

# 2023 年臺灣國際科學展覽會 優勝作品專輯

作品編號	060001
參展科別	植物學
作品名稱	由阿拉伯芥自然族群之環境分佈差異探究新穎 抗旱基因
得獎獎項	一等獎 美國 ISEF 正選代表

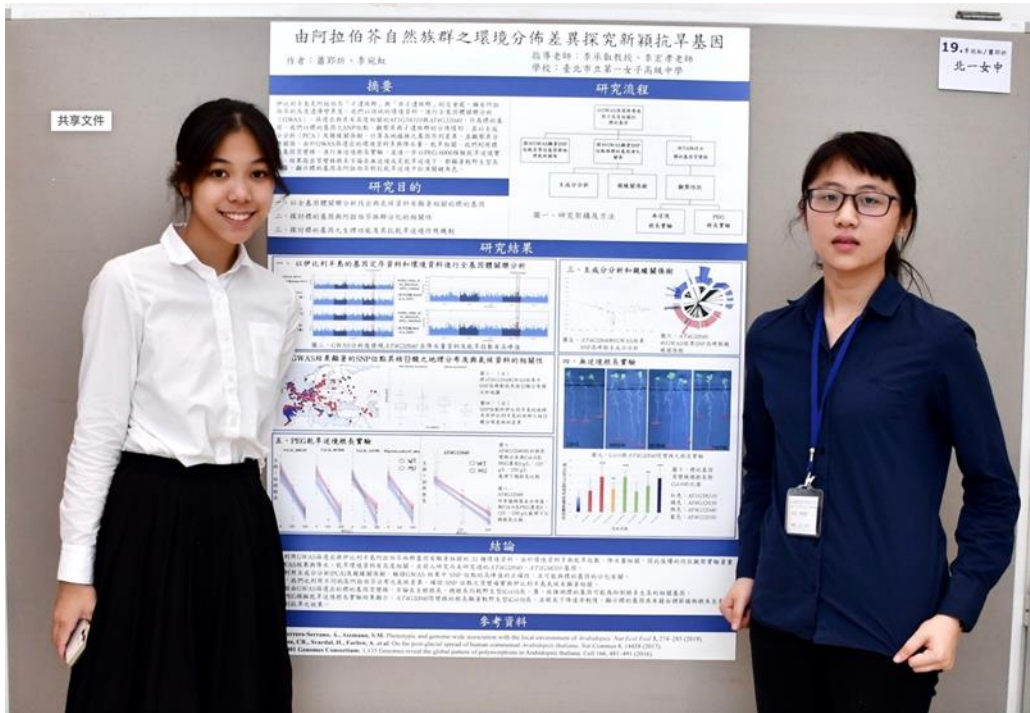
就讀學校 臺北市立第一女子高級中學

指導教師 李宏孝、李承叡

作者姓名 蕭郢炘、李宛虹

關鍵詞 阿拉伯芥、全基因體關聯分析、乾旱逆境

# 作者簡介



大家好，我們是北一女中科學班的蕭郢焢和李宛虹。因應近年來氣候變遷加劇，我們亟欲尋求糧食危機的解方，並下定決心踏上植物學研究之路。感謝李宏孝老師、李承叡教授、羅澄玉學姊一路上的辛勤指導，讓我們能有機會站上國際科展的舞台，也讓我們專研之路擁有許多美好回憶。

## 摘要

目前分布在歐洲地中海沿岸的阿拉伯芥子遺族群曾經廣泛分布於後冰期歐亞大陸，而後被非子遺族群取代成為目前全世界分布的優勢族群。伊比利半島是阿拉伯芥子遺與非子遺族群的交會處，擁有高度遺傳多樣性。我們以該地的環境資料進行全基因體關聯分析，篩選出與降水量、乾旱有高度相關的 *AT1G58310* 與 *AT4G32040* 作為標的基因。我們觀察標的基因之 SNP 位點與子遺族群的分佈情形；並以主成分分析及親緣關係樹，計算各地植株之基因序列差異及其分群關係。接著以標的基因突變株進行無逆境與 PEG-6000 模擬乾旱逆境之根長實驗，結果指出突變株根系都顯著較野生型長、廣，顯示 *AT4G32040* 具有抗乾旱功能。本研究由演化生態學的角度切入，結合遺傳基因體學進行分析，以分子生物學實驗驗證標的基因與植物抗旱能力之關聯。期望透過此研究模式推廣至其他農作物，以應對水資源短缺和糧食生產需求的增長。

## Abstract

*Arabidopsis thaliana* contains genetic groups originating from different ice age refugia, with one particular group comprising over 95% of the current worldwide population. In Europe, relicts of other groups can be found along the Mediterranean Sea. These "relicts" occupied post-glacial Eurasia first and were later replaced by the invading "non-relicts". The Iberian Peninsula is the intersection of the "relict" and "non-relict" populations of *Arabidopsis thaliana* and has a high degree of genetic variability. We performed genome-wide association study (GWAS) based on the environmental data and identified the candidate genes *AT1G58310* and *AT4G32040*, which are highly correlated with precipitation and drought. We investigated the spatial association of the candidate alleles and relict population. Principal component analysis (PCA) and phylogenetic tree of the target genes were used to investigate the among-individual relationship of these candidate genes. Next, we performed root length experiments with target gene mutants and further simulated drought stress with PEG-6000. The results indicated that the root system of the mutant plants was significantly longer and wider than that of the wild type under drought stress or without stress, indicating that the target gene *AT4G32040* plays a key role in *Arabidopsis* resistance to drought stress. This study starts from the perspective of evolutionary ecology, combines genetics and genomics, and uses molecular biology experiments to verify the relationship between target genes and plant drought resistance. Our approach might contribute to the improvement of drought tolerance and water use efficiency of crops to cope with the water shortage and increasing demand for food production in the foreseeable future.

# 壹、前言

## 一、阿拉伯芥在伊比利半島的特殊性及演化的重要性

阿拉伯芥 (*Arabidopsis thaliana*) 是開花植物中被研究最透徹的模式物種，分類上隸屬十字花科(Brassicaceae)，有五對染色體，是最早被完全基因定序的植物。阿拉伯芥之原生地廣佈於歐亞大陸及非洲，擁有高度氣候變異度，為探討植物與環境適應關係之重要的研究對象。

根據前人研究，位於南歐的伊比利半島為阿拉伯芥「孑遺族群」與「非孑遺族群」的交會處，使伊比利半島具有阿拉伯芥遺傳多樣性高的特性(1001 Genomes Consortium, 2016；Lee *et al.* 2017)。伊比利半島地區有較多原生於南歐的孑遺族群，同時也有自法國來的非孑遺族群和此兩種雜交的族群。相對地，西歐法國一帶則多為非孑遺族群。



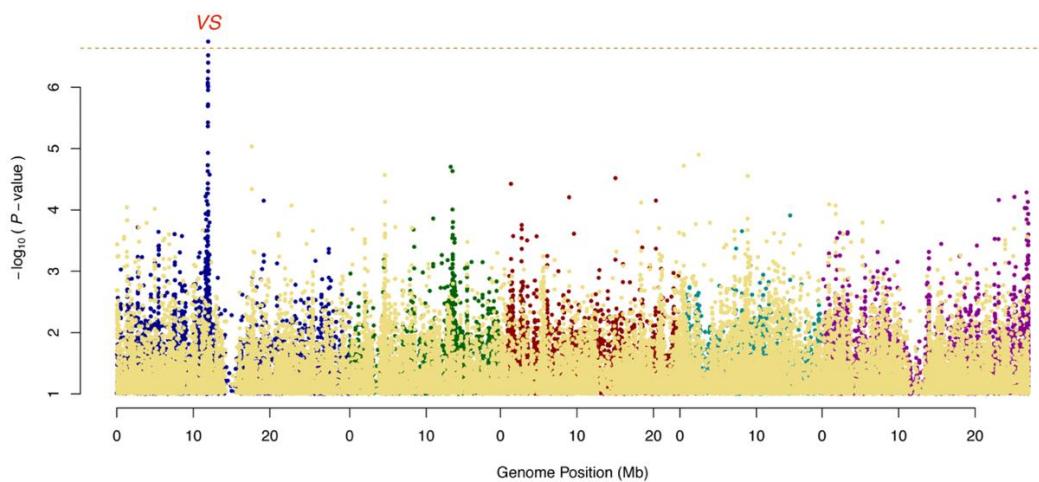
圖一、阿拉伯芥孑遺族群(relict)與非孑遺族群(non-relict)分佈位置

(資料來源：1001 Genomes Consortium, 2016)

另外，伊比利半島氣候環境歧異度高(特別在降雨量方面)：中部為溫帶大陸性氣候，夏季多雨、冬季少雨；南部為地中海氣候，冬季多雨、夏季少雨；北部與西部為溫帶海洋性氣候，全年溫和多雨。有鑒於前人研究多著重在整個物種的分布範圍，本研究決定聚焦於遺傳與氣候均具有高歧異度的伊比利半島。故我們利用南歐伊比利半島的環境資料與不同阿拉伯芥族群基因序列進行全基因體關聯分析，希望能找出與當地環境有高度相關的關鍵基因。

## 二、全基因體關聯分析

全基因體關聯分析(Genome-wide association study, 以下簡稱 GWAS)是遺傳學研究中用於將特定遺傳變異與特定表徵相關聯的方法。從遍布基因體的分子標記(例如: SNPs、RFLPs、STR)中篩選出與待測性狀高度相關者,經過 GWAS 分析後可以得到如圖二結果,圖中橫軸代表的是基因體位置,圖中的每個點都是 SNP 位點,縱軸是每個基因位點與該性狀相關程度之  $p$ -value 取 $-\log$  後之結果,若該基因位點與該性狀的相關程度越顯著,則此數值越高。經過 GWAS 分析後,若有出現明顯、單一的高峰,即代表該高峰下方含有影響性狀之基因多樣性。



圖二、全基因體關聯分析結果示意圖(Manhattan plot) (Xia Shen, 2012)

全基因體關聯分析常用於尋找與「性狀」(如: 開花時間、身高、體重、血糖)有顯著相關的基因。在本研究中,我們以伊比利半島為採集地,並以當地的環境資料數值作為「性狀」,藉以找出和環境適應有顯著相關的基因。

## 貳、研究目的

- 一、以 GWAS 找出與適應氣候有顯著相關的標的基因。
- 二、探討 GWAS 結果顯著 SNP 位點核苷酸地理分布與氣候資料的相關性。
- 三、探討標的基因與阿拉伯芥族群分化之關聯性。
- 四、探討標的基因與阿拉伯芥抗旱能力之關聯性。

## 參、研究設備及器材

### 一、研究器材

試管振盪器(Vortex-Genie 2)、離心機(小)、離心機(大)、滅菌器、磁石攪拌器、pH meter、4°C 冰箱、-20°C 冰箱、植物生長室、核酸增殖器、電泳儀(槽)、電泳掃描儀、生長箱、無菌操作臺、乾浴槽、恆溫槽

### 二、研究材料

本實驗使用阿拉伯芥(*Arabidopsis thaliana*)作為研究對象，所使用的植株有以下幾種：

(一)野生型 (wild type, WT)：Columbia-0 (Col-0)

(二)突變株 (由 ARABIDOPSIS BIOLOGICAL RESOURCE CENTER 訂購)

表一、突變株與其被破壞之基因對照表

Gene	Mutant line
<i>AT1G58310</i>	SALK_094967、SALK_121841、SALK_086130
<i>AT4G32030</i>	SALK_048633、SALK_095164、SALK_007590
<i>AT4G32040</i>	SALK_000339、SALK_067858、SALK_116798、 WiscDsLoxHs147_08A
<i>AT4G32050</i>	SALK_090592、SALK_096859、SALK_141633、 SALK_126745

### 三、實驗藥品

(一)植物培養(培養基成分)：Murashige and Skoog Basal Salt Mixture

(MS salt)、蔗糖 sucrose、2-(N-morpholino)ethanesulfonic acid(MES)、瓊脂 agar、  
RO H<sub>2</sub>O、HCl、KOH

(二)模擬植物乾旱逆境(培養基成分)：MS salt、蔗糖 sucrose、MES、瓊脂 agar、RO  
H<sub>2</sub>O、HCL、KOH、PEG-6000 粉末

(三)種子消毒：75% 酒精、90% 酒精、Triton-X100

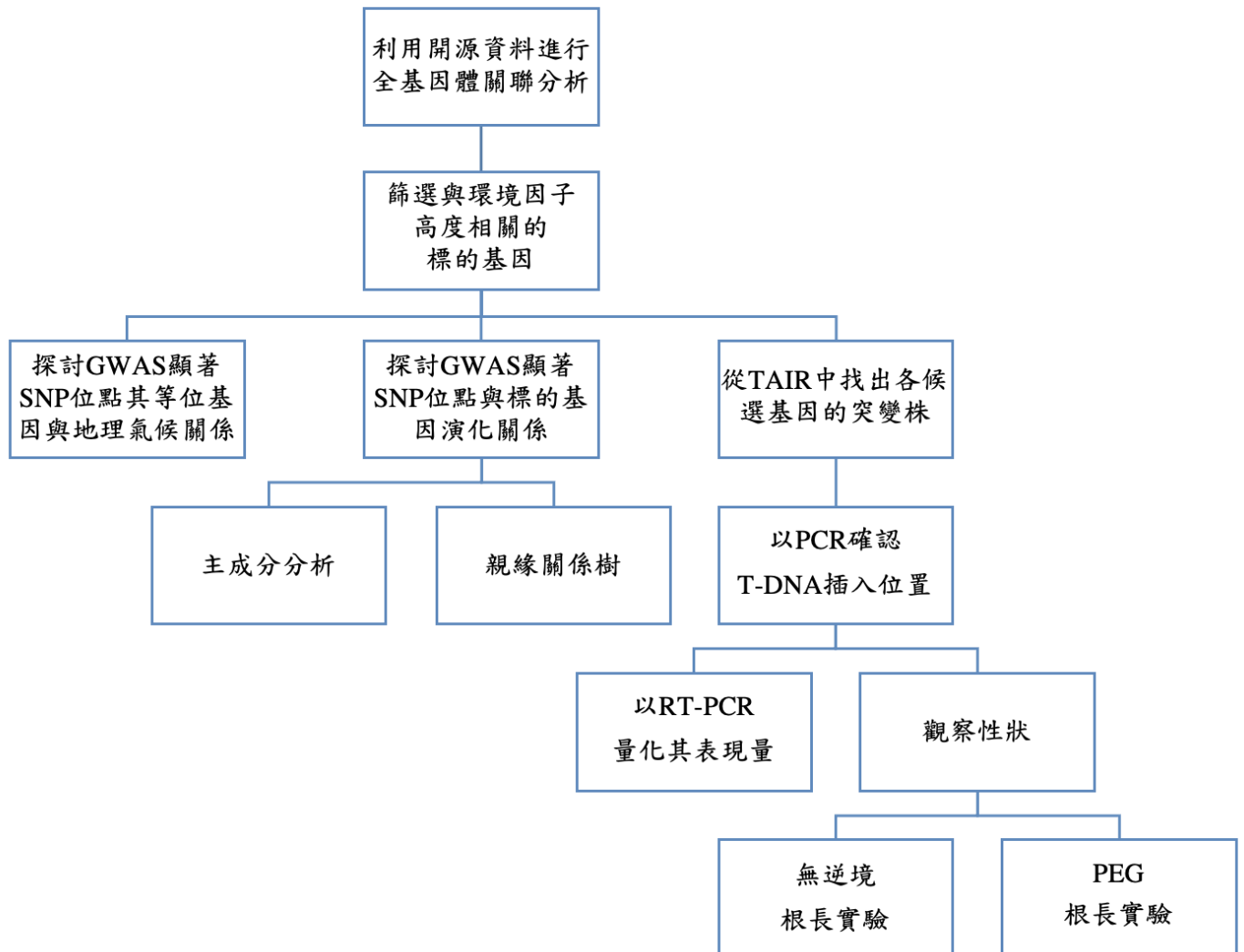
(四)萃取 DNA：DNA extraction buffer、異丙醇 isopropanol

(五)電泳：1X TAE Buffer、SYBR safe DNA Gel Stain、100bp DNA. Ladder(Marker)

## 肆、研究方法或過程

### 一、研究架構

- (一) 利用全基因體關聯分析找出與環境最相關的標的基因及其對應的環境因子
- (二) 找出與環境因子相關性最高的 SNP 位點之不同核苷酸多樣性，並比較其核苷酸在世界分佈情形
- (三) 觀察野生型與標的基因突變株在不同環境因子下之性狀，藉以了解標的基因之生理功能。



圖三、研究流程圖

## 二、研究方法

(一)以全基因體關聯分析找出與環境因子變化有高度相關的基因

1. 資料來源與分析方式：我們從前人研究(Ferrero-Serran *et al.* 2019)所提供的開源資料，取得了伊比利半島各地的 188 種氣候資料，如表二。同時，我們也從前人的研究中取得阿拉伯芥的 1,135 個基因體定序資料，從中挑選出伊比利半島的阿拉伯芥品系資料，並將氣候資料與基因資料放入 GWA—portal 網站中 (Atwell *et al.* 2010) 進行分析。

2. 氣候資料(Ferrero-Serran *et al.* 2019)(共 188 個，以下列舉 10 個為代表)

表二、與「降水量」相關的环境資料

climate data (與降水有關)	Description
Aridity_index_of_De_Martonne_GPCC_Fulldata	乾旱指數
Aridity_index_of_De_Martonne_GPCC_VASclimO	乾旱指數
CHELSA_BIO12	年降水量
CHELSA_BIO13	最濕月降水量
CHELSA_BIO16	最濕季降水量
CHELSA_BIO19	最冷季降水量
CRU_Pre_spring	春季降水量
GPCC Spring	春季降水量
WC2_Pre_spring	春季降水量

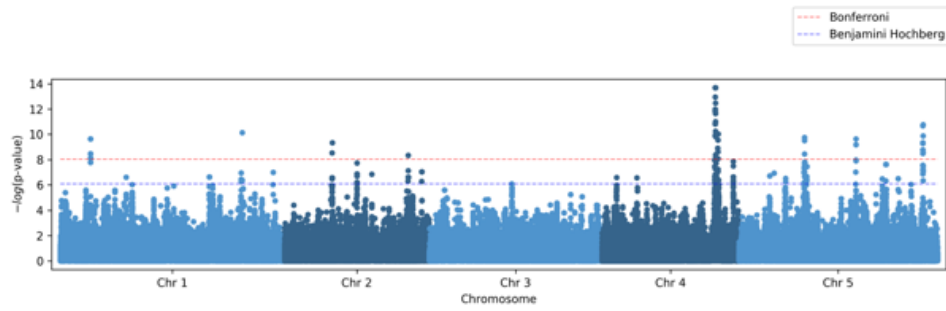
3. 全基因體資料來源——1001genomes project : 1001 基因體計劃於 2008 年初啟動，旨在定序模式植物阿拉伯芥至少 1001 個品系的詳細全基因體序列。此項目的第一個主要階段於 2016 年完成，發布了對 1,135 個基因體的詳細分析。

4. GWA-Portal 網站操作方式與資料判讀：

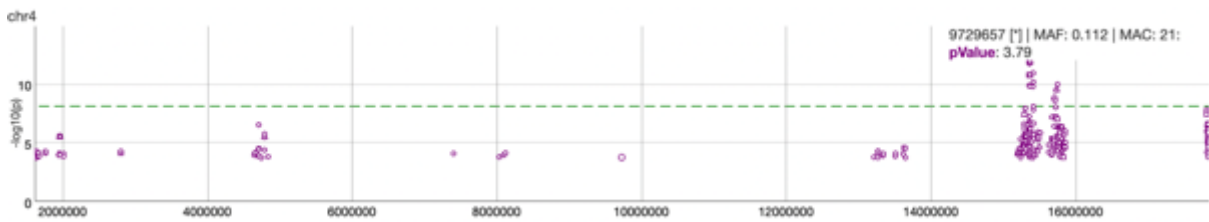
操作步驟：

- (1)上傳 phenotype(氣候資料)
- (2)選擇 genotypes(1001 genomes project)
- (3)性狀數值是否使用統計轉換：NO transformation
- (4)使用分析方法：Accelerated Mixed Model
- (5)分析結果(以環境資料：WC2 BIO9 為例)

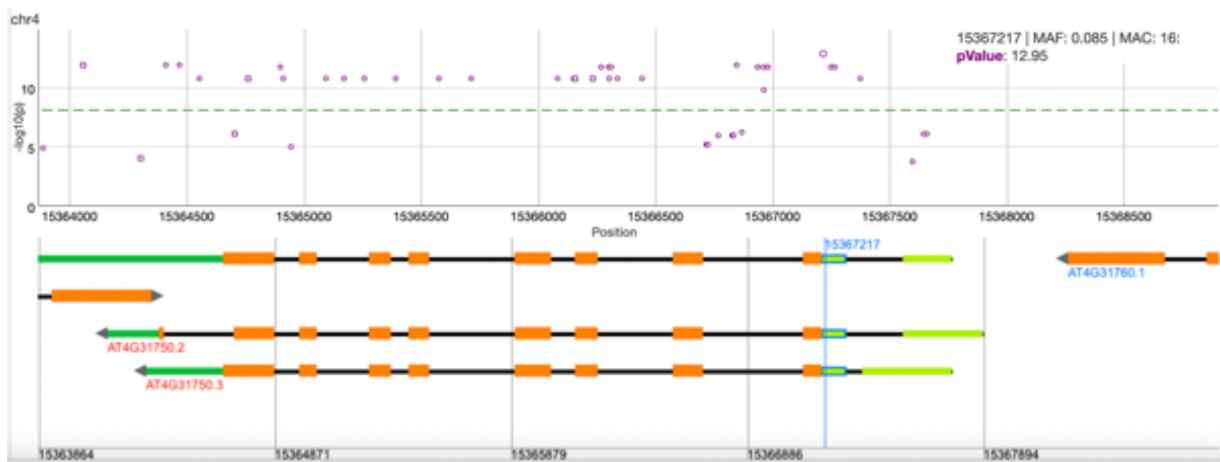




圖四、WC2 BIO9 的 Manhattan plot 結果，此環境資料於第四條染色體有顯著峰值



圖五、將圖四的第 4 條染色體部分放大，顯示顯著高峰。每個點代表一 SNP 位點，橫軸為 SNP 位點在序列上的位置，縱軸為 p 值取 -log



圖六、將圖五中第 4 條染色體的顯著高峰進一步放大，在圖表下方可查出該 SNP 位點的所在基因。

## (二) 觀察 SNP 位點於歐亞非之分佈情形

我們從前人研究(Lee *et al.* 2019) 取得的 1,323 個阿拉伯芥品系資料，利用 VCFtools 篩選出位於伊比利半島的阿拉伯芥 SNP 位點等位基因的資料，每個位點對應一個核苷酸多樣性的兩個核苷酸。我們利用 R studio 將品系依照其核苷酸標示於地圖上：以藍色(佔總核苷酸比例較多的核苷酸)及紅色(佔總核苷酸比例較少的核苷酸)表示，以得出各 SNP 位點在歐亞非三大洲的分佈情形。

### (三)主成分分析(Principal Component Analysis，以下簡稱 PCA)

#### 1. PCA introduction：

主成分分析(Principal Component Analysis, 後簡稱為 PCA)在 100 年前由英國數學家卡爾·皮爾森(Karl Pearson)發明，是一個至今仍在機器學習與統計學領域中被廣泛用以分析資料、降低數據維度的方法。

#### 2. 分析步驟：

- (1) 1,323 個阿拉伯芥品系在目標基因的 SNP 資料  
(原始資料來源：1001 Genomes Consortium, 2016)
- (2) 使用軟體 TASSEL 做出 PCA
- (3) 使用軟體 JMP 畫出 PCA 分群圖

### (四)親緣關係樹(Phylogenetic Tree)

#### 1. Phylogenetic Tree introduction：親緣關係樹(phylogenetic tree)或稱演化樹

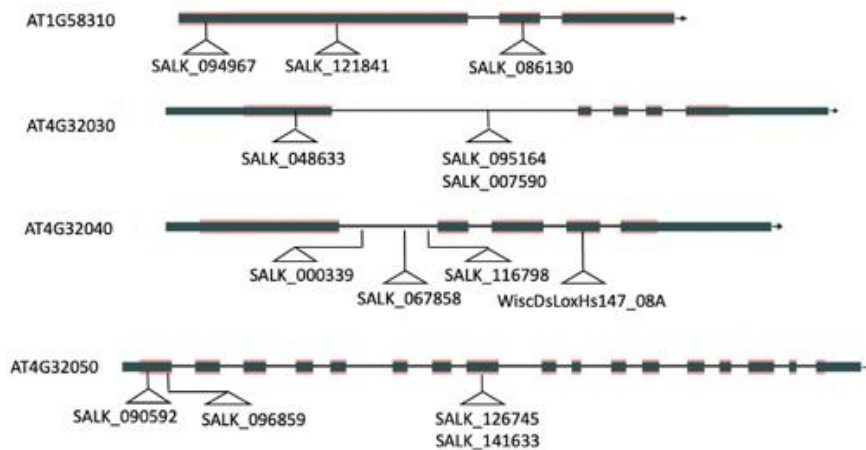
(evolutionary tree)用以表示物種間演化的親疏關係，同時也提供不同物種最小共同祖先的預測方向。

#### 2. 分析步驟：

我們用與 PCA 分析相同的遺傳資料，使用軟體 Tassel 5 計算出該基因的 Neighbor Joining Tree，接著利用 FigTree v1.4.4 將親緣關係數樹依照 SNP 位點的不同 allele 上色。

- (1) input：AT1G58310、AT4G32040 的基因定序資料  
(原始資料來源：Lee *et al.* 2019)
- (2) 使用軟體 Tassel 5 找出該基因的 Neighbor Joining tree
- (3) 將對應的 SNP 位點的核苷酸(每個位點有兩種核苷酸)設定顏色：  
藍色 color=#0000ff(佔該 SNP 位點比例較多的核苷酸)  
紅色 color=#ff0000(佔該 SNP 位點比例較少的核苷酸)  
(SNP、主成份分析、親緣關係樹顏色標示皆是如此)
- (4) 使用 FigTree 軟體上色、畫出親緣關係樹

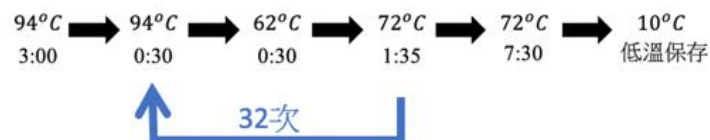
## (五)篩選突變株



圖七、*AT1G58310*、*AT4G32030*、*AT4G32040*、*AT4G32050*  
從 ABRC 訂購的突變種 T-DNA 插入位置圖

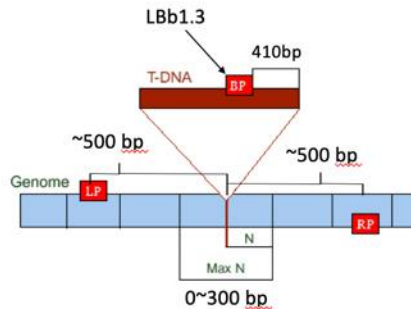
### 1. 流程

- (1) DNA extraction buffer 400 $\mu$ l 萃取葉片 DNA
- (2) 加入 50  $\mu$ l 的 ddH<sub>2</sub>O 回溶，放置 4°C 保存
- (3) 配置好 PCR 所使用的溶液，每 10  $\mu$ l 有: 2x taq master mix 5 $\mu$ l、ddH<sub>2</sub>O 3  $\mu$ l、primer F 0.5  $\mu$ l、primer R 0.5  $\mu$ l、whole genome DNA 1  $\mu$ l
- (4) PCR 模式(約 50 分鐘)



圖八、PCR machine settings

- (5) 配膠(膠是用 TAE buffer 配出 1% 的 agarose 膠)
  - a. Agarose 0.7 g 加入 TAE 70 ml
  - b. 用攪拌棒攪拌和微波爐加熱使 agarose 完全溶解
  - c. 加 EZvision dye 體積比例為 EZvision dye : 膠=1 : 10000
- (6) 在該突變的位點左右約各 500base pair 處設計 primer，以及在該突變位點插入的序列上設計 primer，分別跑 PCR，若植株為異型合子突變株，則兩組 primer 皆會有亮帶。若為同型合子突變株，只有第二組 primer 有亮帶。若植株為非突變株(即野生型)，則第二組 primer 不會出現亮帶。
- (7) 電泳後對照圖十可知基因型，留下同型合子進行後續實驗。



圖九、primer 設計

	wildtype (野生型)	異型合子	同型合子
primer F+R	~1000	~1000	
primer F/R+LB1.3		~500 + 410 ± N	~500 + 410 ± N

圖十、膠圖判讀

## (六)以 RT-PCR 量化突變株之標的基因表現量

此實驗共分為兩組量測，第一組為在無逆境處理下，採集 Col-0 與突變植株葉片以進行 RT-PCR；第二組則是採集種植於 1/2 MS 培養基的 Col-0 與突變植株根系以進行 RT-PCR，並觀測突變植株與 Col-0 在標的基因 *AT4G32040* 表現情形。

### 1. RNA 萃取

- (1) 採收植株樣本最多 100 mg，裝入 1.5 ml 的離心管中，加入滅菌過的砂珠，放入液態氮冷凍，用組織粉碎機打碎
  - a. 第一組：無逆境處理下植株的完全展開葉片
  - b. 第二組：種植於 1/2 MS 培養基的植株根系
- (2) 使用 RNeasyR Plant Mini Kit 萃取樣本 total RNA，以 30µl RNase-free water 回溶
- (3) 使用 Nanodrop 進行 RNA 濃度測定，並保存於-80°C冰箱

### 2. DNase 處理

由於萃取後的 RNA 有混雜 DNA，會影響後續實驗，需進行 DNase 處理以移除 DNA

- (1) 將萃取後的 RNA 定量至總量 10000ng，加入去除 DNA 的 DNase 1 µl、

幫助 DNase 反應的 DNase Buffer 10  $\mu$ l 以及 Nuclease-free water 至體積 100  $\mu$ l

- (2) 在稀釋後的 RNA 中，以 37°C 水浴反應 10 分鐘
- (3) 加入 1  $\mu$ l 的 0.5M EDTA，加入等體積的 LiCl 混合均勻，放過夜(至少 2 小時)沉澱 RNA
- (4) 以 4°C、16000xg 離心 30 分鐘，去除不含 RNA 上清液
- (5) 加入 80% 酒精 500  $\mu$ l，輕拍離心管使含有 RNA 的沉澱懸浮
- (6) 以 4°C、16000xg 離心 10 分鐘
- (7) 加入 100%酒精 500  $\mu$ l，輕拍離心管使含有 RNA 的沉澱懸浮
- (8) 放置通風櫥 5-10 分鐘，加入 20  $\mu$ l nuclease-free water，回溶 RNA
- (9) 使用 Nanodrop 進行 RNA 濃度測定，並保存於-80°C冰箱

### 3.反轉錄 (RT, reverse transcription)

- (1) 混合 RNA 1000ng、50uM oligo dT 1 $\mu$ l、10mM dNTP 1  $\mu$ l
- (2) 放入 PCR machine 65 度 5 分鐘，4 度至少 2 分鐘
- (3) Reaction buffer (5 X) 4  $\mu$ l、dTT 2  $\mu$ l、SSTIII 1  $\mu$ l

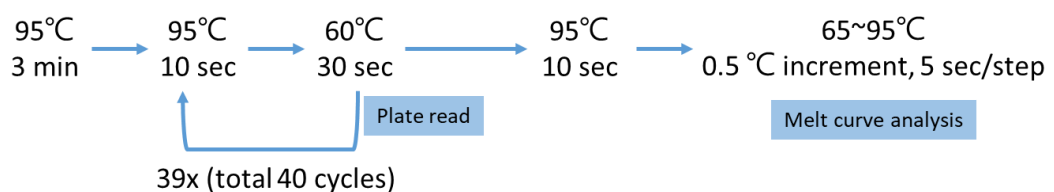
以 PCR machine 37°C 60 分鐘、72°C 15 分鐘、4°C 保存，進行反轉錄反應

- (4) 反轉錄後的 cDNA 進行 10 倍稀釋(20cDNA，180 水)，並放入-20°C保存

### 4.RT-PCR

- (1) 在我們預測的突變株 cDNA 序列上設計一組 primer
- (2) 混合 5  $\mu$ l SYBR Green、1.25  $\mu$ l cDNA、2.75  $\mu$ l ddH<sub>2</sub>O 並加入設計好的 primer 各 0.5 $\mu$ l
- (3) 於核酸增殖器進行 RT-qPCR，觀察熔化曲線、比較 CT 值

PCR 模式(約 1 小時 30 分鐘)



圖十一、PCR 模式圖

### (七)無逆境根長實驗

將篩選出為突變株的種子點在培養基 6 天，選擇植株生長狀況相近的植株移至新的培養基，並將小苗的根尖移至相同起始點，待其生長 7 天後掃描紀錄根長，以觀測並比較突變株與 Col-0 的根系生長情形。

#### 1. 培養基配製：

(1)移苗前的圓形培養基： $\frac{1}{2}$  MS medium 500 ml

(2)方形培養基配製： $\frac{1}{2}$  MS medium 500 ml (同上)

#### 2. 根長實驗流程：

(1) 將篩選出為突變株的種子先點在圓形培養皿中，共 8 個品系

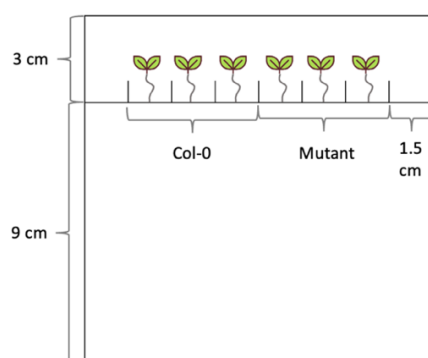
(2) 6 天後，選擇生長狀況相近的小苗，每個品系 9 株

(3) 移苗至方形培養基(規格：12cm x 12cm)

a. 將小苗的根尖對準至相同起始點

b. 一個方形培養基有 3 株突變株、3 株 Col-0，  
一個品系使用 3 個方形培養基。

c. 植株間隔 1.5 cm



圖十二、根長實驗植株與培養基示意圖

(4) 於生長箱生長 7 天(生長環境：溫度 22°C、照光 16 小時/黑暗 8 小時、光合作用光子通量密度：100PPFD)

(5) 掃描培養基記錄根長，並以 ImageJ 測量主根根長、總根長

#### 3. 根長測量

(1) 先在方形培養基旁畫上 1cm 的直線，掃描後使用 ImageJ 軟體的 set scale 功能，制定比例尺。

(2) 從 ImageJ 軟體中直線繪製功能中的折線選項(segmented line)，測量植物根長:扣除植物下胚軸的主根加副根總長

(3) 使用 JMP 軟體做統計分析

## (八)PEG-6000 乾旱逆境根長實驗

藉由調整不同濃度的聚乙二醇(PEG-6000)模擬不同程度的乾旱環境，觀察突變株與 Col-0 之根系生長情形。

### 1. 培養基配製 (Van der Weele *et al.* 2000 ; Verslues *et al.* 2004) :

(1)移苗前的圓形培養基：½ MS medium 500 ml

(2)方形培養基配製：

a. agar media (per liter) :

MS salts 2.2 g

MES buffer (6 mM) 1.2 g

agar 1.5 g

b. PEG added per liter of media to make the overlay solution :

分別為：0 g、125 g、250g

### 2. 實驗材料：

表三、乾旱逆境根長實驗所選用的標的基因及其突變品系

Gene	Mutant line
AT4G32040	SALK_000339、SALK_067858、 SALK_116798、WiscDsLoxHs147_08A

### 3. PEG-6000 根長實驗流程：

(1) 將篩選出為突變株的種子先點在圓形培養皿中，共 4 個品系

(2) 6 天後，選擇生長狀況相近的小苗，每個品系 18 株

(3) 移苗至方形培養基(規格：12cm x 12cm)

a. 將小苗的根尖對準至相同起始點

b. 一個方形培養基有 3 株突變株、3 株 Col-0

c. 各品系使用兩個 0 g、125 g、250g PEG (per liter)培養基

(4) 於生長箱生長 7 天(生長環境：溫度 22°C、照光 16 小時/黑暗 8 小時、光合作用光子通量密度：100PPFD)

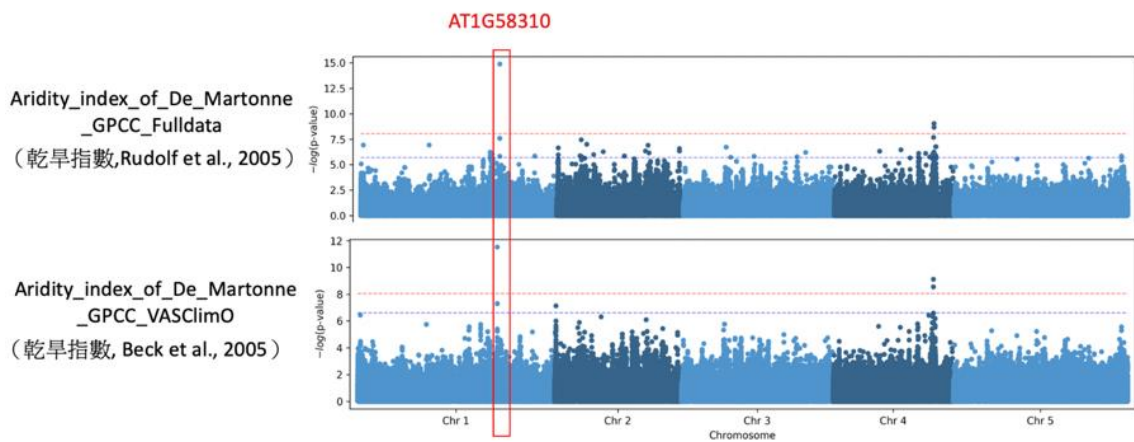
(5) 掃描培養基記錄根長，並以 ImageJ 測量主根根長、總根長

(6) 使用 JMP 軟體進行統計分析

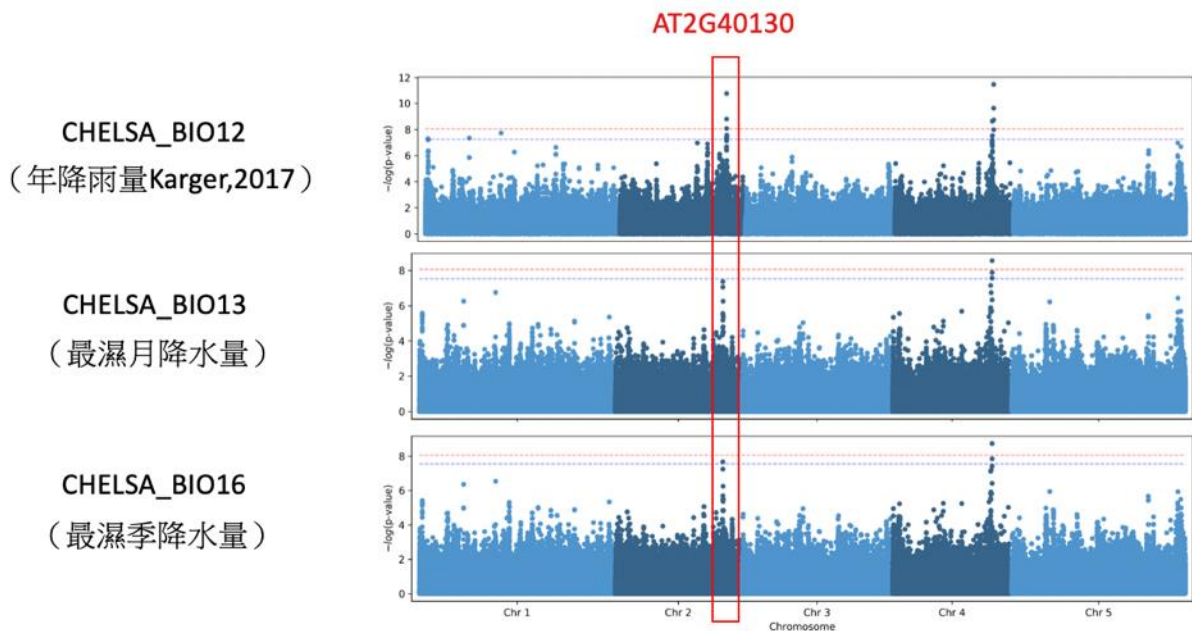
## 伍、研究結果

### 一、以伊比利半島的基因定序資料和環境資料進行全基因體關聯分析

將伊比利半島之阿拉伯芥基因定序資料與該地環境資料進行全基因體關聯分析，發現結果中許多與乾旱指數、降雨量相關的環境資料在 *AT1G58310* 基因、*AT2G40130* 基因和 *AT4G32040* 基因有高峰值，顯示這些基因可能與植物乾旱耐受性有密切相關，如圖十一到圖十四，因此後續將以此三種基因做為研究對象。由於 *AT2G40130* 基因已有前人研究發現其與抗旱功能有關 (Li *et al.* 2020)，因此後續我們將針對 *AT1G58310* 和 *AT4G32040* 進行族群演化和基因功能的研究。



圖十三、GWAS 分析後發現有兩項乾旱指數在 *AT1G58310* 基因有高峰值



圖十四、GWAS 分析後發現有三項降水量資料在 *AT2G40130* 基因有高峰值

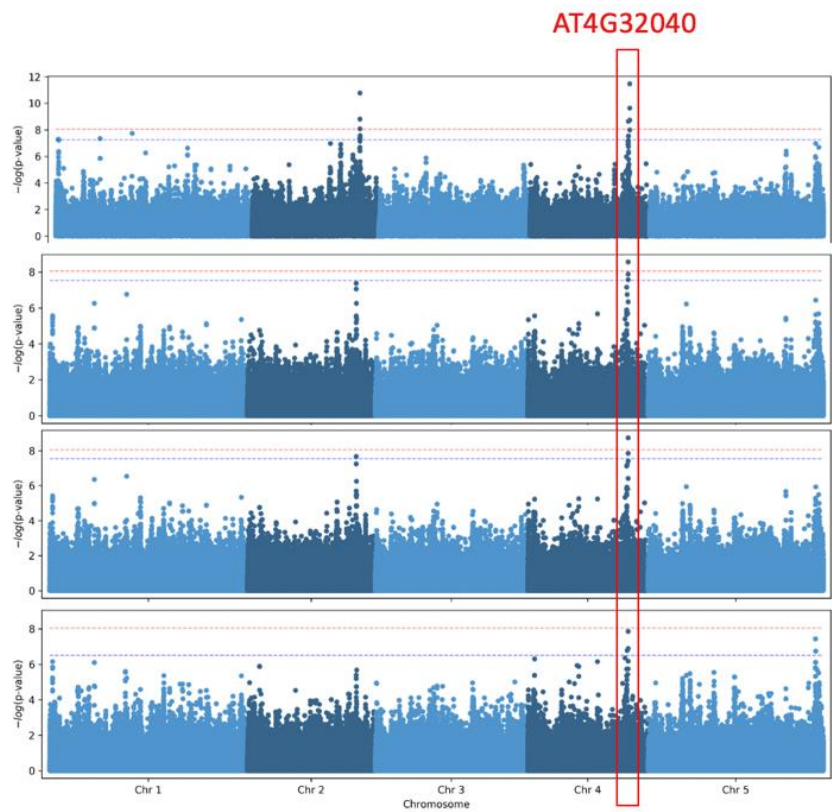


CHELSA\_BIO12  
(年降雨量Karger,2017)

CHELSA\_BIO13  
(最濕月降水量)

CHELSA\_BIO16  
(最濕季降水量)

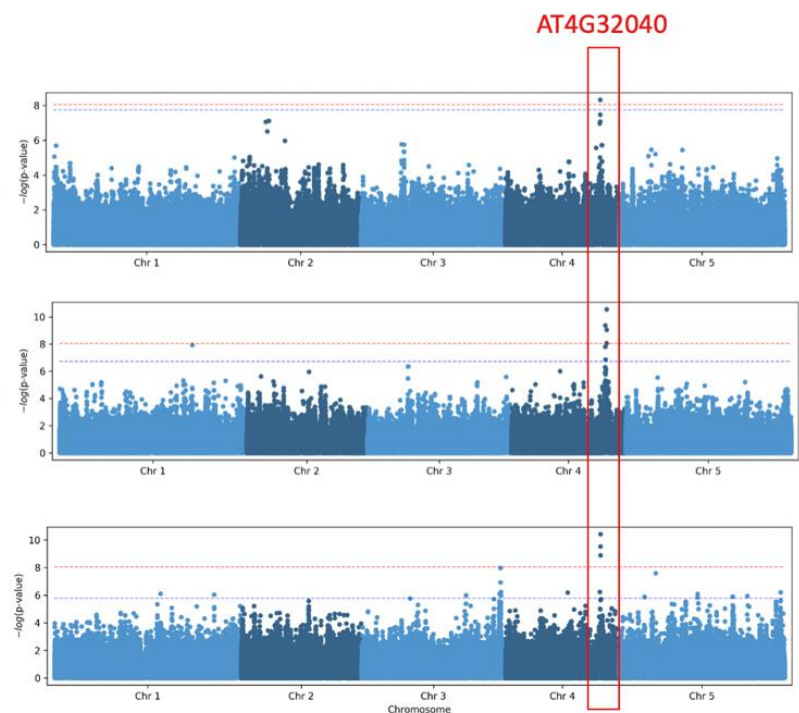
CHELSA\_BIO19  
(最冷季降水量)



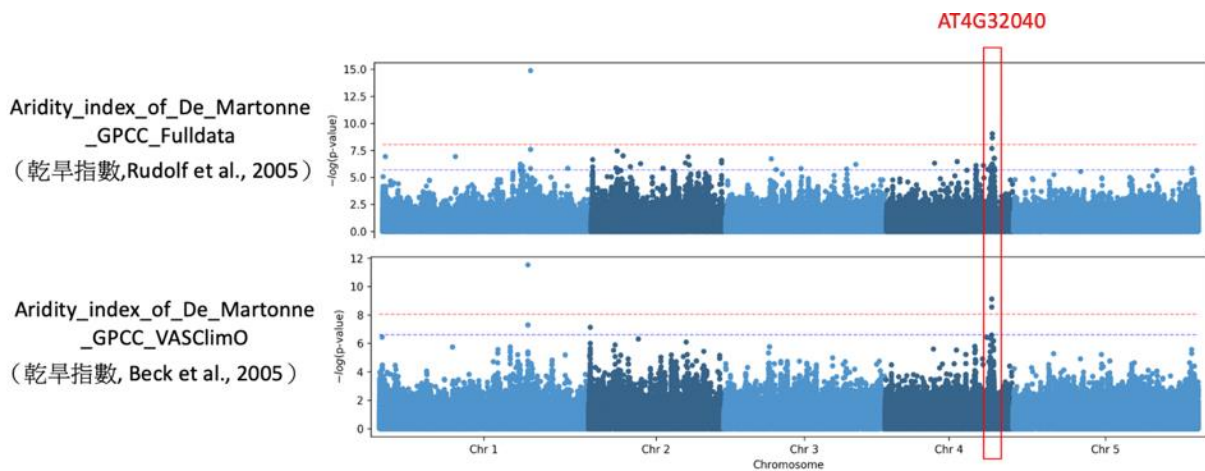
CRU\_Pre\_spring  
(春季降水量Harris. I., 2014)

GPCC Spring  
(春季降水量Schneider, 2011)

WC2\_Pre\_spring  
(春季降水量Fick,2017)



圖十五、GWAS 分析後發現有七項降水量資料在 *AT4G32040* 基因有高峰值



圖十六、GWAS 分析後發現有兩項乾旱指數在 *AT4G32040* 基因有高峰值

## 二、探討全基因體關聯分析結果顯著的SNP位點其核苷酸之地理分布及與氣候資料的相關性

(一)將 *AT1G58310* 和 *AT4G32040* 的 GWAS 結果各四個高峰 SNP 位點的核苷酸標示於地圖上，藉此觀察標的基因於 GWAS 結果出現峰值是否與不同族群分佈有關。

由圖十七、圖十八可知，由這些 SNP 位點之等位基因地理分佈可看出這些位點在整個歐亞大陸尺度確實沒有明顯分布差異。我們這次僅以伊比利半島的資料作分析得以找出這些基因，可能是因為伊比利半島具有的環境梯度與整個歐亞大陸不同，因此無法得到一致的結果。

此外，為了解伊比利半島具有的環境梯度與整個歐亞大陸是否相同，我們將所有植株品系分成伊比利半島、非伊比利半島兩群，再將 *AT1G58310*、*AT4G32040* 單一位點核苷酸的數量分成四群，最後觀察不同地區的不同單一位點核苷酸與其植株品系的氣候資料關係。由於同一基因的不同 SNP 位點之分析結果相近，故以下將以 S1\_21632820 代表 *AT1G58310* 基因、S4\_15496273 代表 *AT4G32040* 基因表示分析結果。



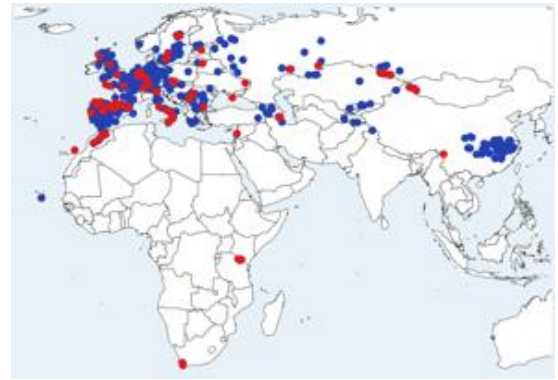
S1\_21631971 A 藍/G 紅



S1\_21632322 G 藍/T 紅



S1\_21632820 A 藍/T 紅

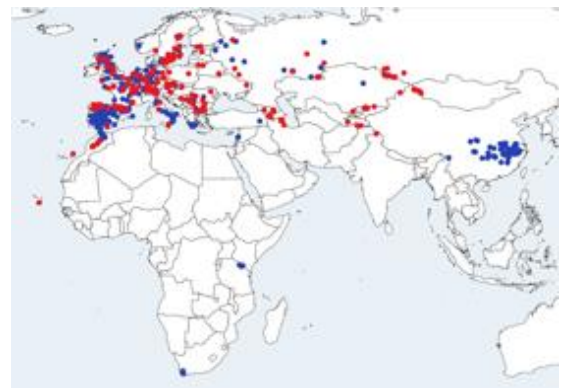


S1\_21632841 G 藍/T 紅

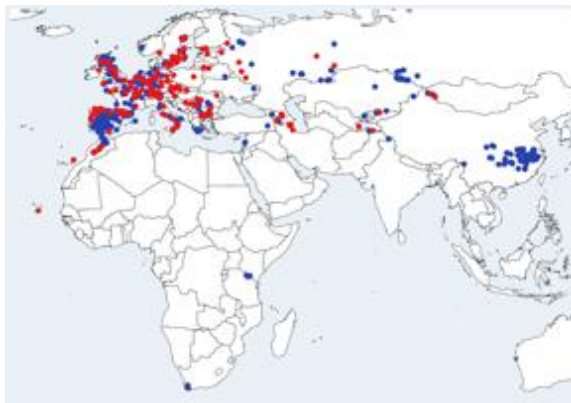
圖十七、將*AT1G58310* GWAS結果 4 個SNP高峰點依其核苷酸分布標示於地圖  
核苷酸於該位點較多者皆標示為藍色，核苷酸較少者皆標示為紅色。



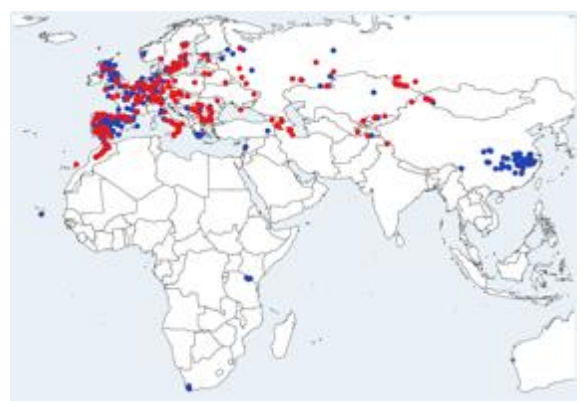
S4\_15495046 T 藍/G 紅



S4\_15495173 T 藍/G 紅



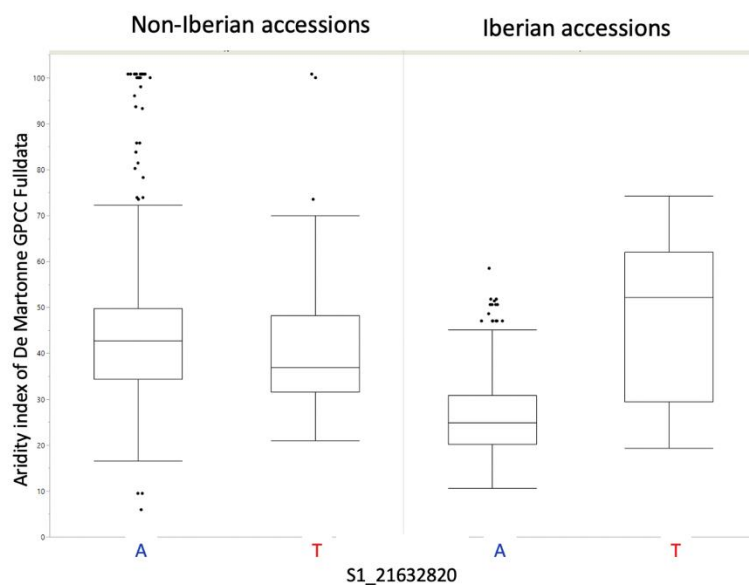
S4\_15496273 T 藍/C 紅



S4\_15499920 C 藍/T 紅

圖十八、將*AT4G32040* GWAS結果 4 個SNP高峰點依其核苷酸分布標示於地圖  
核苷酸於該位點較多者皆標示為藍色，核苷酸較少者皆標示為紅色。

(二)S1\_21632820 的乾旱指數於伊比利半島的族群及非伊比利半島的族群的核苷酸分布比較



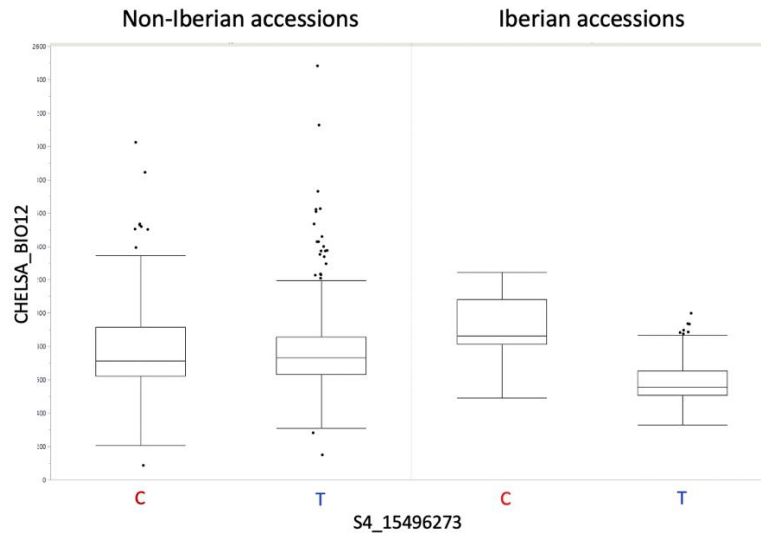
climate data	p-value of two alleles	
	Non-Iberian	Iberian
Aridity_index_of_De_Martonne_GPC_Fulldata	0.2383	<0.0001*
Aridity_index_of_De_Martonne_GPC_VASclimO	0.3650	<0.0001*

圖十九、SNP 位點 S1\_21632820 於伊比利半島的族群及非伊比利半島的族群之核苷酸分佈比較

與乾旱指數環境資料有顯著相關的 SNP 位點 S1\_21632820 於伊比利半島(Iberian)的族群及非伊比利半島(Non-Iberian)的族群的核苷酸分佈比較之整理，由圖可看出在伊比利半島的族群的核苷酸分布有懸殊差距，而在世界各地則無此情況，推測是因伊比利半島基因體之等位基因造成的環境適應優勢在全世界基因體中無法被凸顯。

(三)S4\_15496273 於伊比利半島的族群及非伊比利半島的族群的核苷酸分布比較

S4\_15496273 的降雨量資料於伊比利半島的族群及非伊比利半島的族群的核苷酸分布比較之整理，由圖可看出在伊比利半島的族群的核苷酸分布有懸殊差距，而在世界各地則無此情況。



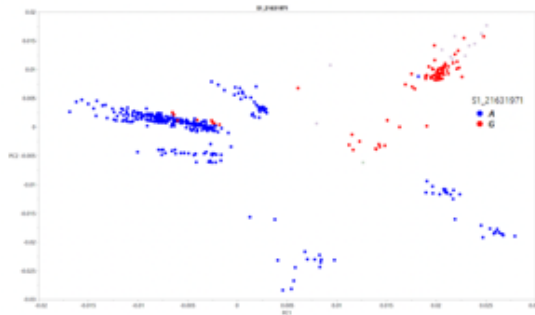
climate data (降水量相關)	p-value of two alleles	
	Non-Iberian	Iberian
CHELSA_BIO12	0.4988	<0.0001*
CHELSA_BIO13	0.3635	<0.0001*
CHELSA_BIO16	0.3444	<0.0001*
CHELSA_BIO19	0.0027*	<0.0001*
CRU_Pre_spring	0.9875	<0.0001*
GPCC Spring	0.1377	<0.0001*
WC2_Pre_spring	0.0335*	<0.0001*

圖二十、SNP 位點 S4\_1549627 於伊比利半島的族群及非伊比利半島的族群之核苷酸分佈差異

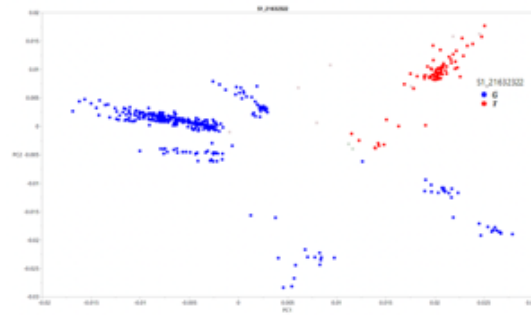
### 三、主成分分析(Principal Component Analysis, PCA)和親緣關係樹

#### (一)主成份分析(Principal Component Analysis, PCA)

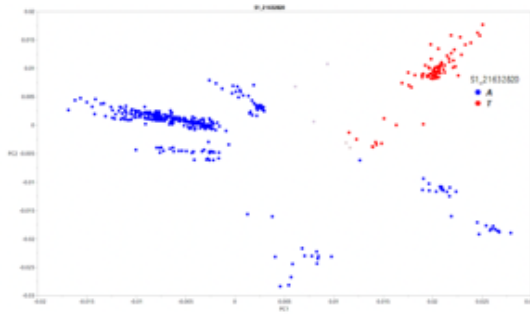
我們利用 *AT1G58310*、*AT4G32040* 的基因定序資料，以 PCA1 解釋各地植株的基因序列變異度，PCA2 解釋第二多的基因序列變異度，在分別計算其 PCA1 和 PCA2 的數值後，將 PCA2 對 PCA1 作圖。圖形中的點的相對距離即可視為彼此兩植株品系在這個基因的遺傳差異，可看出 *AT1G58310* 在遺傳結構上大致上分成五群，紅色為其中一小群，此結果可呼應 *AT1G58310* 的親緣關係樹(圖二十一)，而 *AT4G32040* 在遺傳結構上大致分為二群。



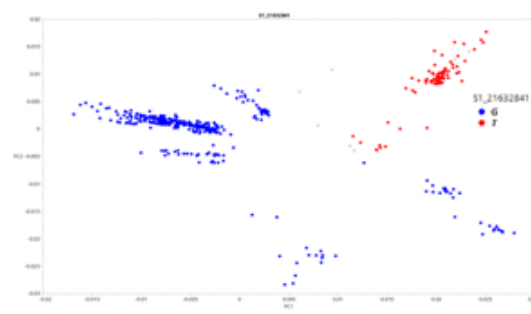
S1\_21631971 A 藍/G 紅



S1\_21632322 G 藍/T 紅



S1\_21632820 A 藍/T 紅

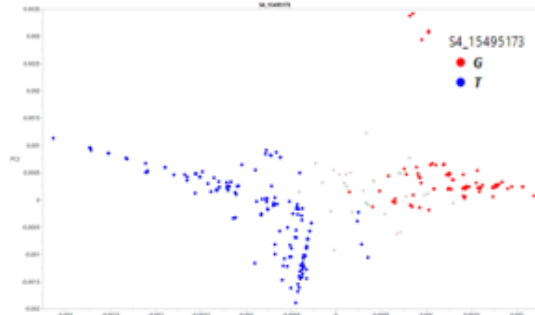


S1\_21632841 G 藍/T 紅

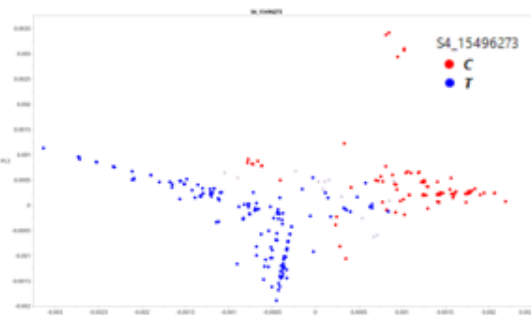
圖二十一、將 *AT1G58310* GWAS 結果中的 4 個 SNP 高峰點進行 PCA 做圖



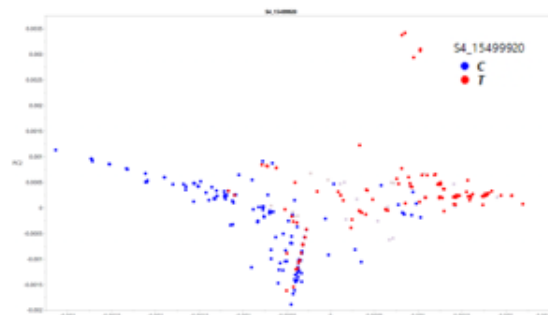
S4\_15495046 T 藍/G 紅



S4\_15495173 T 藍/G 紅



S4\_15496273 T 藍/C 紅

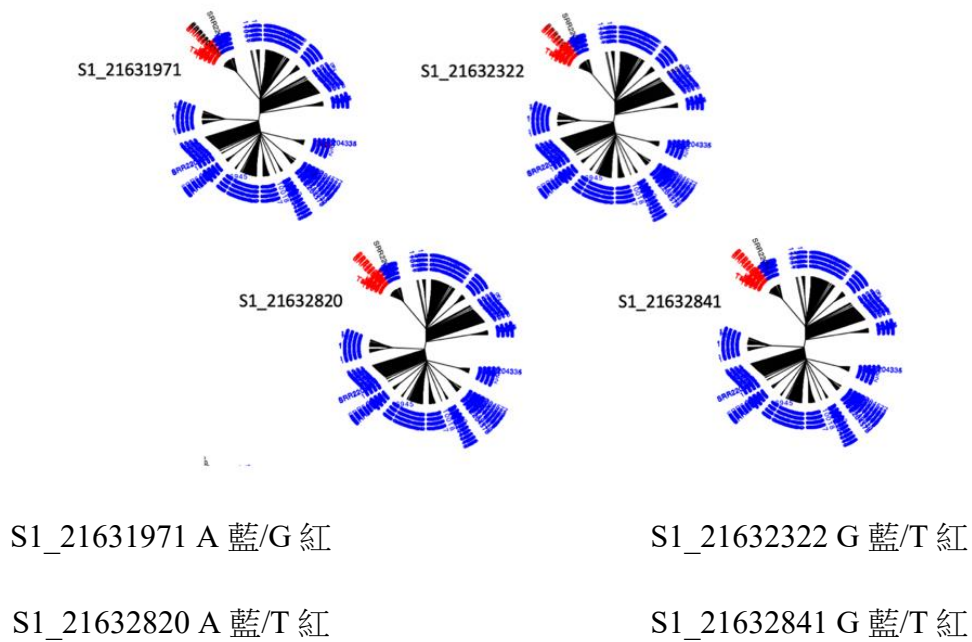


S4\_15499920 C 藍/T 紅

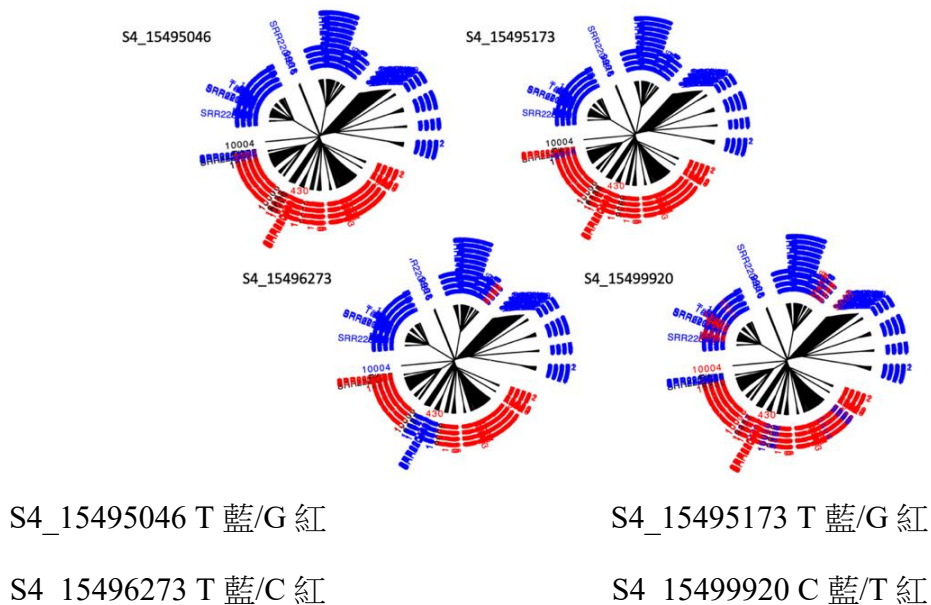
圖二十二、將在 *AT4G32040* GWAS 結果中的 4 個 SNP 高峰點以 PCA 做圖

## (二)親緣關係樹

利用 FigTree v1.4.4 將 *AT1G58310* 與 *AT4G32040* 的親緣關係樹依照不同核苷酸標示以不同顏色後，可觀察到在 *AT1G58310* 中，四個位點的核苷酸分布相似，紅色集中於阿拉伯芥所有族群的一小群，此代表紅色一小群遺傳距離相近且在此位點有相同突變。*AT4G32040* 中，四個位點核苷酸的分布類似，紅藍大約也各佔整個阿拉伯芥族群的一半。在此兩基因中皆有一些藍色中穿插紅色，代表這些位點可能有再次突變，或是這個位點在有些個體有遺傳重組的情況。



圖二十三、將 *AT1G58310* 基因 GWAS 結果中的 4 個 SNP 高峰畫出親緣關係樹



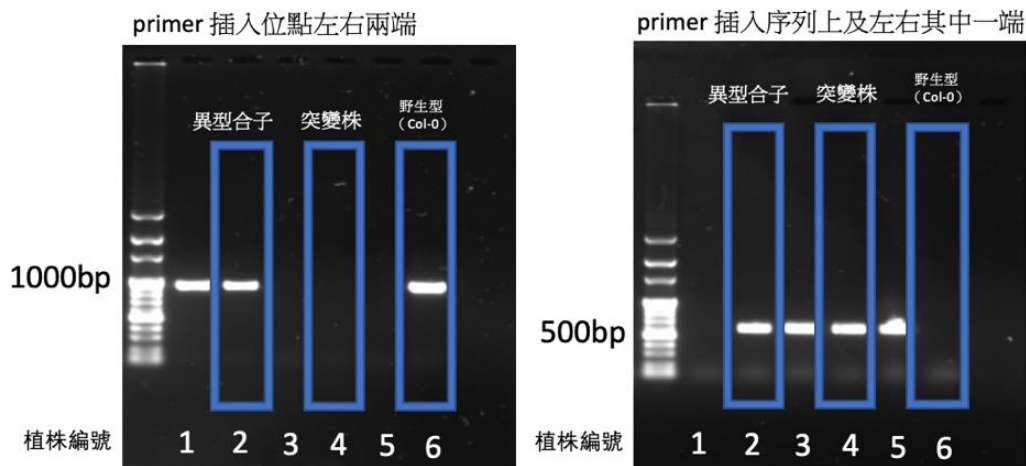
圖二十四、將 *AT4G32040* 基因 GWAS 結果中的 4 個 SNP 高峰畫出親緣關係樹

#### 四、篩選突變株

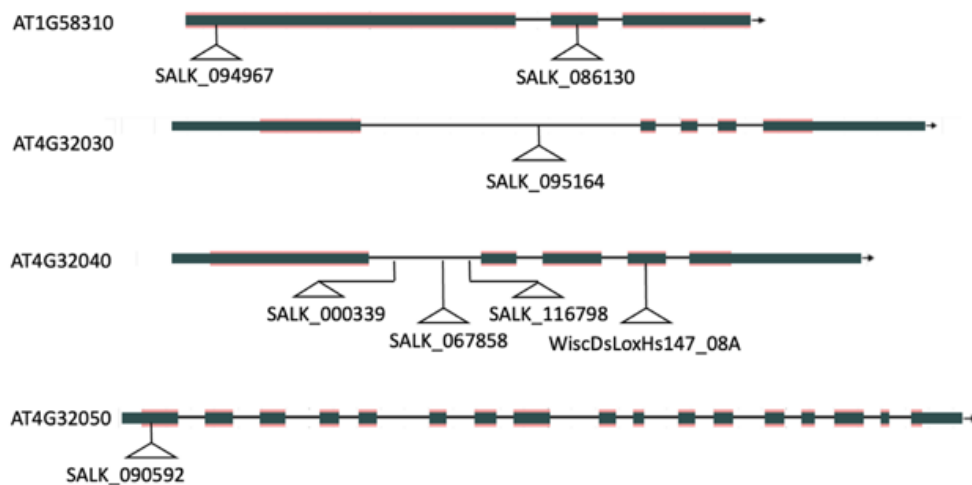
藉由GWAS找到標的基因後，在ABRC網站上購買有T-DNA插入加以破壞、降低標的基因之表現量之突變株。由於無法確認購買的種子是否全都為突變株，故我們將買來的突變株種子生長 25 天後，以PCR篩選出真正為突變株的植株。

接下來，設計該突變位點左右兩端的 primer 及插入序列的 primer 進行 DNA 電泳，膠圖中若左右兩端的 primer 沒有亮帶，而有插入序列的 primer 有亮帶，即可判斷其為突變株。結果如圖二十五所示。我們以篩選出確實有插入標的基因的突變株，接續進行後續的觀察性狀實驗。

以突變株SALK\_094967 的膠圖結果為例：



圖二十五、SALK\_094967 的膠圖結果

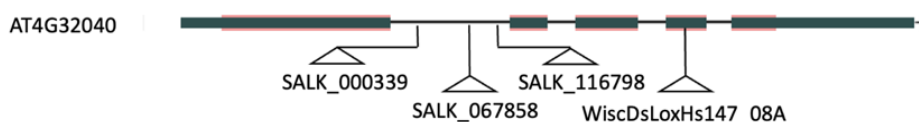


圖二十六、篩選出的突變株及其對應基因



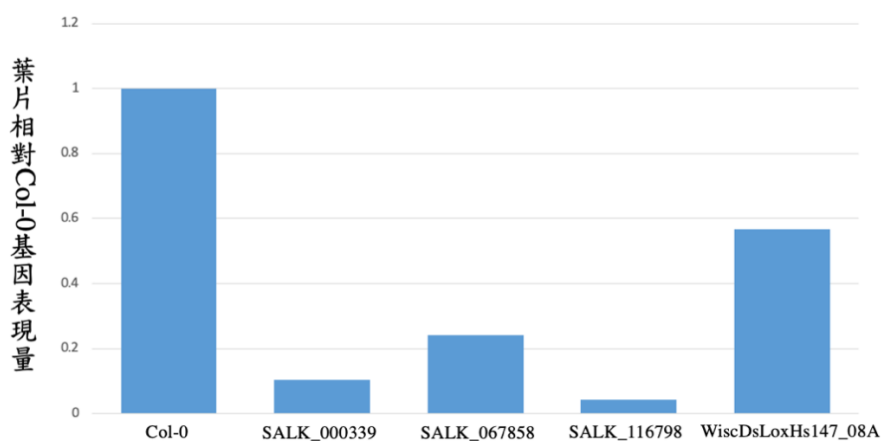
## 五、以 RT-PCR 量化突變植株之標的基因表現量

接著，我們以 RT-PCR 的方式測量標的基因 *AT4G32040* 表現量。圖二十八是在正常情況下突變株葉片相對 Col-0 的標的基因表現量(將 Col-0 表現量定為 1)，圖二十九是在乾旱逆境處理後，突變株葉片相對 Col-0 的標的基因表現量(將 Col-0 表現量定為 1)，結果顯示突變株不論在正常情況或是乾旱逆境處理下，標的基因表現量皆有顯著的下降。透過 genotyping 以及 RT-PCR 實驗，我們確認所選用的突變株確實達到基因減弱的效果。



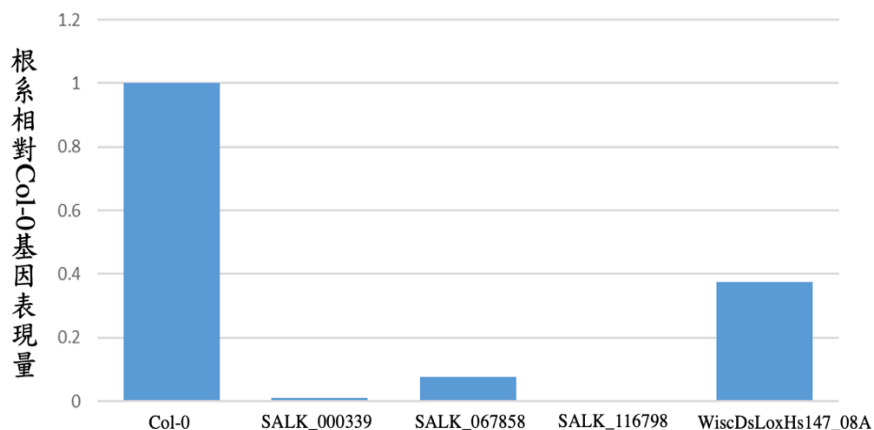
圖二十七、*AT4G32040* 及本實驗取用的四個突變品系

### (一) 量化正常生長情形下突變植株葉片之標的基因表現量



圖二十八、突變品系葉片相對 Col-0 的標的基因表現量(將 Col-0 表現量定為 1)

### (二) 量化種植於 1/2 MS 培養基突變植株根系之標的基因表現量



圖二十九、突變品系根系相對 Col-0 的標的基因表現量(將 Col-0 表現量定為 1)

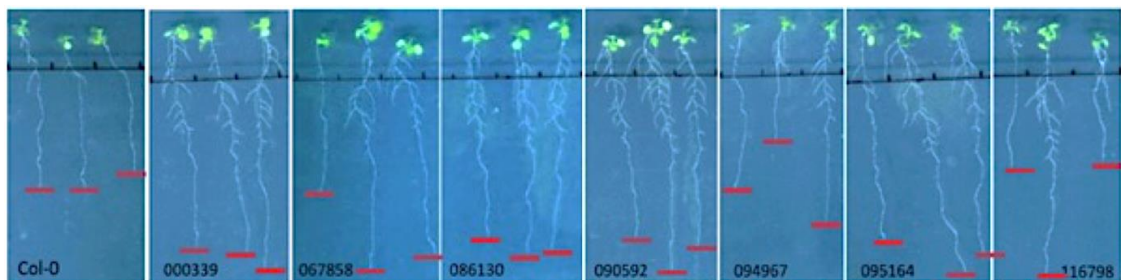
## 六、觀察性狀

由於最初在 GWAS 結果中顯示有峰值的環境資料大多是與「降水量」有關的環境因子，但其相關解釋仍不明確；因此我們選擇由「乾旱逆境」切入，且由於根系生長為植物吸水能力之關鍵因子，我們以 *AT1G58310*、*AT4G32030* 及 *AT4G32050* 突變株進行根長量測實驗，探討標的基因之乾旱逆境耐受力。

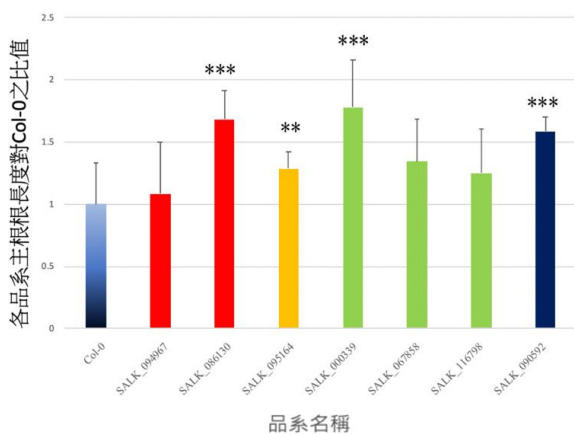
### (一)無逆境根長實驗

將種子泡在 4°C ddH<sub>2</sub>O 中 7 天促進發芽後，於培養基生長六天，將植株移苗使各突變株之根尖有相同起始點；再經 7 天後，量測各突變株的根長長度。

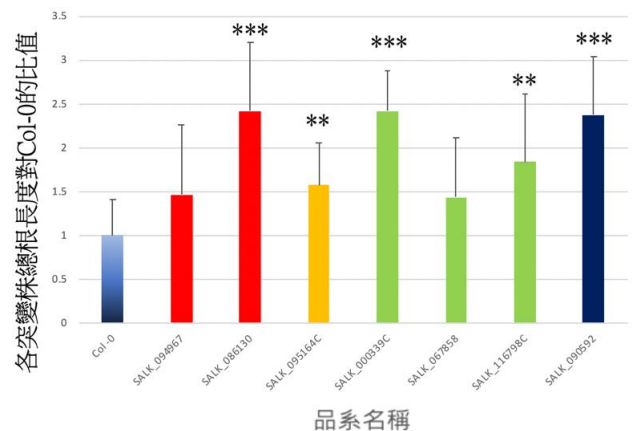
由主根長的實驗結果得知，*AT1G58310* 基因的 SALK\_086130 突變株、*AT4G32030* 基因的 SALK\_095164 突變株、*AT4G32040* 基因的 SALK\_000339C 突變株、*AT4G32050* 基因的 SALK\_090592 突變株與野生型(Col-0)的主根根長有顯著差異(p-value < 0.05)，而 SALK\_086130、*AT1G58310* 基因的 SALK\_086130 突變株、*AT4G32030* 基因的 SALK\_095164 突變株、*AT4G32040* 基因的 SALK\_000339C 和 SALK\_116798C 突變株、*AT4G32050* 基因的 SALK\_090592 突變株在總根長與野生型(Col-0)有顯著差異(p-value < 0.05)。



圖三十、生長於方形培養皿之小苗



圖三十一、各突變株主根根長對 Col-0 之比值



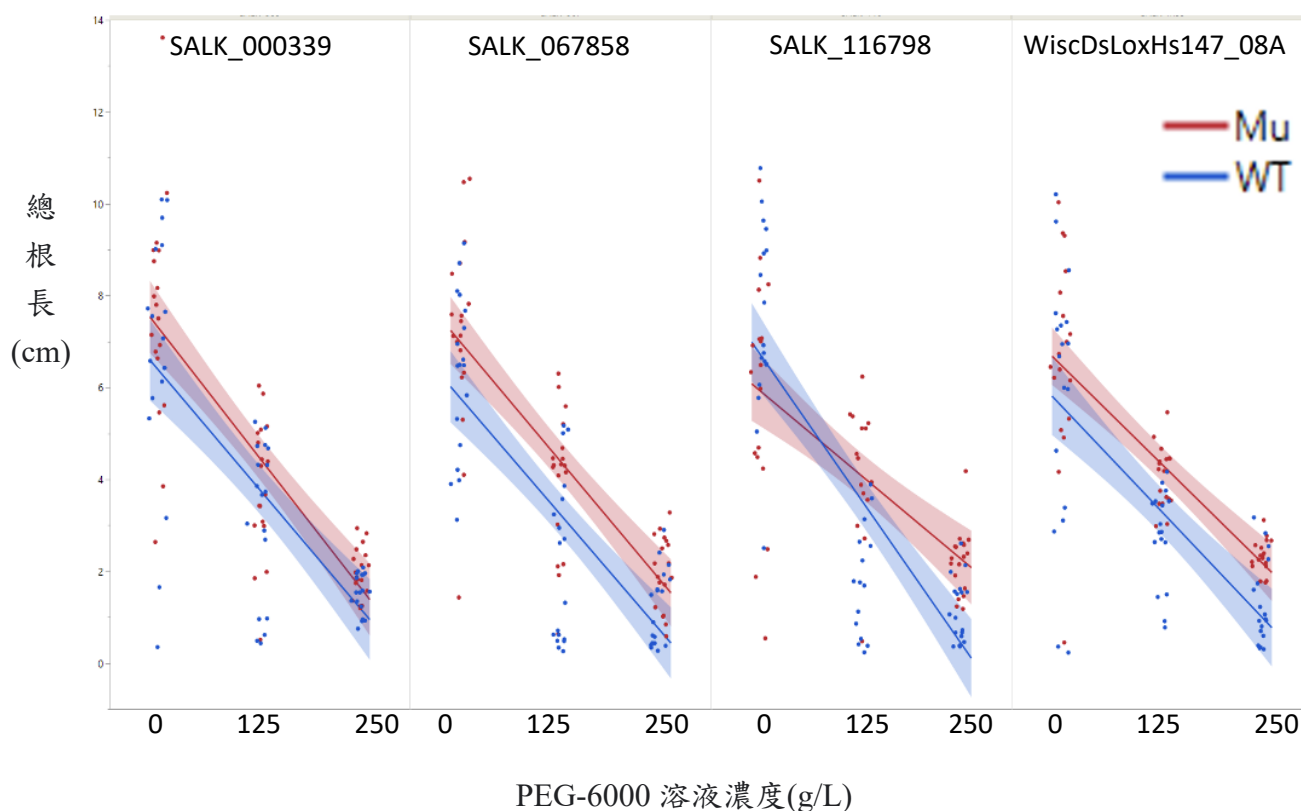
圖三十二、各突變株總根長對 Col-0 之比值

紅色：*AT1G58310* / 黃色：*AT4G32030* / 綠色：*AT4G32040* / 藍色：*AT4G32050*

## (二) PEG-6000 乾旱逆境根長實驗

PEG-6000(Polyethylene Glycol 6000)能有效降低植物生長環境的水勢能，而 PEG-6000 濃度越高，將使植物身處更高程度的乾旱逆境。本實驗藉由調配含有 0g/L、125g/L、250g/L PEG-6000 的培養基，設計了三個不同程度的乾旱逆境，並測量突變株小苗於不同逆境下的總根長。結果如圖三十三所示，SALK\_067858、WiscDsLoxHs147\_08A 突變株與野生型的總根長在不同濃度處理下有顯著差異；而 SALK\_116798 與野生型根長的差異性受 PEG 濃度的影響，在較高濃度處理下，該突變株的抗旱能力始被凸顯，同時也顯示該突變株根長在乾旱逆境下，下降速率顯著較野生型慢。

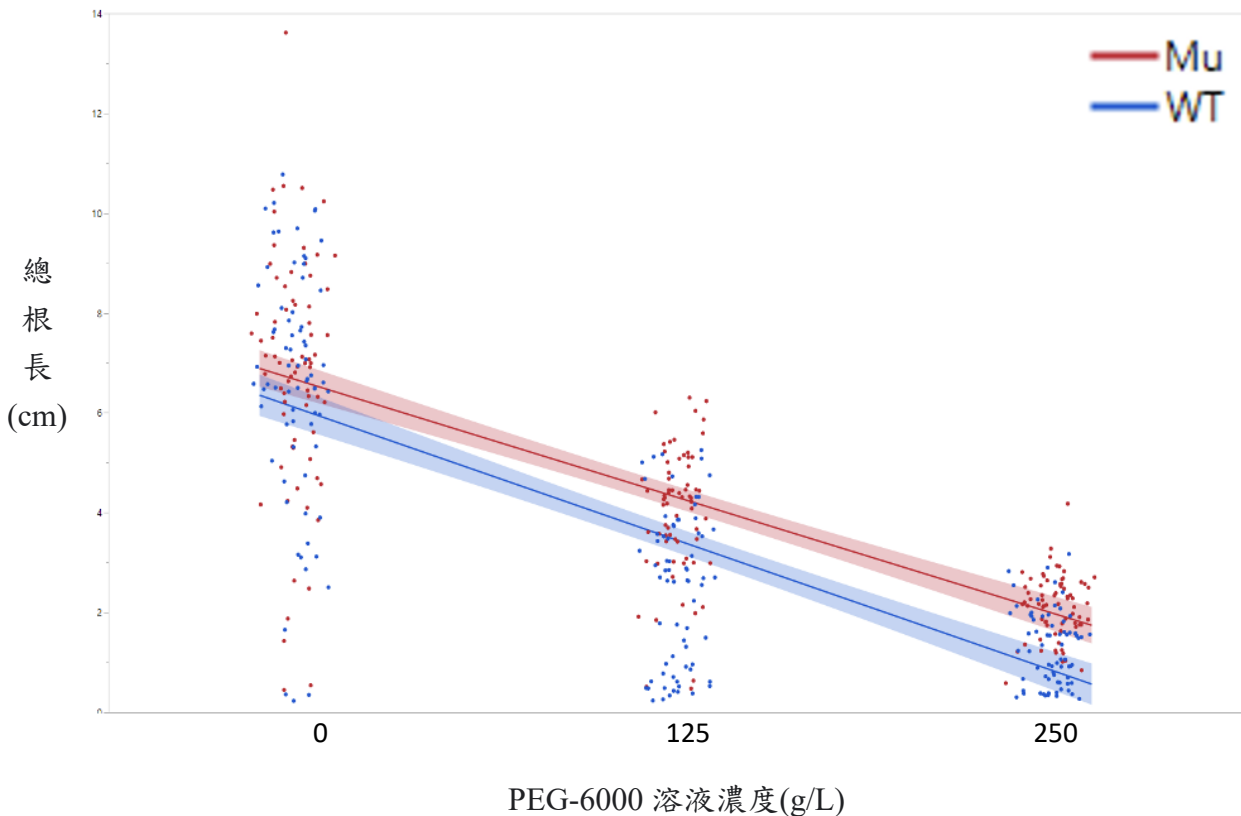
將 *AT4G32040* 的四個突變品系樣本合併分析如圖三十四所示，整體而言，突變株的根長顯著較野生型長，顯示 *AT4G32040* 具有藉由調節植物根系生長達到抗旱之效果。



圖三十三、*AT4G32040* 之四個突變品系：SALK\_000339、SALK\_067858、SALK\_116798、WiscDsLoxHs147\_08A(由左至右)在三種 PEG 濃度處理下的總根長  
紅色：突變株品系總根長對 PEG-6000 濃度之回歸直線  
藍色：Col-0 總根長對 PEG-6000 濃度之回歸直線

表四、 *AT4G32040* 之四個突變品系總根長在 (由上到下) 1. PEG 濃度 2.突變株與野生型比較 3. 濃度與突變株和野生型交互作用關係的 p-value

P-value of each fix effect (by mutant lines)			
SALK_000339		main+lateral	
PEG concentration	<b>1.87E-11</b>	PEG concentration	<b>1.82E-10</b>
Genotype	0.0827	Genotype	0.1352
Concentration*Genotype	0.8768	Concentration*Genotype	<b>0.0014</b>
SALK-067858		main+lateral	
PEG concentration	<b>5.65E-11</b>	PEG concentration	<b>1.96E-09</b>
Genotype	<b>0.0045</b>	Genotype	<b>0.0116</b>
Concentration*Genotype	0.482	Concentration*Genotype	0.863
WiscDsLoxHs147_08A		main+lateral	
PEG concentration	<b>1.96E-09</b>	PEG concentration	<b>1.96E-09</b>
Genotype	<b>0.0116</b>	Genotype	<b>0.0116</b>
Concentration*Genotype	0.863	Concentration*Genotype	0.863



圖三十四、 *AT4G32040* 四個突變品系樣本合併後在三種 PEG 濃度下的總根長回歸直線  
 紅色：所有突變株品系總根長對 PEG-6000 濃度之回歸直線  
 藍色：Col-0 總根長對 PEG-6000 濃度之回歸直線

## 陸、討論

### 一、探討阿拉伯芥與伊比利半島氣候梯度的相關基因

我們由前人研究(1001 Genomes Consortium, 2016)得知，伊比利半島同時具有阿拉伯芥所分化出的不同族群——子遺族群與非子遺族群，因此我們基於該地的演化特殊性，利用該地的阿拉伯芥品系之基因資料、環境資料進行全基因體關聯分析，結果顯示 *AT1G58310* 基因與乾旱相關的環境資料有高度相關性、*AT2G40130* 基因在與降雨量相關的環境資料有高度相關性、*AT4G32040* 基因則在與降雨量、乾旱相關的環境資料兼具高度相關性。此些分析結果在同樣也是進行 GWAS 研究阿拉伯芥的文獻(Ferrero-Serran, 2019)中，並未提及的。我們推測是因該文獻中，阿拉伯芥品系基因資料使用了全世界 1,135 個阿拉伯芥品系的基因體資料，而在本研究中，我們先篩選出了伊比利半島的品系基因體資料，以凸顯該地族群結構和歐洲其他區域的差異，致使前人研究並沒有發現本研究找出的 *AT4G32040*、*AT1G58310* 與乾旱有關。此外，本研究也觀察到 *AT2G40130* 與乾旱環境資料有關，此結果與前人研究相符(Li *et al.* 2020)，代表我們的研究具有一定程度的可驗證性。

### 二、探討 GWAS 結果峰值位點與 *AT1G58310*、*AT4G32040* 分化的關聯性

首先，我們觀察到在 *AT1G58310* 的親緣關係樹中，四個 SNP 位點的核苷酸的分布類似，紅色集中於阿拉伯芥所有族群的一小群，但有一些紅色穿插於藍色中，此代表這些位點可能有再次突變或是這個位點在有些個體有遺傳重組的情況。*AT4G32040* 中，四個 SNP 位點的核苷酸分布類似，紅藍也各佔整個阿拉伯芥族群的一半。

接著，我們以 *AT1G58310*、*AT4G32040* 的基因定序資料計算 PCA。結果顯示 *AT1G58310* 在遺傳結構上大致分成五群，*AT4G32040* 大致分成二群。PCA 紅藍點的分布結果與親緣關係樹結果大致吻合：同個基因上四個 SNP 位點的核苷酸具有類似分布。

由 PCA 與親緣關係樹的共同結果，可驗證經由 GWAS 找到的位點與環境高度相關的正確性，且可能與標的基因的分化相關。

### 三、探討根系性狀與標的基因 *AT4G32040* 功能之關聯性

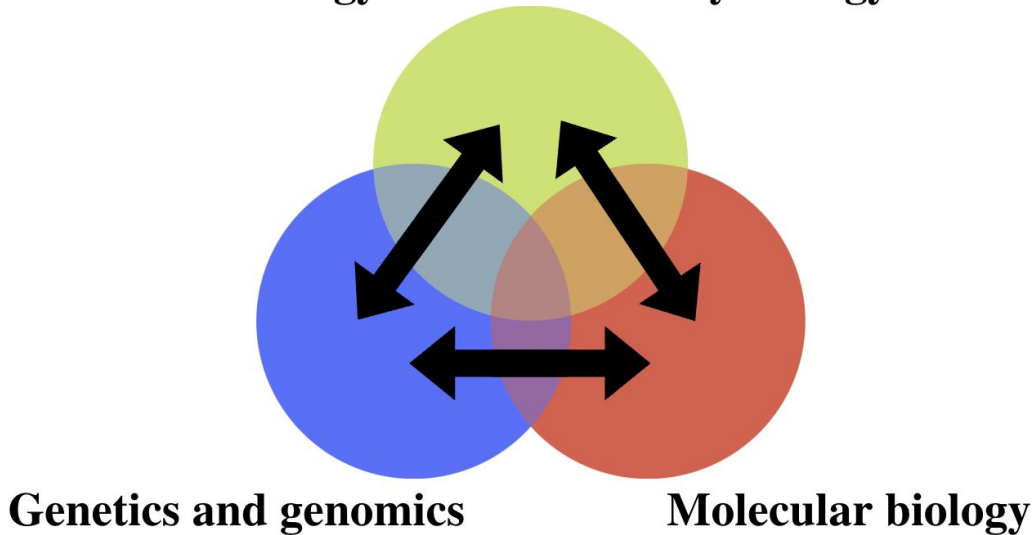
本研究的標的基因 *AT4G32040* (KNAT5)為 KNAT 基因的一支，前人研究(Truernit, 2006)指出，KNAT5 在側根原體(lateral root primordia)的 RNA 表現量顯著較高，且 KNAT5 啟動子在根的延長區(elongation zone)以及分生區(differentiation zone)活性也較高。同樣地，另一研究(Truernit, 2007)也指出，阿拉伯芥的 KNOX class II genes KNAT3, KNAT4 和 KNAT5 透過影響 KNOX II 類蛋白活性，在根系發育過程中發揮不同的作用；推測 *AT4G32040* 在植物根系生長中扮演關鍵角色。而本研究更進一步進行了多項分析及性狀觀察實驗，驗證了 *AT4G32040* 在乾旱逆境下，對於植物根系生長的重大影響。

### 柒、結論與應用

本研究首先藉由全基因體關聯分析(GWAS)，由 188 個環境資料與 1,135 個基因體資料中，篩選出了與降水量、乾旱有顯著相關的 *AT1G58310* 與 *AT4G32040* 基因，並以主成分分析(PCA)以及親緣關係樹，觀察標的基因的分化情形。接下來，我們以分子生物學實驗探討標的基因與植物抗旱能力之關聯，結果顯示標的基因 *AT4G32040* 突變株不論在無逆境根長實驗或是以 PEG-6000 模擬植物乾旱逆境的實驗中，根系都顯著較 Col-0 更長且更廣，得以驗證 *AT4G32040* 對於阿拉伯芥根系生長與抗乾旱逆境之關聯。

本研究運用並探討的領域如圖三十六所示。我們旨在以演化生態學的視野，結合遺傳基因體學進行各項分析(如：GWAS、PCA、與親緣關係樹)，並以分子生物學實驗驗證了標的基因對植物抗旱能力之關鍵影響。期望此研究未來能應用於農作物，使其具備節水、抗旱的能力，以面對全球極端氣候與人口增長所帶來的糧食壓力與危機。

## Ecology and evolutionary biology



圖三十六、本研究涵蓋領域

### 捌、參考文獻

- 一、羅澄玉(2018)。從阿拉伯芥不同族群之性狀差異探討雜草演化。國立臺灣大學生態學與演化生物學研究所。
- 二、Angel, F., & Sarah, M.A. (2019). Phenotypic and genome-wide association with the local environment of *Arabidopsis*. *Nat Ecol Evol* 3, 274–285.
- 三、Atwell, S., Huang, Y., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., Hu, T. T., Jiang, R., Mulyati, N. W., Zhang, Xu., Amer, M.A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., ..., & Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631.
- 四、Lee, C. R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., Alonso-Blanco, C., Weigel, D., & Nordborg, M. (2017). On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nature communications*, 8, 14458.
- 五、1001 Genomes Consortium. (2016). 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166, 481–491.

- 六、Lee, C. R., Lo, C.Y., & Hsu, C.W. (2019). On the postglacial spread of human commensal *Arabidopsis thaliana*: Journey to the East. *New Phytol.* 222, 1447–1457.
- 七、Vashisht D, Hesselink A, Pierik R, Ammerlaan JM, Bailey-Serres J, Visser EJ, Pedersen O, van Zanten M, Vreugdenhil D, Jamar DC, Voeselek LA, & Sasidharan R. (2011) Natural variation of submergence tolerance among *Arabidopsis thaliana* accessions. *New Phytol* 2011;190(2):299-310.
- 八、Harris, I. P. D. J., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—the CRU TS3. 10 Dataset. *International journal of climatology*, 34(3), 623-642.
- 九、Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44, 821–824
- 十、Karger, D. N., Conrad, O., Böhrer, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., ... & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific data*, 4(1), 1-20.
- 十一、Lee, S. C., Mustroph, A., Sasidharan, R., Vashisht, D., Pederson, O., Oosumi, T., Voeselek, L. A. C. J. & Bailey-Serres, J. (2011). Molecular characterization of the submergence response of the *Arabidopsis thaliana* ecotype Columbia. *New Phytologist* 190, 457-471.
- 十二、Bailey-Serres J, & Voeselek LACJ. (2008). Flooding stress: acclimations and genetic diversity. *Annual Review of Plant Biology* 59: 313–339.
- 十三、Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., & Ziese, M. (2011). GPCP full data reanalysis version 6.0 at 0.5: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. *GPCP Data Rep.*, doi, 10, 585.
- 十四、Rudolf, B., Beck, C., Grieser, J., & Schneider, U. (2005). Global precipitation analysis products of the GPCP. *Global Precipitation Climatology Centre Rep*, 8.
- 十五、Beck, C., Grieser, J., & Rudolf, B., (2005) A New Monthly Precipitation Climatology for the Global Land Areas for the Period 1951 to 2000. *Klimastatusbericht 2004*, 181-190, DWD.



- 十六、 Verslues, P. E., & Bray, E. A. (2004). LWR1 and LWR2 are required for osmoregulation and osmotic adjustment in Arabidopsis. *Plant physiology*, 136(1), 2831-2842.
- 十七、 Van der Weele, C. M., Spollen, W. G., Sharp, R. E., & Baskin, T. I. (2000). Growth of Arabidopsis thaliana seedlings under water deficit studied by control of water potential in nutrient-agar media. *Journal of experimental botany*, 51(350), 1555-1562.
- 十八、 Truernit, E., Siemering, K. R., Hodge, S., Grbic, V., & Haseloff, J. (2006). A map of KNAT gene expression in the Arabidopsis root. *Plant molecular biology*, 60(1), 1-20.
- 十九、 Truernit, E., & Haseloff, J. (2007). A role for KNAT class II genes in root development. *Plant signaling & behavior*, 2(1), 10-12.

## 【評語】 060001

一、本研究根據演化生態學的理论作為切入點，利用遺傳基因體學進行系統性的分析。最後並佐以分子生物學實驗進一步證明標的基因與植物抗旱性之關聯，具有相當的新穎性。

二、作者以 GWAS 篩選與降水量、乾早有顯著相關的 AT1G58310 與 AT4G32040 基因，亦做其主成分(PCA)以及親緣關係樹分析。

三、作者更進一步比較 WT 及 knat5 突變株在正常或 PEG6000 處理下的根長，發現 突變株根系都明顯較 WT 更長且更廣，因此認為 KNAT5 具有抗乾旱功能。

四、研究主題及分析方法頗有創意，但仍有問題待釐清。

(1) 突變株根系更長，表示此基因產物存在是不利於根生長及抗旱。在全基因體的分析上，此基因表現是否跟抗旱負相關？

(2) 突變株及 WT 在 PEG6000 處理下，根生長延遲趨勢及比例是否有差別？