



## Article

# 45S rDNA Diversity *In Natura* as One Step towards Ribosomal Heterogeneity in *Arabidopsis thaliana*

Valérie Delorme-Hinoux <sup>1,2,3,4,\*</sup>, Assane Mbodj <sup>1,2,3,5</sup>, Sophie Brando <sup>1,2</sup>, Anne De Bures <sup>1,2</sup>, Christel Llauro <sup>1,2,3</sup>, Fabrice Covato <sup>6</sup>, Joseph Garrigue <sup>6</sup>, Claude Guisset <sup>4,6</sup>, Jacques Borrut <sup>4</sup>, Marie Mirouze <sup>1,2,3,5</sup> , Jean-Philippe Reichheld <sup>1,2</sup> and Julio Sáez-Vásquez <sup>1,2,\*</sup> 

<sup>1</sup> Université de Perpignan Via Domitia (UPVD), Laboratoire Génome et Développement des Plantes (LGDP), UMR 5096, 66860 Perpignan, France

<sup>2</sup> Centre National de la Recherche Scientifique, Laboratoire Génome et Développement des Plantes (LGDP), UMR 5096, 66860 Perpignan, France

<sup>3</sup> EMR LGDP/MANGO, Mechanisms of Adaptation and Genomics, IRD-CNRS-UPVD, 66860 Perpignan, France

<sup>4</sup> Association Charles Flahault, 66350 Toulouges, France

<sup>5</sup> Institut de Recherche pour le Développement (IRD), ECOBIO, 34000 Montpellier, France

<sup>6</sup> FRNC, Fédération des Réserves Naturelles Catalanes, 66500 Prades, France

\* Correspondence: valerie.hinoux@univ-perp.fr (V.D.-H.); saez@univ-perp.fr (J.S.-V.)

**Abstract:** The keystone of ribosome biogenesis is the transcription of 45S rDNA. The *Arabidopsis thaliana* genome contains hundreds of 45S rDNA units; however, they are not all transcribed. Notably, 45S rDNA units contain insertions/deletions revealing the existence of heterogeneous rRNA genes and, likely, heterogeneous ribosomes for rRNAs. In order to obtain an overall picture of 45S rDNA diversity sustaining the synthesis of rRNAs and, subsequently, of ribosomes *in natura*, we took advantage of 320 new occurrences of *Arabidopsis thaliana* as a metapopulation named At66, sampled from 0 to 1900 m of altitude in the eastern Pyrenees in France. We found that the 45S rDNA copy number is very dynamic *in natura* and identified new genotypes for both 5' and 3' External Transcribed Spacers (ETS). Interestingly, the highest 5'ETS genotype diversity is found in altitude while the highest 3'ETS genotype diversity is found at sea level. Structural analysis of 45S rDNA also shows conservation *in natura* of specific 5'ETS and 3'ETS sequences/features required to control rDNA expression and the processing of rRNAs. In conclusion, At66 is a worthwhile natural laboratory, and unraveled 45S rDNA diversity represents an interesting starting material to select subsets for rDNA transcription and alter the rRNA composition of ribosomes both intra- and inter-site.

**Keywords:** rDNA; *Arabidopsis thaliana*; brassicaceae; ribosomal heterogeneity; natural variation



**Citation:** Delorme-Hinoux, V.; Mbodj, A.; Brando, S.; De Bures, A.; Llauro, C.; Covato, F.; Garrigue, J.; Guisset, C.; Borrut, J.; Mirouze, M.; et al. 45S rDNA Diversity *In Natura* as One Step towards Ribosomal Heterogeneity in *Arabidopsis thaliana*. *Plants* **2023**, *12*, 2722. <https://doi.org/10.3390/plants12142722>

Academic Editor: Baohong Zhang

Received: 29 June 2023

Accepted: 13 July 2023

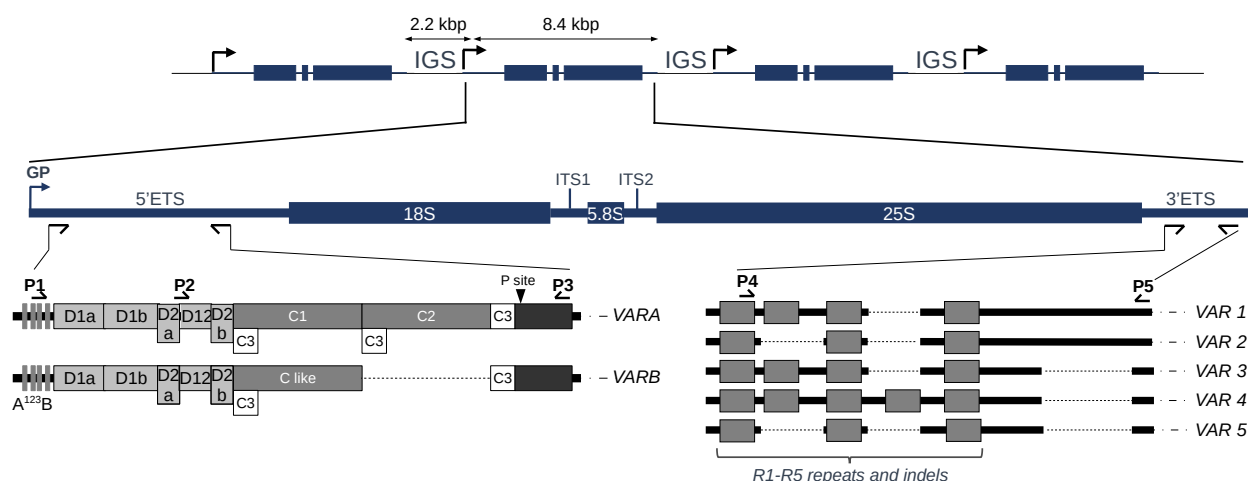
Published: 21 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Ribosomes are responsible for translating the genetic information residing in mRNAs into proteins and are built from four types of ribosomal RNAs (18S, 5.8S, 25S and 5S rRNAs) and approximately 80 ribosomal proteins [1]. Increasing evidence across various model organisms has demonstrated the heterogeneous nature of ribosomes, and *Arabidopsis thaliana* is a highly suitable model to study this phenomenon [2,3]. In *A. thaliana* Col-0 (Figure 1), the 45S rDNA precursor transcript of 8.4 kbp contains structural rRNAs 18S, 5.8S and 25S flanked by External Transcribed Spacers (5'ETS and 3'ETS) separated by Internal Transcribed Spacers (ITS1 and ITS2) [4–7]. In all kingdoms, coding sequences for these three rRNAs are highly conserved, but the sequences that are removed during processing (including the ETS and ITS) show a great extent of variation both in length and organization [4,8,9]. Variation along the 45S rDNA precursor transcript using the available 1135 *A. thaliana* genomes also identified polymorphisms within the highly conserved rRNA genes [10–12], suggesting that ribosome heterogeneity could involve the sequence variation of rRNAs.



**Figure 1.** 45S rDNA organization in the *Arabidopsis thaliana* Col-0 accession. The figure illustrates schematic representations of rDNA units separated by IGSs, 45S rDNA precursor transcript starting at the gene promoter (GP) and 5'ETS (VARA–B) and 3'ETS (VAR1–5) variants. IGS, intergenic spacers; ETS, External Transcribed Spacers; ITS, Internal Transcribed Spacers. The conserved U3 snoRNP binding site A<sup>123</sup>B and the primary processing site P are indicated. Deletions are represented as dotted lines. Primers used for PCR genotyping or variant sequencing are indicated by arrowheads, P1–P3 for 5'ETS and P4–P5 for 3'ETS. The P2 primer was only used for 5'ETS variant sequencing when appropriate.

The *A. thaliana* Col-0 accession counts ~800 copies of 45S rDNA units per haploid genome [12–14]. These copies are localized in tandem arrays in a head-to-tail manner at chromosomal loci known as the Nucleolar Organizer Regions (NOR2 and 4) at the top of chromosomes 2 and 4 [8,12–14]. They are the so-called NORs because transcription and pre-rRNA processing allow the nucleolus to form during the interphase of the cell cycle [15,16]. Both NORs have a similar number of rDNA copies in Col-0 and Ler accessions [14,17] but can also vary in size [12,18], and copy number variation was reported in the range of 500–2500 in Swedish accessions [18,19]. Although rRNA transcripts represent the majority of all transcribed RNAs in a cell, not all 45S rDNA copies are transcribed, and extra copies are silenced by repressive chromatin modifications [10,17,20].

The current Col-0 reference genome assembly based on short reads only contains single rDNA units at the top of chromosomes 2 and 4 [21]. None of the recent assembly approaches performed either with Oxford Nanopore Technologies or PacBio provided contigs with multiple 45S rDNA units in tandem [22,23]. Recently, a combination of long- and short-read sequencing using a BAC-based approach built contigs with seven rDNA units, on average, and the first sequence draft of NOR2 [12]. This study confirmed that NOR2 is organized in distinct rDNA unit clusters, as suggested [14], and the fine-tune transcriptional regulation of rDNA units in a tissue-specific manner. These findings support the concept of specialized ribosome subpopulations that differ in their rRNA composition in order to shape the proteome either in a tissue-specific manner or as a response to environmental stresses.

To discriminate rDNA units and guide the assembly of NOR2, barcode systems included (i) the number of core and spacer promoters, (ii) the length and number of Sall boxes, (iii) the presence of an *Ava*I restriction site in the ITS1 and equal distribution between NORs, (iv) the presence of a specific trinucleotide (CAT) in the ITS2 that strongly enriched on NOR2 and (iv) the length of the 3'ETS [12]. Both 5' and 3'ETS can also guide the classification of rDNA units within one individual (Figure 1).

Comparison of 5'ETS variants between some Brassicaceae species (*Raphanus sativus*, *Brassica oleracea*, *B. rapa*, *B. juncea* and *Arabidopsis thaliana*) revealed the presence of a 1,1 kbp insertion in *Arabidopsis thaliana* [24,25]. This insertion, made of five D and two C repeats, splits away the conserved U3 snoRNP binding site A<sup>123</sup>B and the primary processing site

P [24,25]. Variation at the 5'ETS in the *A. thaliana* Col-0 accession (Figure 1) distinguishes two variants, *VARA* and *VARB*, the first being the major variant [4,9]. Variation at the 3'ETS already distinguishes five variants (*VAR1-5*) based on R1-R5 repeats and indels (Figure 1). Three of these rRNA variants are abundant, *VAR1*, 2 and 3, accounting for ~50%, ~30% and ~20%, respectively. Meanwhile, *VAR4* is relatively rare (1%), and *VAR5* is undetectable by PCR [9,20,26]. Col-0 is one of the most complex out of the 23 ecotypes studied among the 1135 worldwide accessions, which display only one or two variants [17,26]. Accessions bearing only *VAR1* copies, such as Bur-0, are rare worldwide [17,26]. Altogether, these results and observations prompted us to further investigate 45S rDNA organization and, more precisely, variation at the 5' and 3'ETS in *Arabidopsis thaliana* individuals harvested in the eastern Pyrenees in the Occitanie region of France as elevation ranges from sea level up to almost 3000 m of altitude over distances of less than 100 km, displaying a succession of diverse Mediterranean and mountainous habitats. We found major 45S rDNA variations that can be correlated with geographical distribution.

## 2. Results

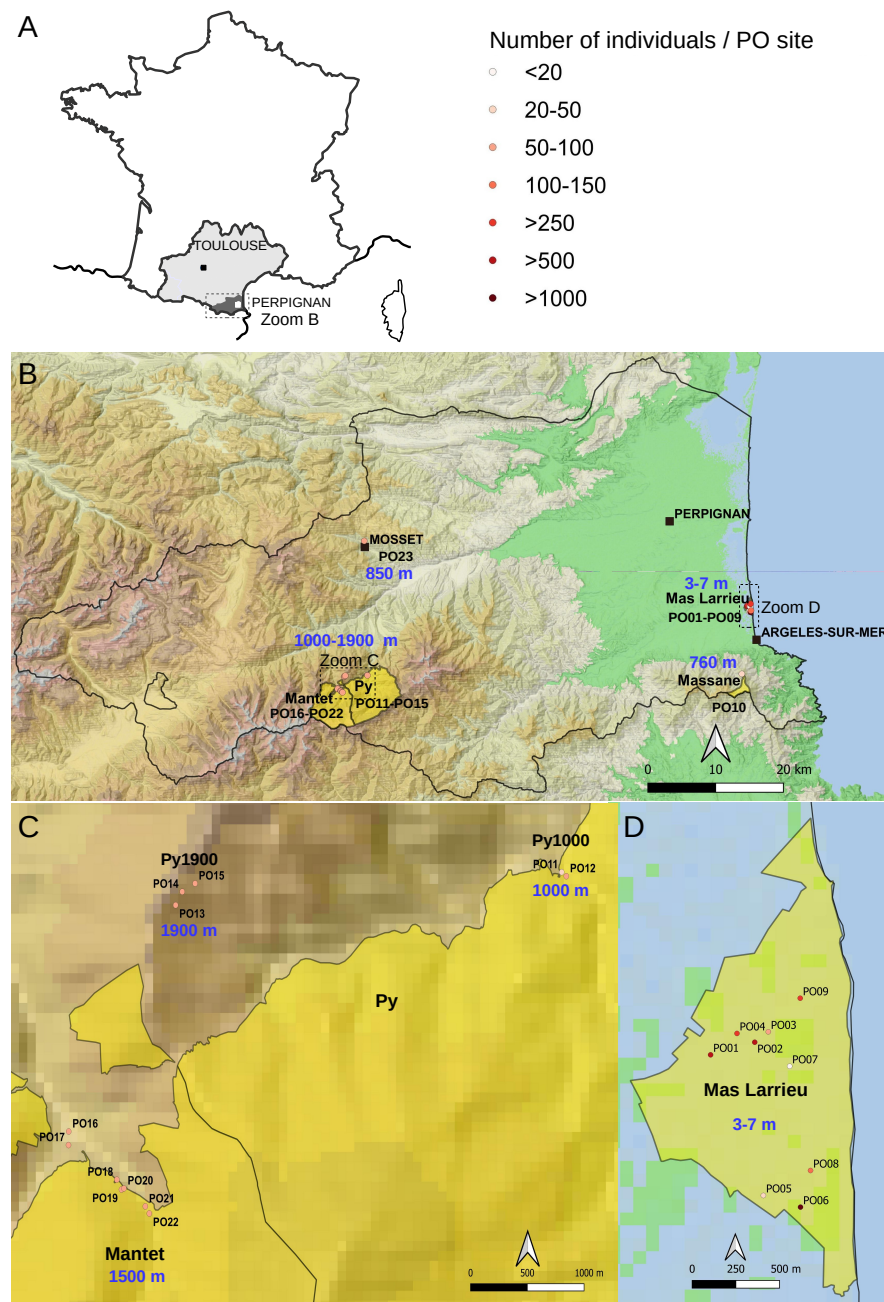
### 2.1. The At66 Metapopulation

The department of Pyrénées-Orientales in the Occitanie region of France has a very interesting topography from the Mediterranean Sea to the Pyrenees, and no natural population of *A. thaliana* was previously sampled there. In this study, we report natural occurrences as a metapopulation named At66 (for *Arabidopsis thaliana* 66, this number being the reference for the Pyrénées-Orientales department (PO) in the French administrative system). At66 ranges from 0 to 1900 m of altitude at a micro-geographic scale as the sites are only distant by several hundreds of meters to less than one hundred kilometers. At66 combines a diversity of habitats and environmental conditions and completes the existing natural populations of the Pyrenees in France [27–29] and across the border between France and Spain [30–32]. At66 consists of a total of 320 individuals harvested from six localities (Mas Larrieu, Massane, Mosset, Py1000, Mantet and Py1900) subdivided into 23 sampling sites (named PO01-23) (Figure 2). The number of sites differs between localities as there is either only one site (Massane and Mosset) or several sites (nine in Mas Larrieu, two at Py1000, seven in Mantet, three at Py1900) per locality. This study took advantage of this new resource *in natura* to genotype copy number variation and variation at the 5' and 3'ETS of 45S rDNA units.

### 2.2. Copy Number Variation in At66

The Col-0 reference genome of *Arabidopsis thaliana* counts ~800 copies of 45S rDNA loci per haploid genome [12–14]. The genome size of worldwide accessions varies by well over 10% [19,33], and copy number variation in 45S rRNA genes has been identified as one of the contributors to this genome size variation [18,19,34,35]. Copy number variation *in natura* was only reported in Sweden, with northern accessions having a particularly large number of copies (~2500 copies for TRÄ-01) compared to southern accessions (~500 for Ale-Stenar-64-24(1002) [18,19]. We calculated rDNA copy number variation (CNV) in At66, estimating the mean read depth of both 18S and 25S rRNA genes divided by the estimate of the mean genome-wide read-depth, selecting the first 10 Mb of chromosome 3 as it displayed a uniform read-depth [11]. Histograms illustrate the frequency distribution of both 18S and 25S CNVs (Figure 3A), whereas box and violin plots depict distributions associated with a statistics summary (Figure 3B and Table S1). A Wilcoxon Test confirmed a true shift between the two distributions ( $W = 12,876$ ,  $p$ -value = 0.0248), and we think that this bias is probably related to the coverage method of the estimation of CNVs knowing that the 25S rRNA gene is longer than the 18S rRNA gene. We found an average of ~600 copies and registered only a few extreme individuals having either a small number of copies (~300) or doubling the average number of copies (~1200). Interestingly, we registered intra-site and inter-site copy number variation but no significant variation along the altitudinal gradient (Figures 3C and S1). This result suggests that individuals can deal with different

numbers of rDNA units to face contrasting environmental conditions and that copy number variation can also distinguish individuals living in the same environmental conditions.



**Figure 2.** Map of the At66 metapopulation. (A) Localization of the Pyrénées-Orientales department (PO and 66) in the Occitanie region of France where the 23 sites, namely PO01-23, were sampled. The Perpignan prefecture is located 200 km from Toulouse. The dotted rectangle indicates the area of magnification shown in (B). (B) Magnification of the department showing the localization of National Nature Reserve (NNR) in yellow and the focus on two localities, PO10 (Massane) and PO23 (Mosset). The border of the Pyrénées-Orientales department (PO and 66) is drawn with a black line. The altitude of each site is indicated in blue. The dotted rectangles indicate the areas of magnification in the mountains (C) and at sea level (D). (C) Magnification of the National Nature Reserve of Py and Mantet in the Pyrenees from 1000 to 1900 m of altitude, focusing on three localities: Py1000 (PO11-PO12), Py1900 (PO13-PO15) and Mantet (PO16-PO22). (D) Magnification of the National Nature Reserve of Mas Larrieu at sea level (PO01-PO09). The sites were georeferenced using QGIS

software 3.10 (<https://www.qgis.org/>, accessed on 12 July 2023), including the base map SRTM, © Jet Propulsion Laboratory 2010. The circle symbol, used to georeference the sites, has a color category as a range indicating the number of individuals per PO site. The scale and north orientation of all maps are specified.



**Figure 3.** Copy number variation in the At66 metapopulation. (A) Frequency histogram for both 18S and 25S CNV estimations. The vertical dashed line indicates the number of copies already known

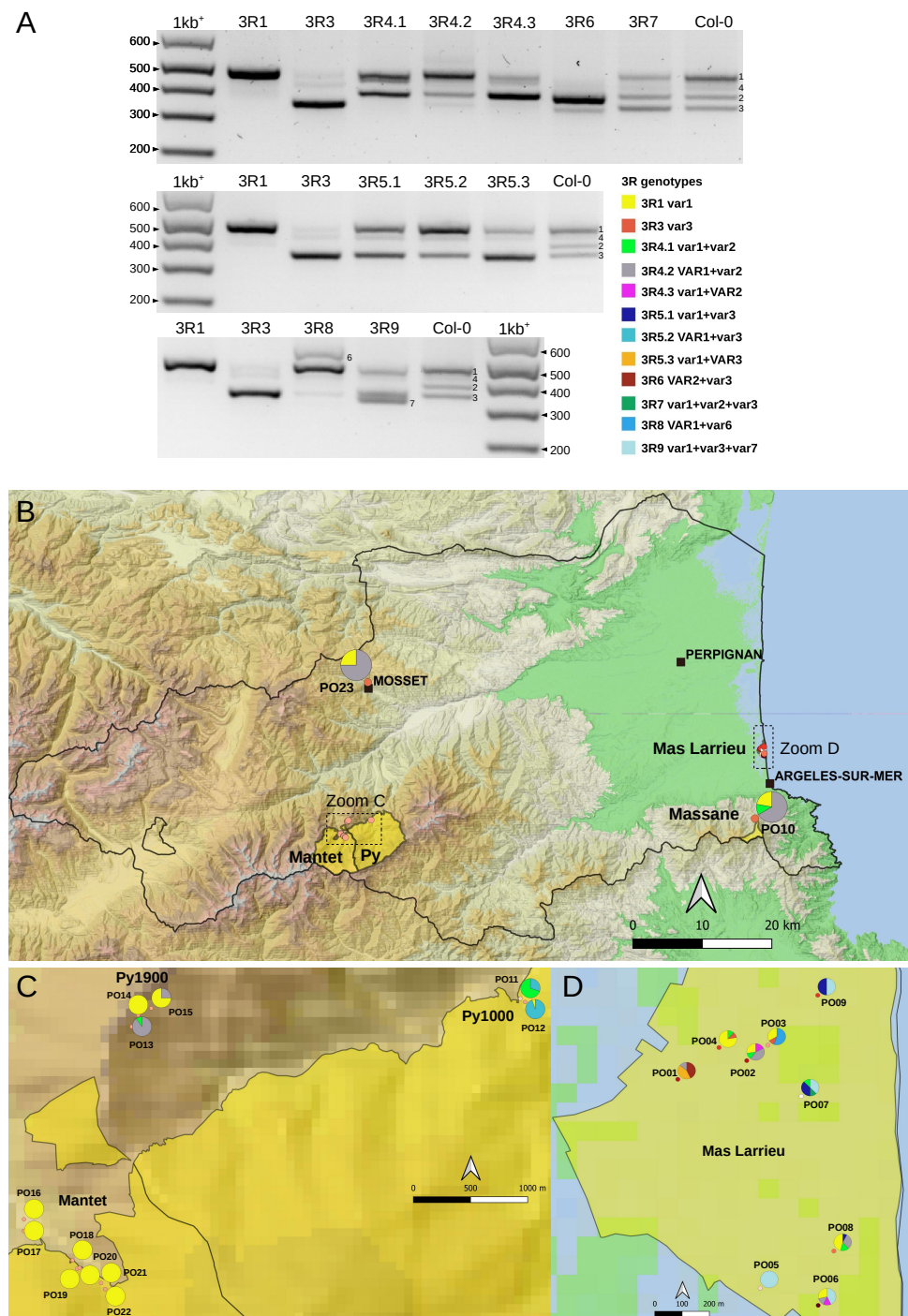
for Col-0 (~800). (B) Box and violin plots for both 18S and 25S CNV estimations. Both statistic summaries of the data distributions and results of the Shapiro-Wilk normality test are indicated in the table. The  $p$ -value is lower than 5%, so we can reject the null hypothesis of the normality of the dataset. (C) 18S copy number variation in At66. Each bar of the bar plot represents one individual of At66, and the color code indicates the PO site of origin. Individuals are ordered along the altitudinal gradient from sea level to altitude, and the six localities are labeled below the x axis. The horizontal red dashed line indicates the number of copies already known for Col-0 (~800).

### 2.3. Variation at the 3'ETS in At66

Small insertions or deletions in the 3'ETS distinguish up to five 45S rDNA variants (*VAR1-5*) in the *A. thaliana* Col-0 accession [9,20,26]. Four of them (*VAR1-4*) are detectable by PCR (Figures 1 and 4A, see Col-0 control lane), and variation at the 3'ETS was previously used to show that nine natural Pyrenean populations were homogeneous [28]. Here, we studied natural variation at the 3'ETS in At66 and identified 12 different 45S rDNA genotypes at the 3'ETS (3R) among 320 individuals (Figure 4A). *VAR4* was not taken into account to define 3R genotypes due to its variability and low frequency. We subdivided 3R4 and 3R5 into three distinct 3R genotypes, 3R4.1-3 and 3R5.1-3, considering the relative abundance of *VAR1-3* copies. For example, we distinguished 3R4.1, 3R4.2 and 3R4.3 to indicate that the abundance of *VAR1* and *VAR2* were equivalent in 3R4.1 and that *VAR1* or *VAR2* were the most abundant in 3R4.2 or 3R4.3, respectively. We identified individuals bearing either only *VAR1* or *VAR3* copies but not only *VAR2* copies (no potential 3R2 genotype). Nearly half of the individuals of At66 (48.1%) have the 3R1 genotype with only *VAR1* copies. The other 3R genotypes, being in the majority, are 3R5.2 (17.2%) and 3R4.2 (15.3%). The minor represented 3R genotypes are 3R4.1 (5.9%), 3R9 (4.7%), 3R4.3 (2.5%), 3R5.1 (1.9%) and 3R8 (1.6%). The 3R3, 3R5.3, 3R6 and 3R7 genotypes are less represented and considered rare (<1%). Pie charts illustrate the proportion of these 3R genotypes for each site georeferenced on the map of the department of Pyrénées-Orientales (Figure 4B–D and Table S2). There is nearly three times more 3R genotypes diversity at sea level (eleven in Mas Larrieu, and only R5.2 is missing out of twelve genotypes) than at altitude (only four in Massane, Mosset, Py and Mantet). The 3R1 genotype, with *VAR1* copies only and present in 154 among 320 individuals, was found all along the altitudinal gradient (0–1900 m). Interestingly, we identified two new 3'ETS rDNA variants in the 3R8 (*VAR6*) and 3R9 (*VAR7*) genotypes (Figure 4A).

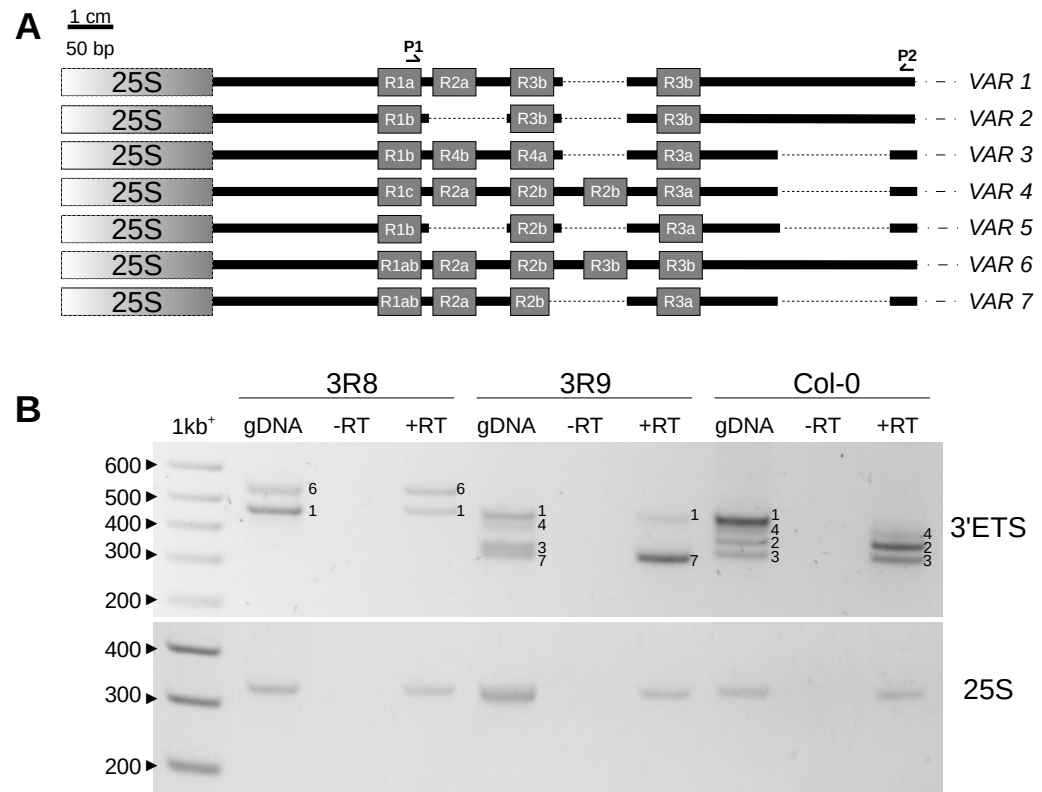
Compared to already known *VAR1-5*, *VAR6* is the longest variant, with five R repeats like *VAR4* but not the deletion shared by *VAR3-5* and *VAR7*. Meanwhile, *VAR7* is similar to *VAR3* with an additional deletion right after the R2b repeat, making it slightly smaller than *VAR3* (schematic representation of 3'ETS variants in Figure 5A, FASTA sequences and multiple alignments in Figure S2). Since only *VAR2-4* copies are expressed in adult *A. thaliana* Col-0 plants, whereas *VAR1* is silenced [20,36], we investigated whether the newly identified *VAR6-7* variants could contribute to the functional subset of 45S rDNA copies (Figure 5B). Interestingly, in the 3R8 genotype (Figure 4A), *VAR6* is much less abundant compared to *VAR1*; however, a similar accumulation of *VAR1* and *VAR6* transcripts is detected (Figure 5B). In contrast, in the 3R9 genotype with a similar abundance of rDNA *VAR1*, 3 and 7 (Figure 4A), mainly transcripts corresponding to *VAR7* are detected (Figure 5B).

Altogether, this analysis reports the geographical distribution of 12 different 3'ETS 45S rDNA genotypes containing previously reported *VAR1-4* and/or novel *VAR6-7* variants that are transcriptionally active. The diversity of genotypes in At66 is three times higher at sea level compared to altitude (760–1900 m). Individuals bearing only *VAR1* copies are more frequent than expected from worldwide studies, as we reported a 50/50 chance of receiving this genotype in At66 all along the altitudinal gradient. This result suggests that individuals can cope with little rDNA diversity at the 3'ETS to face contrasting environmental conditions.



**Figure 4.** 3'ETS variation in the At66 metapopulation. (A) PCR genotyping illustrating the twelve 3R genotypes identified among individuals of At66. There are three panels (high, medium and low), and each panel includes single-variant genotypes (3R1 and 3R3 in the first two lanes) and Col-0 (last lane) as controls to ensure the accuracy of variant assignment. "3R" is a shortcut for "3'ETS 45S rDNA". The capital VAR letters versus the lowercase var letters in the 3R genotypes indicate which variant is the most abundant. When both variants have lowercase letters, this means that they are equally abundant. Lowercase var letters are also used when there is only one variant present. Sizes in base pairs are labeled alongside the lane of the 1kb<sup>+</sup> DNA ladder. (B–D) Map of 3'ETS variation in At66 metapopulation and pie charts recording the proportion of each 3R genotype in each PO site. (B) The magnification of the Pyrénées-Orientales department shows the results of PO10 (National Nature Reserve of Massane) and PO23 (Mosset). The dotted rectangles indicate the areas of magnification in the mountains (C) and at sea level (D). (C) Magnification of the National Nature

Reserve of Py and Mantet in the Pyrenees from 1000 to 1900 m of altitude (PO11-22). (D) Magnification of the National Nature Reserve of Mas Larrieu at sea level (PO01-PO09). The sites were georeferenced using QGIS software 3.28 (<https://www.qgis.org/>, accessed on 21 October 2022) and including the base map SRTM, © Jet Propulsion Laboratory 2010. The color categories for each 3R genotype are indicated. The scale and north orientation of all maps are specified.



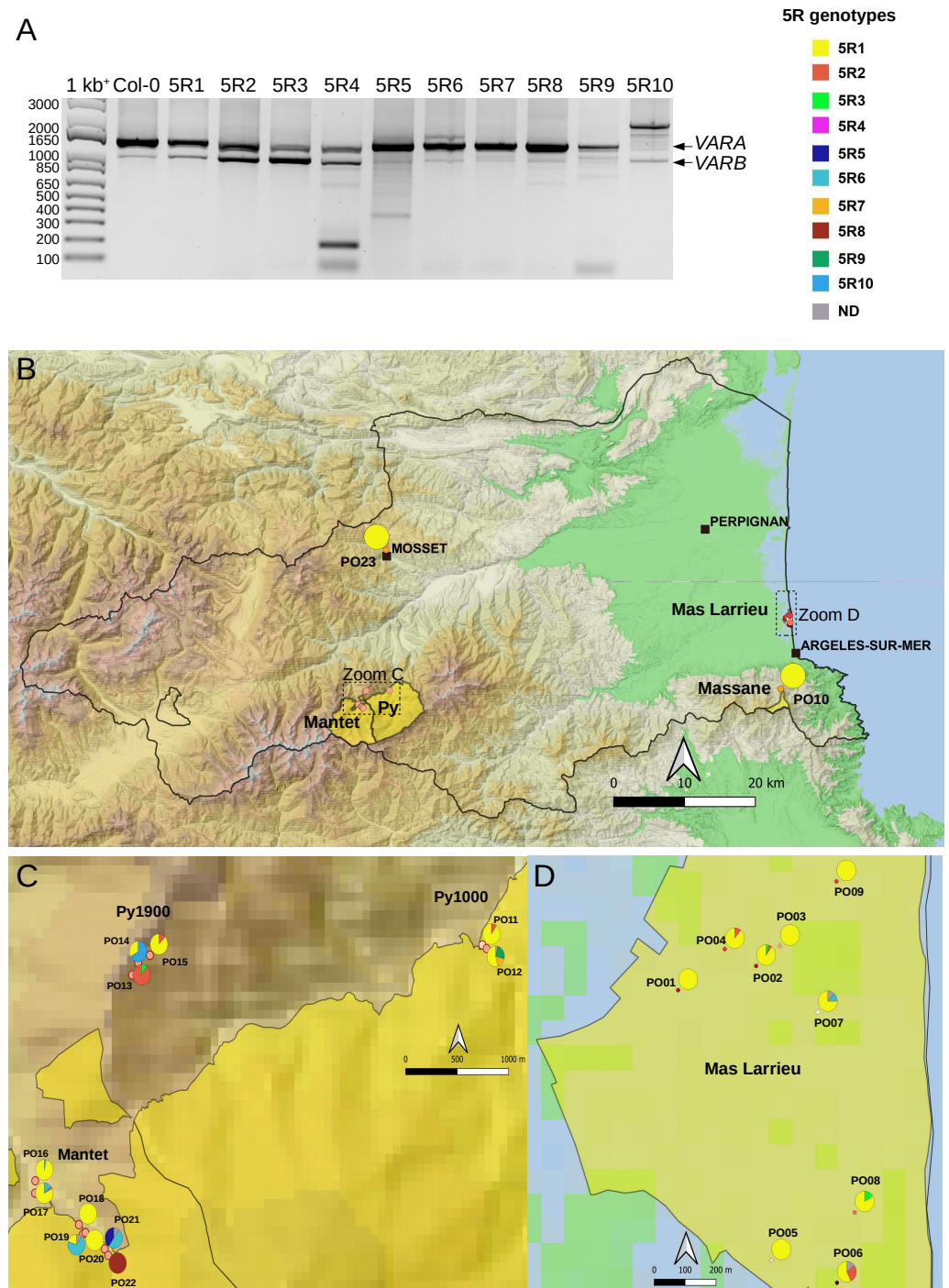
**Figure 5.** New 3'ETS variants identified in the At66 metapopulation. (A) Structure of new VAR6 and VAR7 variants compared to known VAR1-5 variants described in Col-0. (B) RT-PCR analysis carried out for 3R8 and 3R9 genotypes harboring new VAR6 and VAR7 variants compared to VAR1-4 variants known in Col-0. gDNA corresponds to the 3'ETS genotyping, -RT is the control proving that there are no genomic DNA contaminants left in the total RNA sample before reverse transcription, and +RT is the cDNA sample. RT-PCR amplification was performed for 3'ETS and 25S rRNA genes as an internal control of equal cDNA loading. Sizes in base pairs are labeled alongside the lane of the 1kb<sup>+</sup> DNA ladder.

#### 2.4. Variation at the 5'ETS in At66

Processing of the 45S pre-rRNA is a keystone of ribosome biogenesis, and its initial endonucleolytic cleavage in *A. thaliana* is located at the P site in the 5'ETS (Figure 1) [25]. Since the primary cleavage at the P site is impaired upon heat stress, highlighting a functional role of the 5'ETS [37], we decided to study natural variation at the 5'ETS in At66 (Figure 6A–D). Only one 5'ETS genotype was reported to date in *A. thaliana* with two variants, VARA and VARB, VARA being the major variant (Figure 1) [4,9]. The majority of the individuals (71.6%) have only this genotype recorded so far in Col-0 and renamed here 5R1 (Figure 6A and Table S2). However, one-third of the individuals of At66 have a distinct 5'ETS, and we identified nine additional genotypes at the 5'ETS (5R2-10) (Figure 6A and Table S2), eight of which are minorly represented (1.3–5.6%) and one is considered rare (<1%). VARA is often the major variant but can also be as equally present as VARB (5R2, 5R4) or rare (5R10). VARB can be the major variant (5R3). Noteworthily, we identified only one individual among 320 having a 200 bp variant equally present as both VARA



and *VARB* (5R4). We also identified intermediate or longer variants (5R4–5R9), suggesting that length variation of the 5'ETS could be affected between individuals and within rDNA copies of one individual. Pie charts illustrate the proportion of these 5R genotypes for each site georeferenced on the map of the department of the Pyrénées-Orientales (Figure 6B–D and Table S2). There is 2.5 times more 5R genotypes diversity at altitude (ten in Massane, Mosset, Py and Mantet) than at sea level (four in Mas Larrieu). The 5R1 genotype bearing *VARA* and *VARB*, *VARA* being the major variant and present in 229 among 320 individuals, was found all along the altitudinal gradient (0–1900 m).



**Figure 6.** 5'ETS variation in the *At66* metapopulation. (A) PCR genotyping illustrating the ten 5R genotypes identified among individuals of *At66*. Only one 5'ETS genotype was reported for Col-0

with two variants *VARA* and *VARB*, *VARA* being the major one (lane 1) and renamed here 5R1 (lane 2). We identified nine additional 5R genotypes named 5R2-R10 (lanes 3–11), “5R” being a shortcut for “5’ETS 45S rDNA”. Sizes in base pairs are labeled alongside the lane of the 1kb<sup>+</sup> DNA ladder. (B–D) Map of 5’ETS variation in the At66 metapopulation and pie charts recording the proportion of each 5R genotype in each PO site. (B) The magnification of the Pyrénées-Orientales department shows the results of PO10 (National Nature Reserve of Massane) and PO23 (Mosset). The dotted rectangles indicate the areas of magnification in the mountains (C), with the National Nature Reserve of Py and Mantet in the Pyrenees from 1000 to 1900 m of altitude (PO11-22) and at sea level (D) with the National Nature Reserve of Mas Larrieu (PO01-PO09). The sites were georeferenced using QGIS software 3.10 (<https://www.qgis.org/>, accessed on 12 July 2023) and including the base map SRTM, © Jet Propulsion Laboratory 2010. The color categories for each 5R genotype are indicated. The scale and north orientation of all maps are specified.

Altogether, this analysis reports the geographical distribution of 10 different 5’ETS 45S rDNA genotypes, confirming that the only previously reported 5R1 genotype is present in the majority, as we reported a 70/30 chance of receiving this genotype in At66 all along the altitudinal gradient; however, it still identifies nine additional genotypes sustaining ~30% of 5’ETS variation that is unknown so far. This result suggests that individuals can cope with little rDNA diversity at the 5’ETS to face contrasting environmental conditions.

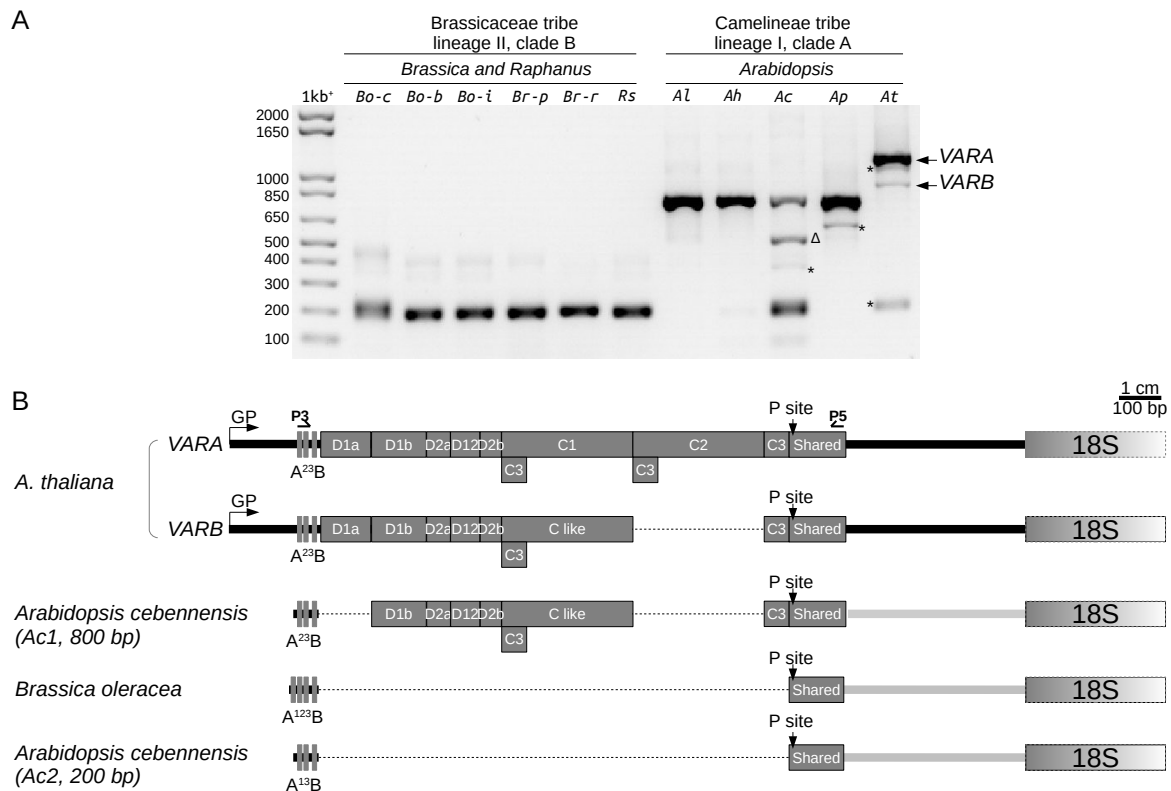
### 2.5. 5’ETS Length Variation in Brassicaceae Species

Since we identified length variation of the 5’ETS between At66 individuals and within rDNA copies of a single individual in At66, we further investigated 5’ETS structural variation in Brassicaceae species. The 5’ETS of the *Arabidopsis thaliana* Col-0 accession is longer compared to Brassicaceae species (*Raphanus sativus*, *Brassica oleracea*, *B. rapa* and *B. juncea*) because a 1.1 kbp insertion made of five D and two C repeats splits away the conserved U3 snoRNP binding site A<sup>123</sup>B and primary processing site P [24,25]. We wondered whether this 1.1 kbp insertion was conserved in the *Arabidopsis* genus, and to answer this question, we combined both 5’ETS genotyping (Figure 7A) and the sequencing of 1–3 independent clones per major variant detected (schematic representation of 5’ETS variants in Figure 7B and FASTA sequences and multiple alignments in Figure S3).

5’ETS PCR genotyping of some species belonging to the Brassicaceae tribe, lineage II, clade B (*Brassica oleracea* var *capitata*, *botrytis* and *italica*, *Brassica rapa* ssp *pekinensis* and *rapa* and *Raphanus sativus*) identified a 200 bp variant (Figure 7A), confirming the absence of the 1.1 kbp insertion [24,25]. In contrast, the *Arabidopsis thaliana* Col-0 has two longer variants (Figure 7A,B), *VARA* and *VARB* of 1260 and 950 kbp, respectively, *VARA* being the major variant [4,9]. *VARA* and *VARB* share the same five D repeats (D1a, D1b, D2a, D12 and D2b) but differ from each other by the deletion of one C repeat of 310 bp long [4,9]. The C repeat left in *VARB* was designed C-like as it likely results from a combination of *VARA* C1 and C2 repeats rather than the selective deletion of either C1 or C2 [9]. All C-repeats (C1, C2 and C-like) begin by a 59 bp C3 repeat [4]. There is an additional third occurrence of the C3 repeat following the C2 or C-like repeat [4]. An additional minor variant slightly smaller than *VARA* was frequently detected while genotyping *A. thaliana* (Figure 7A) but we could not obtain any clone for identification.

5’ETS PCR genotyping in the *Arabidopsis* genus (Camelinae tribe, lineage I, clade A) identified one variant of ~800 bp present in all species (*A. lyrata*, *halleri*, *cebennensis* and *pedemontana*) and a 200 bp variant only identified in *A. cebennensis* and *A. thaliana* (Figure 7A). We were unsuccessful in cloning minor intermediate variants visible in *A. cebennensis* and *A. pedemontana* but found that the major intermediate variant of 550 bp in *A. cebennensis* resulted from an aspecific amplification (Figure 7A). Interestingly, genotyping *A. cebennensis* revealed that a ~800 bp variant similar to the *A. thaliana* *VARB* without the D1a repeat and a 200 bp variant without an insertion like in *Brassica* species can co-exist within a unique individual (Figure 7B). The minor 200 bp variant found in the *A. thaliana* Col-0 accession was only seen once, and we were not successful in amplifying it again

while genotyping hundreds of other Col-0 plants. This result suggests that rDNA units without the 1.1 kbp insertion might also exist in *A. thaliana* but they are rare and frequently below the threshold of detection by genotyping. Noteworthy, only one individual among 320 (5R4 genotype) in the At66 metapopulation has such a 200 bp insertion-free variant equally present as both VARA and VARB (Figure 6A).

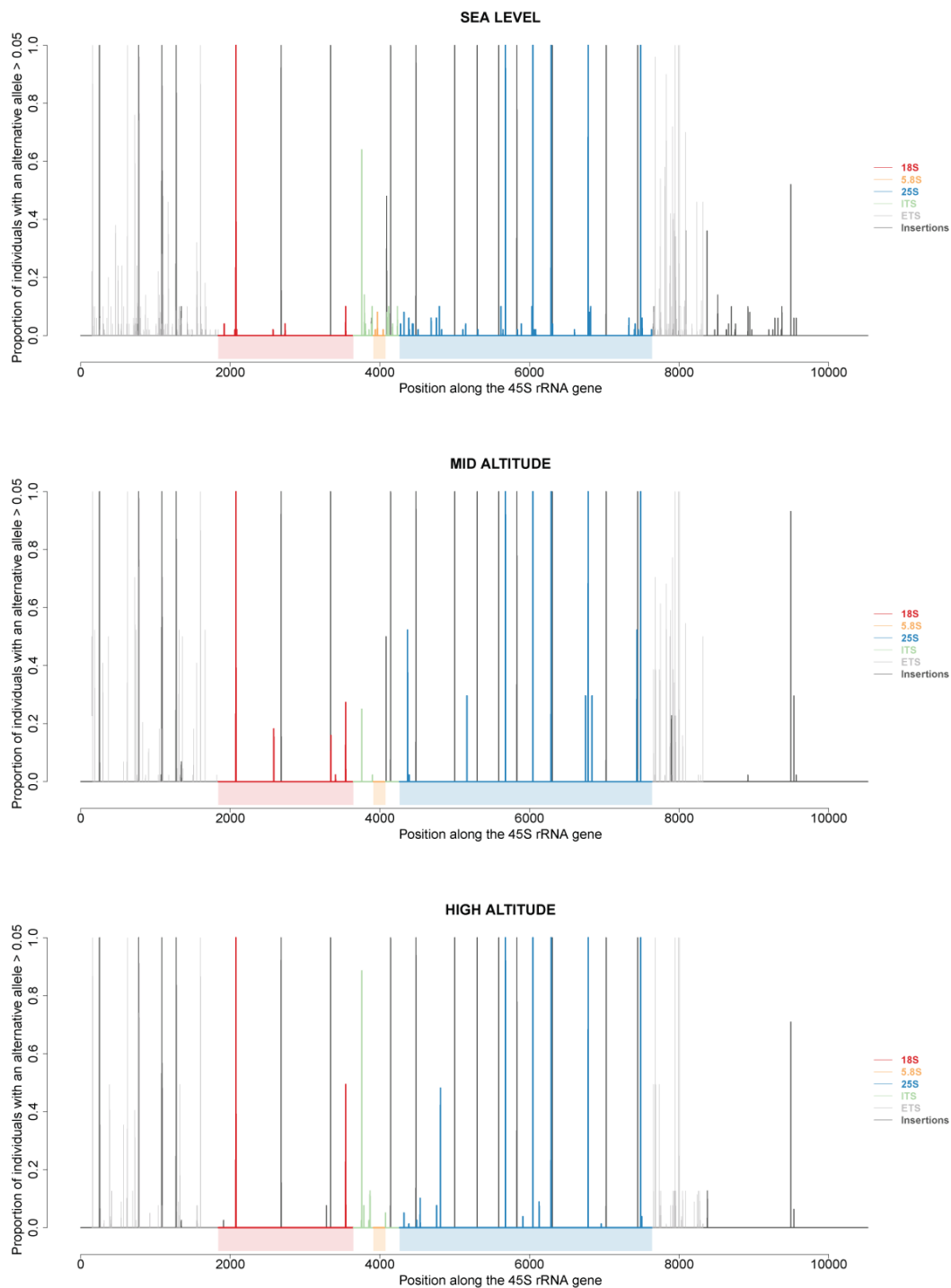


**Figure 7.** 5'ETS variation in Brassicaceae species. (A) 5'ETS genotyping in the genus *Arabidopsis* (Camelineae tribe, lineage I, clade A) compared to *Brassica* and *Raphanus* (Brassicaceae tribe, lineage II, clade B). Bo-c, *Brassica oleracea* var *capitata*; Bo-b, *Brassica oleracea* var *botrytis*; Bo-i, *Brassica oleracea* var *italica*; Br-p, *Brassica rapa* ssp *pekinensis*; Br-r, *Brassica rapa* ssp *rapa*; Rs, *Raphanus sativus*; Al, *Arabidopsis lyrata*; Ah, *Arabidopsis halleri*; Ac, *Arabidopsis cebennensis*; Ap, *Arabidopsis pedemontana*; At, *Arabidopsis thaliana*. \* Uncloned, Δ, aspecific amplified PCR products. Sizes in base pairs are labeled alongside the lane of the 1kb<sup>+</sup> DNA ladder. (B) Schematic representation of *A. thaliana* Col-0 5'ETS variants (VARA and VARB) compared to the 200 bp variant of *Brassica oleracea* without the 1.1 kbp insertion made of five D and three C repeats; splitting away the conserved U3 snoRNP binding site A<sup>123</sup>B from the primary processing site P. *Arabidopsis cebennensis* has both a ~800 bp variant similar to the *A. thaliana* VARB without the D1a repeat and the 200 bp insertion-free variant like *Brassica* species. Dotted lines indicate deletions.

## 2.6. 45S rRNA Gene Polymorphisms in At66

Variations at the 5' and 3'ETS are not directly involved in ribosome heterogeneity since these sequences are removed during the processing of 45S pre-rRNA into mature 18S, 5.8S and 25S rRNAs. However, they clearly helped distinguish different rDNA units within and between individuals of At66, and we wondered whether these variations might be associated with polymorphisms present in specific 18S, 5.8S and 25S rRNA sequences. Using recently developed bioinformatic methods [11], we retrieved the sum of frequencies of all alternative alleles for each position along a ~10.5 kb 45S rDNA unit for each of 173 among 320 individuals of At66. Then, we plotted the proportion of individuals of At66 with an alternative allele for each position along this rDNA unit by studying three cases: firstly, comparing sea level versus mid and high altitude (Figure 8); secondly, by selecting

either the 115 individuals out of 173 sharing the same 5R1 genotype or the 92 individuals out of 173 sharing the same 3R1 genotype (Figure S4A); and thirdly, comparing the six localities (Figure S4B). In the first case studied, true polymorphisms present in 18S, 5.8S and 25S rRNA sequences clearly vary along the altitudinal gradient, and polymorphisms are detected only at sea level for the 5.8S rRNA sequence (Figure 8).



**Figure 8.** 45S rRNA gene polymorphisms at sea level and mid and high altitude. Proportion of individuals of *At66* with an alternative allele (defined as >5% within an individual) along the 45S rRNA gene. Vertical lines represent single nucleotide polymorphisms or deletions in the ETS (gray), 18S (red), ITS (green), 5.8S (orange) and 25S (blue) regions along the 45S rRNA gene. Black vertical lines represent insertions.

In the second case studied, the proportion of individuals sharing alternate alleles vary when fixing either the 5'ETS (same 5R1 genotype and 3'ETS variation) or the 3'ETS (5'ETS variation but same 3R1 genotype). There are additional alternative alleles when fixing the 5'ETS, minorly in 18S and 5.8S rRNA sequences and mostly in the 25S rRNA sequence (Figure S4A). In the third case studied, there is clearly a variation of true polymorphisms within coding (18S, 5.8S and 25S rRNAs) and noncoding ITS sequences between localities. Polymorphisms within the 5.8S rRNA sequence are only detected at Mas Larrieu (sea level). Polymorphisms detected within 18S and 25S rRNA sequences are either specific to one locality (e.g., Massane and 25S) or the proportion of individuals sharing an alternate allele vary between localities (Figure S4B). The most polymorphisms within ITS1 and 2 are detected at Mas Larrieu (sea level) and Mantet (1500 m), whereas there are none at Massane or only a few ITS1 polymorphisms at Mosset, Py1000 and Py1900. Altogether, true SNP polymorphisms detected within coding rRNA sequences in individuals of the At66 metapopulation vary along the altitudinal gradient, between localities and probably depending on variants present at the 5' and 3'ETS.

### 3. Discussion

*In natura*, the number of 45S rDNA loci was previously reported in the range of 500–2500 in Swedish accessions [18,19], and we reported here a 300–1200 copy number variation in At66. We registered both intra-site and inter-site copy number variation along a sharp altitudinal gradient (0–1900 m) and detected 2–2.5-fold changes between individuals within the same site. Interestingly, ten individuals harvested in the PO10 site (Massane locality), namely PO10A–J, are in the range of 320–820 copies, with three individuals having either the upper (PO10H) or the lower number of copies (PO10A and G) and seven individuals in the range of an intermediate number of copies. The estimated kinship coefficient between each pair of these individuals is 0.4, suggesting that this copy number variation is observed among offspring. Moreover, among ten individuals harvested in the PO12 site (Py1000 locality), only one individual has 840 copies (PO12O and the least related one with a negative estimated kinship coefficient in pair comparisons), while the remaining nineteen individuals have 430–510 copies. These two concrete examples from this study suggest that rDNA copy number variation is very dynamic *in natura* and apparently independent of environmental conditions. The first attempt to map copy number variation as a phenotype by means of genome-wide association studies (GWAS) using Swedish accessions did not identify either of the two NORs in a cis-acting locus at chromosome 1 but did in a trans-acting one [19]. Our attempt to map the variation in At66 by means of GWAS conducted by GAPIT version 3 [38] using BLINK [39] or FarmCPU [40] multi-locus models only identified a few SNP outliers shared by both models and both 18S and 25S CNVs, but still apparently not relevant either in cis or trans (Figure S5). Thus, it would confirm the previous statement that the trait would behave like a genotype rather than a phenotype. Indeed, mapping 45S rDNA copy number variation using F2 crosses, recombinant inbred lines, the multiparent advanced generation inter-cross population and mutation accumulation lines showed that both NORs vary in size and that rDNA copy number variation can either behave like a genetic trait heritable in pedigrees or like an unstably inherited trait as it starts to diverge over a timescale of tens of generations [18]. There are two potential mechanisms to promote the dynamics of rDNA loci, recombination and transposition [41]. However, intragenomic fluctuation in rDNA copy number would avoid homologous recombination between NORs on distinct chromosomes, favoring the repair of DNA Strand Breaks by non-homologous end joining (NHEJ), maintaining rDNA integrity during meiosis [42,43]. Therefore, as NORs and neighboring regions are frequent targets for transposable element insertions, transposition would be a good driver to sustain any variation [41,44]. A reduction of the number of 45S rDNA units either in loss-of-function mutants [45,46] or CRISPR-Cas9-induced Double Strand Breaks (DSB) at rDNA loci [43,47] demonstrated that rDNA stability is vital for genome stability. Indeed, affecting the integrity of rDNA loci in loss-of-function mutants [46] or CRISPR-Cas9-induced Double Strand Breaks at rDNA

loci [47] give rise to genome rearrangements such as tandem duplications in direct orientation [46] or chromosome-segment duplication [47]. CRISPR-Cas9-inducing DSB at rDNA loci can trigger seedling lethality or organ/tissue ablation [43] and compromise both ovule fertilization and embryo development [47,48]. Whether impacts on plant development would be direct consequences of modifying rDNA loci number or indirect consequences modifying the integrity of the genome remains to be unraveled. We reported the lowest number of rDNA copies *in natura* so far (300 copies) but the genome seems well-buffered against an artificial loss of rDNA copies occurring randomly at both NORs. Indeed, as little as 10% of the wild-type 800 copies per haploid genome are sufficient for viability in the *A. thaliana* Col-0 accession [47].

Although rRNA transcripts represent the majority of all transcribed RNAs in a eukaryotic cell, only a fraction of available 45S rDNA loci are transcribed [10,20]. Both NORs are transcribed at the onset of meiosis [42], during embryogenesis and the first few days after germination [20,36,49]. In adult somatic tissues, both NORs can remain active in some accessions whereas either of the two NORs can be silenced in other accessions [10,17]. Therefore, if the quantity/quality of available rDNA loci and contribution of NORs vary between individuals *in natura*, this might sustain the variation of expressed subsets of 45S rDNA units as one potential source of ribosomal heterogeneity [2,3,12].

Variations at the 5' and 3'ETS are likely not involved in ribosome heterogeneity since these sequences are removed during the processing of 45S pre-rRNA into mature 18S, 5.8S and 25S rRNAs. However, we demonstrated that studying their variation *in natura* contributes to distinguishing rDNA units both within and between individuals. These variations are also associated with polymorphisms present in specific 18S, 5.8S and 25S rRNA sequences and are potentially different depending on altitude or 5' and 3'ETS variants. Specific nucleotide and/or insertion/deletion polymorphisms in the ETS, ITS and mature rRNA sequences were linked to rDNA variants from active NORs and rRNA assembled into ribosomes [12]. Thus, in a similar way, associating structural variations of pre-rRNA with different 5'ETS and 3'ETS R genotypes should provide information about the impact of these structural variations in controlling the synthesis of rRNAs. Undoubtedly, 5'ETS length might affect the earliest pre-rRNA processing step at the P site and, subsequently, the production of specific rRNAs, while 3'ETS structural variations might affect RNA pol I transcription termination [3].

Our study of the variation at the 3'ETS in 320 individuals of At66 identified 12 different 45S rDNA (R) genotypes and two new expressed variants, VAR6 and VAR7, in addition to already known VAR1-5 [9,20,26]. Still, the question remains open to know whether the newly expressed VAR6 and VAR7 copies are located on the same or different NORs or both NORs. Nearly half of the individuals of At66 have only VAR1 copies and are present from sea level up to 1900 m of altitude. This raises the question of whether these individuals express one or another subset of their VAR1 copies depending on their NOR localization.

This study also investigated variation at the 5'ETS *in natura* and demonstrated that the R1 genotype (VARA much higher than VARB) identified in Col-0 is present in most of the individuals tested in At66 (71.6%), regardless of the 3'ETS. However, this study also identified nine additional R2–10 genotypes, each represented in the order of a small percentage but altogether present in 28.4% of the individuals tested. We do not know if the 45S rDNA genes containing 5'ETS R2–10 sequences are expressed and/or processed into mature rRNAs. This also raises the question of whether rDNA units bearing these variants are expressed or not depending on their localization between or within NORs.

Our investigation of the variation at the 5'ETS in Brassicaceae confirmed a previous result [24,25] showing that the 1.1 kbp insertion made of five D and two C repeats and splitting away the conserved U3 snoRNP binding site A<sup>123</sup>B from the primary processing site P in *A. thaliana* is absent in species belonging to the Brassicaceae tribe, lineage II, clade B. Additionally, 5'ETS genotyping in the Camelinae tribe, lineage I, clade A (*Arabidopsis lyrata*, *halleri*, *cebennensis* and *pedemontana*) demonstrated that the length of the 5'ETS rDNA sequence in the *Arabidopsis* genus depends on the number and integrity of D and C repeats.

The presence of insertion is not only a feature of the *Arabidopsis* genus but was also identified in other tribes of lineage I, clade A (Figure S6). Interestingly, in these additional studied species (*Boechera stricta*, *Chrysochamela velutina*, *Lepidum campestre*, *Cardamine parviflora*, *Descurainia millefolia*), the insertion either partially retains some homologies with some D and C repeats and/or interrupts them by insertions or shares no homology with any of them. Altogether, 5'ETS genotyping in Brassicaceae species highlighted that variants with and without an insertion can co-exist within a single individual (*Arabidopsis cebennensis* and *thaliana*, *Boechera stricta*, *Chrysochamela velutina* and *Lepidum campestre*).

Interestingly in At66, we found the lowest and highest 5'ETS genotype diversity for individuals living, respectively, at sea level and altitude (760–1900 m). Surprisingly, the exact opposite is true for 3'ETS genotype diversity (i.e., highest at sea level and lowest in altitude). In this context, we wondered whether altitude or bioclimatic variables could explain part of the 45S rDNA copy number variation in At66. We standardized the variance for a selection of least correlated variables (altitude and WorldClim bio3, bio4, bio8, bio9, bio13 and bio15) and used a model-averaging approach to answer this question. The best model provided with an  $R^2 = 0.22$  for 18S CNV did not support any effect as the confidence interval was overlapping zero with  $p$ -value  $> 0.05$  for all variables tested (Figure S7) [50–56]. Although there is apparently no correlation between global environmental conditions and copy number variation along a sharp altitudinal gradient, qualitative variation registered for rDNA copies at sea level versus altitude could promote a better response of individuals to local environmental stresses.

At66 is a worthwhile natural laboratory, and the unraveled 45S rDNA diversity represents an interesting starting material to select subsets for rDNA transcription and alter the rRNA composition of ribosomes both intra- and inter-site. It would be interesting to assign rDNA copies and their variants to their NORs of origin as these variants are also reporters of rDNA-cluster-specific expression, but this would only be feasible in recombinant inbred or F2 lines [10,11]. As rDNA stability is vital for genome integrity, it would also be interesting to investigate whether the gain or loss of rDNA copies is associated with polymorphisms in TE insertions or other genome structural variants. Ribosome heterogeneity not only includes the sequence variation of rRNA but also the variation of ribosomal proteins (RP) from the canonical structure, exchange of RP paralogs and post-transcriptional and post-translational modifications of rRNAs or RPs [2,57]. Therefore, plant ribosomes are heterogeneous at multiple levels, and there is still much that remains to be understood regarding the functional consequences of ribosome heterogeneity.

## 4. Materials and Methods

### 4.1. Sampling Individuals of At66 and Growth Conditions

The sampling campaign of 320 individuals of At66 was performed during two consecutive years, 2015–2016, from March to the end of June, depending on the altitude. Sampling was performed in priority within or near National Nature Reserves in collaboration with FRNC (Fédération des Réserves Naturelles Catalanes, Prades, France) to ensure that access to the material would be reliable over time. For all sampling performed *in natural* reserves, special authorizations were granted by the "Direction Départementale des Territoires et de la Mer 66" (DDTM 66). Each individual was collected as an adult plant with matured siliques, haphazardly chosen between multiple patches with a minimal distance of approximately 1 m between individuals to fulfill seed dispersal requirements [58]. Sampling efforts (4–40 individuals) depended on the number of individuals per site and number of patches in order to include the maximum diversity present at a site. Within the same locality, we considered that we had a distinct site any time the habitat changed or when we did not find any individual within a few hundred meters. Field-collected seeds for each individual were reproduced one time in a growth chamber to obtain a batch of seeds with reduced maternal effects on trait expression. Lineages were established by selfing to recover single-seed descendants sharing the same controlled maternal environment. Individuals were named using the two letters PO (for Pyrénées-Orientales) followed by the number of

the site (01–23) and one letter (A–Z) for their identification. The sites were georeferenced using QGIS software 3.28 (<https://www.qgis.org/>, accessed on 21 October 2022), including the base map SRTM, © Jet Propulsion Laboratory 2010, was used to georeference the 23 sites of At66.

All field-collected seeds were surface-sterilized by constant agitation with 70% (*v/v*) ethanol for 10 min, then washed with 95% (*v/v*) ethanol and dried. Seeds were placed on plates containing 50 mL of Murashige and Skoog medium with 0.5 g L<sup>-1</sup> MES and 0.8% (*w/v*) plant agar (M0231; Duchefa) with 1% (*w/v*) sucrose. All plates were first stratified in darkness at 4 °C for 6 days and then incubated for 2 weeks at 20 °C with 160 µE m<sup>-2</sup> s<sup>-1</sup> light intensity and a 16 h light/8 h dark regimen. Seedlings were transplanted in a mixture of soil and vermiculite (3:1, *v/v*), irrigated with water and grown at 22 °C with 70% hygrometry under a 16 h light (200 µE)/8 h dark regimen. Depending on the altitudinal origin of individuals, rosette leaves required vernalization for 4 weeks at 4 °C to induce flowering.

#### 4.2. Additional Plant Material

Seeds from other species of the *Arabidopsis* genus (*lyrata*, *halleri*, *cebnensis* and *pedemontana*) were gently supplied either by Craig Pikaard (Department of Biology, Department of Molecular and Cellular Biochemistry, and Howard Hughes Medical Institute, Indiana University, Bloomington, Indiana 47405) or Julie Jacquemin (Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany) and grown as described above for *Arabidopsis thaliana*.

#### 4.3. Estimating rDNA Copy Number Variation and Detection of rDNA Variants through NGS

A subset of 173 individuals of At66 was selected to perform a whole-genome shotgun sequencing (WGS) subcontracted to Novogene Europe, Cambridge, UK. We used an adapted CTAB method using Carlson buffer [59] to extract genomic DNA from pooled leaf tissue harvested from about 25–50 two-week-old sister plants for each individual single-seed descent lineage. The whole genomic library was prepared and sequenced in paired-end mode (2 × 150 bp, insert size: 350 bp) by Novogene Europe, Cambridge, UK using 1 µg of genomic DNA as input material for library preparation. Sequencing libraries were generated using a NEBNext<sup>®</sup> DNA Library Prep Kit following the manufacturer's recommendation, and indices were added to each sample. Genomic DNA was randomly fragmented to a size of 350 bp by shearing. DNA fragments were end-polished, A-tailed, ligated with NEBNext adapter for Illumina sequencing and further PCR-enriched by P5 and indexed P7 oligos. The PCR products were purified with AMPure XP system and the resulting libraries were analyzed for size distribution by an Agilent 2100 Bioanalyzer and quantified using real-time PCR. Qualified libraries were fed into unspecified Illumina sequencers. The total output of raw data received on 29 September 2019 for 173 individuals of At66 was 585.0 GB (99.65% clean reads), with an average error rate of 0.03%, a QPHRED20 of 97% and 37% GC content. Raw reads were submitted to the European Nucleotide Archive (ENA) with the BioProject accession number PRJEB48282 and the SRA study accession number ERP132629.

FASTAQ raw reads were trimmed to remove adapters and poor-quality sequences using Trimmomatic V0.36 paired-end mode and the following trimming steps: ILLUMINA-CLIP, LEADING, TRAILING AND MINLEN [60]. Duplicated reads were removed using a homemade script (available upon request) in order to improve the yield of recovered sequences compared to SAMtools V0.1.9 RMDUP. Cleaned reads were aligned both to the TAIR-10 Col-0 reference genome and a 45S rDNA reference using Bowtie2 V0.2.3.4 [61]. SAMtools V0.1.9 SORT was used to map the reads by position on the chromosomes [62,63]. The 45S rDNA reference includes the most abundant 5' and 3'ETS variants in Col-0 (*VARA* and *VAR1*), and the annotation of 18S and 25S rRNA genes coordinates, respectively, are 1841–3644 and 4264–7639 (Figure S8). We retrieved per-base read depth for each individual using BEDtools coverage [64] for both 18S and 25S rRNA genes and for the first 10 MB of



chromosome 3 that was free of 45S rDNA copies. Copy number variation was estimated by dividing the average coverage along both 18S and 25S rRNA genes by the average coverage along the first 10Mb of chromosome 3 [11]. Using recently developed bioinformatic scripts [11], we retrieved the sum of frequencies of all alternative alleles for each position along our ~10.5 kb 45S rDNA reference for each of 173 among 320 individuals of At66. Then, we plotted the proportion of individuals of At66 with an alternative allele for each position along this rDNA. The alternative variant must be present in at least 5% of the individual's total 45S rRNA genes (>0.05). All graphs and statistical analyses were performed using the R4.2.2 software [53] in the Rstudio environment [54] and the R package tidyverse [65].

#### 4.4. 5' and 3'ETS Genotyping

5' and 3'ETS rDNA genotyping was performed on rosette leaves while amplifying each individual for At66 or other species of the *Arabidopsis* genus. Brassicaceae leaves were either bought fresh or supplied gently frozen by Martin Lysak (CEITEC, Masaryk University, Brno, Czechia). Genomic DNA was extracted from three-week-old rosette leaves following the method described by Edwards et al. [66]. PCR amplification of the 5'ETS was performed using the forward primer P1 and the reverse primer P3 using GoTaq<sup>®</sup> G2 DNA Polymerase (Promega) and the following PCR program design: 40 cycles of 30 s at 94 °C, 45 s at 55 °C and 2 min at 72 °C. PCR products were diluted 1:5 and resolved by electrophoresis on 1% agarose gels in 0.5× TBE buffer. PCR amplification of the 3'ETS was performed using the forward primer P4 and the reverse primer P5 using GoTaq<sup>®</sup> G2 DNA Polymerase (Promega) and the following PCR program design: 24 cycles of 30 s at 94 °C, 45 s at 55 °C and 45 s at 72 °C. PCR products were diluted 1:5 and resolved by electrophoresis on 1.8% agarose gels in 0.5× TBE buffer running the gel at a maximum of 70 V. A PCR mix control without any DNA template was always included in each PCR amplification to validate the genotypes. When required, PCR amplifications were repeated using the proof-reading DNA polymerase Phusion<sup>™</sup> High-Fidelity (FISHER SCIENTIFIC S.A.S., Illkirch Graffenstaden, France), and the PCR products were A-tailed using GoTaq<sup>®</sup> G2 DNA Polymerase (PROMEGA, Charbonnières-les-bains, France), purified using GeneClean<sup>®</sup> Turbo for PCR Kit (MP BIOMEDICALS, Illkirch, France) and cloned in the pGEM<sup>®</sup>-T Easy Vector Systems (Promega). Plasmid DNA of 1–3 independent clones per identified band was extracted using Wizard<sup>®</sup> Plus SV Minipreps DNA Purification Systems (Promega) and sequenced using the Lightrun service of Eurofins GATC BIOTECH (Eurofins GATC BIOTECH—GMBH, Konstanz, Germany). The additional forward primer P2 was used when appropriate for sequencing the longest 5'ETS variants. Sequence data analysis was conducted using Geneious R11 software (<https://www.geneious.com>, accessed on 12 July 2023), and Multiple Sequence Comparison by Log-Expectation (MUSCLE) multiple alignments were performed using either Geneious or Seaview5 [67]. The QGIS software 3.10 (<https://www.qgis.org/>, accessed on 12 July 2023), including the base map SRTM, © Jet Propulsion Laboratory 2010, was used to georeference the At66 populations and to display associated pie charts illustrating the proportion of each 3'ETS or 5'ETS genotypes.

#### 4.5. Expression of New VAR6 and VAR7 3'ETS Variants

Total RNA was extracted from three-week-old rosette leaves using the Monarch<sup>®</sup> Total RNA Miniprep Kit according to the manufacturer's protocol (NEW ENGLAND Biolabs<sup>®</sup> Inc., Evry, France). DNase treatment was already included in this kit and no additional treatment was required as PCR amplification confirmed that there were no genomic DNA contaminants left in the total RNA. Reverse transcription was carried out using the Invitrogen<sup>™</sup> SuperScript<sup>™</sup> IV Reverse Transcriptase (FISHER SCIENTIFIC S.A.S., Illkirch Graffenstaden, France) according to the manufacturer's protocol using the reverse primer P6 located at the end of the 3'ETS of 45S rRNA. RT-PCR amplification of the 3'ETS was performed using the forward primer P4 and the reverse primer P5 using GoTaq<sup>®</sup> G2 DNA Polymerase (Promega) and the following PCR program design: 30 cycles of 30 s at 94 °C, 45 s at 55 °C and 45 s at 72 °C. RT-PCR amplification of the 25S rRNA gene was

performed using the forward primer P7 and the reverse primer P8 using GoTaq<sup>®</sup> G2 DNA Polymerase (Promega) and the following PCR program design: 16 cycles of 30 s at 94 °C, 45 s at 55 °C and 45 s at 72 °C. The 25S rRNA gene amplification was used as an internal control within 45 rRNAs for equal cDNA loading. PCR products were diluted 1:10 and resolved by electrophoresis on 1.8% agarose gels in 0.5× TBE buffer running the gel at a maximum of 70 V.

#### 4.6. Primer List

P1, 5'-CTTTTCCGGGCACTTTTCCGG-3'  
 P2, 5'-CTTGCGCACGAAATACCGAG-3'  
 P3, 5'-GTTCCGGCGTATGAGTGGTGATCGG-3'  
 P4, 5'-GACAGACTTGTCCAAAACGCCCACC-3'  
 P5, 5'-CTGGTCGAGGAATCCTGGACGATT-3'  
 P6, 5'-CCTCGGACCCGGTAAAC-3'  
 P7, 5'-TGTTACCCACCAATAGGGAA-3'  
 P8, 5'-GGCGGTCCGAACGACCGTTCGGCC-3'

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/plants12142722/s1>, Figure S1: 25S copy number variation in At66; Figure S2: FASTA sequences and multiple alignment of 3'ETS variants; Figure S3: 5'ETS variants identified in Brassicaceae Lineage I, Clade A and Camelinae and Brassicaceae Lineage II, Clade B; Figure S4: 45S rRNA gene polymorphisms in At66; Figure S5: GWAS analysis using 45S rDNA copy number variation as a phenotype; Figure S6: 5'ETS variation in additional species of Brassicaceae Lineage I, Clade A; Figure S7: Model-averaging approach testing the effects of altitude and bioclimatic variables on 18S CNV; Figure S8: FASTA sequence of the 45S rDNA of 10,524 bp used as a reference to estimate 18S and 25S copy number variation and rDNA variants; Table S1: 45S rDNA copy number variation in At66; Table S2: Variation at the 5' and 3'ETS in At66.

**Author Contributions:** Conceptualization, V.D.-H. and J.S.-V.; Data curation, V.D.-H.; Formal analysis, V.D.-H. and A.M.; Funding acquisition, V.D.-H., J.-P.R. and J.S.-V.; Investigation, V.D.-H., S.B., A.D.B. and C.L.; Methodology, V.D.-H. and J.S.-V.; Project administration, V.D.-H., M.M., J.-P.R. and J.S.-V.; Resources, V.D.-H., F.C., J.G., C.G. and J.B.; Software, V.D.-H. and A.M.; Supervision, V.D.-H. and J.S.-V.; Validation, V.D.-H. and J.S.-V.; Visualization, V.D.-H.; Writing—original draft, V.D.-H.; Writing—review and editing, V.D.-H., M.M., J.-P.R. and J.S.-V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Centre National de la Recherche Scientifique (CNRS), the Université de Perpignan Via Domitia (UPVD), the “Laboratoires d’Excellence (LABEX)” TULIP (ANR-10-LABX-41), through Labex AGRO (under I-Site Muse framework) coordinated by the Agropolis Foundation (grant no. Flagship Project 1802-002—CalClim) and by the ANR grants MetRibo 20-CE12-0024 and RiboStress 17-CE12-0026. This study was initiated with the support of a small “Bonus Quality Recherche” project from UPVD and set within the framework of the “Laboratoires d’Excellence (LABEX)” TULIP (ANR-10-LABX-41) and the École Universitaire de Recherche TULIP-GS (ANR-18-EURE-0019).

**Data Availability Statement:** Raw reads are available at the European Nucleotide Archive (ENA) with the BioProject accession number PRJEB48282 and the SRA study accession number ERP132629.

**Acknowledgments:** We thank Martin Lysak and Terezie Mandáková for supplying us with Brassicaceae leaves (Central European Institute of Technology, Masaryk University, 62500 Brno, Czech Republic), Craig Pikaard (Department of Biology and Department of Molecular and Cellular Biochemistry, Howard Hughes Medical Institute, Indiana University, Bloomington, Indiana 47405) and Julie Jacquemin (Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany) for supplying us with seeds of *Arabidopsis* (either *lyrata* and *halleri* or *cebennensis* and *pedemontana*). We thank Marie-Christine Carpentier and Martin Rosalie who give courses and provide training workshops in R language and environment for statistical computing and graphics at UPVD and LGDP. We thank Anaïs Gibert, Pascaline Salvado and Joris Bertrand for their advice in designing R scripts for downloading climatic variables or running GWAS and model

averaging studies. We thank Véronique Suréda, Viviane Jean and Christelle Roigt for their help in growing plants and harvesting seeds. We thank students (Jeanne Aguilo, Marie Morlans, François Ciavatti and Antoine Cousin) who also contributed to 5' and 3'ETS genotyping under the supervision of Valérie Delorme-Hinoux.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Carroll, A.J. The *Arabidopsis* cytosolic ribosomal proteome: From form to function. *Front. Plant Sci.* **2013**, *4*, 32. [[CrossRef](#)] [[PubMed](#)]
2. Martinez-Seidel, F.; Beine-Golovchuk, O.; Hsieh, Y.-C.; Kopka, J. Systematic review of plant ribosome heterogeneity and specialization. *Front. Plant Sci.* **2020**, *11*, 948. [[CrossRef](#)] [[PubMed](#)]
3. Sáez-Vásquez, J.; Delseny, M. Ribosome biogenesis in plants: From functional 45S ribosomal DNA organization to ribosome assembly factors. *Plant Cell* **2019**, *31*, 1945–1967. [[CrossRef](#)] [[PubMed](#)]
4. Gruendler, P.; Unfried, I.; Pascher, K.; Schweizer, D. rDNA intergenic region from *Arabidopsis thaliana*: Structural analysis, intraspecific variation and functional implications. *J. Mol. Biol.* **1991**, *221*, 1209–1222. [[CrossRef](#)] [[PubMed](#)]
5. Gruendler, P.; Unfried, I.; Pointner, R.; Schweizer, D. Nucleotide sequence of the 25S-18S ribosomal gene spacer from *Arabidopsis thaliana*. *Nucleic Acids Res.* **1989**, *17*, 6395–6396. [[CrossRef](#)]
6. Unfried, I.; Gruendler, P. Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*. *Nucleic Acids Res.* **1990**, *18*, 4011. [[CrossRef](#)]
7. Unfried, I.; Stocker, U.; Gruendler, P. Nucleotide sequence of the 18S rRNA gene from *Arabidopsis thaliana* Col0. *Nucleic Acids Res.* **1989**, *17*, 7513. [[CrossRef](#)]
8. Copenhaver, G.P.; Doelling, J.H.; Gens, J.S.; Pikaard, C.S. Use of RFLPs larger than 100 kbp to map the position and internal organization of the nucleolus organizer region on chromosome 2 in *Arabidopsis thaliana*. *Plant J.* **1995**, *7*, 273–286. [[CrossRef](#)]
9. Havlová, K.; Dvořáčková, M.; Peiro, R.; Abia, D.; Mozgová, I.; Vansáčová, L.; Gutierrez, C.; Fajkus, J. Variation of 45S rDNA intergenic spacers in *Arabidopsis thaliana*. *Plant Mol. Biol.* **2016**, *92*, 457–471. [[CrossRef](#)]
10. Rabanal, F.A.; Mandáková, T.; Soto-Jiménez, L.M.; Greenhalgh, R.; Parrott, D.L.; Lutzmayer, S.; Steffen, J.G.; Nizhynska, V.; Mott, R.; Lysak, M.A.; et al. Epistatic and allelic interactions control expression of ribosomal RNA gene clusters in *Arabidopsis thaliana*. *Genome Biol.* **2017**, *18*, 75. [[CrossRef](#)]
11. Rabanal, F.A. Characterization of the different levels of variation in 45S rRNA genes. *Methods Mol. Biol.* **2022**, *2484*, 363–379. [[PubMed](#)]
12. Sims, J.; Sestini, G.; Elgert, C.; von Haeseler, A.; Schlögelhofer, P. Sequencing of the *Arabidopsis* NOR2 reveals its distinct organization and tissue-specific rRNA ribosomal variants. *Nat. Commun.* **2021**, *12*, 387. [[CrossRef](#)]
13. Copenhaver, G.P.; Pikaard, C.S. RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *Plant J.* **1996**, *9*, 259–272. [[CrossRef](#)]
14. Copenhaver, G.P.; Pikaard, C.S. Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J.* **1996**, *9*, 273–282. [[CrossRef](#)]
15. McClintock, B. The relation of a particular chromosomal element to the development of the nucleoli in *Zea mays*. *Z. Zellforsch.* **1934**, *21*, 294–326. [[CrossRef](#)]
16. McStay, B. Nucleolar organizer regions: Genomic ‘dark matter’ requiring illumination. *Genes Dev.* **2016**, *30*, 1598–1610. [[CrossRef](#)] [[PubMed](#)]
17. Chandrasekhara, C.; Mohannath, G.; Blevins, T.; Pontvianne, F.; Pikaard, C.S. Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes Dev.* **2016**, *30*, 177–190. [[CrossRef](#)] [[PubMed](#)]
18. Rabanal, F.A.; Nizhynska, V.; Mandáková, T.; Novikova, P.Y.; Lysak, M.A.; Mott, R.; Nordborg, M. Unstable inheritance of 45S rRNA genes in *Arabidopsis thaliana*. *G3 Genes Genomes Genet.* **2017**, *7*, 1201–1209. [[CrossRef](#)]
19. Long, Q.; Rabanal, F.A.; Meng, D.; Huber, C.D.; Farlow, A.; Platzer, A.; Zhang, Q.; Vilhjálmsón, B.J.; Korte, A.; Nizhynska, V.; et al. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* **2013**, *45*, 884–890. [[CrossRef](#)]
20. Pontvianne, F.; Abou-Ellail, M.; Douet, J.; Comella, P.; Matia, I.; Chandrasekhara, C.; Debures, A.; Blevins, T.; Cooke, R.; Medina, F.J.; et al. Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in *Arabidopsis thaliana*. *PLoS Genet.* **2010**, *6*, e1001225. [[CrossRef](#)]
21. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815. [[CrossRef](#)] [[PubMed](#)]
22. Jiao, W.B.; Schneeberger, K. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat. Commun.* **2020**, *11*, 989. [[CrossRef](#)] [[PubMed](#)]
23. Michael, T.P.; Jupe, F.; Bemm, F.; Motley, S.T.; Sandoval, J.P.; Lanz, C.; Loudet, O.; Weigel, D.; Ecker, J.R. High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* **2018**, *9*, 541. [[CrossRef](#)] [[PubMed](#)]

24. Caparros-Ruiz, D.; Lahmy, S.; Piersanti, S.; Echeverría, M. Two ribosomal DNA-binding factors interact with a cluster of motifs on the 5' external transcribed spacer, upstream from the primary pre-rRNA processing site in a higher plant. *Eur. J. Biochem.* **1997**, *247*, 981–989. [[CrossRef](#)] [[PubMed](#)]
25. Sáez-Vasquez, J.; Caparros-Ruiz, D.; Barneche, F.; Echeverría, M. A plant snoRNP complex containing snoRNAs, fibrillarin, and nucleolin-like proteins is competent for both rRNA gene binding and pre-rRNA processing in vitro. *Mol. Cell. Biol.* **2004**, *24*, 7284–7297. [[CrossRef](#)]
26. Abou-Ellail, M.; Cooke, R.; Sáez-Vásquez, J. Variations in a team: Major and minor variants of *Arabidopsis thaliana* rDNA genes. *Nucleus* **2011**, *2*, 294–299. [[CrossRef](#)]
27. Bartoli, C.; Frachon, L.; Barret, M.; Rigal, M.; Huard-Chauveau, C.; Mayjonade, B.; Zanchetta, C.; Bouchez, O.; Roby, D.; Carrère, S.; et al. In situ relationships between microbiota and potential pathobiota in *Arabidopsis thaliana*. *ISME J.* **2018**, *12*, 2024–2038. [[CrossRef](#)]
28. Duruflé, H.; Ranocha, P.; Mbadinga Mbadinga, D.L.; Déjean, S.; Bonhomme, M.; San Clemente, H.; Viudes, S.; Eljebbawi, A.; Delorme-Hinoux, V.; Sáez-Vásquez, J.; et al. Phenotypic trait variation as a response to altitude-related constraints in *Arabidopsis* populations. *Front. Plant Sci.* **2019**, *10*, 430. [[CrossRef](#)]
29. Frachon, L.; Bartoli, C.; Carrère, S.; Bouchez, O.; Chaubet, A.; Gautier, M.; Roby, D.; Roux, F. A genomic map of climate adaptation in *Arabidopsis thaliana* at a micro-geographic scale. *Front. Plant Sci.* **2018**, *9*, 967. [[CrossRef](#)]
30. Busoms, S.; Teres, J.; Huang, X.Y.; Bomblies, K.; Danku, J.; Douglas, A.; Weigel, D.; Poschenrieder, C.; Salt, D.E. Salinity is an agent of divergent selection driving local adaptation of *Arabidopsis* to coastal habitats. *Plant Physiol.* **2015**, *168*, 915–929. [[CrossRef](#)]
31. Montesinos-Navarro, A.; Wig, J.; Picó, F.X.; Tonsor, S.J. *Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude. *New Phytol.* **2011**, *189*, 282–294. [[CrossRef](#)]
32. Montesinos, A.; Tonsor, S.J.; Alonso-Blanco, C.; Picó, F.X. Demographic and genetic patterns of variation among populations of *Arabidopsis thaliana* from contrasting native environments. *PLoS ONE* **2009**, *4*, e7213. [[CrossRef](#)] [[PubMed](#)]
33. Schmuths, H.; Meister, A.; Horres, R.; Bachmann, K. Genome size variation among accessions of *Arabidopsis thaliana*. *Ann. Bot.* **2004**, *93*, 317–321. [[CrossRef](#)] [[PubMed](#)]
34. Davison, J.; Tyagi, A.; Comai, L. Large-scale polymorphism of heterochromatic repeats in the DNA of *Arabidopsis thaliana*. *BMC Plant Biol.* **2007**, *7*, 44. [[CrossRef](#)]
35. Prokopowich, C.D.; Gregory, T.R.; Crease, T.J. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **2003**, *46*, 48–50. [[CrossRef](#)] [[PubMed](#)]
36. Earley, K.W.; Pontvianne, F.; Wierzbicki, A.T.; Blevins, T.; Tucker, S.; Costa-Nunes, P.; Pontes, O.; Pikaard, C.S. Mechanisms of HDA6-mediated rRNA gene silencing: Suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. *Genes Dev.* **2010**, *24*, 1119–1132. [[CrossRef](#)]
37. Darriere, T.; Jobet, E.; Zavala, D.; Escande, M.L.; Durut, N.; De Bures, A.; Blanco-Herrera, F.; Vidal, E.A.; Rompais, M.; Carapito, C.; et al. Upon heat stress processing of ribosomal RNA precursors into mature rRNAs is compromised after cleavage at primary P site in *Arabidopsis thaliana*. *RNA Biol.* **2022**, *19*, 719–734. [[CrossRef](#)] [[PubMed](#)]
38. Wang, J.; Zhang, Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics. Proteom. Bioinform.* **2021**, *19*, 629–640. [[CrossRef](#)]
39. Huang, M.; Liu, X.; Zhou, Y.; Summers, R.M.; Zhang, Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* **2019**, *8*, giy154. [[CrossRef](#)]
40. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S.; Zhang, Z. Iterative usage of Fixed and Random Effect Models for powerful and efficient Genome-Wide Association Studies. *PLoS Genet.* **2016**, *12*, e1005767. [[CrossRef](#)]
41. Rosato, M.; Kovařík, A.; Garilleti, R.; Rosselló, J.A. Conserved Organisation of 45S rDNA Sites and rDNA Gene Copy Number among Major Clades of Early Land Plants. *PLoS ONE* **2016**, *11*, e0162544. [[CrossRef](#)] [[PubMed](#)]
42. Sims, J.; Copenhaver, G.P.; Schlögelhofer, P. Meiotic DNA repair in the nucleolus employs a Nonhomologous End-Joining Mechanism. *Plant Cell* **2019**, *31*, 2259–2275. [[CrossRef](#)] [[PubMed](#)]
43. Schindele, A.; Gehrke, F.; Schmidt, C.; Rohrig, S.; Dorn, A.; Puchta, H. Using CRISPR-Kill for organ specific cell elimination by cleavage of tandem repeats. *Nat. Commun.* **2022**, *13*, 1502. [[CrossRef](#)] [[PubMed](#)]
44. Datson, P.M.; Murray, B.G. Ribosomal DNA locus evolution in *Nemesia*: Transposition rather than structural rearrangement as the key mechanism? *Chromosome Res.* **2006**, *14*, 845–857. [[CrossRef](#)] [[PubMed](#)]
45. Pavlišťová, V.; Dvořáčková, M.; Jež, M.; Mozgová, I.; Mokroš, P.; Fajkus, J. Phenotypic reversion in fas mutants of *Arabidopsis thaliana* by reintroduction of FAS genes: Variable recovery of telomeres with major spatial rearrangements and transcriptional reprogramming of 45S rDNA genes. *Plant J. Cell Mol. Biol.* **2016**, *88*, 411–424. [[CrossRef](#)] [[PubMed](#)]
46. Picart-Piccolo, A.; Grob, S.; Picault, N.; Franek, M.; Llauro, C.; Halter, T.; Maier, T.R.; Jobet, E.; Descombin, J.; Zhang, P.; et al. Large tandem duplications affect gene expression, 3D organization, and plant-pathogen response. *Genome Res.* **2020**, *30*, 1583–1592. [[CrossRef](#)]
47. Lopez, F.B.; Fort, A.; Tadini, L.; Probst, A.V.; McHale, M.; Friel, J.; Ryder, P.; Pontvianne, F.; Pesaresi, P.; Sulpice, R.; et al. Gene dosage compensation of rRNA transcript levels in *Arabidopsis thaliana* lines with reduced ribosomal gene copy number. *Plant Cell* **2021**, *33*, 1135–1150. [[CrossRef](#)]
48. Lopez, F.B.; McKeown, P.C.; Fort, A.; Brychkova, G.; Spillane, C. The boys are back in town: Rethinking the function of ribosomal DNA repeats in the genomic era. *Mol. Plant* **2023**, *16*, 514–516. [[CrossRef](#)]

49. Durut, N.; Abou-Ellail, M.; Pontvianne, F.; Das, S.; Kojima, H.; Ukai, S.; De Bures, A.; Comella, P.; Nidelet, S.; Rialle, S.; et al. A duplicated NUCLEOLIN gene with antagonistic activity is required for chromatin organization of silent 45S rDNA in *Arabidopsis*. *Plant Cell* **2014**, *26*, 1330–1344. [[CrossRef](#)]
50. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [[CrossRef](#)]
51. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very high-resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **2005**, *25*, 1965–1978. [[CrossRef](#)]
52. Heming, N.M.; Dambros, C.; Gutiérrez, E. ENMwizard: Advanced Techniques for Ecological Niche Modeling Made Easy. 2018. Available online: <https://github.com/HemingNM/ENMwizard/> (accessed on 13 July 2023).
53. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022; Available online: <https://www.R-project.org/> (accessed on 12 July 2023).
54. Posit Team. *RStudio: Integrated Development Environment for R*; Posit Software, PBC: Boston, MA, USA, 2022. Available online: <http://www.posit.co/> (accessed on 12 July 2023).
55. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [[CrossRef](#)]
56. Barton, K. MuMIn: Multi-Model Inference. R Package. 2022. Available online: <https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf> (accessed on 12 July 2023).
57. Azevedo-Favory, J.; Gaspin, C.; Ayadi, L.; Montacé, C.; Marchand, V.; Jobet, E.; Rompais, M.; Carapito, C.; Motorin, Y.; Sáez-Vásquez, J. Mapping rRNA 2'-O-methylations and identification of C/D snoRNAs in *Arabidopsis thaliana* plants. *RNA Biol.* **2021**, *18*, 1760–1777. [[CrossRef](#)]
58. Wender, N.J.; Polisetty, C.R.; Donohue, K. Density-dependent processes influencing the evolutionary dynamics of dispersal: A functional analysis of seed dispersal in *Arabidopsis thaliana* (Brassicaceae). *Am. J. Bot.* **2005**, *92*, 960–971. [[CrossRef](#)]
59. Vaillancourt, B.; Buell, C.R. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. *bioRxiv* **2019**, 783159.
60. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
61. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
62. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup, The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
63. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [[CrossRef](#)]
64. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)] [[PubMed](#)]
65. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemond, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [[CrossRef](#)]
66. Edwards, K.; Johnstone, C.; Thompson, C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* **1991**, *19*, 1349. [[CrossRef](#)] [[PubMed](#)]
67. Galtier, N.; Gouy, M.; Gautier, C. SeaView and Phylo\_win, two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **1996**, *12*, 543–548. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.