michel Zivy[b,c], Josiane Chourré[a], Hélène San Clemente[a], Vincent Burlat[a], Cécile Albenne[a], Sébastien Déjean[d], Elisabeth Jamet[a,*], Christophe Dunand[a,*]

[a] Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, BP42617, 31326 Castanet-Tolosan, France
[b] CNRS, PAPPSO, UMR 0320/UMR 8120 Génétique Végétale, 91190 Gif sur Yvette, France
[c] INRA, PAPPSO, UMR 0320/UMR 8120 Génétique Végétale, 91190 Gif sur Yvette, France
[d] Institut de Mathématique de Toulouse, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

ARTICLE INFO

Keywords:
Temperature acclimation
Arabidopsis thaliana
Cell wall
Cell wall polysaccharide
Gene toolbox
Integrative analysis
Proteomics
RNA seq

ABSTRACT

With the global temperature change, plant adaptations are predicted, but little is known about the molecular mechanisms underlying them. *Arabidopsis thaliana* is a model plant adapted to various environmental conditions, in particular able to develop along an altitudinal gradient. Two ecotypes, Columbia (Col) growing at low altitude, and Shahdara (Sha) growing at 3400 m, have been studied at optimal and sub-optimal growth temperature (22 °C vs 15 °C). Macro- and micro-phenotyping, cell wall monosaccharides analyses, cell wall proteomics, and transcriptomics have been performed in order to accomplish an integrative analysis. The analysis has been focused on cell walls (CWs) which are assumed to play roles in response to environmental changes. At 15 °C, both ecotypes presented characteristic morphological traits of low temperature growth acclimation such as reduced rosette diameter, increased number of leaves, modifications of their CW composition and cuticle reinforcement. Altogether, the integrative analysis has allowed identifying several candidate genes/proteins possibly involved in the cell wall modifications observed during the temperature acclimation response.

1. Introduction

In the global warming context, elevated temperature is considered as the most serious change and it is already observed. The seasons are often altered with changes in temperature and occurrence of freezing stress that can appear without any preceding chilling period [1]. A study has shown that in the main European mountains, climate changes are ongoing and gradually transforming mountain plant communities [2]. As a consequence of this process, the more cold-adapted species are declining whereas the warm-adapted ones are prospering. Global warming is also critical to maintain agricultural productivity in the future [3]. *Arabidopsis thaliana* (L.) Heyhn, a *Brassicaceae* originating from the Eurasian continent [4], is adapted to multiple environmental conditions. This annual self-fertilized plant is a very good model for phenotypic plasticity studies. Furthermore, the large accumulation of genomics, genetics and molecular data regarding this plant is very helpful for the understanding of stress responses at multiple levels [5,6]. Although the variability between *A. thaliana* populations is well-recognized and studied at the genomics level, the molecular mechanisms below are still poorly described [7]. Conversely, phenotypic predictions from genotypes are complex because of epistatic interactions between genes usually controlling responses to environment.

Despite the evidence that adaptation to local climate changes is common in plant populations, a lot of work remains to be done to understand the genetic evolution contributing to climate acclimation. As an example, plant ecotypes are able to specifically develop along an altitudinal gradient and modification of their cell walls (CWs) could be one trait of their adaptation and/or acclimation. Indeed, CWs represent dynamic external physical barriers, the composition and structure of which vary upon developmental and environmental changes [8,9]. They play critical roles in the control of growth, cell shape, and structural integrity [10] thanks to modifications in CW protein (CWP) content which lead to changes in cell wall architecture [11–13].

The large collection of *A. thaliana* ecotypes [14] living in contrasted

Abbreviations: CW, cell wall; CWP, cell wall protein; DIR, dirigent protein; GH, glycoside hydrolase; HG, homogalacturonan; HRGP, hydroxyproline-rich glycoprotein; LTP, lipid transfer protein; PCA, principal component analysis; PME, pectin methylesterase; Prx, class III peroxidase; RGI, rhamnogalacturonan I; RNA seq, RNA sequencing; XG, xyloglucan

* Corresponding authors.
E-mail addresses: jamet@lrsv.ups-tlse.fr (E. Jamet), dunand@lrsv.ups-tlse.fr (C. Dunand).
[1] Co-first authors.

http://dx.doi.org/10.1016/j.plantsci.2017.07.015
Received 19 April 2017; Received in revised form 12 July 2017; Accepted 18 July 2017
Available online 20 July 2017

habitats differing by their climate is a material of choice to look at the physiological responses to environmental constraints [15]. Although there are numerous studies aiming at understanding the plant response to limited-exposure to abiotic stresses (for reviews, see [16,17]), data regarding the physiology of plants grown at sub-optimal temperatures is scarce. Two studies have recently been performed with three ecotypes from Sweden, Poland and Italy corresponding to three different growth temperature regimes [18,19]. Changes in photosynthetic capacities, leaf thickness and morphology of tracheary elements in correlation with transpirational water loss have been observed when these ecotypes were cultivated at low (day at 12.5 °C/night at 8 °C), moderate (25 °C/20 °C) or high temperature (35 °C/25 °C). The results of this phenotyping could be correlated with the latitude and the temperature of the habitat of origin. However, data are still lacking to deeply understand the acclimation mechanisms at the molecular level.

For our study, two ecotypes of *A. thaliana* originating from contrasted natural environments, Columbia (Col) and Shahdara (Sha) have been used [20]. Col initially originating from Poland was adapted to both low altitude and high temperature, whereas Sha is growing at 3400 m in a high valley of Tajikistan. Since, it was not possible to recreate the real ecological environment of Col and Sha, we have focused the study on the temperature effect as a first step towards the understanding of the environment acclimation response. Two different growth conditions have been studied: 22 °C (optimal growth condition for Col) and 15 °C (optimal growth condition for Sha). An integrative approach has been performed, combining macro- and micro-phenotyping, and CW monosaccharide, CW proteomics, and transcriptomics analyses. Statistical analysis of the data has allowed establishing correlations between the different datasets and identifying candidate genes/proteins possibly involved in the temperature acclimation response of Col and Sha ecotypes.

2. Materials and methods

2.1. Plant material

The Col and Sha ecotypes of *A. thaliana* (L.) Heyhn were used (Supplementary Fig. S1). Climatic variables were obtained from WorldClim [21] (www.worldclim.org) at the GPS geographic origins of the Col (52.745416, 15.235557) and Sha (39.250103, 68.249919) ecotypes (publiclines.versailles.inra.fr). Seeds were sown in Jiffy-7® peat pellets (Jiffy International, Kristiansand, Norway). After 48 h of stratification at 4 °C in darkness, plants were grown at a light intensity of 90 µmol.photons/m^2/s, a humidity of 70% and under a 16 h light/8 h dark photoperiod at two different temperatures: 22 °C and 15 °C. Four- or 6-week-old rosettes were collected at the bolting developmental stage after growth at 22 °C or 15 °C, respectively (Supplementary Fig. S1).

2.2. Macrophenotyping

The rosette diameter and mass were measured at the time of sampling, together with the count of leaves. Before storage at −80 °C, pictures were taken to measure rosette areas with the ImageJ software [22]. Leaf density was determined from the leaves of three representative plants per experiment.

2.3. Anthocyanin content

Ground rosette material (0.1 g) was mixed with 1 mL of 95% ethanol/1% HCl and stored at 5 °C for 24 h in darkness. The samples were centrifuged for 10 min at ×1000*g*. The absorbance of the supernatant was measured at 530 and 657 nm. Anthocyanin concentration (µg anthocyanin/mg fresh material) was calculated using the following formula: $(A_{530} - 0.25 \times A_{657})$/mg of fresh material [23].

2.4. Histological staining of cell walls

Whole rosettes were harvested by cutting above the crown and rapidly individually infiltrated under vaccum in 50 mL Falcon tubes with FAA (10% formalin (37% formaldehyde solution, Sigma-Aldrich, Saint-Quentin Fallavier, France); 50% ethyl alcohol; 5% acetic acid; 35% distilled water). They were fixed for 16 h at 4 °C. The dehydration and paraplast infiltration protocol was as previously described [24]. The whole rosettes were processed in individual tubes used as embedding molds in order to keep traces of the phyllotaxy. Twenty µm-thick serial sections were disposed on silane-coated microscope slides (2–4 rosettes per slide). In order to ensure comparison of labelling intensities among the rosettes, a 20 slide-plastic holder and 200 mL staining jars were used [24]. Intracellular material was removed by incubating sections in 2.6% bleach for 20 min [25] followed by thorough washes with distilled water. Sections were then incubated for 5 min in 0.5% safranin red solution (CI 50240; Kuhlmann, Paris, France), washed with distilled water to remove excess staining solution and then incubated for 90 min in 0.005% alcian blue solution (CI 74240, Sigma-Aldrich) [26,27]. Following extensive washing, slides were dried, mounted in Eukitt® (quick-hardening mounting medium, Sigma-Aldrich) and scanned using a NanoZoomer HT scanner (Hamamatsu, Hamamatsu City, Japan). The auto-fluorescence of aromatic compounds was observed using a DAPI filter set (excitation: 387 nm +/− 11 nm; dichroic mirror 405 nm; emission: 440 +/− 40 nm) using a NanoZoomer RS scanner (Hamamatsu).

2.5. Measurement of cuticle permeability

To quantify water-loss, rosettes were excised, gently wiped to remove excess water and put in an oven at 40 °C. The fresh masses were recorded, before and after different times at 40 °C, using a micro-balance. Data were expressed as percentage of fresh rosette mass reduction in reference to the initial fresh mass. Epidermal permeability was also assessed using chlorophyll efflux. Entire rosettes were collected and immersed in 40 mL 80% ethanol/0.7 g of fresh mass and gently shaken at room temperature. Aliquots of 2 µL were sampled at different times after immersion and chlorophyll content was determined by measuring absorption spectra at 664 and 647 nm using a spectrophotometer DS-11 FX (DeNovix, Wilmington, DE, USA). Chlorophyll content was calculated using the following formula: $(A_{663} \times 7.15 + A_{647} \times 18.71)$/mg of fresh material = µg chlorophyll/mg fresh material [28].

2.6. Extraction of proteins from purified cell walls

CW purification was performed using 20 rosettes (about 10 g) for each experiment as described [29]. The sequential extraction of proteins from purified CWs was performed as described [30]. Typically, 0.2 g of lyophilized CWs was used for one extraction and about 500 µg proteins were extracted. The final protein extract was lyophilized. Proteins were quantified with the CooAssay Protein Assay kit (Interchim, Montluçon, France).

2.7. Cell wall monosaccharide analyses

The sequential extraction of CW polysaccharides in four steps was adapted from [31]. One hundred mg of a deproteinized CW fraction (corresponding to 20 mg dry cell walls) were used. Four successive extractions were carried out to obtain extracts enriched in pectins (E1 and E2) and hemicelluloses (E3 and E4). Extractions were performed at room temperature as follows: (i) overnight-incubation with 300 µL of 50 mM diamino-cyclo-hexane-tetra-acetic acid (CDTA), pH 7.5, to obtain E1 extracts; (ii) 3 h-incubation with 300 µL of 50 mM Na_2CO_3, 20 mM $NaBH_4$, to obtain E2 extracts; (iii) 3 h-incubation with 300 µL 20 mM $NaBH_4$, 1 M NaOH and then 4 M NaOH, to obtain E3 and E4 extracts, respectively. The supernatants were recovered by

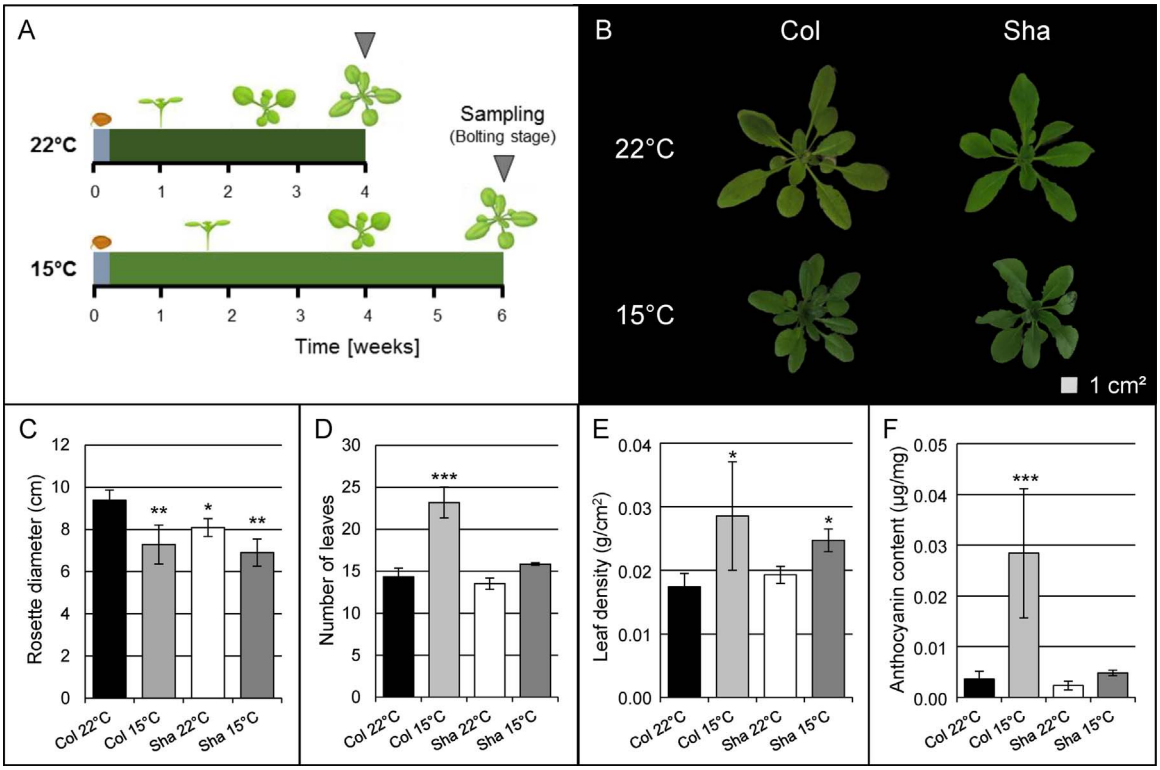


Fig. 1. Macro-phenotyping of Col and Sha plants grown at 22 °C and 15 °C. Experimental design (A), rosettes observed at 22 °C (4 weeks of culture) and 15 °C (6 weeks of culture) in growth chamber conditions (B). Measurement of rosette diameter (C), number of leaves (D), leaf density (E), and anthocyanin content (F). Twenty plants from 3 independent batches have been analyzed (mean ± SD). Statistically significant differences with regard to the reference (Col 22 °C) are shown with asterisks (ANOVA: '***' 0.001; '**' 0.01; '*' 0.05).

centrifugation for 10 min at 2500*g*. Each extract was hydrolysed in 2 N TFA for 1 h at 120 °C. After 10X dilution in UHQ water, monosaccharides were analyzed by High-Performance Anion-Exchange Chromatography coupled to Pulsed Amperometric Detection (HPAEC-PAD; Dionex, Sunnyvale, CA, USA) using a CarboPac PA1 column (Dionex). A first isocratic step using 15 mM NaOH was carried out to separate neutral carbohydrates. Then, the NaOH concentration was set up to 150 mM before the application of a linear gradient of sodium acetate from 0 to 600 mM in 150 mM NaOH in 30 min to enable the elution of acidic monosaccharides. All the runs were performed at a flow rate of 1.0 mL/min.

Standard monosaccharides were used for identification and quantification: L-arabinose (Sigma-Aldrich); L-fucose (Sigma-Aldrich); D-galactose (Sigma-Aldrich); galacturonic acid (Sigma-Aldrich); L-rhamnose (Sigma-Aldrich); D-glucose (Merck, Darmstadt, Germany); and D-xylose (Roche, Mannheim, Germany).

2.8. Identification of proteins by LC–MS/MS

Proteins were separated by a short one-dimensional electrophoresis (5 mm-run) and three samples were cut out of the gel as described [32]. The LC–MS/MS analyses were performed at the PAPPSO proteomics platform (pappso.inra.fr/) essentially as described [32]. A few parameters have been changed for MS data processing in the X!Tandem software (www.thegpm.org/tandem/) and the X!Tandem Pipeline 3.3.4 [33]. Trypsin digestion was declared with no possible miscleavage (Supplementary Table S1). Only proteins identified with at least two different specific peptides in the same sample and found in at least two biological repeats were validated. Furthermore, quantification was performed on peptides with standard deviation retention times lower than 20 s and peak width lower or equal to 100 s.

2.9. Bioinformatic annotation of proteins and statistical analyses of quantitative data

The prediction of sub-cellular localization and of functional domains of proteins was performed with the *ProtAnnDB* tool [34]. A protein was considered as a CWP if (i) at least two bioinformatic programs predicted it as a secreted protein, (ii) no intracellular retention signal was found and (iii) no more than one trans-membrane domain was found [35]. The quantification was only performed for the 379 identified CWPs using the MassChroQ software [36]. Proteomics data and procedure details are available in Supplementary Tables S1 and S2. All the LC–MS/MS data have been deposited at PROTICdb (proteus.moulon.inra.fr/w2dpage/proticdb/angular/) and CWP MS data at *WallProtDB* (www.polebio.lrsv.ups-tlse.fr/WallProtDB/).

2.10. RNA sequencing

For the transcriptomics analysis, frozen samples were ground in liquid nitrogen and total RNAs were extracted using the RNeasy Plant Mini Kit (Qiagen, Courtaboeuf, France). RNA quantification was performed using a spectrophotometer ND-1000 (NanoDrop, Wilmington, DE, USA) and RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Courtaboeuf, France). RNA seq experiments were performed on an Illumina HiSeq 3000 at the GeT-PlaGe platform (get.genotoul.fr, Auzeville, France) according to the standard Illumina protocols. Short pair-end sequencing reads generated were analysed using the commercial CLC Genomic Workbench 8.0 software (CLC bio, Aarhus, Denmark). RNA seq data and procedure details are available in Supplementary Table S3. Sequences are available at NCBI short read archive (SRA, BioProject PRJNA344545).

2.11. Data integration

Data analysis was performed in the R environment (cran.r-

project.org). ANOVA statistical tests were carried out in order to determine differentially expressed genes or accumulated proteins for ecotype, temperature and ecotype × temperature effects. In addition of these tests, the Duncan's multiple range test has been performed on some data to allow a better visualization of the different sets of means. Multivariate analyses (PCA) were achieved using the mixOmics package [37] available in CRAN (cran.r-project.org/package = mixOmics).

3. Results

3.1. Morphological phenotypes of Col and Sha ecotypes depend on temperature growth conditions

Macro-phenotypes of 4-week-old plants grown at 22 °C and 6-week-old plants grown at 15 °C have been analysed (Fig. 1A). In our growth conditions, 4 and 6 weeks corresponded to the bolting time for plants grown at 22 °C and 15 °C, respectively. Rosettes of Col and Sha plants grown at 15 °C showed reduced diameters and an increase in leaf number (Fig. 1B–D), both correlated with an increase in rosette density (Fig. 1E). Col plants presented very contrasted phenotypes between 22 °C and 15 °C and a dramatic increase in anthocyanin content was observed at 15 °C (Fig. 1F). Conversely, Sha plants had traits, such as decrease in leaf number and in anthocyanin content, less modified than Col plants (Fig. 1D, F).

Microscopic analysis of leaf structures was performed on petiole cross-sections of plants cultivated at 22 and 15 °C (Fig. 2). Staining with safranin red and alcian blue of bleached sections allowed distinguishing cell walls containing more hydrophilic polysaccharides (stained in blue) from those containing more hydrophobic compounds (stained in red). Larger petioles and higher safranin staining intensities of cuticle were observed for both Sha and Col plants grown at 15 °C as compared to plants grown at 22 °C. The overall staining with safranin red was the most intense on cross-sections of petioles of Col plants grown at 15 °C and of Sha plants grown at 22 °C. Moreover, lignin auto-fluorescence detection performed on similar cross-sections showed an increase of the number of vessel elements at 15 °C for both ecotypes which could be correlated with the rise of the vascular tissue area vs petiole tissue area ratio (Fig. 2E–F).

In order to see whether the higher safranin staining intensity at low temperature affected permeability properties, water-loss and chlorophyll leaching kinetics were measured on rosettes (Fig. 3). The rosettes of plants grown at 22 °C systematically exhibited a higher percentage of water-loss and leakage of chlorophyll than those of plants grown at 15 °C (Fig. 3). Altogether, these results were consistent with the reinforcement of cuticle hydrophobic properties observed at 15 °C for both ecotypes.

3.2. Cell wall composition is temperature growth condition- and ecotype-dependent

In order to determine if leaf morphological modifications in response to low temperature could be associated with changes in CW carbohydrate composition, polysaccharides were sequentially extracted from rosette CWs. They were submitted to acidic hydrolysis and their monosaccharide content was analysed, except for the cellulose fraction, which could not be hydrolysed in our assay conditions. Four sequential CW extracts were obtained with high quality of fractionation (Table 1). As expected, higher contents in galacturonic acid and xylose residues were respectively found in pectin- (extracts E1-2) and hemicellulose-(extracts E3-4) enriched fractions. Whatever the temperature, Col rosettes contained more fucose, arabinose, galactose, glucose and xylose residues in E3 and E4 extracts than Sha rosettes (Table 1). On the contrary, rosettes of plants grown at 15 °C contained more arabinose, galactose, glucose and galacturonic acid in the E1 and E2 extracts (Table 1). Then, the theoretical CW polysaccharide composition has been reconstructed based on the monosaccharide analyses and the following formula adapted from a previous study [38] (Table 2). The result of this calculation suggested that rosettes of plants grown at 15 °C contained more xyloglucan (XG) compared to those grown at 22 °C (Fig. 4). It also showed that the amounts of rhamnogalacturonan I (RGI) and homogalacturonan (HG) were higher in rosettes of Sha plants grown at 15 °C (Fig. 4).

3.3. Cell wall proteomes are modified at sub-optimal temperature growth

As CW composition depends upon temperature growth conditions, a proteomics approach has been performed to identify CWPs possibly involved in these modifications. The analysis by LC–MS/MS of proteins extracted from purified CWs has led to the identification of 965 different proteins validated in at least two out of three biological replicates (Supplementary Tables S1 and S2). Three hundred and 79 proteins were considered as *bona fide* CWPs, i.e. predicted to be secreted with no intracellular retention signal and at most one trans-membrane domain [35] (Supplementary Table S2A). Altogether, 36 new CWPs have been identified in this study, when compared with previous rosette proteomes [39–43]. Among the 36, 4 were only found in rosettes of Col plants grown at 15 °C, 6 and 13 in rosettes of Sha plants respectively grown at 22 °C and 15 °C [32] (Supplementary Table S2A).

An unsupervised Principal Component Analysis (PCA) of the 379 quantified CWPs (Supplementary Table S2B), showed a clear segregation of the four samples, demonstrating the good quality and repeatability of the analyses (Fig. 5A). The first component has separated the samples according to the temperature (22 °C vs 15 °C) and the second one according to the ecotypes (Col vs Sha). A similar number of CWPs differentially accumulated in each sample, as highlighted by the statistical ANOVA test (Supplementary Table S2B). Thirty nine and 41 out of the 379 CWPs were differentially accumulated, depending on the ecotype or the temperature variables respectively. In addition, 11 CWPs were highlighted by this statistical test for the interaction between the two variables.

Regarding the ecotype specificity, 25 and 14 CWPs were accumulated at higher levels in Col and Sha rosettes, respectively (Table 3). Among the CWPs mainly detected in Col rosettes, there were four proteins acting on polysaccharides (three glycoside hydrolases (GHs), one pectate lyase), two Ser proteases, two GDSL lipase acylhydrolases, two lectins, one purple acid phosphatase and one fasciclin arabinogalactan protein. Among the CWPs accumulated in Sha rosettes, there were five proteins acting on CW polysaccharides (three GHs, one expansin-like A, one pectin methylesterase (PME)), two Asp proteases, one COBRA-like protein, one purple acid phosphatase and one leucine-rich repeat extensin. Regarding the growth temperature, 18 and 11 CWPs were accumulated at higher level respectively at 22 °C and at 15 °C (Table 3). There were four GHs, two proteases and one lipid transfer protein (LTP) accumulated at 22 °C. There were four proteins acting on polysaccharides (two GHs, one α-expansin, one PME) and three proteases (Asp, Cys and Ser proteases) accumulated at 15 °C. The combined effect of ecotype and temperature (interaction) was more subtle because some CWPs accumulated more in Col and Sha rosettes at their optimal growth conditions (e.g. two α-expansins, At3g29030 and At1g26770; one Ser protease, At1g04110), whereas others accumulated more in Col and Sha rosettes at their sub-optimal growth conditions.

Altogether, the quantitative analysis of proteomics data has allowed the identification of CWPs accumulated either in one ecotype, at one growth temperature or in one ecotype/growth temperature combination. Some protein families were represented in only one of these cases, but others like GHs were found in several of them. These results allowed pointing at specific functions for each CWP, due to either specific gene regulation, localization of the protein or substrate/ligand interactions.

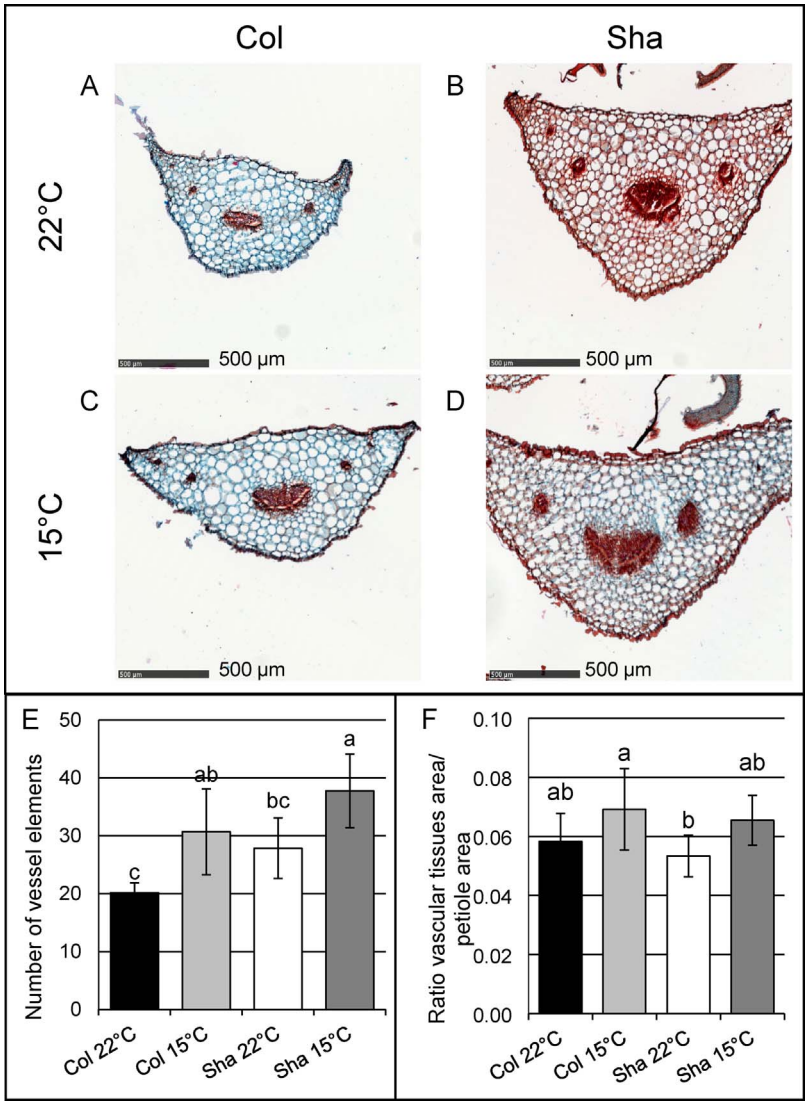


Fig. 2. Phenotyping of petioles of Col and Sha plants grown at 22 °C and 15 °C, a focus on vascular tissues and cuticle. Cross-sections of petioles of Col (A, C) and Sha (B, D) plants grown at 22 °C (A, C) and 15 °C (B, D) were stained with safranin red and alcian blue. Numbers of vessel elements per petiole were counted (E) and ratios between areas of vascular tissues and of petiole were calculated (F). Measurements have been performed at the bolting stage after 4 and 6 weeks of culture at 22 and 15 °C, respectively. Representative plant petioles out of more than 30 petioles analysed are shown for each ecotype grown at 22 °C or 15 °C (see also Supplementary Fig. S2). In each case, all the petioles from 3 plants from 3 independent batches have been analyzed (mean ± SD). Mean values were analysed with one-way ANOVA together with Duncan's multiple range test. Significant differences ($P < 0.1$) are noted with different lowercase letters.

3.4. Regulation of gene expression at the transcriptomic level depends on both growth conditions and ecotypes

To complement this CW proteomics study, RNA seq analyses have been performed leading to the identification of 16,541 genes, among which 2354 showed an ecotype effect, 402 a temperature effect, and 67 an interaction effect (Supplementary Table S3A).

The multivariate analysis of the four "whole transcriptomes" supported the good quality and repeatability of the experiments (Fig. 5B). In contrast to CW proteomics data (Fig. 5A), the PCA restricted to the transcripts encoding the CWPs identified by proteomics ("CWP transcriptomes"; Fig. 5C) showed a better separation according to the ecotype variable, than to the temperature variable: 26% vs 20%, according to the ecotype and 23% vs 38% according to the temperature (Fig. 5A and C). PCA was performed with three additional gene sets: 2206 genes encoding proteins involved in CW biosynthesis or already identified in CW proteomes (Supplementary Table S3I and *WallProtDB*, www.polebio.lrsv.ups-tlse.fr/WallProtDB/); 265 genes encoding proteins involved in generation of precursor metabolites and energy (amigo.geneontology.org/amigo/term/GO:0006091); and 174 genes encoding proteins shown to be involved in thermotolerance (Supplementary Table S3I). As for the "CWP transcriptomes", "CW-related transcriptomes" were well-separated after PCA (Fig. 5D). On the contrary, a weaker separation according to ecotype or temperature variables could be observed for both the "energy transcriptomes" and the "thermotolerance transcriptomes" (Fig. 5E–F). In all cases, Sha rosettes seemed to be less impacted by the growth temperature. Indeed, the range of variation observed between transcripts levels depending of temperatures was smaller for Sha than for Col rosettes (Fig. 5B–E).

Besides, interactions between the ecotype and temperature variables were detected in the PCA done with "CWP transcriptomes" (Fig. 5C). The ratio of genes showing an interaction between the ecotype and the temperature was higher for the "CWP transcriptomes" (0.018%) than for "whole transcriptomes" (0.004%) (Supplementary Table S3A, B).

3.5. An integrative analysis to correlate transcriptomics and proteomics data

A list of 43 CWPs showing the most significant differential accumulation levels between the four samples was established. The corresponding levels of transcripts were also tested for significant differences (Supplementary Table S4). The correlation coefficient between the variations in protein and transcript levels was higher than +0.49 for 23 genes. On the contrary, a negative correlation coefficient (lower than −0.10) was observed for 6 genes. Only 14 genes showed significant quantitative differences (ANOVA test lower than 1E-04) for the transcriptomics data, among which seven showed great differences at both

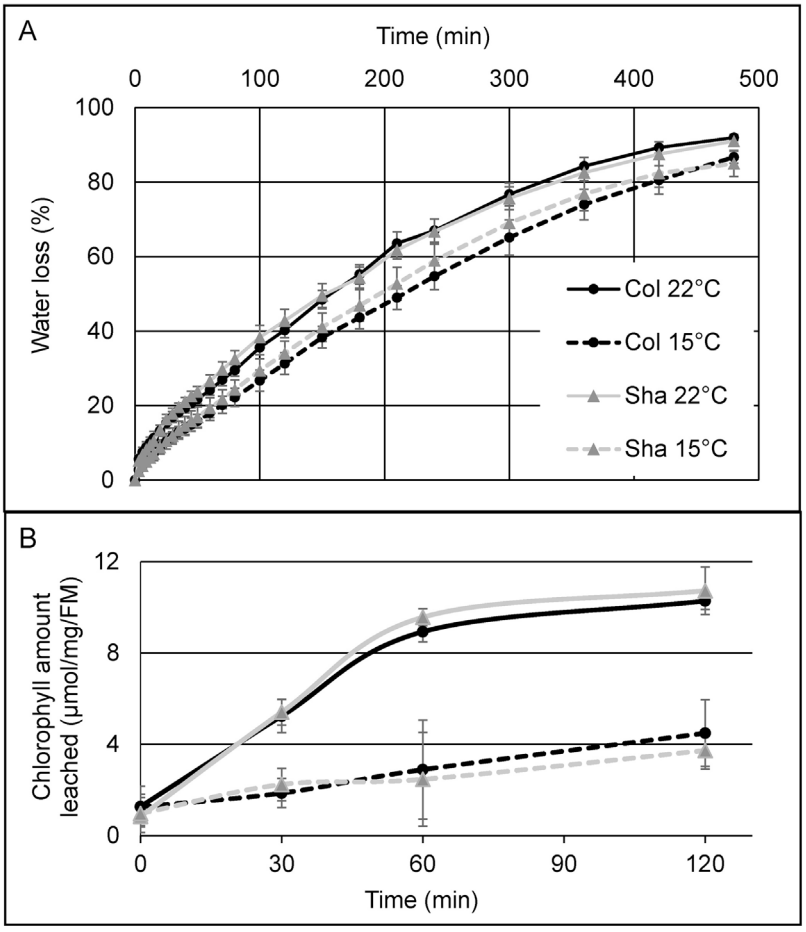


Fig. 3. Cuticle permeability of Col and Sha rosettes grown at 22 °C and 15 °C. Percentage of fresh mass reduction of whole rosettes at 40 °C over time (normalized on the initial fresh mass) (A). Chlorophyll leaching of Col and Sha rosettes of plants grown at 22 °C or 15 °C (B). Measurements have been performed at the bolting stage after 4 and 6 weeks of culture at 22 and 15 °C, respectively (mean ± SD; n = 5–8). Legends are the same for both panels.

the transcriptomics and the proteomics levels: *At2g25510* encoding a protein of unknown function, *At3g14210* encoding a lipase acylhydrolase of the GDSL family, *At4g21650* encoding the AtSBT3.13 Ser protease, *At1g31580* encoding ECS1 involved in resistance mechanisms, *At3g03350* encoding a legume lectin, *At4g22517* encoding a LTP and *At1g58270* encoding a protein of unknown function with a MATH domain. Besides, a gene encoding a GH17 (*At5g56590*) was discriminated for an ecotype effect with transcriptomics data whereas it showed a temperature effect with proteomics data. A gene encoding an α-expansin (*At1g26770*) was discriminated for an ecotype effect with the transcriptomics data and an interaction effect with the proteomics data. Altogether, it was not possible to define a clear correlation between proteomics and transcriptomics data regarding CWPs, suggesting complex regulations of gene expression at post-transcriptomic levels.

Table 1
Monosaccharide composition of rosettes of Col and Sha plants grown at 22 °C or 15 °C. Plants have been harvested at their respective bolting stage. Four cell wall extracts have been obtained (E1–E4) and analyzed by HPAEC Dionex chromatography after acidic hydrolysis to determine their composition in monosaccharides. Means (± SD) of three biological replicates (n = 20 plants per sample) are indicated in µg per 100 mg fresh mass. Values in bold correspond to GalA in E1-2 fractions and Xyl in E3-4 fractions which are specific for the pectin- and hemicelluloses-enriched cell wall extracts. Fucose (Fuc), rhamnose (Rha), arabinose (Ara), galactose (Gal), glucose (Glc), xylose (Xyl) and galacturonic acid (GalA).

Monosaccharide			Fuc	Rha	Ara	Gal	Glc	Xyl	GalA
E1	Col	22 °C	0.8 ± 0.1	1.4 ± 0.3	2.8 ± 0.2	6.6 ± 0.7	3.5 ± 1.2	2.1 ± 0.3	**12.3** ± 0.8
	Col	15 °C	1.3 ± 0.1	1.5 ± 0.3	3.6 ± 0.5	7.4 ± 0.8	15.0 ± 2.9	2.1 ± 0.3	**16.6** ± 2.9
	Sha	22 °C	0.7 ± 0.2	0.9 ± 0.3	1.9 ± 0.4	4.5 ± 0.7	3.7 ± 1.1	1.4 ± 0.1	**9.9** ± 3.8
	Sha	15 °C	1.0 ± 0.0	1.3 ± 0.2	2.8 ± 0.2	5.8 ± 0.3	11.7 ± 4.6	1.7 ± 0.2	**16.0** ± 4.6
E2	Col	22 °C	0.7 ± 0.1	3.4 ± 0.6	3.8 ± 0.6	5.0 ± 1.0	1.3 ± 0.3	1.2 ± 0.3	**66.2** ± 11.7
	Col	15 °C	0.8 ± 0.2	3.3 ± 0.8	5.4 ± 1.0	4.1 ± 0.7	1.5 ± 0.4	1.4 ± 0.2	**71.0** ± 5.7
	Sha	22 °C	0.5 ± 0.1	2.1 ± 0.2	2.3 ± 0.4	3.2 ± 0.5	1.3 ± 0.6	0.8 ± 0.2	**39.5** ± 8.1
	Sha	15 °C	0.7 ± 0.2	3.2 ± 0.6	4.8 ± 0.7	3.8 ± 0.5	2.0 ± 1.0	1.1 ± 0.1	**56.1** ± 10.8
E3	Col	22 °C	1.3 ± 0.1	0.6 ± 0.1	3.8 ± 0.7	13.2 ± 0.8	9.8 ± 0.4	**10.8** ± 1.6	3.3 ± 0.2
	Col	15 °C	1.7 ± 0.2	0.7 ± 0.1	4.7 ± 0.6	13.5 ± 1.2	20.6 ± 2.3	**10.1** ± 1.2	3.2 ± 0.4
	Sha	22 °C	1.1 ± 0.2	0.5 ± 0.1	2.7 ± 0.7	9.6 ± 2.4	9.3 ± 1.5	**10.0** ± 1.9	2.6 ± 0.7
	Sha	15 °C	1.3 ± 0.1	0.9 ± 0.1	3.1 ± 0.1	10.1 ± 1.1	15.7 ± 2.5	**8.8** ± 1.1	2.4 ± 0.5
E4	Col	22 °C	8.0 ± 1.5	0.5 ± 0.1	5.3 ± 1.1	27.7 ± 4.6	64.4 ± 12.8	**62.5** ± 11.2	2.2 ± 0.5
	Col	15 °C	8.3 ± 1.3	0.5 ± 0.1	6.4 ± 1.0	29.2 ± 4.1	72.5 ± 9.8	**59.2** ± 9.6	2.1 ± 0.3
	Sha	22 °C	4.7 ± 1.0	0.3 ± 0.1	3.4 ± 1.0	16.7 ± 3.0	39.1 ± 9.2	**37.9** ± 6.8	1.5 ± 0.6
	Sha	15 °C	5.8 ± 0.4	0.5 ± 0.1	3.8 ± 0.2	19.5 ± 1.7	53.0 ± 5.6	**41.4** ± 3.2	1.3 ± 0.2

Table 2
Reconstruction of cell wall polysaccharide composition from monosaccharide analyses. The formulas were adapted from [38]. The amount of monosaccharide expressed in µg per 100 mg/g fresh material was used to make the calculation (see Table 1). Fucose (Fuc), rhamnose (Rha), arabinose (Ara), galactose (Gal), glucose (Glc), xylose (Xyl), galacturonic acid (GalA), molecular mass of GalA (MGalA), molecular mass of Rha (MRha).

Polysaccharide	Formula
Rhamnogalacturonan I (RGI)	[Rha × (1 + MGalA/MRha)] + Ara + Gal
Homogalacturonan (HG)	GalA − [Rha × (1 + MGalA/MRha)]
Xyloglucan (XG)	Fuc + Glc + Xyl

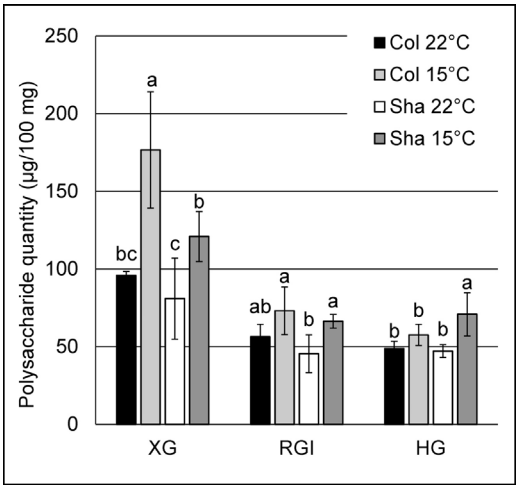


Fig. 4. Reconstruction of three main cell wall polysaccharides from monosaccharide analysis. Monosaccharide composition of Col and Sha plants grown at 22 °C and at 15 °C has been obtained (see Table 1). The formulas used to rebuild the polysaccharides are given in Table 2. Mean values were analysed (mean ± SD) with one-way ANOVA together with Duncan's multiple range test. Significant differences ($P < 0.1$) are marked with different lowercase letters. XG: xyloglucans; RGI: rhamnogalacturonan I; HG: homogalacturonan.

3.6. A focus on six gene sets possibly related to the observed phenotypes

The analysis of the transcriptomes has then been restricted on six gene sets possibly related to the observed phenotypes described above (Supplementary Table S3I): (i) genes encoding annotated CWPs and/or CWPs already identified in CW proteomics studies including predicted plasma membrane receptors (1829 genes) (see *WallProtDB*, www.polebio.lrsv.ups-tlse.fr/WallProtDB/); (ii) genes involved in CW biosynthesis (259 genes); (iii) genes involved in lignin synthesis (176 genes); (iv) genes involved in anthocyanin synthesis (32 genes); (v) genes involved in lipid synthesis (186 genes); and (vi) genes involved in ROS homeostasis (293 genes) [44].

(i) 199 out of the 1859 genes encoding CWPs showed significantly different levels of transcript accumulation (Supplementary Table S3C). Among those, there were 39 genes encoding known or predicted receptor kinases. The ecotype effect on transcriptional activity was predominant with 135 genes more expressed in Sha than in Col rosettes and 49 genes showing the opposite effect. Regarding the temperature effect, only 29 genes were found to be differentially expressed, with 18 at higher level at 15 °C than at 22 °C, and 11 at 22 °C than at 15 °C. However, the differences observed in response to temperature were not as great as those observed between the two ecotypes, with factors of induction lower than 10. The combined effects of the two variables, i.e ecotype and temperature, were observed for 9 genes. Only 5 of the genes showing the highest factors of induction corresponded to proteins having significantly different levels of accumulation (Table 4).

(ii) 22 out of the 259 genes involved in the biosynthesis of CW components had different level of expression in the four samples (Supplementary Table S3D). Among them, three genes involved in the nucleotide-sugar interconversion pathways had higher levels of transcripts in Sha rosettes (*At1g08200*, AXS2; *At5g59290*, SUD2; *At5g39320*, UGD1). This was also the case for two genes involved in the post-translational modifications of hydroxyproline (Hyp)-rich glycoproteins (HRGPs) like extensins or arabinogalactan

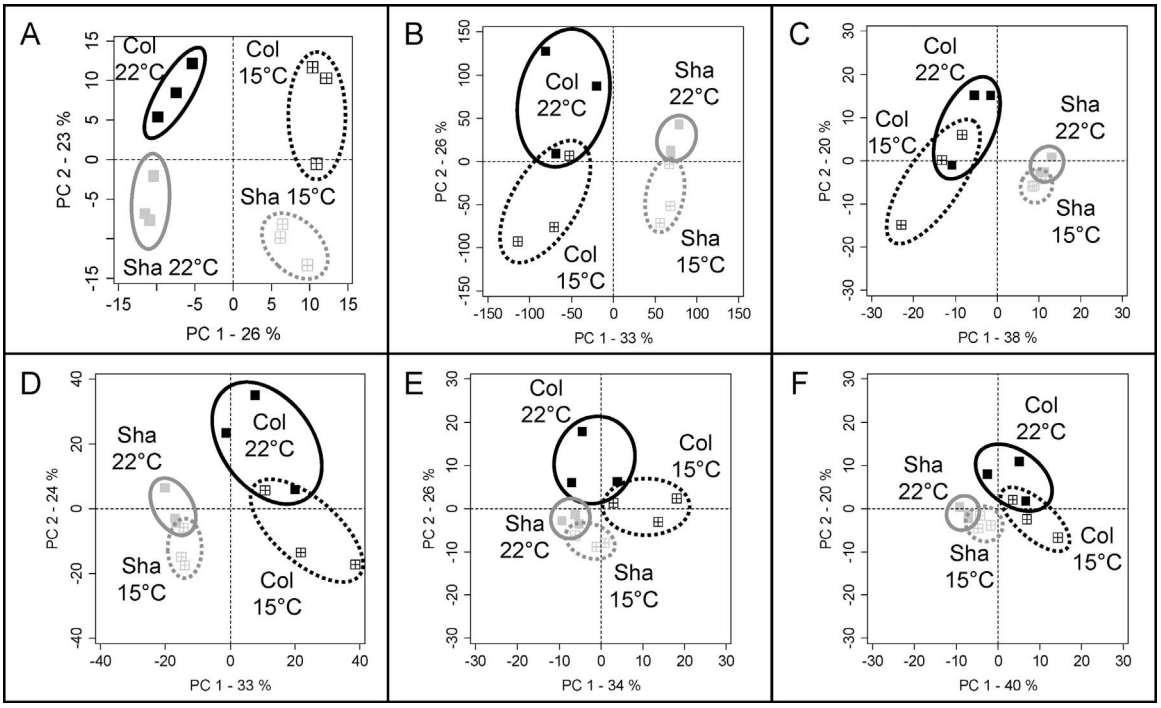


Fig. 5. Overall comparison of cell wall proteomes and transcriptomes of Col and Sha plants grown at 22 °C and 15 °C. Scaled PCA of the proteomics quantitative analysis (A), the whole transcriptomics data (B), the "CWP transcriptomes" (C), the "CW-related-transcriptomes" (D), the "energy transcriptomes" (E) and the "thermotolerance transcriptomes" (F). Values for x and y axes are those of PC1 and PC2, respectively.

Table 3
Synopsis of CWPs preferentially accumulated in one ecotype or at a given growth temperature. The statistical analysis of quantitative proteomics data has been performed with an ANOVA test (see Supplementary Table S2B). Functional annotation of CWPs is given according to *WallProtDB* (www.polebio.lrsv.ups-tlse.fr/WallProtDB/). +means that the CWP was more abundant in one ecotype or at one growth temperature.

AGI code	Putative function	Col	Sha	22 °C	15 °C
At3g57240	GH17 (β-1,3-glucosidase)	+			
At2g28100	GH29 (α-L-fucosidase)	+			
At5g63800	GH35 (β-galactosidase) (AtBGAL6)	+			
At1g67750	PLL1 (pectate lyase) (AtPLL1)	+			
At4g21650	Ser protease (AtSBT3.13)	+			
At2g39850	Ser protease (AtSBT4.1)	+			
At4g11290	class III peroxidase (AtPrx39)	+			
At3g14210	GDSL lipase acylhydrolase (ESM1)	+			
At2g03980	GDSL lipase acylhydrolase	+			
At5g03350	legume lectin	+			
At1g78820	curculin-like lectin	+			
At2g27190	purple acid phosphatase (AtPAP12)	+			
At2g45470	fasciclin arabinogalactan protein (AtFLA8)	+			
At5g48375	GH1 (β-glucosidase)		+		
At3g45940	GH31 (α-glucosidase)		+		
At2g28470	GH35 (β-galactosidase) (AtBGAL8)		+		
At3g45970	expansin-like A (AtEXLA1)		+		
At4g02330	CE8 (pectin methylesterase) (AtPME41)		+		
At5g19100	Asp protease		+		
At5g10760	Asp protease		+		
At3g24480	leucine-rich repeat extensin (AtLRX4)		+		
At4g22517	lipid transfer protein (LTP)		+		
At4g22520	lipid transfer protein (LTP)		+		
At3g07130	purple acid phosphatase (AtPAP15)		+		
At5g60950	COBRA-like protein (AtCOBL5)		+		
At5g56590	GH17 (β-1,3-glucosidase)			+	
At3g55260	GH20 (N-acetyl-β-glucosaminidase)			+	
At3g16850	GH28 (polygalacturonase)			+	
At2g41850	GH28 (polygalacturonase)			+	
At5g43060	Cys protease			+	
At2g22990	Ser carboxypeptidase (AtSCPL8)			+	
At1g27950	lipid transfer protein (AtLTPg4)			+	
At2g41850	GH28 (polygalacturonase)				+
At5g07830	GH79 (*endo*-β-glucuronidase/ heparanase)				+
At1g20190	α-expansin (AtEXPA11)				+
At3g59010	CE8 (pectin methylesterase) (AtPME35)				+
At5g10770	Asp protease				+
At5g45890	Cys protease				+
At5g45650	Ser protease (AtSBT5.6)				+

Table 4
Synopsis of genes having higher levels of transcripts in each ecotype. The statistical analysis of transcriptomics data has been performed with an ANOVA test (see Supplementary Table S3A). Functional annotation of genes is given according to *WallProtDB*. + means that transcripts were at least 10 times more abundant in one ecotype. When the AGI codes are in bold, it means that the proteins they encode were also found to be more abundant in the same sample (see Table 3).

AGI code	Putative function	Col	Sha
At4g21650	Ser protease (AtSBT3.13)	+	
At3g14210	GDSL lipase acylhydrolase	+	
At5g03350	legume lectin	+	
At3g47295	tyrosine-sulfated glycopeptide 2 (AtPSK2)	+	
At5g59670	leucine-rich receptor kinase (LRR I subfamily)	+	
At3g13065	leucine-rich receptor kinase (LRRV subfamily)	+	
At3g21630	LysM receptor-like kinase	+	
At1g14700	purple acid phosphatase (AtPAP3)	+	
At1g31580	ECS1 (role in resistance to *Xanthomonas campestris* strain Xcc75	+	
At2g25510	expressed protein	+	
At5g24460	expressed protein	+	
At1g58270	expressed protein (MATH domain)	+	
At3g09260	GH1 (β-glucosidase)		+
At1g02800	GH9 (endoglucanase) (AtCEL2)		+
At2g03090	α-expansin (AtEXPA15)		+
At5g19100	Asp protease		+
At5g25090	early nodulin (AtEN6)		+
At2g36985	DEVIL16/ROTUNDIFOLIA4 (DVL/ROT)		+
At3g60900	fasciclin-like arabinogalactan protein (FLA10)		+
At1g28290	arabinogalactan protein/PRP (AtAGP31)		+
At3g01700	arabinogalactan protein (AtAGP11)		+
At4g22517	lipid transfer protein (LTP)		+
At3g08770	lipid transfer protein (AtLTP1.6, AtLTP6)		+
At4g30140	GDSL lipase acylhydrolase (AtCDEF1)		+
At4g11190	dirigent protein (AtDIR13)		+
At2g28790	thaumatin (PR5)		+
At1g18250	thaumatin (PR5, ATLP1)		+
At3g15950	expressed protein		+

proteins (AGPs): *At4g21060* (AtGALT2, galactosyl transferase on Hyp residues of AGPs) [45], *At1g75120* (AtRRA1, arabinosyl transferase on Hyp residues of HRGPs) [46]. On the contrary, At2g25300 (AtGALT3) was expressed at a higher level in Col rosettes [47]. A gene encoding a Pro-hydroxylase (*At3g28480*) was induced during growth at 15 °C, as well as in response to anoxia and hypoxia [48]. Besides, the transcript levels of several genes involved in the biosynthesis of CW polysaccharides were higher in Sha rosettes: *At1g02720* (AtGATL5, involved in the synthesis of RGI) [49], *At5g13000* (AtCALS3, involved in the synthesis of callose in plasmodesmata) [50], *At1g02730* (AtCSLD5, involved in xylan and HG synthesis) [51], and At5g22740 (AtCSLA2, involved in mannan synthesis) [52]. Finally, *At2g03220* (MUR2, involved in XG synthesis) was expressed at higher level at 15 °C consistently with the observed higher level of XG in our experiments.

(iii) 14 out of the 176 genes involved in the lignin synthesis pathway displayed differential expression in the four samples. The ecotype effect was predominant with the up-regulation of several genes in Sha rosettes (Supplementary Table S3E): three genes involved in the biosynthesis of monolignols (*At5g38120*, At4CL-like9; *At4g34050*, AtCCoAOMT1; *At2g33590*, AtCCR-like3); four genes encoding class III peroxidases (Prxs) (*At2g22420*; *At3g21770*; *At4g33420*; *At5g64110*); two genes encoding dirigent proteins (*At3g13650*, AtDIR7; *At4g11190*, AtDIR13); and one gene encoding a laccase (*At2g46570*, AtLAC6). The proteins encoded by the latter genes were described as possibly involved in the polymerization of monolignols [53]. Only two genes encoding proteins involved in monolignol synthesis were differentially regulated for the temperature condition, but with opposite effects. At 15 °C, the cinnamoyl-CoA reductase like 5 gene (*At5g58490,* AtCCR-like 5) was up-regulated, whereas the caffeic acid O-methyltransferase gene (*At5g54160*, AtCOMT1) was down-regulated. Finally, *At4g33420* (AtPrx47) was down-regulated at 22 °C. These changes in gene expression could be associated with the observed higher content in hydrophobic compounds observed in Sha petiolesat 15 °C.

(iv) The lignin and anthocyanin biosynthesis pathways share common enzymes at their initial steps, such as AtCOMT1 (*At5g54160*) which was down-regulated at 15 °C. This adjustment could allow favoring the flavonoid biosynthesis pathway compared to the lignin pathway at low temperature. Consistently, a gene encoding a flavonol synthase (*At5g08640*, FLS1) catalysing the formation of flavonols from dihydroflavonols had a much higher level of transcript accumulation in Sha (Supplementary Table S3F).

(v) Several of the genes related to lipid metabolism were up-regulated in Sha plants and especially those encoding GDSL lipases acylhydrolases (10/13) (Supplementary Table S3G). Only *At3g14210* (ESM1) was strongly up-regulated in Col rosettes whereas *At1g33811* and *At4g18970* were down-regulated at 15 °C. Four genes encoding LTPs had higher levels of transcripts in Sha (*At5g64080*, AtLTPg31; *At3g08770*, AtLTP6; *At2g13820*, AtLTPg11; *At4g22490*, LTP type 6). Two genes encoding glycerophosphodiester phosphodiesterases have opposite regulations: *At1g66970* (GDPDL1) and *At5g55480* (GDPDL4) were more

expressed in Col and Sha rosettes, respectively. GDPDL1 was shown to play a role in the rigidity of the cell wall [54]. The strong induction of many of these genes in Sha rosettes could be related to the observed greater thickness of the cuticle.

(vi) Among the 293 genes involved in ROS homeostasis, only 25 were differentially expressed between the four samples, that is less than 10% (Supplementary Table S3H). Among them, four were up-regulated at 15 °C (*At1g65980, At4g11600, At5g0426, At5g40370*) and four were up-regulated at 22 °C (*At5g15350*, *At4g25100*, *At3g62930, At3g06730*) independently of the ecotypes. In addition, one gene was specifically induced in Sha rosettes at 22 °C (*At4g33420*) and two in Col rosettes at 22 °C (*At4g15690*, *At4g15700*). Together with the fact that the four samples could not be distinguished using the thermotolerance gene box (see above), these results suggested that the plants were not submitted to a cold- or a heat-shock stress by growing at sub-optimal temperatures.

4. Discussion

This work has allowed comparing in detail the phenotypes of rosettes of Col and Sha which are two ecotypes of *A. thaliana* initially originating from contrasting environments. Beyond a macro-phenotyping, the aim of this study was to identify molecular players in the acclimation of both ecotypes to sub-optimal growth temperature conditions. Complex responses have been found and are summarized in Fig. 6. Changes observed in Sha compared to Col rosettes as well at 15 °C in either ecotype have been highlighted. The results obtained with transcriptomics and proteomics approaches have been combined, considering that they are complementary even if some discrepancies were observed for genes encoding CWPs. Such finding has already been described and possible explanations have been discussed such as (i) post-transcriptional levels of regulations at the level of translation or protein turn-over and (ii) changes in protein accessibility due to CW polymer rearrangements [55–58]. Altogether, we have observed modifications of molecular mechanisms occurring both inside the plant cell and in the CW. The transcription of genes involved in the biosynthesis of CW components like polysaccharides (callose, hemicelluloses and pectins), structural CWPs (HRGPs) and monolignols is affected. We also observed the accumulation of CWPs involved in the remodeling of CW polysaccharide and protein networks as well as in polymerization of monolignols into lignin. This was correlated to an increase in XG and pectin content of CWs. Finally, consistantly with the increase in the thickness of the cuticle layer and with the decrease of water-loss and chlorophyll leakage, we have observed an increase in the transcription level of genes possibly involved in lipid deposition or polymerization. Remarkably, many genes encoding receptor kinases had modified transcript levels, suggesting that Col and Sha plants perceived their environment in different ways depending on their growth conditions. Besides, the level of accumulation of transcripts encoding genes involved in the biosynthesis of flavonoids was found to be increased in correlation with the increased level of anthocyanin accumulation.

Global comparisons of the quantitative proteomics and transcriptomics data allowed to clearly distinguish the four samples, and especially to highlight the ecotype effect. The importance of the variations observed in the accumulation of transcripts encoding CW-related proteins was a clear outcome of this study. These comparisons also showed that neither the Col plants were submitted to a cold stress when grown at 15 °C, nor the Sha plants to a heat stress when grown at 22 °C. Indeed, the bolting point was the same for the two ecotypes for each growth conditions and the levels of transcripts of a set of thermotolerance genes as well as those of genes involved in ROS homeostasis were not significantly affected. In this work, many genes/CWPs could be identified as candidates playing roles in the acclimation of plants to sub-optimal growth temperatures in laboratory conditions. The discussion will thus be focused on some protein families mainly in relation to cell wall polysaccharides, cuticle and lignin modification.

Morphological traits of low temperature acclimation have been

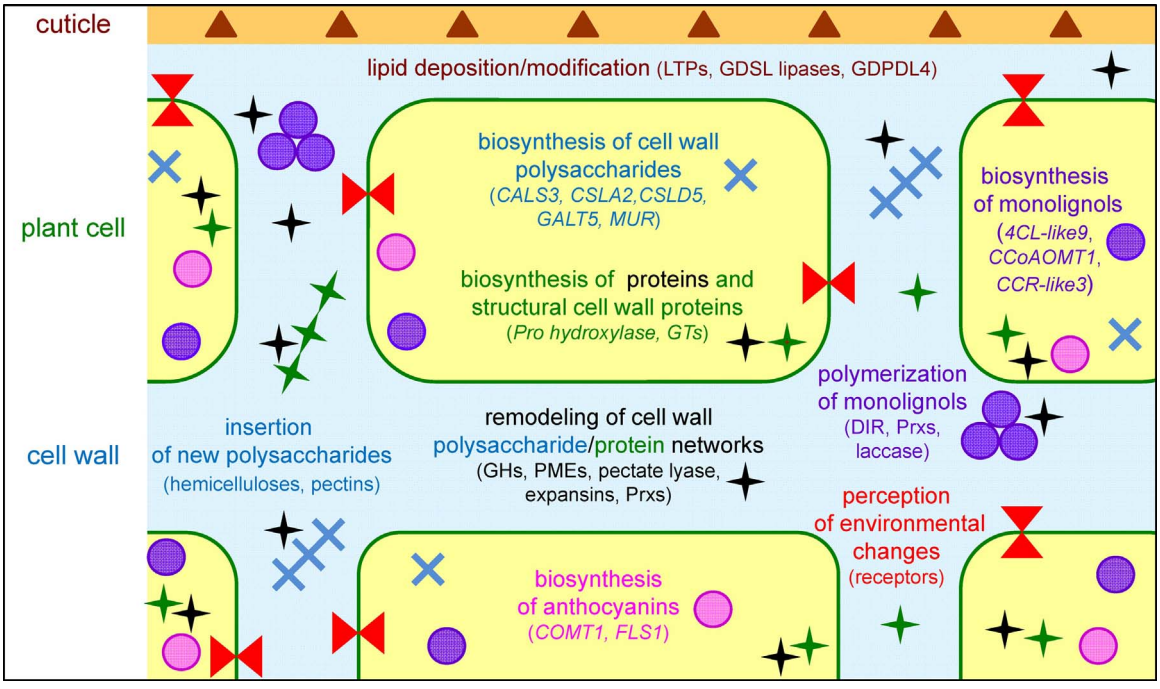


Fig. 6. Overview of the changes observed in rosettes in response to sub-optimal temperature growth conditions: A focus on molecular events occurring at 15 °C and/or in the Sha ecotype. Two compartments are highlighted: intracellular (yellow), extracellular (cell wall proper in blue, cuticle proper in dark yellow). The names of some proteins/genes having increased level of accumulation or of transcripts are given between brackets. The texts and the corresponding symbols are in the same color. Circles represent metabolites, e.g. monolignols or anthocyanins. Blue crosses correspond to polysaccharides. Brown triangles correspond to cuticular lipids. Black stars correspond to secreted proteins acting on cell wall components. Green stars correspond to structural proteins which can be crossed linked in cell walls. Red egg timer-like shapes represent receptors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

detected after growth at 15 °C, and they could be correlated to altitude growth conditions [18,59]. The increase of leaf density, observed for plants grown at 15 °C and more significantly in Col is correlated with (i) an increase of the amount of pectins and XG detected in their CWs, (ii) an increased vessels number and (iii) a reinforced cuticle. Accordingly, a great variety of genes encoding proteins involved in the biosynthesis of CW components and CWPs, which could contribute to the CW integrity and plasticity, had higher levels of transcripts at 15 °C vs 22 °C, but also in Sha vs Col rosettes. Quantitative proteomics has also revealed the differential accumulation of CWPs known to be involved in polysaccharide network remodeling such as GHs, PMEs, a pectate lyase and expansins [11,13,60]. The levels of accumulation of several proteases, such as Ser proteases, Asp proteases, Cys proteases and a Ser carboxypeptidase were also affected in either ecotype or at either growth temperature. Such proteins could participate in protein maturation, protein turnover or signaling [61]. Besides, the genes encoding PR5 or AGP31, known to interact with CW polysaccharides [12,62], were transcribed at higher levels in Sha plants.

The reinforced cuticle layer and the decrease of chlorophyll leaching and water-loss, observed in rosettes of plants grown at 15 °C, are consistent with a modification of cuticular permeability due to changes in structure and chemical composition of the cuticle layer at 15 °C. Variations in the level of accumulation of several CWPs and/or transcripts encoding proteins possibly involved in lipid metabolism have been observed in Sha rosettes. Among them, there are 10 genes encoding proteins of the GDSL lipase acylhydrolase family, four LTPs (AtLTPg31, AtLTP6, AtLTPg11, LTP type 6) and a glycerophosphodiester phosphodiesterase (GDPDL4). Regarding quantitative proteomics data, two LTPs were found to be more abundant in Sha rosettes (At4g22517, LTP type 6). On the contrary, despite the thickening of the cuticle of Col leaves at 15 °C, none of the genes related to lipid metabolism showed modified levels of transcripts at 15 °C compared to 22 °C. However, *GDPDL1* had higher level of transcripts in Col rosettes, and ESM1 (At3g14210) and At2g03980 encoding GDSL lipase acylhydrolases were more abundant in Col rosettes. Members of these three CWP families have been shown to play roles in the cell wall organization and/or biogenesis of the cuticle [54,63–67]. Interestingly, *ECS1*, which belongs to the same co-expression network as *ESM1* according to the ATTED database (atted.jp) [68], had high levels of transcripts and proteins in Col rosettes. *ECS1* is not known to be involved in lipid metabolism, but rather in resistance to *Xanthomonas campestris* (Xcc750) [69]. It could thus play a role in cuticle synthesis or more generally in CW remodelling in response to biotic and abiotic stresses.

Lignin and anthocyanin biosynthesis pathways which share common starting shikimate pathway enzymes were modified at 15 °C whatever the ecotype. Only two genes involved in lignin monomer synthesis were differentially regulated for the temperature conditions (CCR-like 5 and COMT1), but with opposite profiles. The observed variations in transcript accumulation for these two genes at 15 °C may favor the biosynthesis of H monolignol units [53,70]. Genes possibly involved in monolignol polymerization such as Prxs [71] and DIRs [72] were up-regulated in Sha rosettes. This result could be correlated with the higher number of vessel elements detected in the petioles of Sha leaves. Regarding the anthocyanin biosynthesis pathway, *FLS1* had a much higher level of transcripts in Sha than in Col rosettes. The *fls1-3* mutant was shown to accumulate higher anthocyanin levels than wild-type plants [73]. Then, the lower level of expression of *FLS1* in Col plants could explain the observed higher level of anthocyanin accumulation. In a recent study, an increase in anthocyanin accumulation could be correlated in a series of *A. thaliana* accessions grown at 4 °C during 14 days and was assumed to play a role in cold acclimation and freezing tolerance [74].

Altogether, the macro- and micro-phenotyping combined with omics analyses, have shown more differences between ecotypes than between temperature growth conditions, revealing genetic differences between Col and Sha plants. These observations could be related to the original natural environments of both ecotypes. As expected, major changes have been observed regarding the biosynthesis or the remodelling of CW components. In addition to the identification of several candidate genes/proteins of interest, we have also demonstrated that the integrative analysis was complementary to univariate statistical tests. It has allowed validating the candidates provided by simple analyses, but also identifying other candidates. Omics and integrative analyses are extremely powerful to propose new hypotheses, especially with regard to the regulation of whole metabolic pathways such as those leading to CW biosynthesis. Although this study conducted in laboratory conditions does not allow us to conclude on the ability of both ecotypes to adapt to their natural environments, it nevertheless demonstrates that CW remodeling is an important factor influencing growth at sub-optimal growth temperature. It also identified candidates genes of interest for future functional analyses. Studies on a larger set of newly characterized altitudinal ecotypes are underway to evaluate the contribution of the genetic background to the acclimation to different temperature growth conditions.

Acknowledgments

The authors are thankful to the Paul Sabatier-Toulouse 3 University and to the *Centre National de la Recherche Scientifique* (CNRS) for their financial support. This work was also granted by the IdEx project (ANR-11-IDEX-0002-02). HD was supported by a PhD grant from the Midi-Pyrénées Region and the Federal University of Toulouse. The authors are grateful to Prof Philippe Besse (*Institut de Mathématiques*, Toulouse) for his critical advices regarding the integrative analysis, Prof Christophe Roux (LRSV) for his support for the RNA seq analysis and Maria Dauriac for the english editing.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2017.07.015.

References

[1] G.J. Warren, Cold stress: manipulating freezing tolerance in plants, Curr. Biol. 8 (1998) R514–R-516.
[2] M. Gottfried, H. Pauli, A. Futschik, M. Akhalkatsi, et al., Continent-wide response of mountain vegetation to climate change, Nat. Clim. Change 2 (2012) 111–115.
[3] S.B. Gray, S.M. Brady, Plant developmental responses to climate change, Dev. Biol. 419 (2016) 64–77.
[4] M.H. Hoffmann, Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae), J. Biogeogr. 29 (2002) 125–134.
[5] E.M. Meyerowitz, Prehistory and history of Arabidopsis research, Plant Physiol. 125 (2001) 15–19.
[6] B. Müller, U. Grossniklaus, Model organisms—a historical perspective, J. Proteomics 73 (2010) 2054–2063.
[7] C. Trontin, S. Tisné, L. Bach, O. Loudet, What does Arabidopsis natural variation teach us (and does not teach us) about adaptation in plants? Curr. Opin. Plant Biol. 14 (2011) 225–231.
[8] H. Le Gall, F. Philippe, J.M. Domon, F. Gillet, et al., Cell wall metabolism in response to abiotic stress, Plants 4 (2015) 112–166.
[9] K. Houston, M.R. Tucker, J. Chowdhury, N. Shirley, A. Little, The plant cell wall: a complex and dynamic structure as revealed by the responses of genes under stress conditions, Front. Plant Sci. 7 (2016) 984.
[10] R. Sasidharan, L. Voesenek, R. Pierik, Cell wall modifying proteins mediate plant acclimatization to biotic and abiotic stresses, CRC Crit. Rev. Plant Sci. 30 (2011) 548–562.
[11] L. Franková, S.C. Fry, Biochemistry and physiological roles of enzymes that 'cut and paste' plant cell-wall polysaccharides, J. Exp. Bot. 64 (2013) 3519–3550.
[12] M. Hijazi, D. Roujol, H. Nguyen-Kim, L. Del Rocio Cisneros Castillo, et al., Arabinogalactan protein 31 (AGP31), a putative network-forming protein in Arabidopsis thaliana cell walls? Ann. Bot. 114 (2014) 1087–1097.
[13] D.J. Cosgrove, Plant cell wall extensibility: connecting plant cell growth with cell wall structure, mechanics, and the action of wall-modifying enzymes, J. Exp. Bot. 67 (2016) 463–476.
[14] 1001 Genomes Consortium, 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*, Cell 166 (2016) 481–491.
[15] P. Barah, N.D. Jayavelu, S. Rasmussen, H.B. Nielsen, et al., Genome-scale cold stress response regulatory networks in ten *Arabidopsis thaliana* ecotypes, BMC Genomics 14 (2013) 722.

[16] A. Janská, P. Marsík, S. Zelenková, J. Ovesná, Cold stress and acclimation—what is important for metabolic adjustment? Plant Biol. (Stuttg.) 12 (2010) 395–405.
[17] P. Johnová, J. Skalák, I. Saiz-Fernández, B. Brzobohatý, Plant responses to ambient temperature fluctuations and water-limiting conditions: a proteome-wide perspective, Biochim. Biophys. Acta 1864 (2016) 916–931.
[18] J.J. Stewart, B. Demmig-Adams, C.M. Cohu, C.A. Wenzl, et al., Growth temperature impact on leaf form and function in *Arabidopsis thaliana* ecotypes from northern and southern Europe, Plant Cell Environ. 39 (2016) 1549–1558.
[19] W.W. Adams, J.J. Stewart, C.M. Cohu, O. Muller, B. Demmig-Adams, Habitat temperature and precipitation of *Arabidopsis thaliana* ecotypes determine the response of foliar vasculature, photosynthesis, and transpiration to growth temperature, Front. Plant Sci. 7 (2016) 1026.
[20] D. Weigel, R. Mott, The 1001 genomes project for *Arabidopsis thaliana*, Genome Biol. 10 (2009).
[21] R.J. Hijmans, S.E. Cameron, J.L. Parra, P.G. Jones, A. Jarvis, Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol. 25 (2005) 1965–1978.
[22] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (2012) 671–675.
[23] A.L. Mancinelli, A.M. Hoff, M. Cottrell, Anthocyanin production in chl-rich and chl-poor seedlings, Plant Physiol. 86 (1988) 652–654.
[24] E. Francoz, P. Ranocha, C. Pernot, A. Le Ru, et al., Complementarity of medium-throughput in situ RNA hybridization and tissue-specific transcriptomics: case study of Arabidopsis seed development kinetics, Sci. Rep. 6 (2016) 24644.
[25] S.E. Ruzin, Plant Microtechnique and Microscopy, Oxford University Press, New York, 1999.
[26] I. Vazquez-Cooz, R.W. Meyer, A differential staining method to identify lignified and unlignified tissues, Biotech. Histochem. 77 (2002) 277–282.
[27] E. Srebotnik, K. Messner, A simple method that uses differential staining and light microscopy to assess the selectivity of wood delignification by white rot fungi, Appl. Environ. Microbiol. 60 (1994) 1383–1386.
[28] C. Cosio, C. Dunand, Transcriptome analysis of various flower and silique development stages indicates a set of class III peroxidase genes potentially involved in pod shattering in *Arabidopsis thaliana*, BMC Genomics 11 (2010) 528.
[29] L. Feiz, M. Irshad, R.F. Pont-Lezica, H. Canut, E. Jamet, Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from Arabidopsis hypocotyls, Plant Methods 2 (2006) 10.
[30] M. Irshad, H. Canut, G. Borderies, R. Pont-Lezica, E. Jamet, A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers, BMC Plant Biol. 8 (2008) 94.
[31] R. Redgwell, R. Selvendran, Structural features of cell-wall polysaccharides of onion *Allium cepa*, Carbohydr. Res. 157 (1986) 183–199.
[32] V. Hervé, H. Duruflé, H. San Clemente, C. Albenne, et al., An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes, Proteomics 16 (2016) 3183–3187.
[33] O. Langella, B. Valot, T. Balliau, M. Blein-Nicolas, et al., X!TandemPipeline: a tool to manage sequence redundancy for protein inference and phosphosite identification, J. Proteome Res. 16 (2017) 494–503.
[34] H. San Clemente, R. Pont-Lezica, E. Jamet, Bioinformatics as a tool for assessing the quality of sub-cellular proteomic strategies and inferring functions of proteins: plant cell wall proteomics as a test case, Bioinf. Biol. Insights 3 (2009) 15.
[35] C. Albenne, H. Canut, E. Jamet, Plant cell wall proteomics: the leadership of *Arabidopsis thaliana*, Front. Plant Sci. 4 (2013) 111.
[36] B. Valot, O. Langella, E. Nano, M. Zivy, MassChroQ: a versatile tool for mass spectrometry quantification, Proteomics 11 (2011) 3572–3577.
[37] K.A. Lê Cao, S. Boitard, P. Besse, Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems, BMC Bioinf. 12 (2011) 253.
[38] K. Houben, R. Jolie, I. Fraeye, A. Van Loey, M. Hendrickx, Comparative study of the cell wall composition of broccoli, carrot, and tomato: structural characterization of the extractable pectins and hemicelluloses, Carbohydr. Res. 346 (2011) 1105–1111.
[39] G. Boudart, E. Jamet, M. Rossignol, C. Lafitte, et al., Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: identification by mass spectrometry and bioinformatics, Proteomics 5 (2005) 212–221.
[40] V. Hervé, H. Duruflé, H. San Clemente, C. Albenne, et al., An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes, Proteomics 16 (2016) 3183–3187.
[41] R.P. Haslam, A.L. Downie, M. Raveton, K. Gallardo, et al., The assessment of enriched apoplastic extracts using proteomic approaches, Ann. Appl. Biol. 143 (2003) 81–91.
[42] A.R. Trentin, M. Pivato, S.M. Mehdi, L.E. Barnabas, et al., Proteome readjustments in the apoplastic space of *Arabidopsis thaliana ggt1* mutant leaves exposed to UV-B radiation, Front. Plant Sci. 6 (2015) 128.
[43] N. Sultana, H.V. Florance, A. Johns, N. Smirnoff, Ascorbate deficiency influences the leaf cell wall glycoproteome in *Arabidopsis thaliana*, Plant Cell Environ. 38 (2015) 375–384.
[44] N. Fawal, Q. Li, B. Savelli, M. Brette, et al., PeroxiBase: a database for large-scale evolutionary analysis of peroxidases, Nucleic Acids Res. 41 (2013) D441–444.
[45] D. Basu, Y. Liang, X. Liu, K. Himmeldirk, et al., Functional identification of a hydroxyproline-*O*-galactosyltransferase specific for arabinogalactan protein biosynthesis in Arabidopsis, J. Biol. Chem. 288 (2013) 10132–10143.
[46] S.M. Velasquez, M.M. Ricardi, J.G. Dorosz, P.V. Fernandez, et al., *O*-glycosylated cell wall proteins are essential in root hair growth, Science 332 (2011) 1401–1403.
[47] M. Ogawa-Ohnishi, Y. Matsubayashi, Identification of three potent hydroxyproline *O*-galactosyltransferases in Arabidopsis, Plant J. 81 (2015) 736–746.
[48] F. Vlad, T. Spano, D. Vlad, F.B. Daher, et al., Involvement of Arabidopsis prolyl 4 hydroxylases in hypoxia, anoxia and mechanical wounding, Plant Signal. Behav. 2 (2007) 368–369.
[49] Y. Kong, G. Zhou, A.A. Abdeen, J. Schafhauser, et al., GALACTURONOSYLTRANSFERASE-LIKE5 is involved in the production of Arabidopsis seed coat mucilage, Plant Physiol. 163 (2013) 1203–1217.
[50] I. Sevilem, S. Miyashima, Y. Helariutta, Cell-to-cell communication via plasmodesmata in vascular plants, Cell Adh. Migr. 7 (2013) 27–32.
[51] A.J. Bernal, J.K. Jensen, J. Harholt, S. Sørensen, et al., Disruption of *ATCSLD5* results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in Arabidopsis, Plant J. 52 (2007) 791–802.
[52] A.H. Liepman, C.G. Wilkerson, K. Keegstra, Expression of cellulose synthase-like (*Csl*) genes in insect cells reveals that *CslA* family members encode mannan synthases, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 2221–2226.
[53] S. Besseau, L. Hoffmann, P. Geoffroy, C. Lapierre, et al., Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth, Plant Cell 19 (2007) 148–162.
[54] S. Hayashi, T. Ishii, T. Matsunaga, R. Tominaga, et al., The glycerophosphoryl diester phosphodiesterase-like proteins SHV3 and its homologs play important roles in cell wall organization, Plant Cell Physiol. 49 (2008) 1522–1535.
[55] T. Maier, M. Güell, L. Serrano, Correlation of mRNA and protein in complex biological samples, FEBS Lett. 583 (2009) 3966–3973.
[56] A. Battle, Z. Khan, S.H. Wang, A. Mitrano, et al., Impact of regulatory variation from RNA to protein, Science 347 (2015) 664–667.
[57] Z. Minic, E. Jamet, H. San-Clemente, S. Pelletier, et al., Transcriptomic analysis of Arabidopsis developing stems: a close-up on cell wall genes, BMC Plant Biol. 9 (2009) 6.
[58] E. Jamet, D. Roujol, H. San Clemente, M. Irshad, et al., Cell wall biogenesis of *Arabidopsis thaliana* elongating cells: transcriptomics complements proteomics, BMC Genomics 10 (2009) 505.
[59] A. Montesinos-Navarro, J. Wig, F.X. Pico, S.J. Tonsor, *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude, New Phytol. 189 (2011) 282–294.
[60] G. Levesque-Tremblay, J. Pelloux, S.A. Braybrook, K. Müller, Tuning of pectin methylesterification: consequences for cell wall biomechanics and development, Planta 242 (2015) 791–811.
[61] R.A. van der Hoorn, Plant proteases: from phenotypes to molecular mechanisms, Annu. Rev. Plant Biol. 59 (2008) 191–223.
[62] R.I. Osmond, M. Hrmova, F. Fontaine, A. Imberty, G.B. Fincher, Binding interactions between barley thaumatin-like proteins and (1,3)-beta-d-glucans. Kinetics, specificity, structural analysis and biological implications, Eur. J. Biochem. 268 (2001) 4190–4199.
[63] A. Debono, T.H. Yeats, J.K. Rose, D. Bird, et al., Arabidopsis LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface, Plant Cell 21 (2009) 1230–1238.
[64] S.B. Lee, Y.S. Go, H.J. Bae, J.H. Park, et al., Disruption of glycosylphosphatidylinositol-anchored lipid transfer protein gene altered cuticular lipid composition, increased plastoglobules, and enhanced susceptibility to infection by the fungal pathogen *Alternaria brassicicola*, Plant Physiol. 150 (2009) 42–54.
[65] J. Petit, C. Bres, D. Just, V. Garcia, et al., Analyses of tomato fruit brightness mutants uncover both cutin-deficient and cutin-abundant mutants and a new hypomorphic allele of GDSL lipase, Plant Physiol. 164 (2014) 888–906.
[66] T.H. Yeats, K.J. Howe, A.J. Matas, G.J. Buda, et al., Mining the surface proteome of tomato (*Solanum lycopersicum*) fruit for proteins associated with cuticle biogenesis, J. Exp. Bot. 61 (2010) 3759–3771.
[67] A.L. Girard, F. Mounet, M. Lemaire-Chamley, C. Gaillard, et al., Tomato GDSL1 is required for cutin deposition in the fruit cuticle, Plant Cell 24 (2012) 3119–3134.
[68] Y. Aoki, Y. Okamura, S. Tadaka, K. Kinoshita, T. Obayashi, ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression, Plant Cell Physiol. 57 (2016) e5.
[69] W. Aufsatz, D. Amry, C. Grimm, The *ECS1* gene of Arabidopsis encodes a plant cell wall-associated protein and is potentially linked to a locus influencing resistance to *Xanthomonas campestris*, Plant Mol. Biol. 38 (1998) 965–976.
[70] H. Ji, Y. Wang, C. Cloix, K. Li, et al., The Arabidopsis RCC1 family protein TCF1 regulates freezing tolerance and cold acclimation through modulating lignin biosynthesis, PLoS Genet. 11 (2015) e1005471.
[71] N. Demont-Caulet, C. Lapierre, L. Jouanin, S. Baumberger, V. Méchin, Arabidopsis peroxidase-catalyzed copolymerization of coniferyl and sinapyl alcohols: kinetics of an endwise process, Phytochemistry 71 (2010) 1673–1683.
[72] H. Shi, Z. Liu, L. Zhu, C. Zhang, et al., Overexpression of cotton (*Gossypium hirsutum*) dirigent1 gene enhances lignification that blocks the spread of *Verticillium dahliae*, Acta Biochim. Biophys. Sin. (Shanghai) 44 (2012) 555–564.
[73] N.H. Nguyen, J.H. Kim, J. Kwon, C.Y. Jeong, et al., Characterization of *Arabidopsis thaliana FLAVONOL SYNTHASE 1* (*FLS1*)—overexpression plants in response to abiotic stress, Plant Physiol. Biochem. 103 (2016) 133–142.
[74] E. Schulz, T. Tohge, E. Zuther, A.R. Fernie, D.K. Hincha, Natural variation in flavonol and anthocyanin metabolism during cold acclimation in *Arabidopsis thaliana* accessions, Plant Cell Environ. 38 (2015) 1658–1672.

L'analyse de macro- et micro-phénotypage confrontée à des données omiques, a permis d'identifier plusieurs gènes / protéines candidats susceptibles d'être impliqués dans les modifications de la paroi cellulaire observées lors de la réponse d'acclimatation face à des conditions de température sub-optimale. L'approche simple d'analyses de statistique a permis de valider des candidats tandis que l'approche intégrative omiques a été utile pour proposer de nouvelles hypothèses concernant la régulation de voies métaboliques entières. L'analyse statistique intégrative n'a ici été que partiellement développée du fait notamment du petit nombre de variables étudiées. Les quatre variables (2 écotypes x 2 températures) se discriminant très bien dans chacune des analyses effectuées (ACP), les statistiques intégratives ne fournissaient alors que des informations redondantes (Fig. 18).

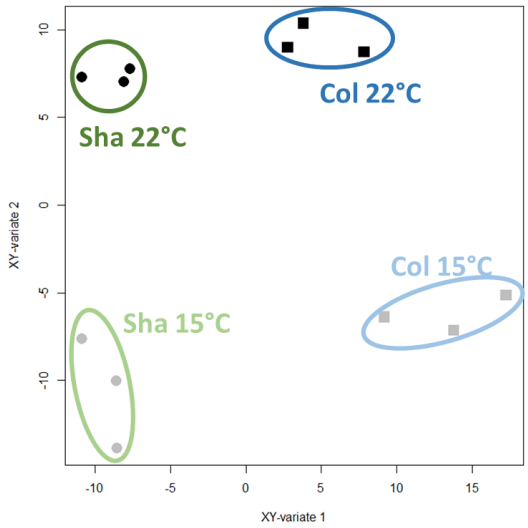


Figure 18 : *Représentation graphique d'une PLS entre les données de protéomiques pariétales et la transcriptomiques pariétales des deux écotypes Col et Sha cultivées à deux températures de culture (15 et 22°C).*

Les réponses plus marqués de l'écotype Col face à des températures sub-optimals confirme que les écotypes offrent différentes plasticité phénotypique mis notamment en évidence dans la partie III. Ainsi, la plasticité de l'écotype Col pourrait être liée à sa capacité d'adaptation à des conditions de croissance extrêmes dans son milieu d'origine.

Afin d'évaluer la contribution du fond génétique à l'acclimatation à différentes conditions de température de croissance d'*A. thaliana*, le projet *WallOmics* a porté sur l'étude d'un grand nombre d'écotypes régionaux et provenant d'un gradient altitudinales *via* une approche intégrative de données omiques hétérogènes.

IV.2 WallOmics: an integrative study of cell wall adaptation to sub-optimal growth conditions of natural population of Arabidopsis

L'objectif principal du projet WallOmics était d'évaluer les réponses des plantes face au réchauffement climatique en se focalisant sur la plasticité pariétale de variants naturels d'*A. thaliana* par une approche intégrative innovante. Pour sa réalisation, plusieurs études ont été préalablement réalisées :

1. La production de données omiques hétérogènes effectuée sur deux organes (détaillée partie II): les rosettes et les hampes florales.
2. L'approche multi-blocks (détaillée en partie II. 7) pour permettre de trier l'information de chaque bloc ainsi que d'identifier les gènes / protéines candidats liés à des traits d'intérêt.
3. L'étude sur les populations (détaillée en partie III) a permis de mieux appréhender la variabilité naturelle d'*A. thaliana* dans la chaine pyrénéenne.
4. L'étude préliminaire détaillée dans la partie IV. 3 nous a permis une première analyse sur des écotypes de référence (Col et Sha) provenant d'altitudes très contrastés.

Ainsi, quatre écotypes (Roch, Grip, Hern et Hosp) ont répondu à plusieurs critères nécessaire pour le bon déroulement du projet WallOmics ; leurs homogénéités génétiques intra population, leurs différences génétiques inter populations ainsi que leurs altitudes contrastées d'origine. L'étude portant sur l'acclimatation à différentes conditions de température de croissance utilisera ces quatre écotypes pyrénéens accompagnés de l'écotype de référence Col.

Ce projet est détaillé dans la partie ci-après et fera l'objet d'une publication qui sera soumise *back to back* avec un article de méthodologie décrivant les techniques de biologie intégrative utilisées au cours de ce travail (Partie II.7).

WallOmics: an integrative study of cell wall adaptation to sub-optimal growth conditions of natural population of *Arabidopsis*

Harold Duruflé[1], Philippe Ranocha[1], Sébastien Déjean[2], Monique Burrus[2], Thierry Balliau[3,4], Michel Zivy[3,4], Cécile Albenne[1], Vincent Burlat[1], Elisabeth Jamet[1], Christophe Dunand[1]

[1] Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, BP42617, 31326 Castanet Tolosan, France.

[2] Institut de Mathématique de Toulouse, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

[3] CNRS, PAPPSO, UMR 0320 / UMR 8120 Génétique Végétale, 91190 Gif sur Yvette, France

[4] INRA, PAPPSO, UMR 0320 / UMR 8120 Génétique Végétale, 91190 Gif sur Yvette, France

[*] Corresponding authors: Christophe Dunand (dunand@lrsv.ups-tlse.fr; +33 (0)5 34 32 38 58)

Abbreviations:

CW: cell wall; CWP: cell wall protein; DIR: dirigent protein; M: Miscellaneous; OR: Oxido-reductases; P: Proteases; PAC: Proteins acting on cell wall polysaccharides; LM: Proteins related to lipid metabolism; ID: Proteins with interaction domains (with proteins or polysaccharides); S: Signaling; SP: Structural proteins; UF: Unknown function

INTRODUCTION

Local adaptation marks the interactions of natural populations with its specific environmental conditions. In the global warming climate context, temperature increase will be a major concern, but the fluctuations will be also a serious environmental changes (Meehl *et al.* 2007). Freezing stress events, without any preceding chilling period, associated to contrasted elevation of temperature between regions, would be critical indeed for plant development. Thus, one of the challenge in the near future is to maintain agricultural productivity by doing a selection of warm-adapted and cold resistant species. Providing study of natural variations of wild species can be helpful to elucidate molecular bases of the phenotypic differentiation (Alonso-Blanco *et al.* 2005). This kind of study needs to consider the species adaptation with their natural environments (Mitchell-Olds & Schmitt 2006).

Natural abiotic gradients, such as mountains, provide an ideal setting to study how species adapt to contrasting environmental factors (Reich *et al.* 2003). For example, plants endure important environmental variations like the diminution of temperature and humidity, combined to the rise of UV radiations with the elevation of the altitude in temperate climates (Körner 2007). In heterogeneous habitats, plants are adapted to specific prevailing conditions (Brousseau *et al.* 2016; Byars *et al.* 2007; Gonzalo-Turpin & Hazard 2009), and / or respond to contrasted environmental conditions thank to their phenotypic plasticity (Auge *et al.* 2017; Sultan). Many studies have shown that plants change their phenotype along an environmental gradients such as biomass, height, reproductive rate, flowering phenology or the number of leaves (Fischer *et al.* 2011; Hamann *et al.* 2017; Völler *et al.* 2017) but the molecular mechanisms associate are still poorly described (Delker & Quint 2011; Trontin *et al.* 2011).

The well-known model plant *Arabidopsis thaliana* (L.) Heyhn. (*Brassicaceae*) is a self-fertilizing and annual species with a worldwide distribution. Several researches about the natural genetic and phenotypic variation of *A. thaliana* along altitudinal gradients have been published (Botto 2015; Günther *et al.* 2016; Luo *et al.* 2015; Montesinos *et al.* 2009; Pico 2012; Suter *et al.* 2014; Tyagi *et al.* 2015; Tyagi *et al.* 2016). Its capacities to adapt to multiple environmental conditions make it a very appropriate model for phenotypic plasticity studies (Duruflé *et al.* 2017b).

Cell walls (CWs) represent a dynamic external physical barrier that contributes to modify the cell and the plant shape at any moment (Braidwood *et al.* 2014).CWs vary their composition and structure upon developmental and environmental conditions (Le Gall *et al.*

2015). Through CW proteins (CWPs), that drive the modification of the CWs architecture (Franková & Fry 2013), is essential in the control of growth and structural integrity (Cosgrove 2016).

In order to evaluate the contribution of the genetic background to the acclimation of the CW plasticity to different temperature growth conditions, four wild ecotypes of the model species of *A. thaliana* from contrasted natural environments have been analysed. In order to simulate the temperature effect in laboratory conditions, two different growth conditions have been studied: 22°C and 15°C. To highlights the complex's mechanisms of the acclimation to sub-optimal temperature, a biological system approach combining phenomics, metabolomics, proteomics, and transcriptomics analyses was used. This project was focused on the two major organs of the plants: the rosette and the floral stems. The statistical integrative study analysis of the data has allowed to identifying new genes / proteins candidates possibly involved in the temperature acclimation response of *A. thaliana*. Studying the natural diversity of *A. thaliana* by an integrative study using heterogeneous omics data sets can provide extremely powerful tool to understand the phenotypic plasticity in response to climate change.

MATERIALS AND METHODS

Plant material

Five ecotypes of the annual plant *A. thaliana* were used; Grip, Hosp, Hern, Roch living between 700m to 1400m in the Pyrenees Mountain (Duruflé, *in review*) and Columbia (Col) as reference from Poland living at 200m altitude. Seeds were planted in a Jiffy-7® peat pellet. After 48h of stratification at 4°C in darkness, all plant were grown at two different conditions, 22 and 15°C at light intensity 90µmol photons m^{-2} s^{-1}. Plants were grown in a long-day conditions (16 h photoperiod) with 70% humidity. Four- or six-week-old rosettes were collected at the bolting developmental stage after growth respectively at 22°C or 15°C. Flowers stems were collected at 6, 7 and 8 week respectively for Col, Roch and Grip and, Hosp and Hern at 22°C to be at the same stage. They have been collected two weeks later at 15°C.

Macrophenotyping

Rosettes grown at 22°C and 15°C were analysed at the time of sampling. Diameter and fresh weight were measured and the number of leaves were also counted. Before frozen, pictures were taken to measure the rosette areas with the *ImageJ* software (Schneider *et al.*

2012). Concerning the floral stems, before freezing, the length, the number of cauline leaves, the length and the diameter at the base of the floral stem were also measured.

Climatic data

Climatic variables were obtained from (Duruflé, *in review*).

Histological staining of cell walls

Whole rosettes and base of floral stems were harvested and rapidly infiltrated with 50 ml Falcon tubes with FAA (10% formalin (37% formaldehyde solution, Sigma-Aldrich, Saint-Quentin Fallavier, France); 50% ethyl alcohol; 5% acetic acid; 35% distilled water) under vacuum and were fixed for 16 h at 4 °C. The dehydration and paraplast infiltration protocol was described in (Francoz *et al.* 2016). The whole rosettes were processed was described in (Duruflé *et al.* 2017b) to keep intact the phylotaxy. Twenty µm-thick serial sections of rosettes or floral stems were disposed on silane-coated microscope slides. To ensure comparison of the labelling intensities among sections, a 20-slides plastic slide holder and 200 mL staining jars were used (Francoz *et al.* 2016). Intracellular material was removed by incubation in 2.6% bleach for 20 min (Ruzin 1999) followed by conscientious washes with distilled water. Sections were then incubated for 1 min in 0.5% safranine red (CI 50240; Kuhlmann, Paris, France) solution, washed with distilled water and then incubated for 30 sec in 0.05% alcian blue (CI 74240, Sigma-Aldrich) solution, adapted from (Srebotnik & Messner 1994; Vazquez-Cooz & Meyer 2002). Following extensive washing, slides were dried, mounted in Eukitt® (quick-hardening mounting medium, Sigma-Aldrich) and scanned using a NanoZoomer HT scanner (Hamamatsu, Hamamatsu City, Japan). The auto-fluorescence of aromatic compounds was observed using a DAPI filter set (excitation: 387 ± 11 nm; dichroic mirror 405 nm; emission: 440 ± 40 nm) using a NanoZoomer RS scanner (Hamamatsu, Hamamatsu City, Japan).

Extraction of proteins from purified cell walls

Cell wall purification was performed using 20 plants for each experiment as described (Feiz *et al.* 2006). The sequential extraction of proteins from purified cell walls was performed as described (Irshad *et al.* 2008). Typically, 0.2 g of lyophilized cell walls was used for one extraction and about 500 µg proteins were extracted. The final protein extract was lyophilized. Proteins were quantified with the CooAssay Protein Assay kit (Interchim, Montluçon, France).

Sequential CW polysaccharides extraction and identification

The protocols of the sequential extraction of cell wall polysaccharides in four steps was details in (Duruflé *et al.* 2017b). In summary, one hundred mg of a deproteinized cell wall fraction were used. Four successive extractions were carried out to obtain extracts enriched in pectins (E1 and E2) and hemicelluloses (E3 and E4). Each extract was hydrolysed in 2 N TFA for 1 h at 120°C. After 10X dilution in UHQ water, monosaccharides were analysed by High-Performance Anion-Exchange Chromatography coupled to Pulsed Amperometric Detection (HPAEC-PAD; Dionex, Sunnyvale, California, USA) using a CarboPac PA1 column (Dionex). The steps of runs were details in (Duruflé *et al.* 2017b).

Standard monosaccharides were used for identification and quantification: L-Fuc, L-Rha, L-Ara, D-Gal and GalA (Sigma-Aldrich); D-Glc (Merck, Darmstadt, Germany); D-Xyl (Roche, Mannheim, Germany).

Identification of proteins by LC-MS/MS

Shotgun protein analysis was performed directly after their extraction from purified cell walls as described (Duruflé *et al.* 2017a). The LC-MS/MS analyses were performed at the PAPPSO proteomics platform (pappso.inra.fr/) essentially as described in (Duruflé *et al.* 2017a). Parameters for MS data processing in the X!Tandem software (www.thegpm.org/tandem/) and the X!Tandem Pipeline 3.3.4 (Langella *et al.* 2017) was available in (Hervé *et al.* 2016). Trypsin digestion was declared with no possible miscleavage. Only proteins identified with at least two different specific peptides in the same sample and found in at least two biological repeats were validated. Furthermore, quantification was performed on peptides with standard deviation retention times lower than 20 s and peak width lower or equal to 100 s.

Bioinformatic annotation of proteins and quantification

The prediction of sub-cellular localization and functional domains of CWPs was performed with the ProtAnnDB tool (San Clemente *et al.* 2009). Protein was considered as a CWP if two bioinformatics programs predicted it as secreted, no intracellular retention signal was found and no more than one trans-membrane domain was found as described in (Albenne *et al.* 2013). Quantification was only operate for CWPs identified using the MassChroQ software (Valot *et al.* 2011). Proteomics data and procedure details are detailed in (Duruflé *et al.* 2017a). Functional annotation of CWPs is given according to WallProtDB (www.polebio.lrsv.ups-tlse.fr/WallProtDB/). LC-MS/MS data have been deposited at

PROTICdb (proteus.moulon.inra.fr/w2dpage/proticdb/angular/) and CWP MS data at WallProtDB (www.polebio.lrsv.ups-tlse.fr/WallProtDB/).

RNA Sequencing

Protocols of the transcriptomics analysis was details in (Duruflé *et al.* 2017b). Briefly, frozen samples were ground in liquid nitrogen and total RNAs were extracted using the RNeasy Plant Mini Kit (Qiagen, Courtaboeuf, France). RNA quantification was performed using the spectrophotometer ND-1000 (NanoDrop, Wilmington, Delaware, USA) and RNA quality was assessed on the Agilent 2100 Bioanalyzer (Agilent Technologies, Courtaboeuf, France). RNA seq experiments were performed on an Illumina HiSeq 3000 at the GeT-PlaGe platform (get.genotoul.fr, Toulouse, France) according to the standard Illumina protocols. Short pair-end sequencing reads generated were analysed using the commercial CLC Genomic Workbench 6.0 software (CLC bio, Aarhus, Denmark). RNA seq data and procedure details are available in Supplementary Table S2.

Statistical analysis and data integration

Data analysis and methodology were achieved as previously described (Duruflé, *in preparation*). Multi block analysis were provided with the sparse version (MB-PLS-DA) including all the phenomic and metabolomic variables associated to a restrictive pool of 40 CW proteins and 40 genes encoding CWPs identified in the proteomics studies. To improve the visualisation, the clustered images map were done with only the 20 best proteins and genes and all the variables of the phenomic and metabolomic blocks.

Accession numbers

Resulting sequences are available at NCBI short read archive (SRA, BioProject PRJNA344545). LC-MS/MS MS data have been deposited at PROTICdb (http://proteus.moulon.inra.fr/w2dpage/proticdb/angular/) and CWP mass spectrometry data have been made publicly available at WallProtDB (http://www.polebio.lrsv.ups-tlse.fr/WallProtDB/).

RESULTS and DISCUSSION

To work in evolutionary biology context, this study allowed to compare in detail natural collections of wild populations originated from altitudinal gradients to make further progress in the comprehension of the cell wall plasticity in response to the environment. This subject is considered with an integrative approach using heteronomous omics dataset to evaluate sub-optimal growth condition on *A. thaliana*. Each omics dataset (Phenomics, Metabolomics, Proteomics and Transcriptomics) are called blocks thereafter.

Four individuals of *A. thaliana* from the Pyrenees Mountains was chosen (Roch, Grip, Hern and Hosp) as natural populations originated from various altitudes (Tab. 1). These four populations have been selected within 30 new populations previously identified (Duruflé, *in review*), in order to verify if a relation existed between genetic, altitude and the cell wall plasticity, Roch and Grip belong to the same genetic cluster and Hern and Hosp to a second one (Duruflé, *in review*). In addition, within each genetics attribution, one population is originated from middle altitude (Roch and Hern) and another from high altitude (Grip and Hosp). The environmental characteristics of the four populations can be illustrate with the value of principal component analyse (Climate PC) useful to summarize all the combine environmental parameters (Wolfe & Tonsor 2014).

Table 1: Environmental parameters of* A. thaliana *populations. *Climate PC1, localisation and the genetic attribution from the molecular phylogenetic analysis was detailed in (Duruflé,* in review*)*

Accession name	Altitude (masl)	Genetics atribution	Climate pc1
Col	200	A	3
Roch	696	B	0.7
Grip	1190	B	-1.4
Hern	780	C	0.7
Hosp	1424	C	-1.6

To observe and study the CW plasticity, plants were grown at two contrasted temperature conditions: 22°C, the optimal growth condition for *A. thaliana* (Boyes *et al.* 2001), and 15°C a sub-optimal temperature (Duruflé *et al.* 2017b). This low temperature was used to simulate the higher altitude temperatures. Each biological repetition represent a pool of twenty

individual that allows i) to verify the repeatability of the phenotype traits into each biological repetitions and ii) to have sustainable biological material for all the analyses. Biological interpretation defined as follows was provided using the statistical analysis detailed in the article (Duruflé, *in preparation*). Because this project aims to make an integrate multi-blocks analyse, each block was presented but not described in detail. Examples were provided to understand the complexity and the global results that could be simplify the final interpretation of this integrative study.

Phenomics: Morphological phenotypes of the ecotypes depend of temperature growth conditions and of the organs studied

Macro and Micro phenotypes, make in rosettes (Fig. 1A) and floral stems of plants grown at 22°C and 15°C, have been analysed. Rosettes were collected at 4 and 6 weeks corresponding to the bolting time (stage 5.10 (Boyes *et al.* 2001)) for Col grown at 22°C and 15°C, respectively. As expected, the impact of low temperature in rosettes is striking, such as the increase of leaf number for all the ecotypes grown at 15°C (Fig. 1B). Interestingly, the values of leaf number are contrasted between ecotypes, like Col grown at 15°C which has the same number of leaves that Roch at 22°C and Hern at 15°C. The leaf number, combined with the weight, the diameter, the density and the projected area of the rosettes were integrated to the following statistical study (Fig. S1). No ecotypes expose a strong phenotype or a clear specificity on this organ. The tendency or tricky change of each ecotype in response to the temperature may be reveal by the integrative study. The reproducibility of the data, illustrate by the close points, will help to the integration study giving more weight of the observations.

On contrary to the rosette, the floral stems was collected at different times to be at the same developmental stage (first flower, stage 6 (Boyes *et al.* 2001)), due to differential stems growth speed. All the ecotypes manifest a global temperature affect such as the variation of cauline leaves number (Fig 1C). Otherwise, excepted Hern, the four other ecotypes studied, shown a diminution of cauline leaves at low temperature. Hern has a tendency to show an opposite behaviour as compared to the others regarding floral stem phenotypic. Likewise rosette phenotypic traits, the information about the weight, the diameter and the length of the floral stems at the harvest time were integrated to the statistical study (Fig. S2). But in summary, floral stems appeared greater and heavier at low temperature that is expected in view of the former study (Duruflé, *in review*).

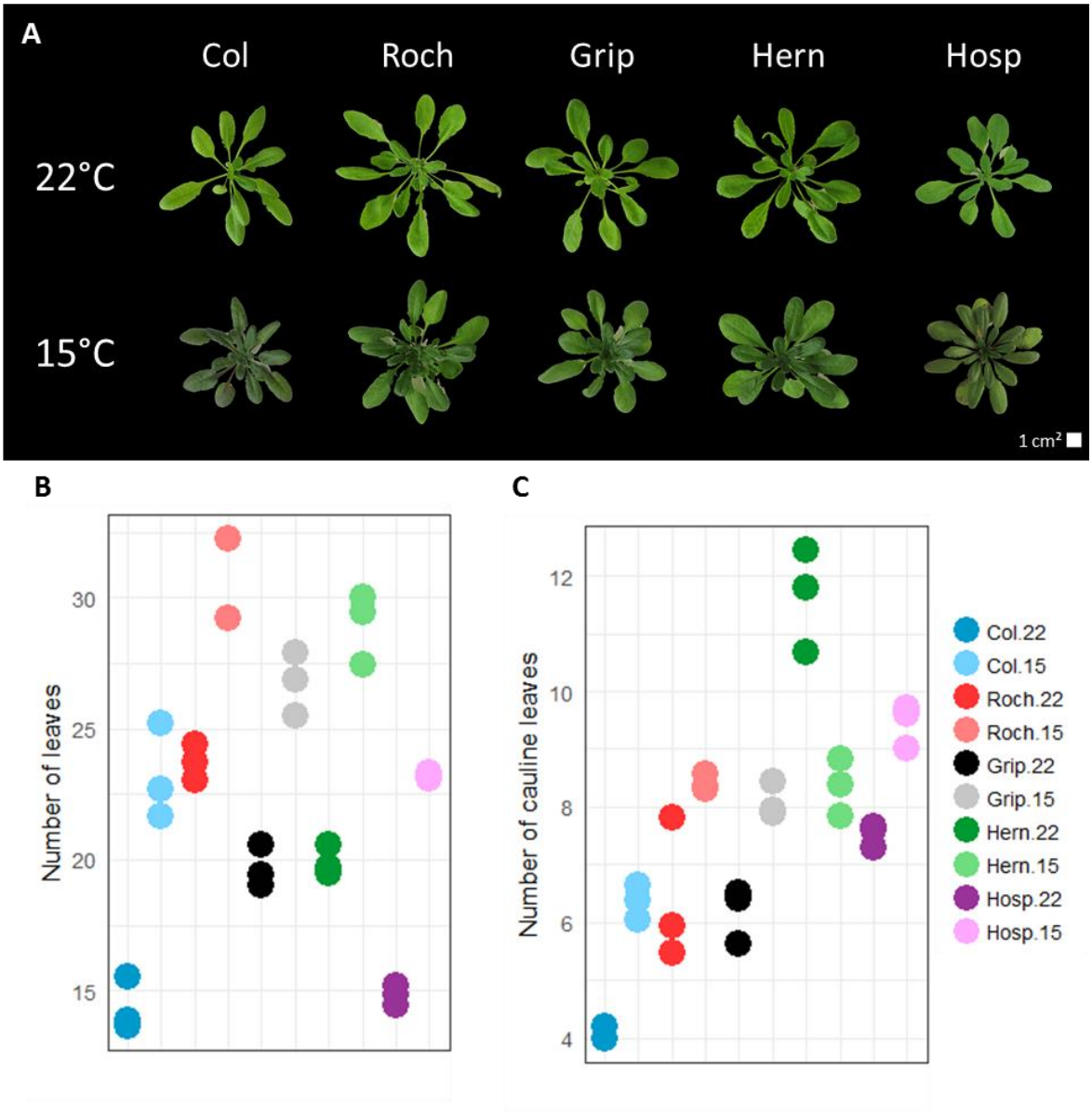


Figure 1. Macro-phenotyping of rosettes the ecotypes grown at 22°C and 15°C. *Rosettes observed at 22°C and 15°C in growth chamber conditions (A). Number of leaves per rosettes (C), number of cauline leaves count in the floral stems. Twenty plants from 3 independent batches have been analysed and each dot stand for the mean of one batch.*

Macro-phenotype highlight the ubiquity effect of the temperature. As expected, development of these two organs of all the ecotypes were impacted by the contrasted growth condition and confirmed the previews analyse make on rosette (Duruflé, *in review*) and brought new information for the floral stem. In spite of the use of biological materials, our phenotypic analysis were highly reproducible, which will allow to link them with the other data in the integrative analysis to understand the underlying molecular mechanisms.

To complete this macro analyse, a microscopic analysis of the leaf and floral stems structures was performed on petiole and base of the stems cross-sections of plants cultivated at 22 and 15°C (Fig. 2)

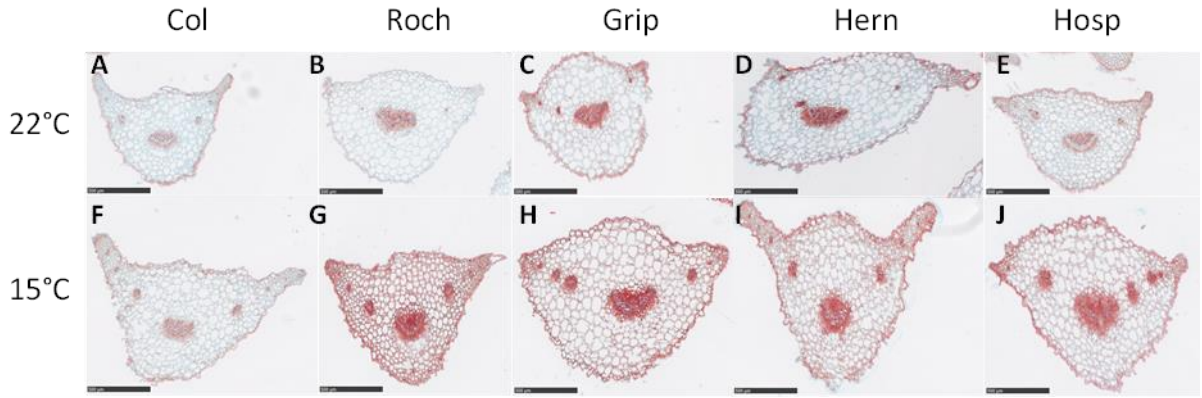


Figure 2. Phenotyping of petioles of plants grown at 22°C (A-E) and 15°C (F-J). *Cross-sections were stained with safranin red and alcian blue. Representative plant petioles of more than 30 petioles from independent batches are shown for each ecotype grown at 22°C or 15°C. Each scale bar represent 500µm.*

Staining sections with safranin red and alcian blue previously bleached allowed to distinguish cell walls enriched in hydrophilic polysaccharides (colored in blue) from those containing more hydrophobic compounds (colored in red). Larger petioles and higher overall safranin red staining were observed on cross-sections from plants grown at 15°C as compared to plants grown at 22°C.

Specificities of the CW composition can be observed for Grip and Hern at 22°C. These two ecotypes presented higher staining of hydrophobic compounds corresponding to cuticle. Moreover, at 15°C higher red staining was observed for the whole petiole cross section. Col has less hydrophobic compounds in there CW that the other ecotypes. A former study, between Col and another reference ecotype Sha (growing at 3400 m in a high valley of Tajikistan), demonstrate a higher cuticle at 15 °C in Sha. This result confirm the interesting response of the natural ecotype to the low temperature compare to the well-known ecotype col.

Unfortunately, these qualitative results could not be used for the integrative analyse but will be important for the future interpretation. Indeed, it is important to investigate with different approaches of analyse to better understanding the CW plasticity of these ecotypes. In order to support the modification of the CW in response to temperature growth variations, the CW composition was also analysed.

Metabolomics: The cell wall composition of the ecotypes vary differentially depending of temperature growth condition

To determine if the morphological modifications in response to the sub-temperature could be due to CW composition modification, polysaccharides were sequentially extracted from the rosettes and the floral stems. After an acidic hydrolysis, their monosaccharide content was analysed, except for the cellulose fraction, which could not be hydrolysed in our assay conditions. Theoretical CW polysaccharide composition has been rebuild based on the monosaccharide analyses and following formula adapted from a previous study (Duruflé *et al.* 2017b; Houben *et al.* 2011) (Table 2).

Table 2. Calculation of the polysaccharides reconstruction and the ratio based on sugar composition data ratio adapted to Houben, 2011. *Fucose (Fuc), Rhamnose (Rha), Arabinose (Ara), Galactose (Gal), Glucose (Glc), Xylose (Xyl) and Galacturonic Acid (GalA); MGalA (Molecular Weight of the Galacturonic Acid), "MRha" (Molecular Weight of the Rhamnose).*

Property	Sugar
Pectin RGI	$(\text{Rha} \times (1 + \frac{\text{MGalA}}{\text{MRha}})) + \text{Ara} + \text{Gal}$
Pectin HG	$\text{GalA} - (\text{Rha} \times (1 + \frac{\text{MGalA}}{\text{MRha}}))$
Xyloglucans	Fuc + Glc + Xyl
Linearity of pectin	$(\text{GalA} - \text{Rha}) / ((\text{Rha} \times (1 + \frac{\text{MGalA}}{\text{MRha}})) + \text{Ara} + \text{Gal})$
Contribution of RG to pectin population	$(\text{Rha} \times (1 + \frac{\text{MGalA}}{\text{MRha}})) / (\text{GalA} - \text{Rha})$
Branching of RGI	$(\text{Ara} + \text{Gal}) / (\text{Rha} \times (1 + \frac{\text{MGalA}}{\text{MRha}}))$

The rosettes and floral stems contained largely lower ratio of linear pectin at 15°C (Fig 3). It also showed that rosettes have a ratio higher than floral stems. Furthermore, the ration in rosette of Hosp was largely effected by the temperature compared to the other ecotypes. On the contrary, Hern show low difference of composition in linearity of pectin between the two temperature growths conditions for the floral stems. This observing can be enlarged to the other reconstruction (Fig. S3).

Like the other qualitative phenomics data, the 6 metabolomics data obtained from polysaccharides reconstruction were integrated into the statistical study.

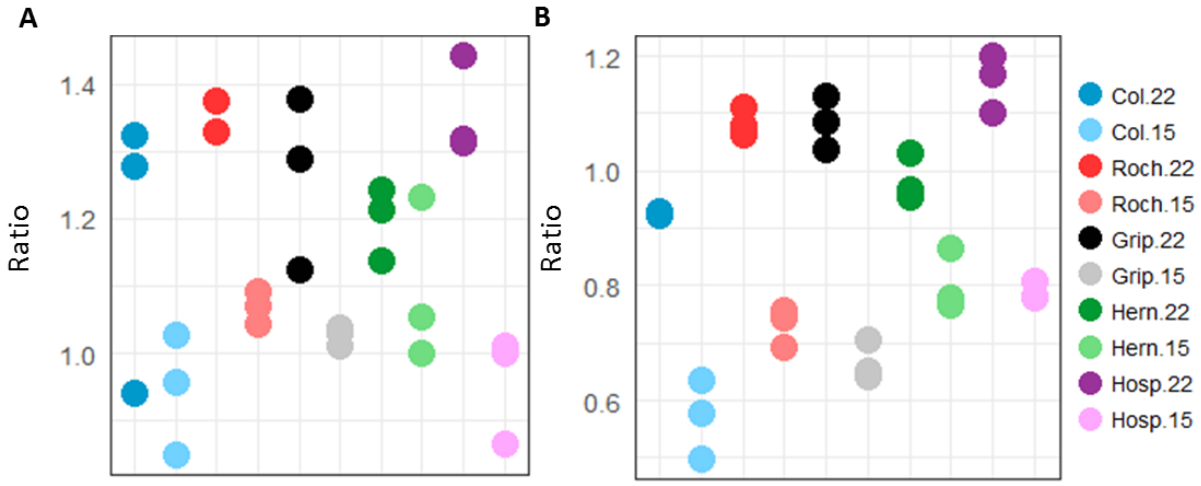


Figure 3: Example of reconstruction of the main cell wall polysaccharides from monosaccharide analysis*. Ratio of the linearity of pectin in rosettes (A) and floral stems (B) from plants grown at 22 °C and at 15 °C. The ratio were obtained from the formula used to rebuild the polysaccharides given in Table 2. Mean values were analysed from tree independent batch. XG: xyloglucans; RGI: rhamnogalacturonan I; HG: homogalacturonan.*

To summarize, plants grown at low temperature rosettes shown more XG and a higher ratio of branching RGI that has 22°C. *In contrario*, the floral stems show more XG at 22°C. Moreover, the HG of pectins were less abundant at low temperature in both organs.

Proteomics: Ecotype modified differently there cell wall proteomes at sub-optimal temperature and between organs

CWPs mainly contribute to the cell wall dynamic thank to their interaction with the polysaccharides. Analyse and integration of the CWP quantification are necessary to understand the cell wall modification.

In this study, the LC-MS/MS analysis of the proteins extracted from purified CWs has allowed to identify 364 and 414 CWPs in the rosettes and the floral stems, respectively (Supplementary Table S1). These CWPs were detected and validated in at least two out of three biological replicates. At the end, 18 and 29 new CWPs have been identified in reference to previous Col rosette proteomes (Hervé *et al.* 2016) and Col floral stems (Duruflé *et al.* 2017a).

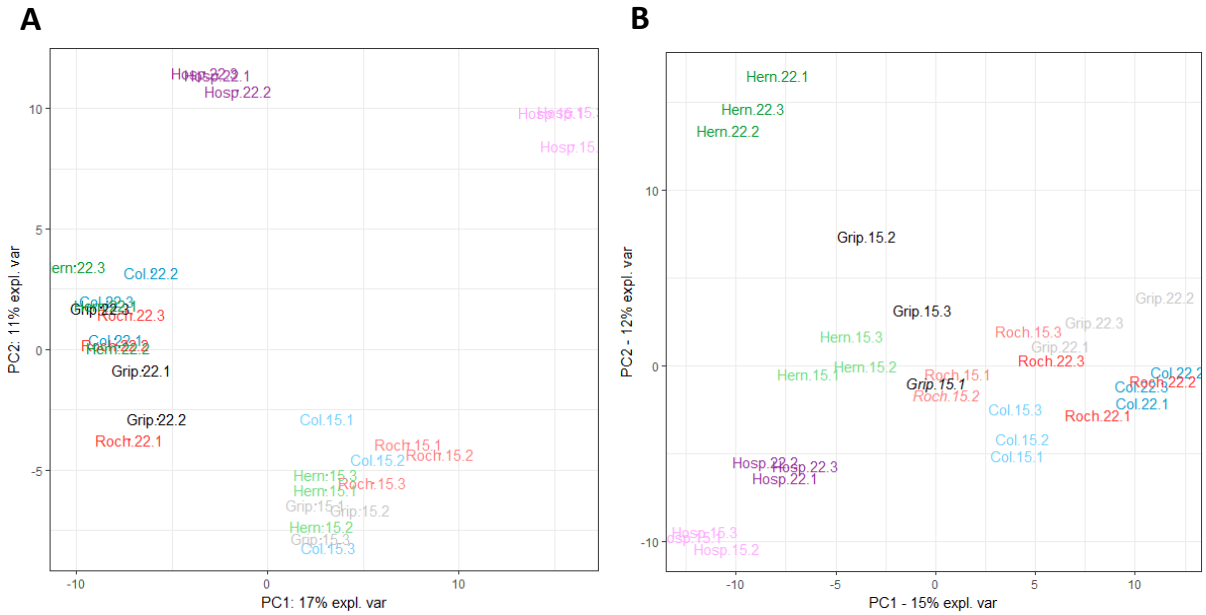


Figure 4. Overall comparison of cell wall proteomes of the plants grown at 22°C (bright colour) and 15°C (pale colour). *Scaled PCA of the proteomics quantitative analysis of the rosettes (A) and the floral stems (B). Values for x and y axes are those of PC1 and PC2, respectively.*

The scaled PCA of the quantified CWPs in the rosettes and the floral stems show the good quality and repeatability of the analyses and a clear segregation of the samples (Fig. 4). The analysis of these two organs demonstrate a specific response of each ecotype at the proteome level depending of growth temperature condition.

In the rosette analyse, the first component (PC1) separate the samples according to the temperature (22°C vs 15°C) and the second one (PC2) according to the specificity of Hosp.

Interestingly, in the floral stems this multivariate analysis discriminated all the samples according to the genetic (A (Roch/Grip) vs B (Hern/Hosp)). In this organ, the temperature affect more the genetic group A than the others (B and C). Indeed, ecotypes of the group B were more discriminated by the temperature condition than the other ecotypes. By the way, all the sample of the plants grown at 22°C and at 15°C were located at right and left, respectively. Surprisingly, only Hern show off a reverse location versus the PC2.

This quantitative analysis of proteomics data could be allowed the identification of CWPs specifically accumulated in on ecotype/growth temperature. But, the proteomics data available in the Supplementary Table S1will not be discussed in details. On the other hand, the integrative analysis and the transcriptomics (last block), could help to sort all the information and highlighted the variables that will better explain the cell wall plasticity.

Transcriptomics: Regulation of gene expression are modified by both growth conditions and ecotypes

To complete this study, RNA seq analyses have been performed and allowed to obtain expression data from 19,763 and 22,570 genes in rosettes and floral stems respectively (Supplementary Table S2A and S3A). The multivariate analysis of the "whole transcriptomes" of the rosettes (Fig. 5A) and the floral stems (Fig. 5B) expound the good quality and repeatability of the data by the clustering of the samples.

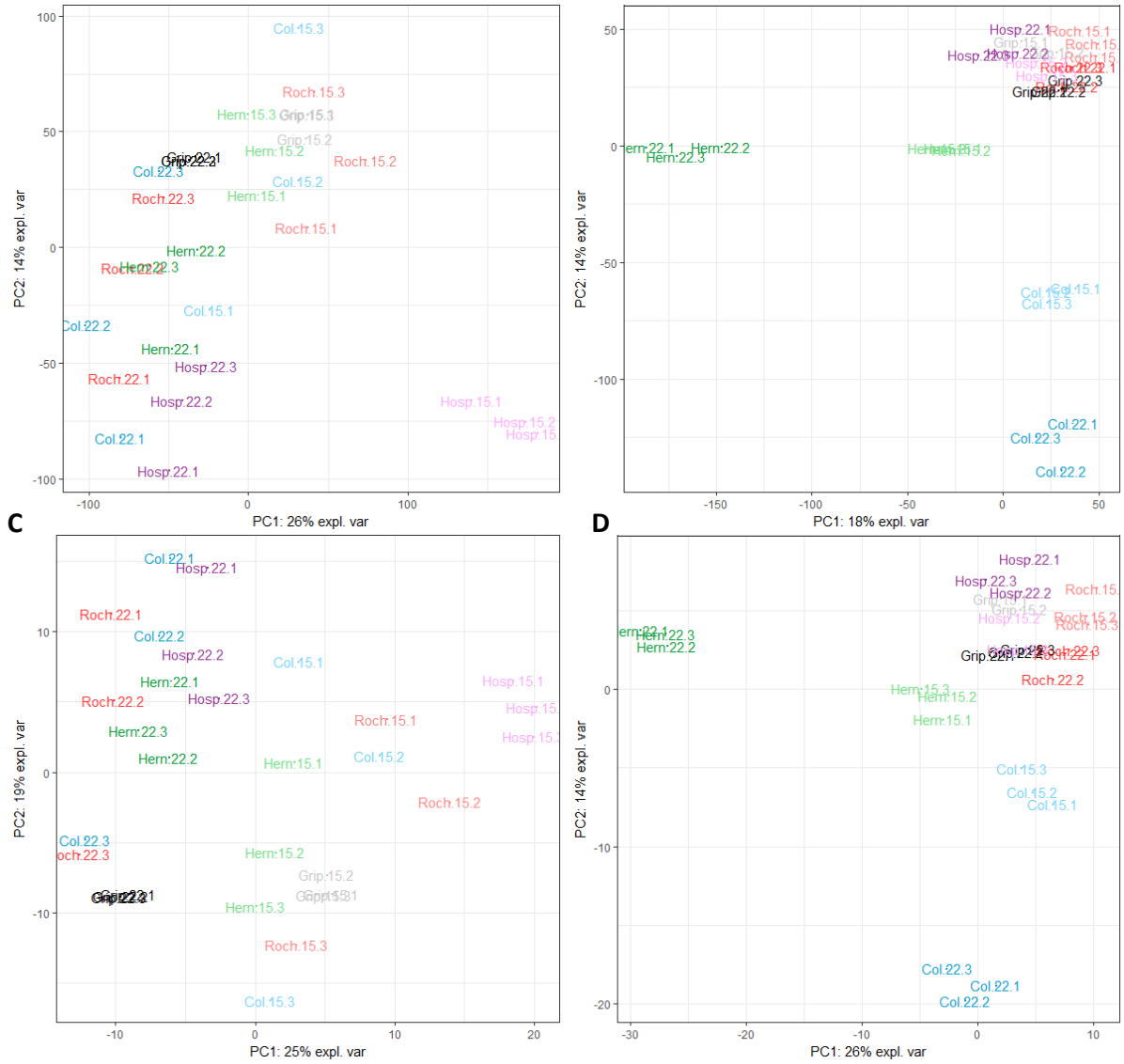


Figure 5. Overall comparison of the ecotypes transcriptomes grown at 22°C (bright colour) and 15°C (pale colour). *Scaled PCA of the whole transcriptomics data of the rosettes (A) and the floral stems (B) and, the "CWP transcriptomes" of the rosettes (C) and the floral stems (D). Values for x and y axes are those of PC1 and PC2, respectively.*

In order to integrate both proteomic and transcriptomic data, the transcripts encoding the CWPs identified by the previous proteomics analyse have been extracted (Supplementary Table S2B and S3B). PCAs restricted to the transcripts showed approximatively the same separation than the whole transcriptomes for these two organs (Fig. 5C-D). This ascertainment known in previews study (Duruflé *et al.* 2017b) allows to extrapolate the comportment of the CW like the whole transcripts.

The PCA of the whole transcriptomics of the rosettes discriminate clearly the ecotype Hosp grown at 15°C. Despite the specific profile of this ecotype, the CWP transcriptomes brings out the temperature effect for the rest of the samples in the PC1.

In the floral stems, this multivariate analyse highlight the specific behaviour of Hern and Col. CWP transcriptomes remains substantially the same profil. Hern kept the same interesting profil than for the proteomics.

To summarise, the impact of the temperature on gene expression was observed in these two organs for all ecotypes. This strong effect still allows to discriminate the ecotype that have specific comportment. This different comportment and the gene / protein candidates associated could be higlighted by an integrative analyse.

Integrative analysis: A powerful tool to understand complex mechanism

Integrative analysis of multi-blocks allows to link each blocks to another. This type of statistical analysis, detailed in the article (Duruflé, *in preparation*) is a powerful tool to help to understand complex mechanisms. To describe the temperature effect on natural ecotype by an integrative study, two analyses were detailed thereafter. Firstly, the different responses of the rosettes and the floral stems to the temperature and, in a second part, the ecotype responses specificity were developed. In order to find the most interesting gene / protein candidates, all the multi block analyses were provided with the sparse version including all the phenotypes and metabolomes data but associated to a restrictive pool of CW proteins and the expression level of genes encoding CWPs identified in the proteomics studies.

Rosette and the floral stems respond differently to the temperature growth conditions

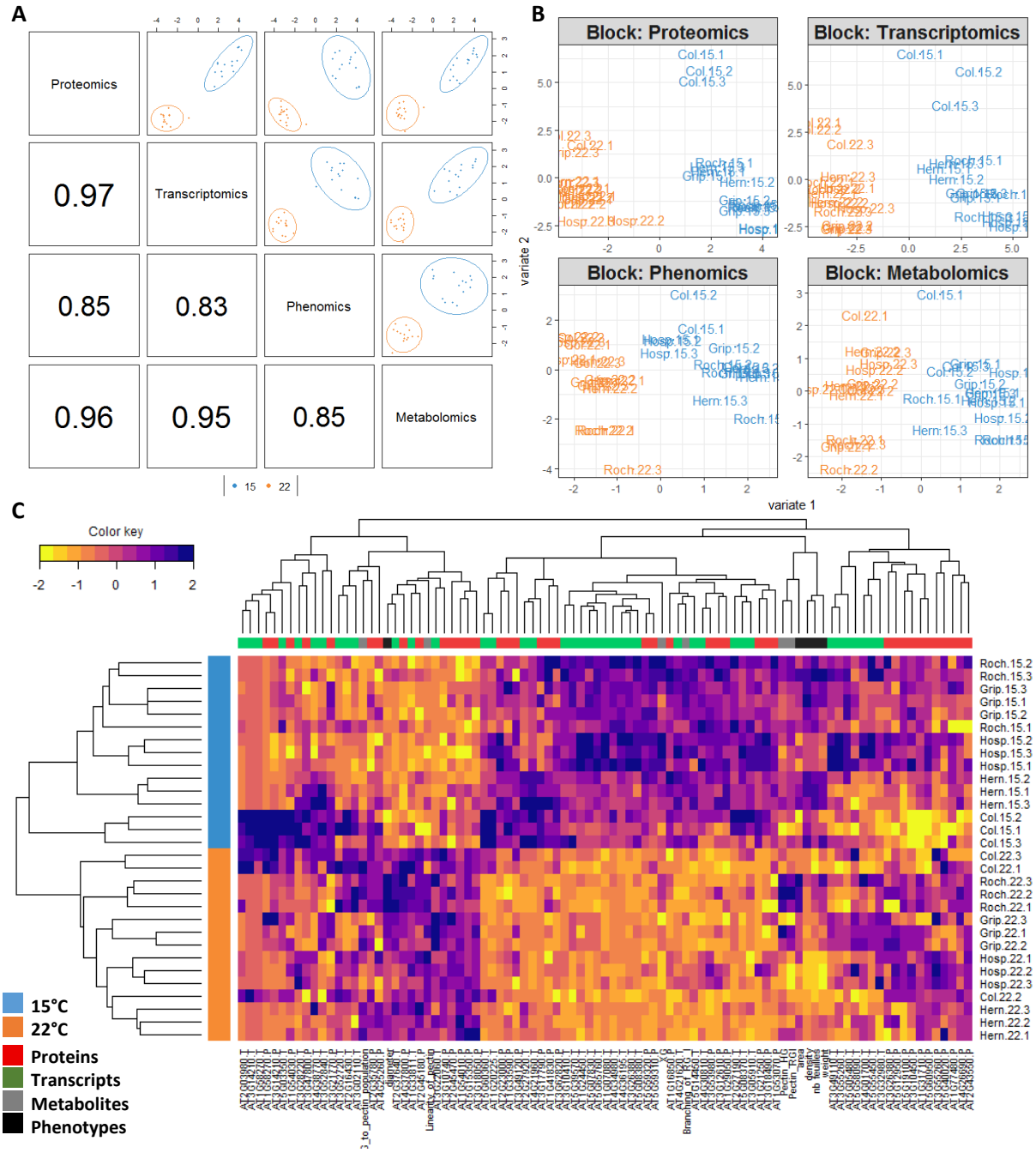


Figure 6: Graphical representation of sparse MB-PLS-DA analyses that discriminate the samples of the rosette by the temperature *A) plotDIABLO show the correlation between each block, B) Individual plots projects the samples for the four blocks. For A and B, each dot corresponds to all the samples of one block. C) Clustered image map representing the multi-omics profiles for each sample discriminated by the temperature: the levels of discriminating variables belonging to the four blocks.*

Table 3. Synopsis of CWPs or transcripts differentially present at a given growth temperature in rosettes. *The statistical analysis of the data have been performed with a sparse MB-PLS-DA with the 5 ecotypes. + corresponds to CWPs or genes having specific profiles (increase or a decrease compared to other samples) at one growth temperature. M: Miscellaneous; OR: Oxido-reductases; P: Proteases; PAC: Proteins acting on cell wall polysaccharides; LM: Proteins related to lipid metabolism; ID: Proteins with interaction domains (with proteins or polysaccharides); S: Signaling; SP: Structural proteins; UF: Unknown function.*

AGI code	Functional classes	Putative function	Proteins	transcrits	22°C	15°C
AT2G45180	LM	homologous to non-specific lipid transfer protein	+		+	
AT1G54010	LM	lipase acylhydrolase (GDSL family)	+		+	
AT5G15350	OR	early nodulin AtEN22 homologous to blue copper binding protein	+		+	
AT4G37800	PAC	GH16 (endoxyloglucan transferase) (At-XTH7)	+		+	
AT3G18050	UF	expressed protein	+		+	
AT3G10740	PAC	GH51 (alpha-arabinofuranosidase)	+		+	
AT2G45470	S	fasciclin-like arabinogalactan protein (FLA8)	+		+	
AT5G59320	LM	non-specific lipid transfer protein (AtLTP1.2, AtLTP3)	+			+
AT5G59310	LM	non-specific lipid transfer protein (AtLTP1.11, AtLTP4)	+			+
AT1G16850	UF	expressed protein	+			+
AT2G23000	P	Ser carboxypeptidase (AtSCPL10)	+			+
AT1G29050	PAC	homologous to *A. thaliana* PMR5 (carbohydrate acylation)	+			+
AT1G21250	S	receptor kinase (AtWAK1, WALL-ASSOICATED KINASE 1)	+			+
AT1G53070	ID	lectin (legume lectin domain)	+			+
AT3G18490	P	Asp protease (pepsin family) (ASPG1)	+			+
AT1G41830	OR	multicopper oxidase (AtSKS6, homologous to SKU5)	+			+
AT3G53980	LM	non-specific lipid transfer protein (AtLTPd7)	+			+
AT1G33590	ID	expressed protein (LRR domains)	+			+
AT3G12610	ID	expressed protein (LRR domains)	+			+
AT3G62820	ID	plant invertase/ pectin methylesterase inhibitor (PMEI)	+			+
AT1G33811	LM	lipase acylhydrolase (GDSL family)		+	+	
AT3G22060	UF	expressed protein		+	+	
AT2G37640	PAC	alpha-expansin (ATHEXP ALPHA 1.9) (AtEXPA3)		+	+	
AT3G26380	PAC	GH27 (alpha-galactosidase/melibiase)		+		+
AT4G21620	UF	expressed protein		+		+
AT4G36195	P	Pro-Xaa carboxypeptidase (Peptidase family S28.A26, MEROPS)		+		+
AT1G07080	OR	thiol reductase (GILT family)		+		+
AT5G65760	P	Pro-Xaa carboxypeptidase (Peptidase family S28.A02, MEROPS)		+		+
AT5G14450	LM	lipase acylhydrolase (GDSL family)		+		+
AT5G08370	PAC	GH27 (alpha-galactosidase/melibiase)		+		+
AT5G39570	UF	expressed protein		+		+
AT5G08380	PAC	GH27 (alpha-galactosidase/melibiase)		+		+
AT4G34980	P	AtSBT1.6 (Peptidase family S08.A39, MEROPS)		+		+
AT5G07830	PAC	GH79 (endo-beta-glucuronidase/heparanase)		+		+
AT3G05910	PAC	CE13 (pectin acylesterase - PAE) (AtPAE12)		+		+
AT2G27190	M	purple acid phosphatase (AtPAP12)		+		+
AT3G10410	P	AtSCPL49 (Peptidase family S10.A45, MEROPS)		+		+
AT4G21960	OR	peroxidase (AtPrx42)		+		+
AT1G24450	M	ribonuclease III		+		+
AT4G30810	P	AtSCPL29 (Peptidase family S10.A32, MEROPS)		+		+

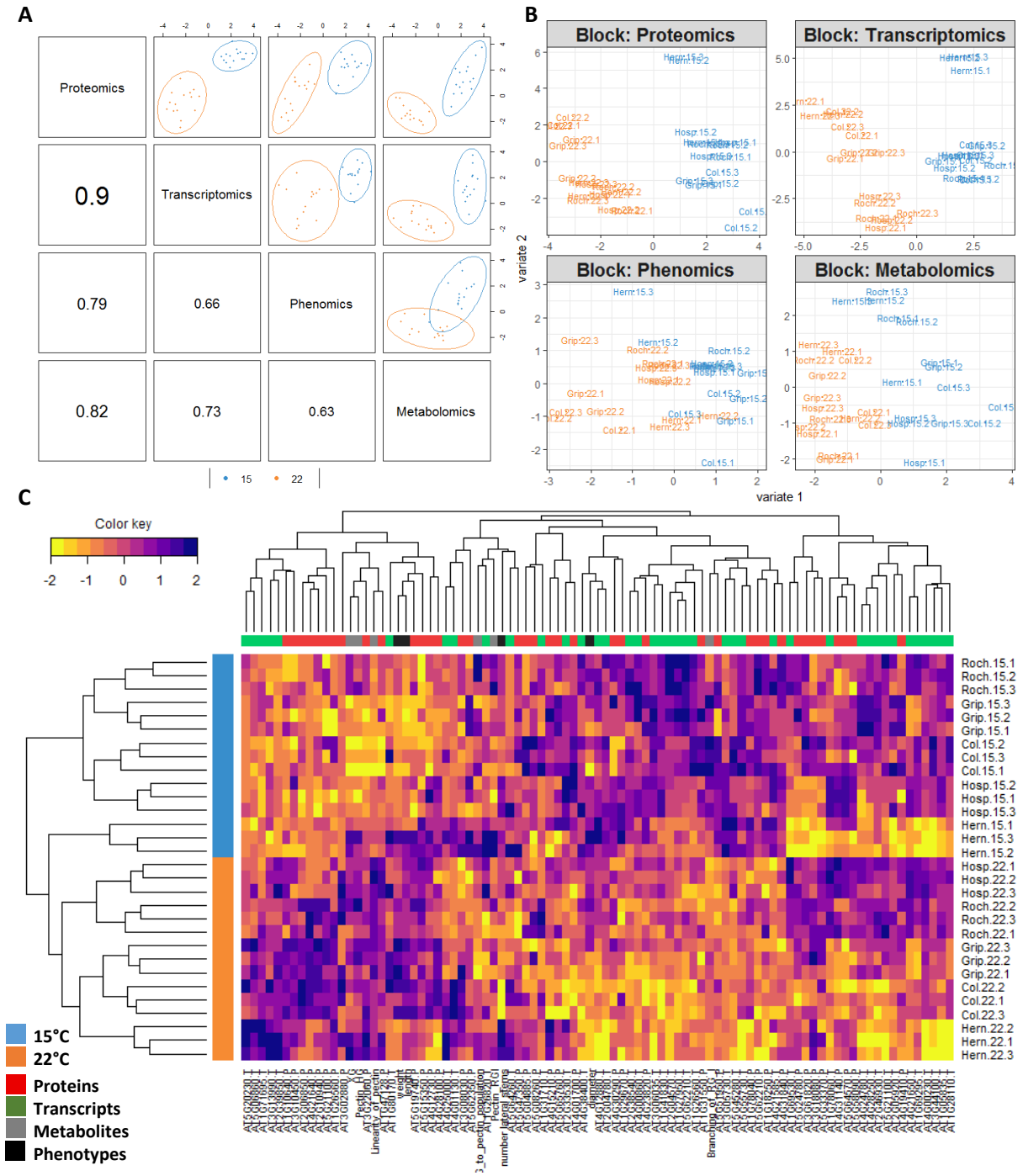


Figure 7: Graphical representation of sparse MB-PLS-DA analyses that discriminate the samples of the floral stems by the temperature *A) plotDIABLO show the correlation between each block, B) Individual plots projects the samples for the four blocks. For A and B, each dot corresponds to all the samples of one block. C) Clustered image map representing the multi-omics profiles for each sample discriminated by the temperature: the levels of discriminating variables belonging to the four blocks.*

Table 4. Synopsis of CWPs or transcripts differentially present at a given growth temperature in floral stems. *The statistical analysis of the data have been performed with a sparse MB-PLS-DA for the five ecotypes. + corresponds to CWPs or genes having specific profiles (increase or a decrease compared to other samples) at one growth temperature. AGI codes in bold means that the protein and the gene were both differentially present. M: Miscellaneous; OR: Oxido-reductases; P: Proteases; PAC: Proteins acting on cell wall polysaccharides; LM: Proteins related to lipid metabolism; ID: Proteins with interaction domains (with proteins or polysaccharides); S: Signaling; SP: Structural proteins; UF: Unknown function.*

AGI code	Functional classes	Putative function	Proteins	transcrits	22°C	15°C
AT1G10640	PAC	GH28 (polygalacturonase)	+		+	
AT3G22060	UF	expressed protein (Gnk2-homologous domain)	+		+	
AT1G47128	P	Cys protease (papain family) (Peptidase family C01.064) (RD21A)	+		+	
AT4G21640	P	AtSBT3.15 (Peptidase family S08.A42, MEROPS)	+		+	
AT1G26560	PAC	glycoside hydrolase family 1 - GH1 (beta-glucosidase) (AtBGLU40)	+		+	
AT3G10450	P	AtSCPL7 (Peptidase family S10.A15, MEROPS)	+		+	
AT2G06850	PAC	GH16 (endoxyloglucan transferase) (At-XTH4)	+		+	
AT5G21100	OR	multicopper oxidase	+		+	
AT5G19740	P	peptidase M28 (peptidase family M28.A02, MEROPS)	+		+	
AT5G15350	OR	early nodulin AtEN22 homologous to blue copper binding protein	+		+	
AT3G02880	S	leucine-rich repeat receptor protein kinase (LRR III subfamily)	+		+	
AT2G10940	LM	homologous to non-specific lipid transfer protein	+		+	
AT5G62210	LM	expressed protein (lipase/lipooxygenase domain, PLAT/LH2)	+			+
AT4G01700	PAC	glycoside hydrolase family 19 - GH19	+			+
AT4G23820	PAC	glycoside hydrolase family 28 - GH28 (polygalacturonase)	+			+
AT5G51950	OR	expressed protein (GMC oxido-reductase domain)	+			+
AT1G18250	M	thaumatin (PR5, ATLP1)	+			+
AT1G31550	LM	lipase acylhydrolase (GDSL family)	+			+
AT4G31840	OR	early nodulin AtEN13 homologous to blue copper binding protein	+			+
AT1G29670	LM	lipase acylhydrolase (GDSL family)	+			+
AT2G39850	P	AtSBT4.1 (Peptidase family S08.A46, MEROPS)		+	+	
AT3G13990	OR	multicopper oxidase (AtSKS11, homologous to SKU5)		+	+	
AT5G06860	ID	PGIP1 (LRR domains)		+	+	
AT5G20230	OR	stellacyanin AtSTC1, BCB (blue copper binding protein)		+	+	
AT1G71695	OR	peroxidase (AtPrx12)		+	+	
AT4G21585	LM	phospholipase C/P1 nuclease		+		+
AT5G55730	S	fasciclin-like arabinogalactan protein (FLA1)		+		+
AT3G06035	UF	expressed protein		+		+
AT5G45280	PAC	CE13 (pectin acylesterase - PAE) (AtPAE11)		+		+
AT4G00860	UF	expressed protein (DUF1138)		+		+
AT2G38540	LM	non-specific lipid-transfer protein (AtLTP1.5, AtLTP1)		+		+
AT1G27950	LM	non-specific lipid transfer protein (AtLTPg4, LTPG1)		+		+
AT1G41830	OR	multicopper oxidase (AtSKS6, homologous to SKU5)		+		+
AT2G04780	S	fasciclin-like arabinogalactan protein (FLA7)		+		+
AT4G12880	OR	early nodulin AtEN20 homologous to blue copper binding protein		+		+
AT3G05730	UF	expressed protein		+		+
AT5G62210	LM	expressed protein (lipase/lipooxygenase domain, PLAT/LH2)		+		+
AT4G38400	PAC	expansin-like A (ATHEXP BETA 2.2) (AtEXLA2)		+		+
AT2G33530	P	AtSCPL46 (Peptidase family, S10.A42, MEROPS)		+		+
AT2G04570	LM	lipase acylhydrolase (GDSL family)		+		+

At the rosette level, correlation between samples from each block has been close to 1 (Fig. 6A). The high level of correlation was due to the well separation of the two temperatures between each blocks. The highest correlation was between the proteomics and transcriptomics blocks (0.97) and the lowest between the transcriptomics and the phenomics blocks (0.83). This result could be due to the good discrimination for the transcriptomic and proteomic blocks regarding the individual plots (Fig. 6B). All the blocks show a clear segregation between the temperatures in the first component that confirm the global effect of the temperature at multi-level.

The identifiable variables, from each block involved in the discrimination according to the temperature was represented in the figure 6C (see also Supplementary Table S4A). The clustered image map includes also the results of a hierarchical clustering performed jointly on the variables and the samples. The list of the proteins and genes are detailed in the table 3. The samples are clearly discriminate according to the two growth conditions.

On this set of protein/gene candidates, 23 % encode proteins acting on cell wall polysaccharides (PAC) and 18 % for proteins related to lipid metabolism (LM). This enrichment support the impact of the temperature in the CW and could allow to determine the actors of the CW modifications. Most of the PAC (6 / 9) are from the family of the glycoside hydrolase that could have a role in the CW response to cold stress. The *AT5G08380* encoding for an alpha-galactosidase, was up-regulated by low temperature in our study in which was in agreement with previews study performed on another *Brassicaceae*, *Thlaspi arvense* (Oono *et al.* 2006; Sharma *et al.* 2007). That result open research of the role of this family in the cold acclimation.

In the LM proteins family, three LTPs were also found more abundant at 15°C. In this set of proteins, the LTP3 (*AT5G59320*) has been shown to be involved in the freezing stress in *A. thaliana* (Guo *et al.* 2013). This proteins related to lipid metabolism was close to the LTP4 in the dendrogram (correlation > 0.9). The co-induction of these two genes in rosettes at low temperature could have a redundant function and be responsible of the higher level of hydrophobic compounds found in the CW related to the red staining observe in the micro-phenotyping study (Fig. 2). In this line, the analysis of the mutant *ltp3* that did not show cuticular phenotype. No study relative to the *LTP4* are known. Analyse of the double mutant *ltp3* and *ltp4* could be help to better understand the lipid transport and the impact in the cold acclimation.

On the other hand, floral stems show less correlation between each block (Fig. 7A). As well for the rosette analyses, the highest correlation (0.9) was observed between the proteomics

and transcriptomics and the lowest (0.63) between the phenomics and the metabolomics blocks. This latter result could be due to the lower segregation between the temperatures in the first component in the phenomics and the metabolomics blocks. It was probably amplified by the specific phenotype of Hern at 15°C that was different to the other ecotypes (Fig. 7B). In contrast to the rosettes, this analyse highlight the differential effect of the temperature at multi-level.

The important effects of the temperature and the specificity of Hern in the floral stems were also highlighted by the hierarchical clustering of the samples in the clustered image map of the variables (Fig. 7C). The dendrogram discriminate sample between growth temperatures: Roch, Grip Col, Hosp grown at 15°C clustered together and all the samples grown at 22°C with Hern at 15°C belonged to the same cluster. It clearly highlighted the specificity of Hern floral stem development at low temperature.

The characteristics of Hern has marked by the lower levels of six transcripts including three proteases: *AT1G28110, AT4G00230,* and *AT4G21650* revealed by the dendrogram. The latter, was up-regulated by drought and heat stress (Rizhsky *et al.* 2004). The specific acclimation of this ecotype to low temperature may be due to the low level of this family of proteases.

More generally in the floral stems, the majority of gene and protein candidates at 15°C, detailed in the table 4 (see also Supplementary Table S5A), was on the LM class (34 %; 8 / 23). A previous analyse from the epidermal peels of *A. thaliana* stems had also highlighted an enrichment in genes encoding proteins predicted to be in lipid metabolism and allowed to identified new proteins involved in the biosynthesis of waxes and cutin (Suh *et al.* 2005). In our study, the gene *LTPG1* (*AT1G27950*) and the genes and the protein LH2 (*AT5G62210*) emerged in condition of low temperature. The cuticle of the floral stems forms a vital hydrophobic barrier, providing mechanical strength and viscoelastic properties as well protecting the plant from the environmental stresses. Interestingly, the gene *CAPIP2* which exhibited 41 % homology with *LH2* (Lee *et al.* 2006) was known as a defence protein against abiotic stress stresses (Majid *et al.* 2017). These two proteins, involved in the lipid transfer, may be play a major role in stress resistance.

At 22°C the situation was different, the candidates differentially detected was mainly represented by proteases (P; 30%; 5 / 17) and oxido-reductases (OR; 30 %; 5 / 17) classes. This result may be play a part in the higher ratio of linearity pectin and xyloglucan observed at this temperature (Fig. 3). This change of polysaccharides compounds may be including a necessity of specific protein represented by these classes. The proteases may be participate to the

maturation of parietal proteins while the OR regulates the cellular elongation by forming covalent bonds between structural proteins.

To help the interpretation, an MB-PLS-DA that discriminate the samples following the ecotypes will help to better understanding the temperature effect of the different ecotypes.

Specific behaviour of the ecotype to temperature effect between different organs

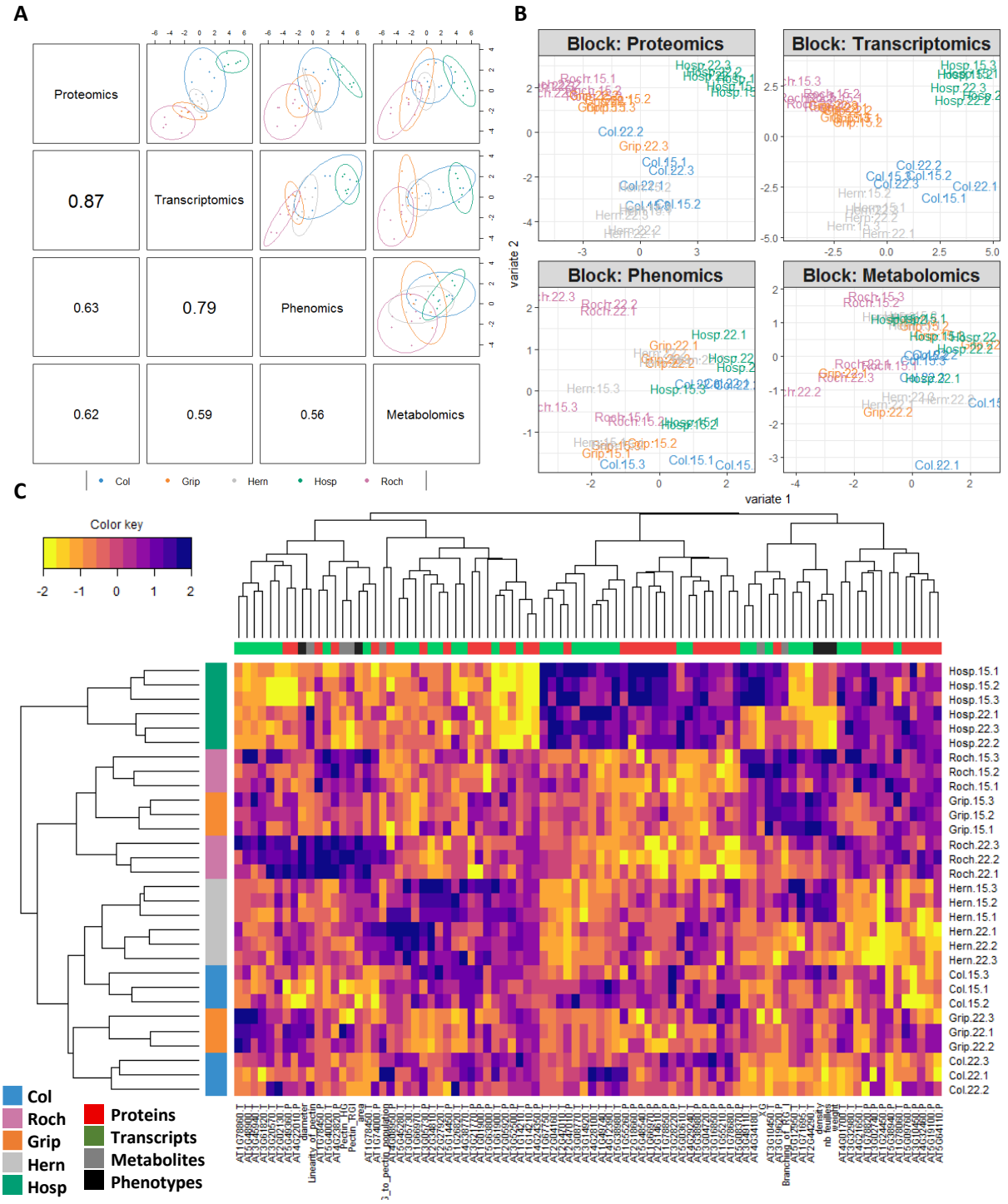


Figure 8: Graphical representation of sparse MB-PLS-DA analyses that discriminate the samples of the rosette by the ecotype *A) plotDIABLO show the correlation between each block, B) Individual plots projects the samples for the four blocks. For A and B, each dot corresponds to all the samples of one block. C) Clustered image map representing the multi-omics profiles for each sample discriminated by the ecotype: the levels of discriminating variables belonging to the four blocks.*

Table 5. Synopsis of CWPs or transcripts differentially present in rosettes of the ecotype Hosp. *The statistical analysis of the data has been performed with a sparse MB-PLS-DA. + corresponds to CWPs or genes having specific profiles in this ecotype (increase or a decrease compared to other samples). AGI codes in bold means that the protein and the gene were both differentially present. M: Miscellaneous; OR: Oxido-reductases; P: Proteases; PAC: Proteins acting on cell wall polysaccharides; LM: Proteins related to lipid metabolism; ID: Proteins with interaction domains (with proteins or polysaccharides); S: Signaling; SP: Structural proteins; UF: Unknown function.*

AGI code	Functional classes	Putative function	Proteins	transcrits
AT1G78850	ID	lectin (curculin-like)	+	
AT5G08370	PAC	GH27 (alpha-galactosidase/melibiase)	+	
AT2G14610	M	(PR) 1 / Cys-rich secretory protein (SCP)	+	
AT5G38980	UF	expressed protein	+	
AT3G04720	ID	expressed protein (Barwin domain, defense protein)	+	
AT1G66970	LM	GDPDL1, GDPL3	+	
AT3G28220	UF	expressed protein (MATH domain)	+	
AT5G48540	UF	expressed protein	+	
AT1G78680	P	peptidase C26 (peptidase family C26.003, MEROPS)(GGH2)	+	
AT1G55260	LM	AtLTPg6	+	
AT1G55210	M	AtDIR20	+	
AT3G16850	PAC	GH28 (polygalacturonase)	+	
AT2G18660	PAC	EXR3	+	
AT5G38980	UF	expressed protein		+
AT1G49740	LM	expressed protein (phospholipase C)		+
AT4G29240	ID	expressed protein (LRR domains)		+
AT4G28100	UF	expressed protein		+
AT4G12390	ID	plant invertase/pectin methylesterase inhibitor (PMEI-like)		+
AT2G47010	UF	expressed protein		+
AT5G03610	LM	lipase acylhydrolase (GDSL family)		+
AT3G14920	PAC	Peptide-N4-(N-acetyl-beta-glucosaminyl) asparagine amidase A		+
AT3G08770	LM	AtLTP1.6, AtLTP6		+

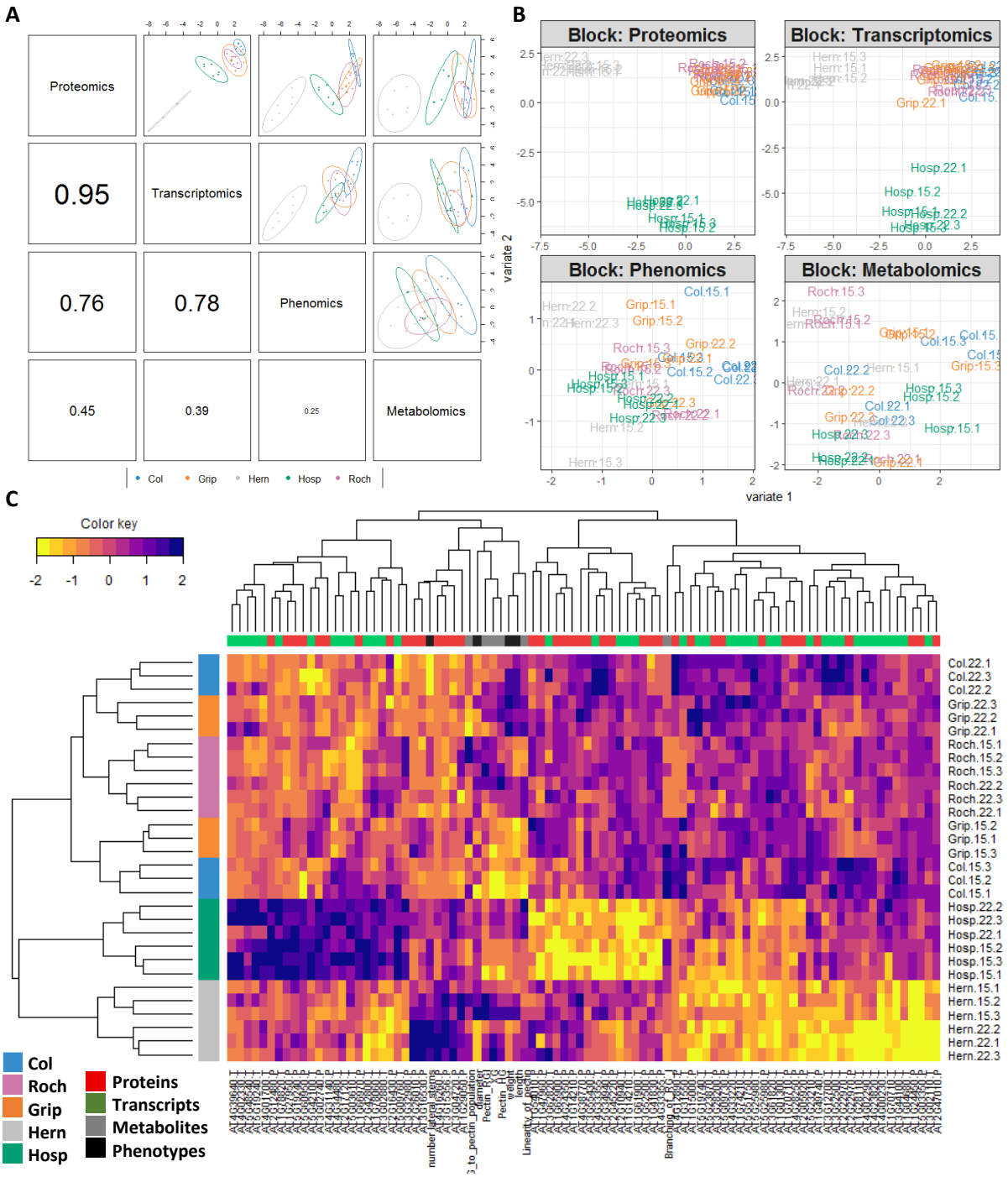


Figure 9: Graphical representation of sparse MB-PLS-DA analyses that discriminate the samples of the floral stems by the ecotype *A) plotDIABLO show the correlation between each block, B) Individual plots projects the samples for the four blocks. For A and B, each dot corresponds to all the samples of one block. C) Clustered image map representing the multi-omics profiles for each sample discriminated by the ecotype: the levels of discriminating variables belonging to the four blocks.*

Table 6. Synopsis of CWPs or transcripts differentially present in floral stems of the ecotypes Hosp and Hern. *The statistical analysis of the data has been performed with a sparse MB-PLS-DA. + corresponds to CWPs or genes having specific profiles in this ecotype (increase or a decrease compared to other samples). AGI codes in bold means that the protein and the gene were both differentially present. M: Miscellaneous; OR: Oxido-reductases; P: Proteases; PAC: Proteins acting on cell wall polysaccharides; LM: Proteins related to lipid metabolism; ID: Proteins with interaction domains (with proteins or polysaccharides); S: Signaling; SP: Structural proteins; UF: Unknown function.*

AGI code	Functional classes	Putative function	Proteins	transcrits	Hosp	Hern
AT2G12480	P	AtSCPL43	+		+	
AT5G19240	UF	expressed protein	+		+	
AT3G02740	P	Asp protease (pepsin family)	+		+	
AT1G66970	LM	GDPDL1, GDPL3	+		+	
AT1G27950	LM	AtLTPg4, LTPG1	+		+	
AT4G31140	PAC	GH17 (beta-1,3-glucosidase)	+		+	
AT5G60950	S	AtCOBL5	+		+	
AT3G32980	OR	peroxidase (AtPrx32) (ATP16A)	+		+	
AT3G15356	ID	lectin (legume lectin domain)	+			+
AT4G16260	PAC	GH17 (beta-1,3-glucosidase)	+			+
AT3G16530	ID	lectin (legume lectin domain)	+			+
AT2G26010	M	homologous to gamma thionin (defensin)	+			+
AT3G04720	ID	expressed protein (Barwin domain, defense protein)	+			+
AT1G29050	PAC	PMR5	+			+
AT1G78820	ID	lectin (curculin-like)		+	+	
AT4G02330	PAC	CE8 (AtPME41)		+	+	
AT5G19240	UF	expressed protein		+	+	
AT5G09760	PAC	CE8 (AtPME51)		+	+	
AT5G48540	UF	expressed protein		+	+	
AT1G78060	PAC	GH3 (beta-xylosidase) (AtBXL7)		+	+	
AT4G39640	M	gamma glutamyltranpeptidase (GGT2)		+	+	
AT4G34480	PAC	GH17 (beta-1,3-glucosidase)		+	+	
AT4G01700	PAC	GH19		+	+	
AT2G42800	ID	expressed protein (LRR domains)		+	+	
AT5G03610	LM	lipase acylhydrolase (GDSL family)		+	+	
AT2G17120	ID	expressed protein (LysM domain)		+	+	
AT3G02880	S	LRR III subfamily		+	+	
AT5G42100	PAC	GH17 (beta-1,3-glucosidase)		+	+	
AT3G44100	LM	MD-2-related lipid-recognition (ML) domain		+		+
AT1G01300	P	Peptidase family A01.A05, MEROPS)		+		+
AT4G00230	P	AtSBT4.14, XSP1		+		+
AT3G61820	P	Peptidase family A01.A13		+		+
AT3G02110	P	AtSCPL25		+		+
AT4G02290	PAC	GH9 (endo-1,3(4)-beta-glucanase)		+		+
AT3G52500	P	Peptidase family A01.A47, MEROPS		+		+
AT4G23820	PAC	GH28 (polygalacturonase)		+		+
AT5G23870	PAC	CE13 (pectin acylesterase - PAE) (AtPAE9)		+		+
AT2G22970	P	SCPL11		+		+
AT5G25980	PAC	AtBGLU37		+		+
AT3G12700	P	Peptidase family A01.A30		+		+
AT1G28110	P	AtSCPL45		+		+
AT4G33220	PAC	CE8 (AtPME44)		+		+
AT5G19740	P	peptidase M28		+		+
AT5G51750	P	AtSBT1.3		+		+
AT1G04680	PAC	PL1 (pectate lyase) (AtPLL26)		+		+
AT1G70710	PAC	GH9 (AtCEL1)		+		+
AT5G23210	P	AtSCPL34		+		+
AT1G31550	LM	lipase acylhydrolase (GDSL family)		+		+

The ecotypes used in this study were originated from contrasted environments and they all shown different response amplitude to low temperature. The correlation between components from each block varied from 0.87 when the proteomics and transcriptomics blocks were compared to 0.56 when the metabolomics and the phenomics blocks were compared (Fig. 8A). Regarding the individual plots, each blocks discriminate different ecotypes (Fig. 8B). Thus, i) the proteomic blocks clustered Roch with Grip, Col with Hern and discriminated Hosp, ii) the transcriptomic block exhibited the same profile but with a better segregation ii) the phenomic block allowed only to separate Roch at 22°C and the metabolomics block did not allowed to clusterize one condition. These differences of profiles were induced by contrasted expression of variables represented in the figure 8C.

The clustered image map of the samples from each block involved in the discrimination according to the ecotypes show a segregation in three groups (Fig. 8C). The first one contained only Hosp, the second combined Roch and Grip at 15°C and the last one was composed by Hern, Col and Grip at 22°C. The specific profile of the Hosp was supported by a pool of genes and proteins sort in the table 5 (see also Supplementary Table S4B). A majority of these candidates has unknown function (27%) like the expressed protein AT5G38980 found at both protein and transcript levels. Studying specifically this gene could reveal one major actor of the CW plasticity. The second functional class represented in this ecotype is the LM. Two non-specific lipid transfer protein were found in this set of candidates. The protein LTPG6 (AT1G55260) was more abundant and, the gene *LTP6* (*AT3G08770*) was up-regulated in Hosp at 15°C. This results, coupled to the *LTP3* and *LTP4* found more abundant at low temperature may be help to understanding the higher level of hydrophobic compounds found in the CW of Hosp compared to Col observe in the micro-phenotyping study. Based on these results, the cold acclimation may be eased by the lipid transport. Furthermore, some variables of the phenomic (density, number of leaves and the weight), the metabolomic blocs (XG contents) and genes / proteins were clustered on the dendogram. This grouping was mainly due to the higher values of Grip and Hern at 15°C, Roch at both temperatures and the lower values of Hosp (Fig. 8C). Indeed, this group contains six transcripts including two Prxs (*AT3G49110*, *AT1G71695*), one LTP (AT2G44290) and also one protein (Glycosyl hydrolase; AT3G19620). These list, which clustered with the variables from other blocs could increase the list of candidates implicated on the modification of the CW by the XG and it implication in cold acclimation.

The behaviour of the ecotypes to the temperature effect were different at the floral stem. Correlations between components from each block varied more strongly passing from 0.95

between the proteomics and transcriptomics to 0.25 between the metabolomics and the phenomics blocks (Fig. 9A). This large range of correlation was due to better grouping of the samples observed in the proteomics and transcriptomics blocks and worst the metabolomics and the phenomics blocks (Fig. 9B). This observation was confirmed by the strong isolation of Hern in the first axe and of Hosp in the second one regarding the proteomics and transcriptomics blocks on the individual plots (Fig. 9B). On the other hand, only the ecotype Hern can be separated between the two growth temperature conditions with the phenomic block but not as clearly observed with the two previous blocks. All samples are well separated by their specific variability of a few number of variable viewable with the clustered image map (Fig. 9C). This analyse has allow to segregated the samples in three major groups: Hern, Hosp and the others ecotypes. By this sparse approach, a pool of genes and proteins differentially present in floral stems of the ecotypes Hosp and Hern were sort in the table 6 (see also Supplementary Table S5B). This analyse, discriminating for the ecotype variable did not provide any common candidates with this analyse focus to the temperatures, because the specificity of each qualitative blocs appears too important to allow to isolate the same candidate for the temperature and the ecotype analysis.

As previously mentioned, the floral stems of Hern appeared clearly different from the others ecotypes. This specificity could be noticed by the cluster grouping of all the phenomics and metabolomics variables, except the ratio of branching of RGI. Co-expressed with these blocks, 6 proteins were more abundant in this ecotype among them, 3 belonged to the interaction domains class (ID) and 2 to the PAC class. Among the ID, the two lectins associated (*AT3G15356, AT3G16530*) were known to be sensible to abiotic stress and were found up-regulated with hormonal responses (Lyou *et al.* 2009; Wong *et al.* 2006). The other ID protein, *PR4* (*AT3G04720*), was also positively regulates to the jasmonate and the ethylene (Catinot *et al.* 2015). The β-1,3-endoglucanase proteins of the PAC class (AT4G16260) was also an ethylene-regulated gene (Zhong *et al.* 2003). These results highlight the link between the phenotype and the hormones on the stems of Hern. Hormones were not analysis in this study but could play a major role to the phenotype of this organ. Interestingly, at the gene expression level, the proteases, largely present in this set of genes (11 / 20) and the PAC (7 / 20), are down regulated in Hern. Proteases play a role in the generation of signals involved in the development and also contribute to the turnover of CWPs (Feiz *et al.* 2006). This process, still poorly understood may have a role in the reduction of the diameter and the length the floral stems of

Hern at low temperature compared to the other ecotypes. Hern is an interesting population to studying this class of proteins and their contribution of CW plasticity in floral stems.

The ecotype Hosp shown also a specific profile at multi-level. The relative specificity of this ecotype was only viewable at the proteomic and transcriptomic level. With less contrasted altered phenotypic than Hern, Hosp demonstrated fewer representativeness of a specific class of genes and proteins. But, some candidates appears as specific to this ecotype like the serine AT2G12480 and the aspartic proteases AT3G02740 proteins. And, *in contrario* to the candidates of Hern down-regulated, the genes highlighted by this analyse appears as specifically up-regulated in Hosp. As example, the lectin (*AT1G78820*) belonging to the PAC class was induced more than 2 times compare to the ecotype Col. Moreover, two candidates already found in the stem epidermis (Suh *et al.* 2005) specific to the ecotype Hosp: the protein LTPG1 (AT1G27950) more abundant at 15°C and the transcript of the *PME41* (*AT4G02330*). The protein LTPG1 is known to be associated to LTPG2 inducing a cuticular wax accumulation (Kim *et al.* 2012). And, *PME41* are sensitive to chilling stress and regulated by brassinosteroids (Qu *et al.* 2011). This PME could be play a role in plant chilling tolerance by modifying the mechanical properties of the CW. The viewable phenotype of Hosp may be determined at the epidermis. Furthermore, the possible regulation of the PMEs by hormones like the brassinosteroids lead to investigate their actions in the plants cold acclimation.

CONCLUSIONS

This study has compared at multi-level natural ecotypes initially originated from contrasted environments in the Pyrenees. The aims of this study was to identify new molecular players of the CW plasticity to acclimation to sub-optimal growth temperature conditions by an integrative study. The ecotypes showed different phenotypes depending to the growth temperature and the organs studied. The complex response have been highlighted at different level. For example, the increase of the XG in rosettes and decrease in floral stems at 15°C illustrate the contrasted CW composition. The differential CW plasticity mechanisms between organs can be also illustrated by the CW proteomes and transcriptomes modified differently depending to the ecotype and the temperature. The integrative analysis approach is an effective tool to sort and highlighted complex and specific mechanisms of the ecotypes to sub-optimal conditions.

The global integration of all blocks allowed to found new candidates in the LTP family to better understand the establishing of the cuticle layer. The difference observe in the rosette or floral stems coupled to the natural variability of response like the ecotype Hosp gave a powerful tool to studying the cuticle. The specificity of each ecotypes related to genes and proteins may be important to identified new candidates, new hypotheses or new regulations of the CW plasticity.

This project revealed that cold acclimation affects the expression of several genes and proteins with little or unknown biological functions, thereby providing new putative candidates targets for future research. Most importantly, this powerful integrative analysis has allowed to identify some interesting candidates for exploring the adaptation of natural populations to cold acclimation, further studies using mutants with altered expression of these targets are necessary to elucidate their biological significance. Although the amount of work may grow exponentially, hypothesis-based approaches exploring understudied traits can provide extremely interesting new insights and new questions in *A. thaliana* research and by transfer of knowledge to agronomical plants.

Acknowledgment

The authors are thankful to the Paul Sabatier Toulouse 3 University and to the *Centre National de la Recherche Scientifique* (CNRS) for granting their work. This work was also supported by the French Laboratory of Excellence project "TULIP" (ANR-10-LABX-41; ANR-11-IDEX-0002-02). HD is supported by a Midi Pyrenees Region and the federal university of Toulouse. The authors are grateful to Jean-Malo Couzigou for his support for the RNA extraction.

Supplementary data

Figure S1. Morphological impacts of the rosettes at contrasted growing conditions.

Figure S2. Morphological impacts of the floral stems at contrasted growing conditions

Figure S3. Reconstruction of the main cell wall polysaccharides from monosaccharide analysis

Table S1. Statistical analysis of proteomic data.xls

Table S2. Statistical analysis of rosettes transcriptomics data.xls

Table S3. Statistical analysis of floral stems transcriptomics data.xls

Table S4. Statistical results of the sparse MB-PLS-DA of rosettes.xls

Table S5. Statistical results of the sparse MB-PLS-DA of floral stems.xls

REFERENCES

Albenne C, Canut H, Jamet E (2013) Plant cell wall proteomics: the leadership of *Arabidopsis thaliana*. *Front Plant Sci* **4**, 111.

Alonso-Blanco C, Mendez-Vigo B, Koornneef M (2005) From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *Int J Dev Biol* **49**, 717-732.

Auge GA, Leverett LD, Edwards BR, Donohue K (2017) Adjusting phenotypes via within- and across-generational plasticity. *New Phytol*.

Botto JF (2015) Plasticity to simulated shade is associated with altitude in structured populations of *Arabidopsis thaliana*. *Plant Cell Environ* **38**, 1321-1332.

Boyes DC, Zayed AM, Ascenzi R, *et al.* (2001) Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* **13**, 1499-1510.

Braidwood L, Breuer C, Sugimoto K (2014) My body is a cage: mechanisms and modulation of plant cell growth. *New Phytol* **201**, 388-402.

Brousseau L, Postolache D, Lascoux M, *et al.* (2016) Local Adaptation in European Firs Assessed through Extensive Sampling across Altitudinal Gradients in Southern Europe. *PLoS One* **11**, e0158216.

Byars SG, Papst W, Hoffmann AA (2007) Local adaptation and cogradient selection in the alpine plant, *Poa hiemata*, along a narrow altitudinal gradient. *Evolution* **61**, 2925-2941.

Catinot J, Huang JB, Huang PY, *et al.* (2015) ETHYLENE RESPONSE FACTOR 96 positively regulates *Arabidopsis* resistance to necrotrophic pathogens by direct binding to GCC elements of jasmonate - and ethylene-responsive defence genes. *Plant Cell Environ* **38**, 2721-2734.

Cosgrove DJ (2016) Plant cell wall extensibility: connecting plant cell growth with cell wall structure, mechanics, and the action of wall-modifying enzymes. *J Exp Bot* **67**, 463-476.

Delker C, Quint M (2011) Expression level polymorphisms: heritable traits shaping natural variation. *Trends Plant Sci* **16**, 481-488.

Duruflé H, Clemente HS, Balliau T, *et al.* (2017a) Cell wall proteome analysis of *Arabidopsis thaliana* mature stems. *Proteomics* **17**.

Duruflé H, Hervé V, Ranocha P, *et al.* (2017b) Cell wall modifications of two *Arabidopsis thaliana* ecotypes, Col and Sha, in response to sub-optimal growth conditions: an integrative study. *Plant Science.*

Duruflé H, Selmani M, Ranocha P, *et al.* (*in preparation*) A framework for omics data analysis: from univariate statistics to multi-block integrative analysis.

Duruflé H, Ranocha P, *et al. (in review)* Phenotypic trait variation in genetically distinct *Arabidopsis thaliana* populations from the Pyrenees Mountains highlight acclimation to environmental constraints.

Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E (2006) Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis* hypocotyls. *Plant Methods* **2**, 10.

Fischer M, Weyand A, Rudmann-Maurer K, Stöcklin J (2011) Adaptation of *Poa alpina* to altitude and land use in the Swiss Alps. *Alpine botany* **121**, 91.

Francoz E, Ranocha P, Pernot C, *et al.* (2016) Complementarity of medium-throughput in situ RNA hybridization and tissue-specific transcriptomics: case study of *Arabidopsis* seed development kinetics. *Sci Rep* **6**, 24644.

Franková L, Fry SC (2013) Biochemistry and physiological roles of enzymes that 'cut and paste' plant cell-wall polysaccharides. *J Exp Bot* **64**, 3519-3550.

Gonzalo-Turpin H, Hazard L (2009) Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. *Journal of Ecology* **97**, 742-751.

Guo L, Yang H, Zhang X, Yang S (2013) Lipid transfer protein 3 as a target of *MYB96* mediates freezing and drought stress in *Arabidopsis*. *J Exp Bot* **64**, 1755-1767.

Günther T, Lampei C, Barilar I, Schmid KJ (2016) Genomic and phenotypic differentiation of *Arabidopsis thaliana* along altitudinal gradients in the North Italian Alps. *Mol Ecol* **25**, 3574-3592.

Hamann E, Scheepens JF, Kesselring H, Armbruster GFJ, Stöcklin J (2017) High intraspecific phenotypic variation, but little evidence for local adaptation in *Geum reptans* populations in the Central Swiss Alps. *Alpine Botany*, 1-12.

Hervé V, Duruflé H, San Clemente H, *et al.* (2016) An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes. *Proteomics* **16**, 3183–3187.

Houben K, Jolie R, Fraeye I, Van Loey A, Hendrickx M (2011) Comparative study of the cell wall composition of broccoli, carrot, and tomato: Structural characterization of the extractable pectins and hemicelluloses. *Carbohydrate Research* **346**, 1105-1111.

Irshad M, Canut H, Borderies G, Pont-Lezica R, Jamet E (2008) A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers. *BMC Plant Biol* **8**, 94.

Kim KW, Moinuddin SG, Atwell KM, *et al.* (2012) Opposite stereoselectivities of dirigent proteins in *Arabidopsis* and *schizandra* species. *J Biol Chem* **287**, 33957-33972.

Körner C (2007) The use of 'altitude' in ecological research. *Trends Ecol Evol* **22**, 569-574.

Langella O, Valot B, Balliau T, *et al.* (2017) X!TandemPipeline: A Tool to Manage Sequence Redundancy for Protein Inference and Phosphosite Identification. *J Proteome Res* **16**, 494-503.

Le Gall H, Philippe F, Domon JM, *et al.* (2015) Cell wall metabolism in response to abiotic stress. *Plants* **4**, 112-166.

Lee SC, Kim SH, An SH, Yi SY, Hwang BK (2006) Identification and functional expression of the pepper pathogen-induced gene, CAPIP2, involved in disease resistance and drought and salt stress tolerance. *Plant Mol Biol* **62**, 151-164.

Luo Y, Widmer A, Karrenberg S (2015) The roles of genetic drift and natural selection in quantitative trait divergence along an altitudinal gradient in *Arabidopsis thaliana*. *Heredity (Edinb)* **114**, 220-228.

Lyou SH, Park HJ, Jung C, *et al.* (2009) The *Arabidopsis AtLEC* gene encoding a lectin-like protein is up-regulated by multiple stimuli including developmental signal, wounding, jasmonate, ethylene, and chitin elicitor. *Mol Cells* **27**, 75-81.

Majid MU, Awan MF, Fatima K, *et al.* (2017) Genetic resources of chili pepper (*Capsicum annuum* L.) against *Phytophthora capsici* and their induction through various biotic and abiotic factors. *Cytology and Genetics* **51**, 296-304.

Meehl GA, Stocker TF, Collins WD, *et al.* (2007) Global climate projections.

Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* **441**, 947-952.

Montesinos A, Tonsor SJ, Alonso-Blanco C, Picó FX (2009) Demographic and genetic patterns of variation among populations of *Arabidopsis thaliana* from contrasting native environments. *PLoS One* **4**, e7213.

Oono Y, Seki M, Satou M, *et al.* (2006) Monitoring expression profiles of *Arabidopsis* genes during cold acclimation and deacclimation using DNA microarrays. *Funct Integr Genomics* **6**, 212-234.

Pico FX (2012) Demographic fate of *Arabidopsis thaliana* cohorts of autumn- and spring-germinated plants along an altitudinal gradient. *Journal of Ecology* **100**, 1009-1018.

Qu T, Liu R, Wang W, *et al.* (2011) Brassinosteroids regulate pectin methylesterase activity and *AtPME41* expression in *Arabidopsis* under chilling stress. *Cryobiology* **63**, 111-117.

Reich PB, Wright IJ, Cavender-Bares J, *et al.* (2003) The evolution of plant functional variation: traits, spectra, and strategies. *International Journal of Plant Sciences* **164**, S143-S164.

Rizhsky L, Liang H, Shuman J, *et al.* (2004) When defense pathways collide. The response of *Arabidopsis* to a combination of drought and heat stress. *Plant Physiol* **134**, 1683-1696.

Ruzin SE (1999) *Plant microtechnique and microscopy* Oxford University Press New York.

San Clemente H, Pont-Lezica R, Jamet E (2009) Bioinformatics as a tool for assessing the quality of sub-cellular proteomic strategies and inferring functions of proteins: plant cell wall proteomics as a test case. *Bioinform Biol Insights* **3**, 15-28.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675.

Sharma N, Cram D, Huebert T, Zhou N, Parkin IA (2007) Exploiting the wild crucifer *Thlaspi arvense* to identify conserved and novel genes expressed during a plant's response to cold stress. *Plant Mol Biol* **63**, 171-184.

Srebotnik E, Messner K (1994) A simple method that uses differential staining and light microscopy to assess the selectivity of wood delignification by white rot fungi. *Appl Environ Microbiol* **60**, 1383-1386.

Suh MC, Samuels AL, Jetter R, *et al.* (2005) Cuticular lipid composition, surface structure, and gene expression in *Arabidopsis* stem epidermis. *Plant Physiol* **139**, 1649-1665.

Sultan SE (2004) Promising directions in plant phenotypic plasticity. *Perspectives in Plant Ecology, Evolution and Systematics* **6**, 227-233.

Suter L, Rüegg M, Zemp N, Hennig L, Widmer A (2014) Gene regulatory variation mediates flowering responses to vernalization along an altitudinal gradient in *Arabidopsis*. *Plant Physiol* **166**, 1928-1942.

Trontin C, Tisné S, Bach L, Loudet O (2011) What does *Arabidopsis* natural variation teach us (and does not teach us) about adaptation in plants? *Curr Opin Plant Biol* **14**, 225-231.

Tyagi A, Singh S, Mishra P, *et al.* (2015) Genetic diversity and population structure of *Arabidopsis thaliana* along an altitudinal gradient. *AoB Plants* **8**.

Tyagi A, Yadav A, Tripathi AM, Roy S (2016) High light intensity plays a major role in emergence of population level variation in *Arabidopsis thaliana* along an altitudinal gradient. *Sci Rep* **6**, 26160.

Valot B, Langella O, Nano E, Zivy M (2011) MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* **11**, 3572-3577.

Vazquez-Cooz I, Meyer RW (2002) A differential staining method to identify lignified and unlignified tissues. *Biotech Histochem* **77**, 277-282.

Völler E, Bossdorf O, Prati D, Auge H (2017) Evolutionary responses to land use in eight common grassland plants. *Journal of Ecology*.

Wolfe MD, Tonsor SJ (2014) Adaptation to spring heat and drought in northeastern Spanish *Arabidopsis thaliana*. *New Phytol* **201**, 323-334.

Wong CE, Li Y, Labbe A, *et al.* (2006) Transcriptional profiling implicates novel interactions between abiotic stress and hormonal responses in *Thellungiella*, a close relative of *Arabidopsis*. *Plant Physiol* **140**, 1437-1450.

Zhong GY, Zhong GV, Burns JK (2003) Profiling ethylene-regulated gene expression in *Arabidopsis thaliana* by microarray analysis. *Plant Mol Biol* **53**, 117-131.

Figure S1. Phenotype of rosettes in contrasted growth conditions. A, Rosette diameter; B, Rosette Mass; C, Rosette Density ; D, Projected rosette area.
Measurements were done during the harvest. For each condition, 20 plants from 3 independent batches have been analyzed. Mean values are represented.

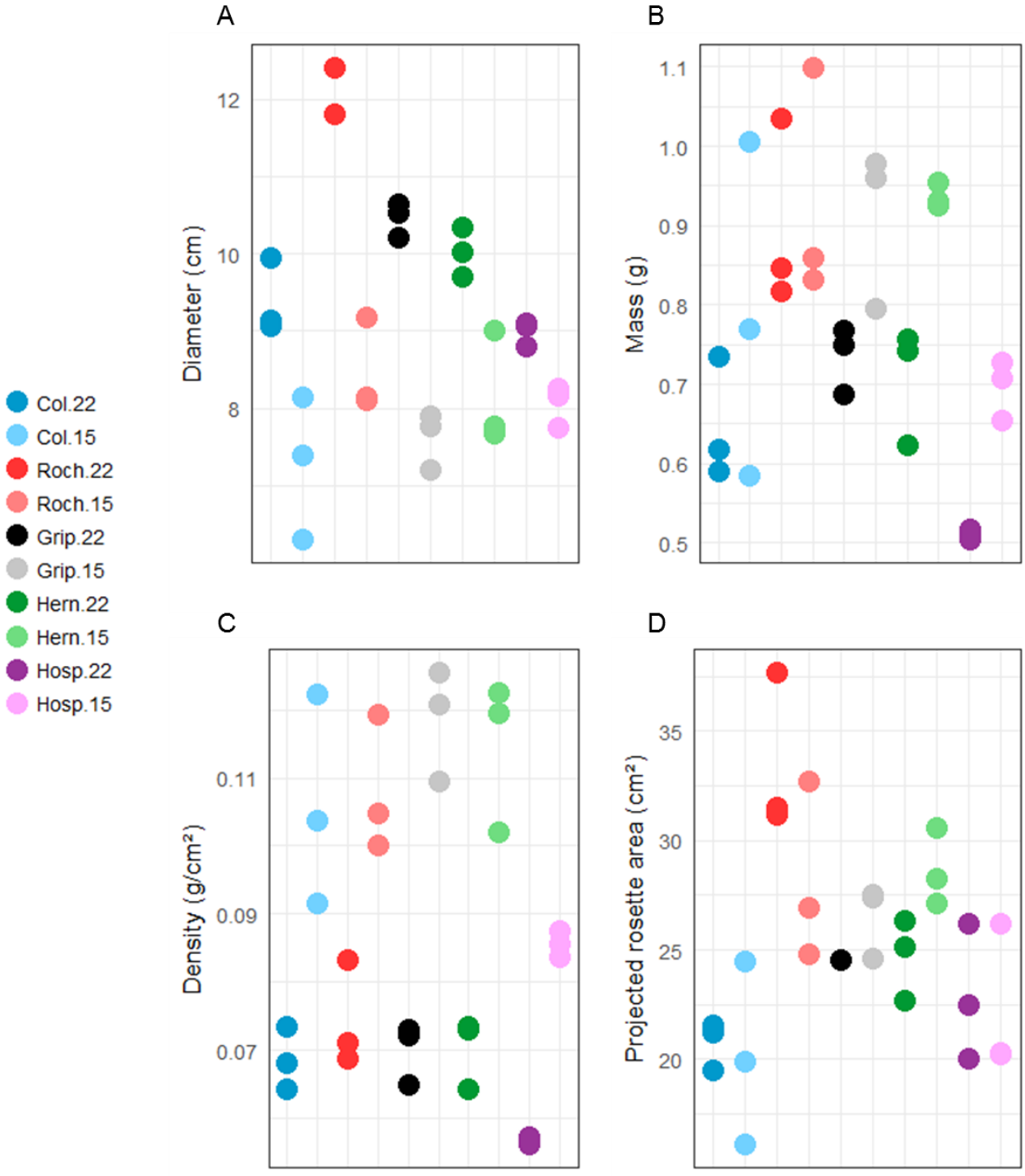


Figure S2. Phenotype of floral stems in contrasted growth conditions. A, Mass; B, Diameter; C, Length.
Measurements were done before the harvest. For each condition, 20 plants from 3 independent batches have been analyzed. Mean values are represented.

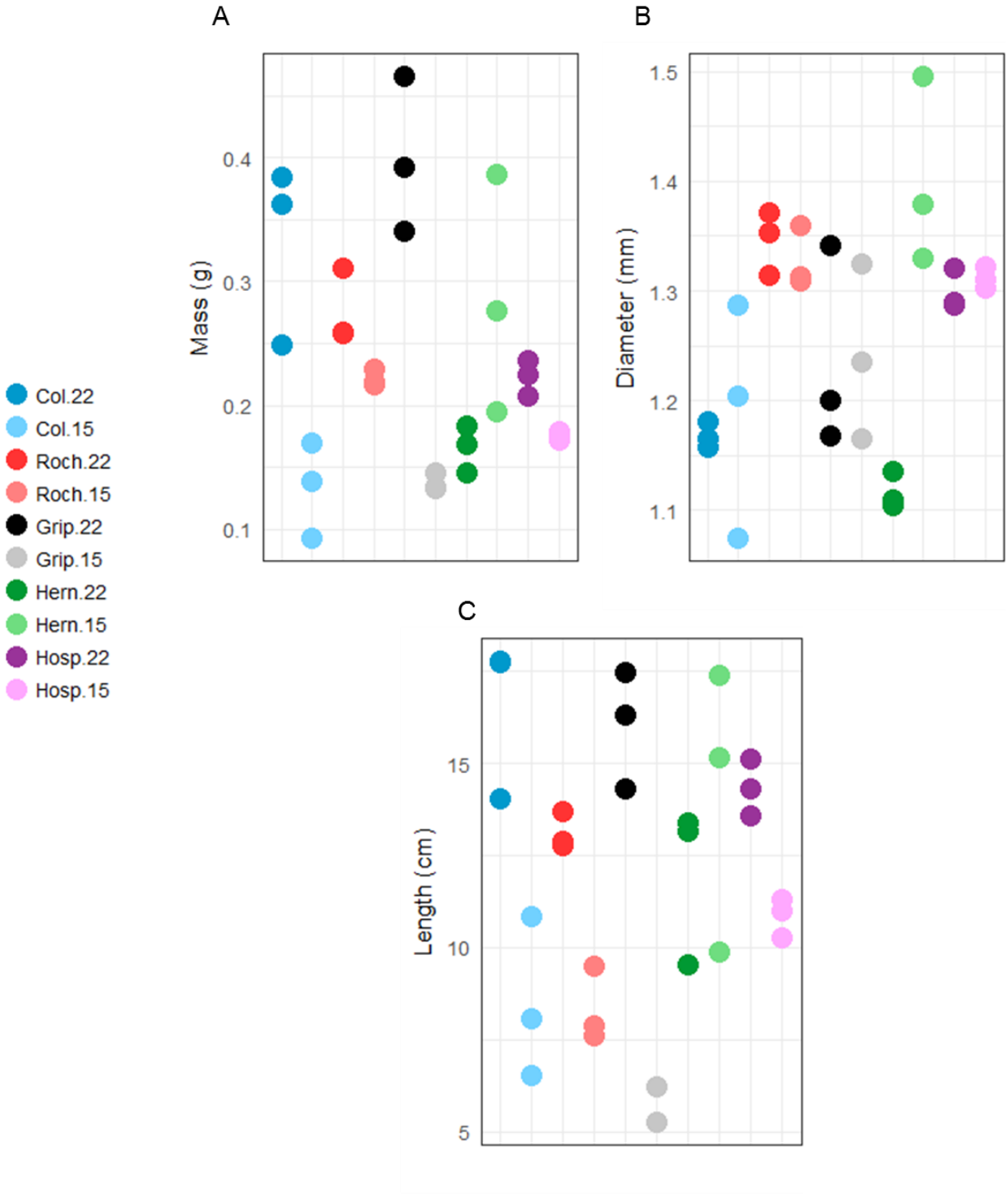


Figure S3: Reconstruction of the main cell wall polysaccharides from monosaccharide analysis. Monosaccharide composition rosettes (A) and floral stems (B) plants grown at 22 °C and at 15 °C has been obtained. The formulas used to rebuild the polysaccharides are given in Table 2. Mean values were analyzed from tree independent batch. XG: xyloglucans; RGI: rhamnogalacturonan I; HG: homogalacturonan.

L’analyse intégrative de données hétérogènes ou multi-blocs a été utilisé. En effet, ce projet regroupe quatre blocs quantitatifs de données (phénomique, métabolomique, protéomique et transcriptomiques) et 20 variables différentes (5 écotypes et 2 températures). L’analyse systématiques de chaque bloc de données de façon indépendante est envisageable mais la mise en relation des blocs de données entre eux étant inconcevable, l’approche statistique a permis de trier les différentes informations ainsi que de faire ressortir les gènes et protéines les plus représentative d’une condition. L’interprétation biologique de tous les résultats n’est pas encore finaliser au vu du nombre d’illustration graphique et de listes de gènes/protéines candidats obtenues dans ce projet. Cependant, les premiers résultats présentés ici sont prometteur.

L’utilisation de la diversité naturelle est un atout majeur pour étudier et appréhender la capacité d’adaptation et d’acclimatation des plantes à leurs milieux naturels. Leurs différences génétiques et leurs altitudes d’origines contrastées nous ont permis également d’appréhender l’évolution de la plasticité phénotypique de cette espèce au changement climatique.

Enfin, les études et l’analyse de grands jeux de données hétérogènes ont permis de d’identifier des protéines et des gènes dont les fonctions restent inconnues dans le remodelage pariétal. Ces nouveaux candidats ouvrent la voie à des études fonctionnelles qui permettront de mieux comprendre les mécanismes induisant des modifications des parois face à des températures de croissance sub-optimales.

V. Conclusion générale et perspectives

Pour aboutir, ce projet a dû mettre en lien des disciplines diverses : l'écologie, la biologie et la statistique. De par la complexité de l'interdisciplinarité, la stratégie adoptée fût de réaliser plusieurs études préliminaires. Ces pré-projets ont été nourris par de riches discussions pour définir un vocabulaire, des méthodologies ou encore les concepts compris de chaque communauté. Ainsi, les deux études en protéomique ont permis l'élaboration de nouveaux modèles de modification post-traductionnelle, l'étude génétique des populations naturelles a été complétée par des analyses intégratives couplant des données génétiques, phénotypiques et environnementales. Enfin, l'aboutissement d'une première étude intégrative combinant des données multi-omiques nous a permis dans un premier temps, sur un petit nombre de variables (deux écotypes à deux températures), d'avoir un premier retour sur les difficultés rencontrées lors de la gestion de grands jeux de données.

L'approche innovante du projet *WallOmics* pour l'analyse d'un grand nombre de jeux de données a fait l'objet de trois années de discussion et d'interaction entre des biologistes et des statisticiens. Cette interaction a abouti à la rédaction d'une publication de méthodologie statistique qui rendra, nous l'espérons, plus faciles et plus sereines les futures analyses intégratives. De plus, l'utilisation et le choix des représentations graphiques associées ont participé à la compréhension et à l'interprétation de chaque résultat d'analyse. Appliqué dans le projet *WallOmics*, cette méthodologie d'analyse bloc par bloc puis *via* l'intégration de chacun d'entre eux a été extrêmement efficace. Elle a notamment permit de faire ressortir des protéines et des transcrits ayant un profil spécifique d'une condition ou d'un écotype donné, essentiel pour de futures analyses fonctionnelles.

L'identification de ces nouveaux candidats va permettre l'orientation de futures recherches. Ainsi, ils pourraient par exemple servir de marqueurs de stress dans un organe donné. Certains candidats révélés par ces analyses peuvent également donner lieu à de nouvelles pistes d'étude de mécanisme moléculaire chez *A. thaliana* lors de recherches menées sur des réponses aux variations de la température. Ces analyses fonctionnelles pourront également être élargies à d'autres organes de la plante.

Pour faire suite à ce projet portant sur les organes aériens d'Arabidopsis et d'élargir le champ de recherche à un nouvel organe, des analyses phénotypiques sur les racines ont été entreprises. Dans la volonté d'observer la réponse de nos populations naturelles aux changements de température, les analyses ont été réalisées à 15, 22 et 28°C. Cette dernière température permettra d'observer l'impact des températures modérément hautes sur la croissance racinaire. De plus, l'environnement naturel d'une plante n'étant pas restreint qu'à

des variables climatiques, des analyses pédologiques sur les lieux de récolte des populations pourraient être envisagées. Ces nouveaux paramètres permettrons d'enrichir notre vue sur l'adaptation des plantes à leurs environnements naturels. Dans cet objectif, des analyses préliminaires de phénotypage racinaire ont également débuté en condition de carence en phosphate et stress salin.

Pour finir, ce projet transdisciplinaire a généré de riches interactions qui ont pu être valorisées *via* des publications mais outre cela, il a aussi permis le développement d'un réseau scientifique offrant la possibilité de mener des études complexes par des approches innovantes sur des populations naturelles. Ce projet ouvre la voie à des études combinant des données multi-omiques sur des espèces d'intérêt agronomique et pouvant porter sur des combinaisons de stress abiotique et biotique.

VI. Références

Albenne C, Canut H, Boudart G, *et al.* (2009) Plant cell wall proteomics: mass spectrometry data, a trove for research on protein structure/function relationships. *Mol Plant* **2**, 977-989.
Barker M, Rayens W (2003) Partial least squares for discrimination. *Journal of chemometrics* **17**, 166-173.
Beniston M (2003) Climatic change in mountain regions: a review of possible impacts. *Climatic change* **59**, 5-31.
Berry J, Bjorkman O (1980) Photosynthetic response and adaptation to temperature in higher plants. *Annual Review of plant physiology* **31**, 491-543.
Boudart G, Jamet E, Rossignol M, *et al.* (2005) Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: identification by mass spectrometry and bioinformatics. *Proteomics* **5**, 212-221.
Boyes DC, Zayed AM, Ascenzi R, *et al.* (2001) Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* **13**, 1499-1510.
Braidwood L, Breuer C, Sugimoto K (2014) My body is a cage: mechanisms and modulation of plant cell growth. *New Phytol* **201**, 388-402.
Bro R, Smilde AK (2014) Principal component analysis. *Analytical Methods* **6**, 2812-2831.
Burghardt LT, Metcalf CJ, Wilczek AM, Schmitt J, Donohue K (2015) Modeling the influence of genetic and environmental variation on the expression of plant life cycles across landscapes. *Am Nat* **185**, 212-227.
Carroll JD (1968) Generalization of canonical correlation analysis to three or more sets of variables **3**, 227-228.
Chawla K, Barah P, Kuiper M, Bones AM (2011) Systems biology: a promising tool to study abiotic stress responses. *Omics and Plant Abiotic Stress Tolerance*, 163-172.
Cheng CY, Krishnakumar V, Chan AP, *et al.* (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J* **89**, 789-804.
Cosgrove DJ (2001) Wall structure and wall loosening. A look backwards and forwards. *Plant Physiol* **125**, 131-134.
Cosgrove DJ (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol* **6**, 850-861.
Domon JM, Baldwin L, Acket S, *et al.* (2013) Cell wall compositional modifications of *Miscanthus* ecotypes in response to cold acclimation. *Phytochemistry* **85**, 51-61.
Duruflé H, Clemente H. S, Balliau T, Zivy M, Dunand C, Jamet E (2017). Cell wall proteome analysis of *Arabidopsis thaliana* mature stems. *Proteomics*, 17(8).
Déjardin A, Laurans F, Arnaud D, *et al.* (2010) Wood formation in Angiosperms. *C R Biol* **333**, 325-334.
El-Lithy ME, Clerkx EJ, Ruys GJ, Koornneef M, Vreugdenhil D (2004) Quantitative trait locus analysis of growth-related traits in a new *Arabidopsis* recombinant inbred population. *Plant Physiol* **135**, 444-458.
Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E (2006) Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis* hypocotyls. *Plant Methods* **2**, 10.
Francoz E, Ranocha P, Nguyen-Kim H, *et al.* (2015) Roles of cell wall peroxidases in plant development. *Phytochemistry* **112**, 15-21.
Franková L, Fry SC (2013) Biochemistry and physiological roles of enzymes that 'cut and paste' plant cell-wall polysaccharides. *J Exp Bot* **64**, 3519-3550.
Ghosh D, Xu J (2014) Abiotic stress responses in plant roots: a proteomics perspective. *Front Plant Sci* **5**, 6.
Gonzalo-Turpin H, Hazard L (2009) Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. *Journal of Ecology* **97**, 742-751.

González I, Déjean S, Martin P, *et al.* (2009) Highlighting relationships between heterogeneous biological data through graphical displays based on regularized canonical correlation analysis. *Journal of Biological Systems* **17**, 173-199.
Günther OP, Shin H, Ng RT, *et al.* (2014) Novel multivariate methods for integration of genomics and proteomics data: applications in a kidney transplant rejection study. *Omics: a journal of integrative biology* **18**, 682-695.
Hamann T (2012) Plant cell wall integrity maintenance as an essential component of biotic stress response mechanisms. *Front Plant Sci* **3**, 77.
Haslam RP, Downie AL, Raveton M, *et al.* (2003) The assessment of enriched apoplastic extracts using proteomic approaches. *Annals of Applied Biology* **143**, 81--91.
Houben K, Jolie R, Fraeye I, Van Loey A, Hendrickx M (2011) Comparative study of the cell wall composition of broccoli, carrot, and tomato: Structural characterization of the extractable pectins and hemicelluloses. *Carbohydrate Research* **346**, 1105-1111.
Huala E, Dickerman AW, Garcia-Hernandez M, *et al.* (2001) The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res* **29**, 102-105.
Irshad M, Canut H, Borderies G, Pont-Lezica R, Jamet E (2008) A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers. *BMC Plant Biol* **8**, 94.
Jacq A, Pernot C, Martinez Y, *et al.* (2017) The Arabidopsis Lipid Transfer Protein 2 (*AtLTP2*) Is Involved in Cuticle-Cell Wall Interface Integrity and in Etiolated Hypocotyl Permeability. *Frontiers in Plant Science* **8**.
Jiang Z, Liu X, Peng Z, *et al.* (2011) AHD2.0: an update version of *Arabidopsis* Hormone Database for plant systematic studies. *Nucleic Acids Res* **39**, D1123-1129.
Kettenring JR (1971) Canonical analysis of several sets of variables. *Biometrika* **58**, 433-451.
Krämer U (2015) Planting molecular functions in an ecological context with *Arabidopsis thaliana*. *Elife* **4**.
Kumar M, Atanassov I, Turner S (2017) Functional Analysis of Cellulose Synthase (CESA) Protein Class Specificity. *Plant Physiol* **173**, 970-983.
Körner C (2003) Alpine plant life: functional plant ecology of high mountain ecosystems Springer Science & Business Media.
Körner C (2007) The use of 'altitude' in ecological research. *Trends Ecol Evol* **22**, 569-574.
Le Gall H, Philippe F, Domon JM, *et al.* (2015) Cell wall metabolism in response to abiotic stress. *Plants* **4**, 112-166.
Lebart L, Piron M, Morineau A (2006) Statistique exploratoire multidimensionnelle: visualisation et inférences en fouilles de données.
Leucci MR, Lenucci MS, Piro G, Dalessandro G (2008) Water stress and cell wall polysaccharides in the apical root zone of wheat cultivars varying in drought tolerance. *J Plant Physiol* **165**, 1168-1180.
Lê Cao KA, González I, Déjean S (2009) integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* **25**, 2855-2856.
Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of statistical software* **25**, 1-18.
Meehl GA, Stocker TF, Collins WD, *et al.* (2007) Global climate projections.
Minic Z, Jamet E, Négroni L, *et al.* (2007) A sub-proteome of *Arabidopsis thaliana* mature stems trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. *J Exp Bot* **58**, 2503-2512.
Mitchell-Olds T, Schmitt J (2006) Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* **441**, 947-952.

Pérez-Enciso M, Tenenhaus M (2003) Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Human genetics* **112**, 581-592.
Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol* **133**, 1051-1071.
Rajasundaram D, Selbig J (2016) More effort - more results: recent advances in integrative 'omics' data analysis. *Curr Opin Plant Biol* **30**, 57-61.
Roland JC, Vian B (1979) The wall of the growing plant cell: its three-dimensional organization. *International Review of Cytology* **61**, 129-166.
Saha P, Sade N, Arzani A, *et al.* (2016) Effects of abiotic stress on physiological plasticity and water use of *Setaria viridis* (L.). *Plant Sci* **251**, 128-138.
Saporta G (2006) Probabilités, analyse des données et statistique Editions Technip.
Seren Ü, Grimm D, Fitz J, *et al.* (2017) AraPheno: a public database for *Arabidopsis thaliana* phenotypes. *Nucleic Acids Res* **45**, D1054-D1059.
Singh A, Gautier B, Shannon CP, *et al.* (2016) DIABLO-an integrative, multi-omics, multivariate method for multi-group classification. *bioRxiv*, 067611.
Somerville C, Bauer S, Brininstool G, *et al.* (2004) Toward a systems approach to understanding plant cell walls. *Science* **306**, 2206-2211.
Sticklen MB (2008) Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nat Rev Genet* **9**, 433-443.
Streb P, Aubert S, Bligny R (2003) High temperature effects on light sensitivity in the two high mountain plant species *Soldanella alpina* (L.) and *Rannunculus glacialis* (L.). *Plant Biology* **5**, 432-440.
Swarbreck D, Wilks C, Lamesch P, *et al.* (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res* **36**, D1009-1014.
Tenenhaus A, Tenenhaus M (2011) Regularized generalized canonical correlation analysis. *Psychometrika* **76**, 257-284.
Tenenhaus A, Tenenhaus M (2014) Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *European Journal of operational research* **238**, 391-403.
Tenhaken R (2014) Cell wall remodeling under abiotic stress. *Front Plant Sci* **5**, 771.
Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.
Tomasi P, Wang H, Lohrey GT, *et al.* (2017) Characterization of leaf cuticular waxes and cutin monomers of *Camelina sativa* and closely-related *Camelina* species. *Industrial Crops and Products* **98**, 130-138.
Trentin AR, Pivato M, Mehdi SM, *et al.* (2015) Proteome readjustments in the apoplastic space of *Arabidopsis thaliana ggt1* mutant leaves exposed to UV-B radiation. *Front Plant Sci* **6**, 128.
Van Eeuwijk FA, Bustos-Korts DV, Malosetti M (2016) What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype× Environment Interactions? *Crop Science* **56**, 2119-2140.
Van Norman JM, Benfey PN (2009) *Arabidopsis thaliana* as a model organism in systems biology. *Wiley Interdiscip Rev Syst Biol Med* **1**, 372-379.
Voragen AGJ, Coenen GJ, Verhoef RP, Schols HA (2009) Pectin, a versatile polysaccharide present in plant cell walls. *Structural Chemistry* **20**, 263.
Weigel D, Mott R (2009) The 1001 Genomes Project for *Arabidopsis thaliana*. *Genome Biol* **10**.
Wolf S, Hématy K, Höfte H (2012) Growth control and cell wall signaling in plants. *Annu Rev Plant Biol* **63**, 381-407.

Wood IP, Pearson BM, Garcia-Gutierrez E, *et al.* (2017) Carbohydrate microarrays and their use for the identification of molecular markers for plant cell wall composition. *Proc Natl Acad Sci U S A* **114**, 6860-6865.
Zhang W, Zhang H, Ning L, Li B, Bao M (2016) Quantitative Proteomic Analysis Provides Novel Insights into Cold Stress Responses in Petunia Seedlings. *Front Plant Sci* **7**, 136.

VII. Annexe

frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 October 2017
doi: 10.3389/fpls.2017.01802

Proline Hydroxylation in Cell Wall Proteins: Is It Yet Possible to Define Rules?

Harold Duruflé[1†], Vincent Hervé[1,2†], Thierry Balliau[3], Michel Zivy[3], Christophe Dunand[1] and Elisabeth Jamet[1]*

[1] Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Toulouse, France, [2] INRS – Institut Armand Frappier, Laval, Canada, [3] PAPPSO, GQE Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France

Cell wall proteins (CWPs) play critical and dynamic roles in plant cell walls by contributing to developmental processes and response to environmental cues. Since the CWPs go through the secretion pathway, most of them undergo post-translational modifications (PTMs) which can modify their biological activity. Glycosylation is one of the major PTMs of CWPs and refers to *N*-glycosylation, *O*-glycosylation and glypiation. Each of these PTMs occurs in different amino acid contexts which are not all well defined. This article deals with the hydroxylation of Pro residues which is a prerequisite for *O*-glycosylation of CWPs on hydroxyproline (Hyp) residues. The location of Hyp residues is well described in several structural CWPs, but yet rarely described in other CWPs. In this article, it is studied in detail in five *Arabidopsis thaliana* proteins using mass spectrometry data: one of them (At4g38770, AtPRP4) is a structural CWP containing 32.5% of Pro residues arranged in typical motifs, the others are either rich (27–28%, At1g31580 and At2g10940) or poor (6–8%, At1g09750 and At3g08030) in Pro residues. The known rules of Pro hydroxylation allowed a good prediction of Hyp location in AtPRP4. However, they could not be applied to the other proteins whatever their Pro content. In addition, variability of the Pro hydroxylation patterns was observed within some amino acid motifs in all the proteins and new patterns of Pro hydroxylation are described. Altogether, this work shows that Hyp residues are present in more protein families than initially described, and that Pro hydroxylation patterns could be different in each of them. This work paves the way for completing the existing Pro hydroxylation code.

Keywords: *Arabidopsis thaliana*, cell wall protein, hydroxyproline, mass spectrometry, proline hydroxylation, proline-rich protein, post-translational modification

OPEN ACCESS

Edited by:
Ján A. Miernyk, Agricultural Research Service (USDA), United States

Reviewed by:
Ian S. Wallace, University of Nevada, Reno, United States
Li Tan, University of Georgia, United States

***Correspondence:**
Elisabeth Jamet
jamet@lrsv.ups-tlse.fr

†Co-first authors

Specialty section:
This article was submitted to Plant Physiology, a section of the journal Frontiers in Plant Science

Received: *20 June 2017*
Accepted: *04 October 2017*
Published: *17 October 2017*

Citation:
Duruflé H, Hervé V, Balliau T, Zivy M, Dunand C and Jamet E (2017) Proline Hydroxylation in Cell Wall Proteins: Is It Yet Possible to Define Rules? Front. Plant Sci. 8:1802. doi: 10.3389/fpls.2017.01802

INTRODUCTION

Cell wall proteins (CWPs) are important players in plant cell walls, otherwise mainly constituted of polysaccharides, and eventually of phenolic compounds around differentiated lignified cells (Carpita and Gibeaut, 1993). They have been involved in the remodeling of cell wall polymers networks by hydrolysing covalent bounds, inserting newly synthesized polysaccharides, cross-linking together structural proteins, proteins and polysaccharides or polysaccharides and phenolic compounds (Franková and Fry, 2013; Cosgrove, 2015). Together with extracellular peptides, some CWPs have also been involved in signaling, thus allowing cell-to-cell communication

(Matsubayashi and Sakagami, 2006; Lamport and Várnai, 2013; Tavormina et al., 2015). Altogether, CWPs and peptides contribute to both developmental processes and response to environmental cues (Tenhaken, 2015; Borassi et al., 2016). Most of them undergo post-translational modifications (PTMs) during their transport through the secretion pathway (Faye et al., 2005; Kim and Brandizzi, 2016) which can modify their conformation, their biological activity and/or their ability to interact with cell wall components (Lannoo and Van Damme, 2015; Baer and Millar, 2016; Strasser, 2016). As an example, the site-directed mutagenesis of the *N*-glycosylation motifs of a class III peroxidase was shown to reduce its thermal stability, its catalytic activity and to modify its conformation (Lige et al., 2001).

During recent years, proteomics has facilitated a better knowledge of the plant cell wall proteome by increasing its coverage thanks to the design of specific strategies able to recover protein extracts enriched in extracellular proteins from organs of several model plants and crops (Lee et al., 2004; Albenne et al., 2013; Komatsu and Yanagawa, 2013; Rodríguez-Celma et al., 2016). Beyond the identification of proteins, technological advances have also permitted description of their PTMs. Various methods have been developed to address this particular question. In particular, immobilized affinity chromatography (IMAC) and lectin-affinity chromatography have allowed studying protein phosphorylation and glycosylation, respectively (Ytterberg and Jensen, 2010; Nakagami et al., 2012; Ruiz-May et al., 2014; Canut et al., 2016).

Glycosylation is one of the major PTMs of CWPs and refers to *N*-glycosylation, *O*-glycosylation and addition of glycophosphatidylinositol (GPI)-anchors, also named glypiation (Faye et al., 2005). Each of these PTMs occurs on specific amino acid sequences. *N*-glycosylation is the best described. It occurs on Asn residues in Asn-X-Ser/Thr motifs, where X cannot be a Pro residue. In these motifs, the hydroxyl functional group of Ser/Thr residues was shown to be required in the transglycosylation reaction on the Asn residue (Bause and Legler, 1981), whereas the presence of a Pro residue modifies the local conformation of the protein, thus preventing its *N*-glycosylation (Bause, 1983). The different structures of *N*-glycans are well-known, thus allowing systematic search in mass spectrometry (MS) data obtained in conditions preserving glycan-peptide bonds (Ruiz-May et al., 2012). GPI-anchors are transferred by a transamidase to a carboxy-terminal GPI-attachment signal peptide which can be predicted by bioinformatics (Eisenhaber et al., 2003). Several targeted proteomic studies have contributed to the identification of GPI-anchored proteins and some of them could be released from plasma membrane fractions by a phospholipases C or D which cleave GPI-anchors (Borner et al., 2003; Elortza et al., 2003; Elortza et al., 2005). *O*-glycosylation is the most complex type of glycosylation. In plant CWPs, it can occur on Ser and hydroxyproline (Hyp) residues (Faye et al., 2005). Galactose can be linked to Ser and Hyp residues whereas arabinose can only be linked to Hyp residues (Canut et al., 2016). According to the so-called Hyp contiguity hypothesis initially proposed for hydroxyproline-rich proteins (HRGPs), contiguous Hyp residues are arabinosylated and clustered non-contiguous Hyp residues are galactosylated (Shpak et al., 2001). Then, glycosyltransferases can extend the *O*-glycans in different ways depending on the initial pattern of Pro hydroxylation. The correct *O*-glycosylation of HRGPs was shown to be required for their conformation or their biological activity (Stafstrom and Staehelin, 1986; Velasquez et al., 2011).

Pro hydroxylation is a major step for *O*-glycosylation, but it is still difficult to predict in which amino acid context it occurs. In a previous review, we have proposed an extended Pro hydroxylation code (Canut et al., 2016), based on both (i) the initial Pro hydroxylation code (Kieliszewski and Lamport, 1994) and (ii) additional experimental LC-MS/MS and Edman sequencing data. The extended code has taken into account more protein and peptide families than the former one, including structural proteins like HRGPs and Pro/Hyp-rich proteins, solanaceous lectins, allergens, systemins and CLE peptides. Briefly, Pro residues could be hydroxylated when they are located after Ala, Gln, Hyp, Pro, Ser, Thr, and Val residues, whereas the first Pro residue following the other amino acids could not be hydroxylated (**Figure 1**). Only little information is available regarding Trp and Met residues. In a large mutagenesis screen performed on the amino acids surrounding the only Pro residue of sporamin shown to be hydroxylated, Trp and Met were not shown to favor the hydroxylation of the following Pro residue (Shimizu et al., 2005).

In this article, our aim was to test the extended Pro hydroxylation code on a new set of CWPs including non-structural CWPs. We have thus selected five CWPs with various contents in Pro residues. Three of them were rich in Pro residues among which AtPRP4 which is a structural CWP (Fowler et al., 1999) and two of them were poor in Pro residues. We have performed a deep data mining on two recent cell wall proteomic studies performed on rosettes and stems of *Arabidopsis thaliana* (Hervé et al., 2016; Duruflé et al., 2017). From the fine analysis of MS data, we have compared the observed patterns of Pro/Hyp location to the predicted ones according to the Pro hydroxylation extended code. The limits of the existing extended Pro hydroxylation code are discussed and new motifs are described.

RESULTS

Mapping of Hyp Residues

For this analysis, we have taken advantage of two cell wall proteomics studies which have lead to the identification of numerous CWPs, 361 in rosettes and 302 in stems, i.e., 397 different CWPs (Hervé et al., 2016; Duruflé et al., 2017). This body of data corresponded to three independent experiments (two for rosettes and one for stems), each of them including three biological replicates. The parameters used for peptide identification included a possible mass delta of 15.99 Da for each Pro residue, corresponding to its hydroxylation. As an example, in one of the rosette experiment, 79% of the identified CWPs were predicted to be *N*-glycosylated (presence of the PS00001 PROSITE motif), whereas 17.5% had at least one peptide carrying a Hyp residue. Among the latters and in addition to the

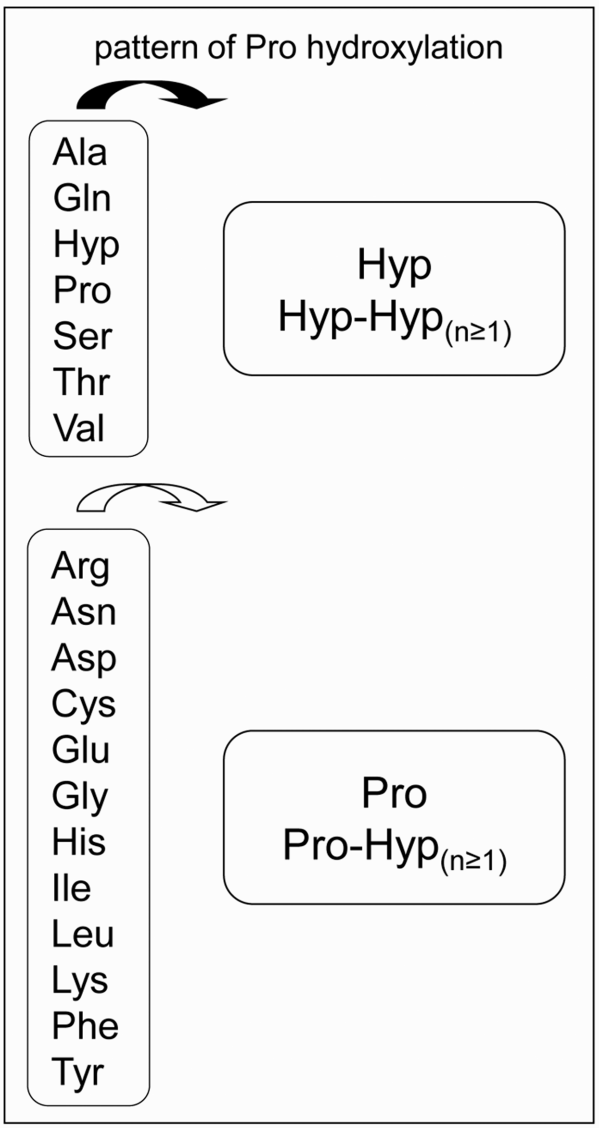


FIGURE 1 | The proposed Pro-hydroxylation code for plant cell wall proteins. On the left side, amino acids preceding Pro residues. On the right side, patterns of Pro hydroxylation (Canut et al., 2016).

proteins described below, there were lectins, Asp proteases, lipases acylhydrolases of the GDSL family, and class III peroxidases.

Five CWPs were selected on the basis of the following criteria: (i) their abundance in these aerial organs, as shown by the high number of sequenced peptides for each of them (from 89 to 533, depending on the protein); and (ii) a high sequence coverage (from 26 to 72% of the mature protein) (**Table 1**). The MS data corresponding to these five proteins are given in Supplementary Table S1. In addition, none of them has already been shown to contain Hyp residues. At4g38770 (AtPRP4) is a Pro-rich protein and its gene was shown to be expressed in aerial organs (Fowler et al., 1999). At1g09750 is a predicted Asp protease. At1g31580 (ECS1/CXc750) was assumed to be involved in resistance mechanisms (Aufsatz et al., 1998). At3g08030 (AthA2-1) has a predicted DUF642 domain (Vázquez-Lobo et al., 2012). Finally, At2g10940 is a protein showing homology to non-specific plant lipid transfer proteins. Three out of these CWPs have amino acid sequences rich in Pro residues as calculated from their mature sequence: AtPRP4 (32.5% Pro), At1g31580 (27.9%), and At2g10940 (27.3%). The two others, At1g09750 and At3g08030, are poor in Pro residues (7.3 and 6.4%, respectively). Contrarily to the three former proteins which exhibit Pro-rich motifs, the latter ones have dispersed Pro residues.

The extended Pro hydroxylation code was applied to predict the location of Hyp residues in the five amino acid sequences and to compare them to the observed ones. All details are given in Supplementary Figure S1, and simplified views of AtPRP4 and At2g10940 are shown in **Figures 2** and **3**. The sequences including the predicted Hyp residues are shown on the left and the observed ones are framed on the right of each figure. No obvious difference could be found between the three datasets regarding the frequency of occurrence of the different peptide variants according to the location of Pro and Hyp residues (Pro/Hyp peptide variants). In particular, no difference could be found between the rosette and the stem samples. The MS/MS data were manually checked as shown in Supplementary Figure S2 for two peptides of AtPRP4. All the Pro/Hyp locations were confirmed with the exception of two motifs in the amino acid sequence of At1g09750 (GPM and LPM). We could not discriminate between the hydroxylation of a Pro residue and the oxidation of a Met residue (Supplementary Table S2B and Figure S1). Thus, we did not retain the hypothesis of a Pro hydroxylation in these motifs in the following.

Several observations could be done. (i) A very high proportion of Pro residues was hydroxylated in the three Pro-rich proteins (69/92 in AtPRP4, 8/10 in At1g31580, 36/49 in At2g10940). (ii) Only a few Hyp residues could be found in proteins poor in Pro residues, and rarely at predicted positions (none in At1g09750; only at two out of nine predicted possibilities, and two at unexpected positions, but at a very low frequency, in At3g08030). (iii) For a given peptide, several variants could be observed. For example, two variants of GFDHPFPLPPPLELPPFLK and three variants of YSPPVEVPPPVPVYEPPPKK were found in AtPRP4: GFDHPF**P**LPOOLELPOFLK as predicted, and GFDHPF**O**LPOOLELPOFLK; YSOOVEVOOOVOVYE**POO**KK as predicted, YSOOVEVOOOVOVYE**POP**KK, and YSOOVE VOOOVOVYE**OOO**KK (Supplementary Figure S2). (iv) The observed discrepancies between the predicted and the observed Pro/Hyp locations could be either a Pro instead of a Hyp residue or vice-versa. (v) Some discrepancies could be systematically observed, as the F**O**OR motif in At1g31580 instead of the predicted F**P**OR motif.

This survey has allowed the fine mapping of Pro/Hyp residues in the five selected CWPs. The next issue was to know how efficiently the extended Pro hydroxylation code could predict their location.

TABLE 1 | Sequence coverage and number of peptides analyzed by LC-MS/MS for each protein.

	Size of the protein (number of amino acids)	Average percentage of sequence coverage	Total number of sequenced peptides
At4g38770 (AtPRP4)	419	51	131
At1g09750	428	26	450
At1g31580	68	60	89
At3g08030	353	72	533
At2g10940	264	60	250

The list of the sequenced peptides is provided in Supplementary Table S1. The percentage of sequence coverage is calculated for the mature protein sequence and takes the results of the three experiments into account.

FIGURE 2 | Hydroxylation of Pro residues in the amino acid sequence of At4g38770 (AtPRP4) encoding a Pro-rich protein. The amino acid sequence of AtPRP4 is written from left to right and from top to bottom. The predicted peptide signal is indicated in light blue. The Pro-rich domain is displayed in order to emphasize repetitive sequences and tryptic peptides (one per line). On the left side, predicted Pro (P) and Hyp (O) residues are in pink and green, respectively. On the right side, observed Pro and Hyp residues at unexpected positions are underlined. For each peptide, the number between brackets corresponds to its frequency of occurrence, expressed as a ratio between the number of observed peptides and the total number of sequenced peptides. The numbers inside stars allow the comparison between the predicted/observed peptides (on the left side) and the corresponding observed Pro/Hyp peptide variants (on the right side).

Efficiency of the Prediction of the Location of the Pro/Hyp Residues

For each protein sequence, the total number of Pro and Hyp positions was recorded and compared to the number of correct predictions (**Table 2** and Supplementary Table S2). The percentage of mis-predictions was found to range from 5.1 to 46.7%. The best prediction was obtained for AtPRP4 which is a HRGP, i.e., a canonical protein with regard to the proposed rule. Except one motif in peptide 3 (**Figure 2** and Supplementary Figure S1A), V**P**OOV instead of the predicted V**O**OOV, all the other predicted motifs were found at least once. Some variability was observed within 12 motifs located in seven peptides (numbered 1, 2, and 4–8 on **Figure 2**). The KPPPK motif was the most variable one, with the following variants: KPOOK as predicted (peptides 3–5 and 7), KP**P**OK (peptides 4, 5, and 7), K**OPP**K (peptides 5 and 7), and KP**PP**K (peptide 7). Other motifs including three Pro/Hyp residues were also variable,

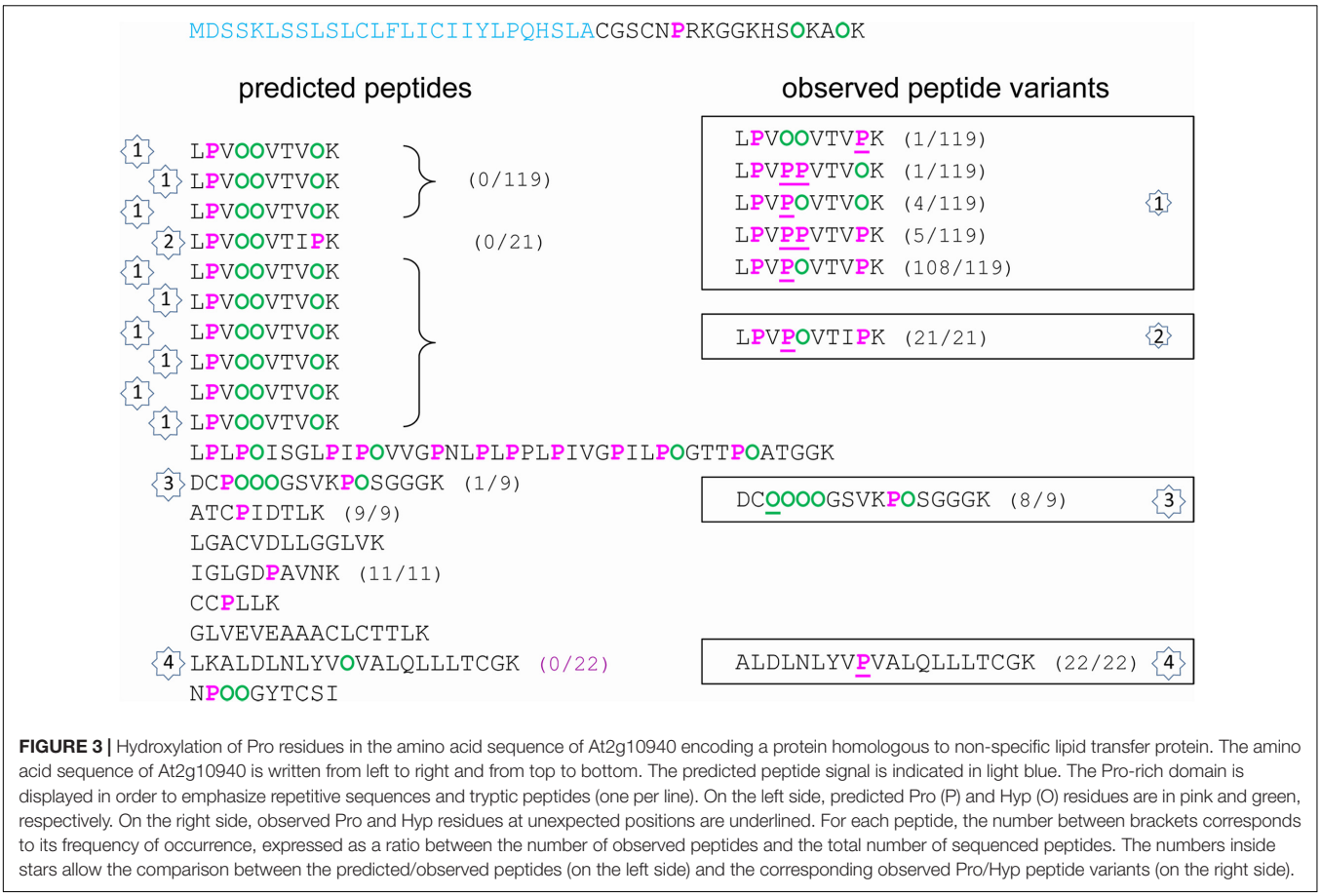


FIGURE 3 | Hydroxylation of Pro residues in the amino acid sequence of At2g10940 encoding a protein homologous to non-specific lipid transfer protein. The amino acid sequence of At2g10940 is written from left to right and from top to bottom. The predicted peptide signal is indicated in light blue. The Pro-rich domain is displayed in order to emphasize repetitive sequences and tryptic peptides (one per line). On the left side, predicted Pro (P) and Hyp (O) residues are in pink and green, respectively. On the right side, observed Pro and Hyp residues at unexpected positions are underlined. For each peptide, the number between brackets corresponds to its frequency of occurrence, expressed as a ratio between the number of observed peptides and the total number of sequenced peptides. The numbers inside stars allow the comparison between the predicted/observed peptides (on the left side) and the corresponding observed Pro/Hyp peptide variants (on the right side).

TABLE 2 | Efficiency of the prediction of Pro/Hyp location in CWP amino acid sequences according to the proposed rules.

Accession number	Number of analyzed Pro/Hyp positions	Number of unexpected location of Pro/Hyp residues	Percentage of mis-predictions
At4g38770	868	44	5.1
At1g09750	343	104	30.3
At1g31580	439	154	35.1
At3g08030	502	225	44.8
At2g10940	602	281	46.7

Five CWP sequences were analyzed by LC-MS/MS (Supplementary Table S1) and the percentage of mis-predictions was calculated (Supplementary Table S2).

such as EPPPK in peptide 2 (EPOOK as predicted, EPO**P**K and E**O**OOK), HPPPV in peptide 4 (HPOOV as predicted and H**OP**O), CPPPV in peptide 8 (CPOOV as predicted and CP**PP**V). The other cases of variability concerned shorter motifs such as FPL (F**O**L in peptide 1), VPV/I (V**P**V in peptide 5 and V**P**I in peptide 8), KPPT/V (K**O**PT in peptide 6, KP**P**V in peptide 8). However, all these Pro/Hyp peptide variants were not the prevailing forms of the motifs (see Supplementary Table S2).

The prediction of Pro and Hyp location was much less efficient for the four other proteins, irrespectively, to their Pro content. Regarding the two proteins with a low Pro content, the percentage of mis-prediction of Pro/Hyp location was very high (30.3% for At1g09750 and 44.8% for At3g08030). Hyp was only found in At3g08030, but in solely three motifs (VOF, G**O**H, and L**O**L) and at a low frequency (**Table 3**). Regarding the two proteins with a high percentage of Pro residues (At1g31580 and At2g10940), the situation was very different. Although they both exhibited a high percentage of Pro residues, the proposed rules did not allow reaching a high level of correct prediction of Pro/Hyp location. For At1g31580, peptide variants could mostly be observed for short motifs like RPI/R/T in peptides 1 and 3 (RPI/R/T as predicted and R**O**I/R/T as observed) and VPI/G in peptides 1 and 2 (VOI/G as predicted and V**P**I/G) (Supplementary Figure S2C). Two larger motifs were variable: FPPR in peptide 2 (FPOR as predicted and F**O**OR); LPPY in peptide 3 (LPOY as predicted

TABLE 3 | Exceptions to the proposed rules of Pro/Hyp location.

Accession number	Predicted location of Hyp residues	Observed motifs	Frequency of occurrence
At4g38770	KPOOK	KPOOK	19/29
		KP**P**OK	5/29
		K**OPP**K	4/29
		KP**PP**K	1/29
At1g09750	QOV	QOV	0/73
		Q**P**V	73/73
	VON	VON	0/21
		V**P**N	21/21
At1g31580	RPI	RPI	34/49
		R**O**I	15/49
	RPT	R**O**TRPT	23/3411/34
	RPR	RPR	11/34
		R**O**R	23/34
	VOG	VOG	33/55
		V**P**G	22/55
	VOI	VOI	52/55
		V**P**I	3/55
	FPOR	FPOR	55/55
		F**O**OR	55/55
	LPOY	LPOY	9/34
		LP**P**Y	23/34
At3g08030	VOS	VOS	0/21
		V**P**S	21/21
	VOF	VOF	3/63
		V**P**F	60/63
	VOH	VOH	0/20
		V**P**H	20/20
	LPL	LPL	18/21
		L**O**L	3/21
	SOL	SOL	0/48
		S**P**L	48/48
	SOG	SOG	0/49
		S**P**G	49/49
At2g10940	VOOV	VOOV	1/140
		V**P**OV	133/140
		V**PP**V	6/140
	VOK	VOK	5/119
		V**P**K	114/119
	VOV	VOV V**P**V	0/22 22/22

Only motifs observed at most 20 times are considered in this table.

and LP**P**Y). With the exception of the FPPR motif in which Pro residues were always both hydroxylated and the VPI motif which was found in the VOI form in all but three cases out of 55, all the other variants were found in the one third/two third proportion between predicted and mis-predicted variants or vice versa. For At2g10940, variability was observed in four motifs (**Figure 3** and Supplementary Figure S2E): VPPV in peptides 1 and 2 (VOOV as predicted and V**P**OV); VPK in peptides 1 and 2 (VOK as predicted and V**P**K); VPV in peptide 4 (VOV as predicted and V**P**V); and CPPPPG in peptide 3 (CPOOOG as predicted and C**O**OOOG). A very high proportion of the observed motifs did not follow the extended Pro hydroxylation code for their first Pro residue.

DISCUSSION

This work has allowed mapping the Pro/Hyp residues in five CWPs and comparing the prediction of Pro/Hyp location according to a previously proposed extended Pro hydroxylation code (Kieliszewski and Lamport, 1994; Canut et al., 2016).

Among these five proteins, only AtPRP4 was assumed to contain Hyp residues because it is known as a structural CWP with a high content of Pro residues and canonical amino acid motifs such as KKPCPP (7 occurrences) and PPV (14 occurrences) (Showalter et al., 2010). However, to our knowledge, its pattern of Pro hydroxylation has not yet been described. Regarding the other four proteins, our results have allowed enlarging the number of protein families possibly modified at the post-translational level by the hydroxylation of Pro residues. They have also shown that the Pro hydroxylation patterns can be variable at a given amino acid position.

The prediction of Hyp residue location could be done with a high level of confidence in one of the so-called HRGPs using the extended Pro hydroxylation code probably because this code was designed from the analysis of such protein sequences (Kieliszewski and Lamport, 1994; Canut et al., 2016). They include proteins rich in Pro, Ala, Ser and Thr residues such as (i) extensins with repetitive $S(P)_{n \geq 2}$ and YXY motifs, (ii) arabinogalactan proteins (AGPs) with AP/PA/SP/TP repeats and Pro-rich proteins (PRPs) with PPVX[KT], KKPCPP and PPV motifs (Showalter et al., 2010) or chimeric proteins containing a Pro-rich domain with XP_nY motifs (Hijazi et al., 2012; Canut et al., 2016). However, although prediction of the location of Pro/Hyp residues in the AtPRP4 sequence was very efficient (94.9% of successful predictions), some variability could be observed at a low frequency, particularly in the KPPPK motif, with 19 canonical Pro hydroxylation patterns (KPOOK) out of the 29 observed patterns and 10 variants (KPPOK, KOPPK, and KPPPK).

For the other proteins rich in Pro residues, many exceptions to the extended Pro hydroxylation code could be observed. In particular, in At1g31580, only about one third of the three amino acid-motifs had Hyp at the predicted location in RPI/T/R and VPG/I motifs and the FPPR motifs was systematically found with two Hyp residues (55/55). Besides, only one predicted VOOV motif could be recorded in At2g10940 out of the 140 observed peptides. The major variant was VPOV (133/140). A similar situation was found for the predicted VOK and VOV motifs (5/119 and 0/22 observations, respectively). For proteins having a low content in Pro residues, the prediction of Pro/Hyp location was also inefficient. Only a few Hyp residues could be observed and at a low frequency. Unexpected Hyp residues were also found in a previous study focused on class III peroxidases (Nguyen-Kim et al., 2016). For these proteins, a few Hyp residues were observed in CPN/Q/R, DPA, GPS/N, HPD, IPD, and LPA/Q/S motifs whereas some Pro residues were observed in APF/A, VPT, SPT/D, and TPG/L motifs. These results suggest that the extended Pro hydroxylation code cannot be used for such proteins. They also show that the rule established for the hydroxylation of the Pro residue within the EPA motif of sporamin cannot be applied (Shimizu et al., 2005). Based on mutagenesis of the surrounding amino acids, it was shown that the hydroxylation of the Pro residue required the following environment in tobacco BY-2 cells: [AVSTG]-P-[AVSTGA]-[GAVPSTC]-[APSDE].

Our results raise the question of the specificity of prolyl-4 hydroxylases (P4Hs) which have to recognize some features on the target protein at the level of its primary amino acid sequence or the secondary/tertiary structure. The specificity of three out of the 13 P4Hs of *A. thaliana* was characterized. P4H1 was shown to preferentially hydroxylate the second Pro residue in PPG motifs (Hieta and Myllyharju, 2002). All the peptides hydroxylated by P4H2 have at least three consecutive Pro residues and the third of them is preferentially hydroxylated (Tiainen et al., 2005). Finally, P4H5 was shown to hydroxylate Pro residues in SP_4 motifs in a sequential way, but never on the fourth Pro residue (Velasquez et al., 2015). Besides, P4H2 and P4H13 were assumed to complement the Pro hydroxylation pattern of SP_4 motifs in extensins (Velasquez et al., 2015). The characterization of additional P4Hs will give clues to understand this process which is probably tightly regulated because of its importance for biological activity. Indeed, this PTM is the first step prior to *O*-glycosylation: (i) poly-arabinosylation in extensins; or (ii) complex *O*-glycans like type II arabinogalactans (AGs), type III AGs or peanut agglutinin (PNA) AGs in AGPs, allergens or AtAGP31, respectively (Hijazi et al., 2014). None of the five CWPs analyzed in this work is known to be *O*-glycosylated. However, a previous proteomic study based on affinity chromatography with PNA, a lectin specific for galactose residues, has allowed identifying a protein of the same family as At3g08030 (Zhang et al., 2011). The next step will consist in correlating the presence of Hyp residues to *O*-glycosylation. Finally, the presence of Pro/Hyp peptide variants raises the question of the role of Hyp residues as previously discussed for class III peroxidases (Nguyen-Kim et al., 2016). This variability could be incidental or contribute to the regulation of the biological activity of CWPs.

Altogether, Pro hydroxylation events are probably more abundant in CWPs than initially thought, but the precise rules of this PTM need additional experiments to be fully described. Some clues can be proposed from our results. For example, a Hyp residue is found in the VPX motifs of AtPRP4 (77 Pro hydroxylations among the 82 observed VPX motifs) and At1g31580 (85 out of 110), whereas it is mainly a Pro residue in the three other proteins (only 15 Pro hydroxylations out of the 406 observed VPX motifs). The Pro residues in the 97 observed SPX motifs of At3g08030 were never hydroxylated whereas only a few lack of Pro hydroxylation have been described in cell wall Pro-rich proteins (Canut et al., 2016). A systematic mining of MS data is now required to permit the identification of other CWPs or secreted peptides containing Hyp residues and to map them. This task is challenging because, as mentioned above, 17.5% of the CWPs identified in one of our rosette experiments had at least one peptide carrying a Hyp residue. However, the amount of MS data corresponding to all these proteins was not sufficient to perform a relevant statistical analysis and to propose yet a further expanded Pro hydroxylation code. Finally, such a code should probably take into account tissue-specific patterns as for *O*-glycosylation (Estevez et al., 2006) and protein families. This work paves the way for a better description of Pro hydroxylation patterns in CWPs.

MATERIALS AND METHODS

Extraction of Proteins from Cell Walls

Arabidopsis thaliana plants were cultivated in growth chambers at 22°C with a photoperiod of 16 h light/8 h dark. Rosettes and mature stems were collected after 4 and 6 weeks, respectively. The detailed description of the experiments is given in our previous articles (Hervé et al., 2016; Duruflé et al., 2017). Three biological replicates were performed for each experiment. Briefly, cell walls were purified as described (Feiz et al., 2006). Proteins were extracted from lyophilized cell walls in four steps using a 5 mM acetate buffer pH 4.6 complemented with 0.2 M $CaCl_2$ (two successive extractions) or 2 M LiCl (*ditto*) (Irshad et al., 2008). The four protein extracts were combined prior to further analysis.

Analysis of Proteins by LC/MS-MS and Bioinformatics

The same amount of each protein extract (40 μg) was analyzed by LC-MS/MS. In the case of rosettes, two types of analysis were performed: (i) the first one after separation of proteins by a short 1D-electrophoresis in three fractions prior to *in gelo* tryptic digestion, (ii) the second one by shotgun analysis of the extracted proteins after tryptic digestion (Hervé et al., 2016). In the case of stems, only the second method was used (Duruflé et al., 2017). LC-MS/MS analyses were performed with a Q-exactive instrument (Thermo Fisher Scientific, Villebon-sur-Yvette, France) as described (Feiz et al., 2006; Hervé et al., 2016). All the MS/MS data were made publicly available in the PROTICdb[1] and *WallProtDB* databases[2]. The following modifications were taken into account for peptide identification: Met oxidation, Pro hydroxylation, N-ter acetylation, N-ter deamidation of Glu, N-ter deamidation of Cys and loss of H_2O on N-ter Glu. The lists of peptides allowing the identification of the five CWPs studied in detail in this article are given in Supplementary Table S1. The search for *N*-glycosylation motifs was performed with PROSITE[3].

[1] http://proteus.moulon.inra.fr/w2dpage/proticdb/angular/

[2] http://polebio.lrsv.ups-tlse.fr/WallProtDB/

[3] http://prosite.expasy.org/

AUTHOR CONTRIBUTIONS

HD and VH performed the protein extractions from purified cell walls and contributed to the analyses of results. TB and MZ did the MS/MS analyses. CD and EJ initiated the research, designed the study and discussed the results. EJ coordinated the analysis of the results and the writing of the manuscript. All authors read and approved the final manuscript.

FUNDING

The authors are thankful to Université Paul Sabatier (Toulouse, France) and CNRS for supporting their research work. HD is granted by the Toulouse University and the Occitanie region. This work was also supported by the French Laboratory of Excellence project entitled "TULIP" (ANR-10-LABX-41; ANR-11-IDEX-0002-02).

ACKNOWLEDGMENTS

LC-MS/MS analyses were performed at the PAPPSO proteomics facility (pappso.inra.fr). The authors wish to thank Dr. Hervé Canut for stimulating discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2017.01802/full#supplementary-material

REFERENCES

Albenne, C., Canut, H., and Jamet, E. (2013). Plant cell wall proteomics: the leadership of *Arabidopsis thaliana*. *Front. Plant Sci.* 4:111. doi: 10.3389/fpls.2013.00111

Aufsatz, W., Amry, D., and Grimm, C. (1998). The ECS1 gene of Arabidopsis encodes a plant cell wall-associated protein and is potentially linked to a locus influencing resistance to *Xanthomonas campestris*. *Plant Mol. Biol.* 38, 965–976. doi: 10.1023/A:1006028605413

Baer, B., and Millar, A. (2016). Proteomics in evolutionary ecology. *J. Proteomics* 135, 4–11. doi: 10.1016/j.jprot.2015.09.031

Bause, E. (1983). Structural requirements of *N*-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* 209, 331–336. doi: 10.1042/bj2090331

Bause, E., and Legler, G. (1981). The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the *N*-glycosylation step during glycoprotein biosynthesis. *Biochem. J.* 195, 639–644. doi: 10.1042/bj1950639

Borassi, C., Sede, A., Mecchia, M., Salgado Salter, J., Marzol, E., Muschietti, J., et al. (2016). An update on cell surface proteins containing extensin-motifs. *J. Exp. Bot.* 67, 477–487. doi: 10.1093/jxb/erv455

Borner, G. H., Lilley, K. S., Stevens, T. J., and Dupree, P. (2003). Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. *Plant Physiol.* 132, 568–577. doi: 10.1104/pp.103.021170

Canut, H., Albenne, C., and Jamet, E. (2016). Post-translational modifications of plant cell wall proteins and peptides: a survey from a proteomics point of view. *Biochim. Biophys. Acta* 1864, 983–990. doi: 10.1016/j.bbapap.2016.02.022

Carpita, N., and Gibeaut, D. (1993). Structural models of primary cell walls in flowering plants, consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* 3, 1–30. doi: 10.1111/j.1365-313X.1993.tb00007.x

Cosgrove, D. (2015). Plant cell wall extensibility: connecting plant cell growth with cell wall structure, mechanics, and the action of wall-modifying enzymes. *J. Exp. Bot.* 67, 463–476. doi: 10.1093/jxb/erv51

Duruflé, H., San Clemente, H., Balliau, T., Zivy, M., Dunand, C., and Jamet, E. (2017). Cell wall proteome analysis of *Arabidopsis thaliana* mature stems. *Proteomics* doi: 10.1002/pmic.201600449 [Epub ahead of print].

Eisenhaber, B., Wildpaner, M., Schultz, C., Borner, G., Dupree, P., and Eisenhaber, F. (2003). Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for

Arabidopsis and rice. *Plant Physiol.* 133, 1691–1701. doi: 10.1104/pp.103.023580

Elortza, F., Nühse, T., Foster, L., Stensballe, A., Peck, S., and Jensen, O. (2003). Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins. *Mol. Cell. Proteomics* 2, 1261–1270. doi: 10.1074/mcp.M300079-MCP200

Elortza, F., Shabaz, M., Bunkenborg, J., Foster, L., Nühse, T., Brodbeck, U., et al. (2005). Modification-specific proteomics of plasma membrane proteins: identification and characterization of glycosylphosphatidylinositol-anchored proteins released upon phospholipase D treatment. *J. Proteome Res.* 5, 935–943. doi: 10.1021/pr050419u

Estevez, J. M., Kieliszewski, M. J., Khitrov, N., and Somerville, C. (2006). Characterization of synthetic hydroxyproline-rich proteoglycans with arabinogalactan protein and extensin motifs in Arabidopsis. *Plant Physiol.* 142, 458–470. doi: 10.1104/pp.106.084244

Faye, L., Boulaflous, A., Benchabane, M., Gomord, V., and Michaud, D. (2005). Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. *Vaccine* 23, 1770–1778. doi: 10.1016/j.vaccine.2004.11.003

Feiz, L., Irshad, M., Pont-Lezica, R. F., Canut, H., and Jamet, E. (2006). Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from Arabidopsis hypocotyls. *Plant Methods* 2:10. doi: 10.1186/1746-4811-2-10

Fowler, T. J., Bernhardt, C., and Tierney, M. L. (1999). Characterization and expression of four proline-rich cell wall protein genes in Arabidopsis encoding two distinct subsets of multiple domain proteins. *Plant Physiol.* 121, 1081–1092. doi: 10.1104/pp.121.4.1081

Franková, L., and Fry, J. (2013). Biochemistry and physiological roles of enzymes that 'cut and paste' plant cell-wall polysaccharides. *J. Exp. Bot.* 64, 3519–3550. doi: 10.1093/jxb/ert201

Hervé, V., Duruflé, H., San Clemente, H., Albenne, C., Balliau, T., Zivy, M., et al. (2016). An enlarged cell wall proteome of *Arabidopsis thaliana* rosettes. *Proteomics* 16, 3183–3187. doi: 10.1002/pmic.201600290

Hieta, R., and Myllyharju, J. (2002). Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor alpha-like peptides. *J. Biol. Chem.* 277, 23965–23971. doi: 10.1074/jbc.M201865200

Hijazi, M., Durand, J., Pichereaux, C., Pont, F., Jamet, E., and Albenne, C. (2012). Characterization of the arabinogalactan protein 31 (AGP31) of *Arabidopsis thaliana*: new advances on the Hyp-*O*-glycosylation of the Pro-rich domain. *J. Biol. Chem.* 287, 9623–9632. doi: 10.1074/jbc.M111.247874

Hijazi, M., Velasquez, S., Jamet, E., Estevez, J., and Albenne, C. (2014). An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. *Front. Plant Sci.* 5:395. doi: 10.3389/fpls.2014.00395

Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R., and Jamet, E. (2008). A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: confirmed actors and newcomers. *BMC Plant Biol.* 8:94. doi: 10.1186/1471-2229-8-94

Kieliszewski, M. J., and Lamport, D. T. A. (1994). Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J.* 5, 157–172. doi: 10.1046/j.1365-313X.1994.05020157.x

Kim, S.-J., and Brandizzi, F. (2016). The plant secretory pathway for the trafficking of cell wall polysaccharides and glycoproteins. *Glycobiology* 26, 940–949. doi: 10.1093/glycob/cww044

Komatsu, S., and Yanagawa, Y. (2013). Cell wall proteomics of crops. *Front. Plant Sci.* 4:17. doi: 10.3389/fpls.2013.00017

Lamport, D. T., and Várnai, P. (2013). Periplasmic arabinogalactan glycoproteins act as a calcium capacitor that regulates plant growth and development. *New Phytol.* 197, 58–64. doi: 10.1093/aob/mcu161

Lannoo, N., and Van Damme, E. (2015). Review/N-glycans: the making of a varied toolbox. *Plant Sci.* 239, 67–83. doi: 10.1016/j.plantsci.2015.06.023

Lee, S. J., Saravanan, R. S., Damasceno, C. M., Yamane, H., Kim, B. D., and Rose, J. K. (2004). Digging deeper into the plant cell wall proteome. *Plant Physiol. Biochem.* 42, 979–988. doi: 10.1016/j.plaphy.2004.10.014

Lige, B., Shengwu, M., and Van Huystee, R. (2001). The effects of the site-directed removal of *N*-glycosylation from cationic peanut peroxidase on its function. *Arch. Biochem. Biophys.* 386, 17–24. doi: 10.1006/abbi.2000.2187

Matsubayashi, Y., and Sakagami, Y. (2006). Peptide hormones in plants. *Annu. Rev. Plant Biol.* 57, 649–674. doi: 10.1146/annurev.arplant.56.032604.144204

Nakagami, H., Sugiyama, N., Ishihama, Y., and Shirasu, K. (2012). Shotguns in the front line: phosphoproteomics in plants. *Plant Cell Physiol.* 53, 118–124. doi: 10.1093/pcp/pcr148

Nguyen-Kim, H., San Clemente, H., Balliau, T., Zivy, M., Dunand, C., Albenne, C., et al. (2016). *Arabidopsis thaliana* root cell wall proteomics: Increasing the proteome coverage using a combinatorial peptide ligand library and description of unexpected Hyp in peroxidase amino acid sequences. *Proteomics* 16, 491–503. doi: 10.1002/pmic.201500129

Rodríguez-Celma, J., Ceballos-Laita, L., Grusak, M., Abadía, J., and López-Millán, A. (2016). Plant fluid proteomics: delving into the xylem sap, phloem sap and apoplastic fluid proteomes. *Biochim. Biophys. Acta* 1864, 991–1002. doi: 10.1016/j.bbapap.2016.03.014

Ruiz-May, E., Hucko, S., Howe, K., Zhang, S., Sherwood, R., Thannhauser, T., et al. (2014). A comparative study of lectin affinity based plant N-glycoproteome profiling using tomato fruit as a model. *Mol. Cell. Proteomics* 13, 566–579. doi: 10.1074/mcp.M113.028969

Ruiz-May, E., Thannhauser, T. W., Zhang, S., and Rose, J. (2012). Analytical technologies for identification and characterization of the plant N-glycoproteome. *Front. Plant Sci.* 3:150. doi: 10.3389/fpls.2012.00150

Shimizu, M., Igasaki, T., Yamada, M., Yuasa, K., Hasegawa, J., Kato, T., et al. (2005). Experimental determination of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. *Plant J.* 42, 877–889. doi: 10.1111/j.1365-313X.2005.02419.x

Showalter, A. M., Keppler, B., Lichtenberg, J., Gu, D., and Welch, L. R. (2010). A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol.* 153, 485–513. doi: 10.1104/pp.110.156554

Shpak, E., Barbar, E., Leykam, J. F., and Kieliszewski, M. J. (2001). Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. *J. Biol. Chem.* 276, 11272–11278. doi: 10.1074/jbc.M011323200

Stafstrom, J. P., and Staehelin, L. A. (1986). The role of carbohydrate in maintaining extensin in an extended conformation. *Plant Physiol.* 81, 242–246. doi: 10.1104/pp.81.1.242

Strasser, R. (2016). Plant protein glycosylation. *Glycobiology* 26, 926–939. doi: 10.1093/glycob/cww023

Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., and Cammue, B. (2015). The plant peptidome: an expanding repertoire of structural features and biological functions. *Plant Cell* 27, 2095–2118. doi: 10.1105/tpc.15.00440

Tenhaken, R. (2015). Cell wall remodeling under biotic stress. *Front. Plant Sci.* 5:771. doi: 10.3389/fpls.2014.00771

Tiainen, P., Myllyharju, J., and Koivunen, P. (2005). Characterization of a second *Arabidopsis thaliana* prolyl 4-hydroxylase with distinct substrate specificity. *J. Biol. Chem.* 280, 1142–1148. doi: 10.1074/jbc.M411109200

Vázquez-Lobo, A., Roujol, D., Zuñiga-Sánchez, E., Albenne, C., Piñero, D., Gamboa de Buen, A., et al. (2012). The highly conserved spermatophyte cell wall DUF642 protein family: phylogeny and first evidence of interaction with cell wall polysaccharides *in vitro*. *Mol. Phylogenet. Evol.* 63, 510–520. doi: 10.1016/j.ympev.2012.02.001

Velasquez, S., Ricardi, M., Poulsen, C., Oikawa, A., Dilokpimol, A., Halim, A., et al. (2015). Complex regulation of prolyl-4-hydroxylases impacts root hair expansion. *Mol. Plant* 8, 734–746. doi: 10.1016/j.molp.2014.11.017

Velasquez, S. M., Ricardi, M. M., Dorosz, J. G., Fernandez, P. V., Nadra, A. D., Pol-Fachin, L., et al. (2011). O-glycosylated cell wall proteins are essential in root hair growth. *Science* 332, 1401–1403. doi: 10.1126/science.1206657

Ytterberg, A., and Jensen, O. (2010). Modification-specific proteomics in plant biology. *J. Proteomics* 73, 2249–2266. doi: 10.1016/j.jprot.2010.06.002

Zhang, Y., Giboulot, A., Zivy, M., Valot, B., Jamet, E., and Albenne, C. (2011). Combining various strategies to increase the coverage of the plant cell wall glycoproteome. *Phytochemistry* 72, 1109–1123. doi: 10.1016/j.phytochem.2010.10.019

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR : Harold DURUFLÉ
TITLE : Study of the cell wall plasticity in various Pyrenean altitudinal *Arabidopsis thaliana* ecotypes.

PhD SUPERVISORS : Pr. Christophe DUNAND and Pr. Philippe BESSE

ABSTRACT

Global warming is a current issue of great concern because of its potential effects on biodiversity and the agricultural sector. Better understanding the adaptation of plants to this recent phenomenon is therefore a major interest for science and society.

The study of natural populations from an altitude gradient allows correlating a set of climatic conditions (temperature, humidity, radiation, etc...) with phenotypic traits. These different populations are considered as adapted to their climatic conditions *in natura*. By cultivating these plants under standardized laboratory conditions (light intensity, substrate, temperature, watering, etc.), the observed phenotypic variability, is essentially due to the genetic variability intrinsic to each genotype. The growth of these same plants by changing a single variable, for example temperature, makes possible to highlight a characteristic phenotype. This phenotype may be an acclimation response of a relevant genome. The WallOmics project aims at characterizing the adaptation of plants to altitude by studying natural populations of *Arabidopsis thaliana* from the Pyrenees.

The molecular actors of the adaptation of plants are still poorly described, but it appears that the plant cell wall could play an important role in this process. Indeed, it represents the skeleton of plants and gives them rigidity while representing a dynamic and sensitive external barrier to environmental changes. Its structure and composition can be modified at any time. It is also possible to say that the plant cell wall gives the general shape of the plant (size, shape, density, etc.), that is its observable phenotype. This project will focus mainly on the study of the plant cell wall.

New technologies have enabled the emergence of the so-called "omics" data, large sets of data at multiple biological levels, such as ecological, phenotypic, metabolomic, proteomic, transcriptomic and genomic data. The study and the links between these data have favoured the development of integrative approaches aimed at establishing a response at several scales. It is precisely by this type of non-mechanistic approach that the WallOmics project has contributed to establish the molecular players of plant cell wall modifications in the global warming context.

KEYWORDS: *Arabidopsis thaliana*, cell wall, integrative study, natural variation, temperature acclimation

TOPIC: Plant development

NAME AND ADDRESS OF THE LABORATORY: Plant Science Research Laboratory (LRSV), UMR 5546 UPS/CNRS, 24 chemin de Borde Rouge - Auzeville, BP 42617, 31326 Castanet-Tolosan Cedex, France

AUTEUR : Harold DURUFLÉ

TITRE : Production et traitement de données omiques hétérogènes en vue de l'étude de la plasticité de la paroi chez des écotypes de la plante modèle *Arabidopsis thaliana* provenant d'altitudes contrastées.

DIRECTEURS DE THESE : Pr. Christophe DUNAND et Pr. Philippe BESSE

LIEU ET DATE DE SOUTENANCE : Auzeville, le 20 Octobre 2017

RESUME

Le réchauffement climatique constitue une problématique d'actualité très préoccupante en raison de ses effets potentiels sur la biodiversité et le secteur agricole. Mieux comprendre l'adaptation des plantes face à ce phénomène récent représente donc un intérêt majeur pour la science et la société.

L'étude de populations naturelles provenant d'un gradient d'altitude permet de corréler l'impact d'un ensemble de conditions climatiques (température, humidité, radiation, etc.) à des traits phénotypiques. Ces différentes populations sont dites adaptées à leurs conditions climatiques *in natura*. En cultivant ces plantes dans des conditions standardisées de laboratoire (intensité lumineuse, substrat, température, arrosage, etc...), la variabilité phénotypique observée, est alors due essentiellement à la variabilité génétique intrinsèque à chaque plante, donc à son génotype. La mise en culture de ces mêmes plantes en changeant une seule variable, par exemple la température, permet de mettre en évidence un phénotype caractéristique. Ce phénotype observé peut être une réponse d'acclimatation d'un génome adapté. Le projet WallOmics vise à caractériser l'adaptation des plantes à l'altitude par l'étude de populations naturelles d'*Arabidopsis thaliana* provenant des Pyrénées.

Les acteurs moléculaires de l'adaptation des plantes au climat sont encore mal connus mais il apparaît que la paroi des cellules végétales pourrait avoir un rôle important dans ce processus. En effet, celle-ci représente le squelette des plantes et leur confère une rigidité tout en représentant une barrière externe sensible et dynamique face aux changements environnementaux. Sa structure et sa composition peuvent être modifiées à tout moment. Il est d'ailleurs possible de dire que cette paroi végétale donne la forme générale de la plante (taille, forme, densité, etc…), son phénotype observable. Ce projet se consacrera principalement à l'étude des parois des cellules végétales.

Les nouvelles technologies ont permis l'émergence des données dites «omiques», c'est-à-dire de vastes ensembles de données provenant de niveaux biologiques multiples, comme des données écologiques, de phénotypages, biochimiques, protéomiques, transcriptomiques et génomiques. L'étude et la mise en relation de ces données ont favorisé le développement d'approches globales qui visent à établir une réponse à plusieurs échelles. C'est justement par ce type d'approche non mécanistique que le projet WallOmics a contribué à établir les bases moléculaires des modifications des parois face aux changements climatiques.

MOTS CLEFS : *Arabidopsis thaliana,* paroi cellulaire, analyse intégrative, variation naturelle, acclimatation à la température

DISCIPLINE : Développement des plantes

INTITULE ET ADRESSE DU LABORATOIRE : Laboratoire de Recherche en Sciences Végétales (LRSV), UMR 5546 UPS/CNRS, 24 chemin de Borde Rouge - Auzeville, BP 42617, 31326 Castanet-Tolosan Cedex, France