Current Applied Science and Technology Vol. 22 No. 2 (March-April 2022)

Research article

Geospatial Paradigm of Nucleotide Substitution within the Gene Family Encoding Plasma Membrane H^+-ATPase Pumps in *Arabidopsis thaliana*

Mahmood Ahmadi-kordshooli, Hassan Pakniyat* and Abbas Alemzadeh

Department of Plant Production and Genetics, School of Agriculture, Shiraz University, Shiraz, Iran

Received: 30 October 2020, Revised: 5 June 2021, Accepted:16 July 2021

DOI: 10.55003/cast.2022.02.22.003

Abstract

Keywords

1001 genomes project;
adaptation;
chi-square;
latitude interval;
nucleotide substitution;
PM H+-ATPase

The PM H^+-ATPase gene family consists of 12 members in *Arabidopsis thaliana* denoted to Auto-inhibited H^+-ATPase 1-12 (*AHA1-AHA12*). This gene family performs crucial roles in many aspects of plant growth and development such as ion and metabolite transportation, pollen development, stomatal opening, signalling process, and adaptation to abiotic stresses especially thermal changes. The H^+-ATPase action dynamically fluctuates during thermal changes in order to enhance adaptation strategies in plants. Based on the fact that surface temperature on average is typically warmer in lower latitudes and cooler in higher latitudes, the association of PM H^+-ATPase genetic variants (SNPs) with latitude intervals was studied using chi-square testing of 1001 Genomes project data. The results indicated a significant association between nucleotide substitution and latitude interval, and therefore a geospatial paradigm of PM H^+-ATPase nucleotide substitution was established. A>T and T>A substitutions showed a close positive association with higher latitudes (cold climate) and negative association with lower latitudes (warm climate), and vice versa for A>G. Most of the nucleotide substitutions were associated with extreme cold and warm climates while cold climate had more impact. The most significant signal of PM H^+-ATPase genes strategy for adaptation to cold climate was A>T nucleotide substitution. Also, the results showed that the most high-impact SNPs were identified in coding regions and related to premature stop codons.

*Corresponding author: Tel.: (+98) 7132287118
E-mail: pakniyat@shirazu.ac.ir

1. Introduction

The P-type superfamily of ion pumps can be divided into six subfamilies: P1B, P2A, P2B, P3A, P4 and P5, based on their ion transportation [1, 2]. Among them, P3A (Plasma membrane H^+- ATPase (PM H^+-ATPase)) pumps, which are membrane-localized proteins and plant cell master enzymes, play an important role in both the physiology and biochemistry of plant [3]. In this way, H^+-ATPase pumps extrude H^+ ions from the cell providing the driving force for secondary transporters to uptake and efflux ions and metabolites across the plasma membrane [4]. Based on the importance of H^+-ATPase activity in generating a proton electrochemical gradient and a membrane potential [5-8], it is expected that plants cope with the environmental changes via PM H^+-ATPase modulation [9, 10].

PM H^+-ATPase genes are found in plants, fungi, protista, and archaea, but not in animals [11], and the activity of the genes is regulated by the post translational phosphorylation of various amino acid residues [12]. P3A group in *Arabidopsis thaliana* consists of 11 very closely related members (*AHA*1-*AHA*11) and a truncated gene, which might be a pseudogene or might be one that has alternative functions (*AHA*12) [13]. Ueno *et al*. [14] displayed eight expressing *AHA* genes in green leaves (*AHA1*, *2*, *3*, *5*, *7*, *8*, *10*, and *11*), eight ones in roots (*AHA1*, *2*, *3*, *4*, *7*, *8*, *10*, and *11*), and three ones in all tissues (*AHA1*, *2* and *11*) using RT-PCR. Also, the roles for *AHA1* in steroid signaling, *AHA2* in iron transport, *AHA3* in pollen development, *AHA4* in salt stress, and *AHA10* in vacuole biosynthesis, have been reported using genetic analysis [15-17]. Therefore, from the physiological and biochemical points of view, PM H^+-ATPase performs a crucial role in nutrient uptake, stomatal opening, cell growth as well as plant adaptation to stresses [18]. PM H^+-ATPase resistance mechanisms can be activated in response to various abiotic stress conditions such as temperature, light, salt, and so on [18].

Many studies have shown the gene expression changes of the PM H^+-ATPase in response to low and high temperatures. For example, according to White *et al*. [19], PM H^+-ATPase activity increased at low temperature, followed by an increase in unsaturation fatty acid, particularly linoleic acid in rye roots. Similarly, in chilling-tolerant plant *Oryza sativa,* PM H^+-ATPase activity increased at low temperature whereas a slight decrease in enzyme activity was observed in chilling-sensitive plants [20]. Moreover, Martz *et al*. [21] indicated that the activity of PM H^+-ATPase and fatty acid composition increased more than twofold following cold acclimation in *Pinus resinosa*. The activity of the fatty acid-regulated plasma membrane proton pump is involved in the cellular response underlying cold acclimation and de-acclimation as a resistant response. On the other hand, the activity of PM H^+-ATPase pump changes in exposure to cold temperature based on the length of imposed cold stress. Short time cold decreases the pump activity while longer time cold increases the activity [22, 23]. For high temperature, Liu *et al*. [24] found that heat (38°C) increased the accumulation of the PM H^+-ATPase transcript and protein after 1 h of high-temperature treatment in pea leaves, and then after 3 h, the amount reduced. Therefore, the low/high temperature ranges during plant growth can lead to dynamic changes the components of the cell membrane in order to increase adaptation strategies [25]. By virtue of plants constant nature, the evaluation of their adaptation mechanism is ideal for researchers to study or improve plant characteristics via genetic information or geospatial genetic variation studies.

In the study of genetic variation, single nucleotide polymorphisms (SNPs) are an important source of genetic information which is used in phylogenetics [26], cancer and other human disease identification [27], climatology and geography [28, 29], marker study [30] and many other areas of biology. SNPs are ubiquitous genomic variants that result from the substitution of only a single base (nucleotide substitution) in DNA [31]. A study of SNPs in *A. thaliana*, as the most important model organism [32] and the first completely sequenced plant, caused researchers to shift towards identifying genetic variants that contributed to phenotypic variation in different accessions worldwide and to survey changes in genomes in terms of their spatial adaptation. For this purpose,

1001 Genomes project was developed. It provides a rich genetic resource that includes a whole-genome map of SNPs in *A. thaliana* which included 1,135 genomes of naturally inbred lines (accession) from North America, Asia, North Africa and Europe, from 15.1 to 68.8 degree of latitude and -123.2 to 136.3 degree of longitude [33].

Despite the knowledge that environmental factors such as the temperature gradient effect in the dynamics of PM H^+-ATPase pumps are enhancers of adaptation strategies and plant acclimation, studies of the effects of different latitude on nucleotide substitution within this plasma membrane gene family in *A. thaliana* are scarce. Therefore, the present study was conducted to understand the effect of latitude on nucleotide substitution in PM H^+-ATPase family on account of thermal gradients that coordinate with global latitudinal position using the accessions of the 1001G Project as the geo-references. The results of such studies would be the first report that provides new insights into the finding of a plausible signal of adaption at the level of nucleotide substitution.

2. Materials and Methods

In order to conduct a survey of an association between nucleotide substitution in PM H^+-ATPase genes and latitude intervals, genomic annotation of the PM H^+-ATPase gene family in *A. thaliana* was retrieved from Araport11 database [6]. SNP sites were identified within PM H^+-ATPase genes based on *A. thaliana* Col-0 reference genome via HDF5 file from the 1001 Genomes Project using Python interpreter in 1135 strains (accessions). The type of SNPs was determined using "Polymorph 1001" online variant browser (http://tools.1001genomes.org/polymorph/). All types of nucleotide substitutions (12 types) were studied; denoted as A>C (reference A that altered to C), A>G, A>T, C>A, C>G, C>T, G>A, G>C, G>T, T>A, T>C and T>G. Moreover, high-impact SNPs were identified utilizing Polymorph 1001.

Subsequently, the geographic coordinate (longitude and latitude) data for each accession was obtained from the 1001 genome data base (https://1001genomes.org/). In spite of the fact that the SNP data of all 1135 accessions were available, five accessions were filtered out due to the lack of geographical information. Latitudes were divided by 5 degrees into 12 intervals from 15 to 70 degree in the Northern Hemisphere, and then the relevant accessions were spatially identified in latitude intervals (Table 1). Based on the geographical distribution of 1001 genome accessions (Figure 1), down-stream analysis was performed with six latitude intervals including 35-40°, 40-45°, 45-50°, 50-55°, 55-60° and 60-65°, each of which contained more than 30 accessions. The other intervals with less than 30 accessions were discarded (Table 1). Random sampling was conducted in terms of the disparate number of accessions 91, 355, 222, 173, 205 and 73 in latitude intervals of interest. Random sampling was carried out in size of 73 without replacement in five replications in latitude intervals by R data processing. Hence, all the accessions in a latitude interval were allowed to have an equal chance of selection. The reason for selecting this sample size was that the interval latitude 60-65° which was important with 73 accessions as the candidate for high-latitude interval and cold climate zone and therefore, it was preferred.

Afterwards, three Pearson's chi-square tests of independence were implemented considering the total length (both coding and non-coding), coding region and non-coding (Introns and Utrs) region of the genes. The tests were performed with two variables in R programming language. Nucleotide substitution and latitude intervals were denoted as variables with 12 and 6 categories, respectively. Null hypothesis (H_0) of the test was that the nucleotide substitutions within the H^+-ATPase gene family occur independently in each accession inside different latitude

Table 1. Count of accessions that were associated with SNPs occurrence in the genes encoding PM H^+-ATPase pumps in *A. thaliana* varied in different latitude intervals

Latitude interval	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70
Count of Accessions	1	0	1	8	91	355	222	173	205	73	1
Sample Size	-	-	-	-	73	73	73	73	73	73	-

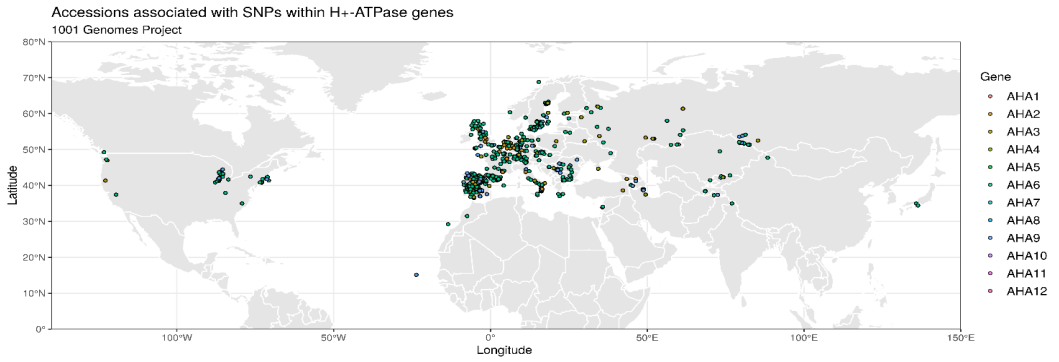

Figure 1. Geospatial mapping of accessions that were associated with SNPs occurrence in PM H^+-ATPase genes in *A. thaliana* from 1001 Genomes project

intervals. The mean count of SNPs within the H^+-ATPase gene family obtained from random sampling (5 replications without replacement) of accession inside the latitude intervals were considered as observed values (Table 2). The expected value for each cell of Table 2 was calculated using the following formula (equation 1) [34]:

$$E = (MR \times MC)/N \quad (1)$$

where E is expected value of the cell, MR and MC are row marginal and column marginal of the respected cell, respectively and N denotes the total size.

For the purpose of testing the null hypothesis, the chi-square statistic (χ^2) was calculated with the following formula [34], where O is observed value and E is expected value.

$$\chi^2 = \sum(O-E)^2 / E \quad (2)$$

In order to interpret the association between latitude interval and nucleotide substitution, the residuals of Pearson's chi-square (R) were calculated as in the formula below (equations 3) [35]. The R value assesses the difference between observed (O) and expected (E) variables.

$$R = (O-E) / \text{sqtr}(E) \quad (3)$$

Table 2. Contingency table of observed data: Each main cell presents the mean count of SNPs within the H^+-ATPase gene family which obtained from random sampling of accessions in latitude intervals (sample size =73) with 5 replications without replacement. Marginal cells present the total mean counts of rows and columns

	A>C	A>G	A>T	C>A	C>G	C>T	G>A	G>C	G>T	T>A	T>C	T>G	Sum
Lat60_65	1175.0	1646.0	1427.0	556.0	582.0	1838.0	1838.0	758.0	835.0	874.0	2835.0	654.0	15018.0
Lat55_60	1084.4	1642.8	1556.4	596.6	607.2	1682.4	1731.2	752.8	711.2	855.6	2581.8	670.2	14472.6
Lat50_55	877.8	1425.8	973.2	411.8	501.0	1373.8	1451.2	586.8	616.0	619.8	2242.0	546.6	11625.8
Lat45_50	868.4	1324.8	1023.6	446.4	457.6	1454.4	1382.4	609.8	596.8	618.8	2065.2	530.4	11378.6
Lat40_45	938.6	1580.6	1165.2	523.4	516.0	1478.6	1507.8	644.6	690.4	583.0	2371.8	571.6	12571.6
Lat35_40	948.4	1544.6	1025.0	510.8	500.0	1495.8	1602.2	642.8	711.2	636.2	2283.6	599.0	12499.6
Sum	5892.6	9164.6	7170.4	3045.0	3163.8	9323.0	9512.8	3994.8	4160.6	4187.4	14379.4	3571.8	**77566.2**

Then, contribution of each latitude interval and nucleotide substitution to the total chi-square statistic was calculated to uncover the nature of dependency between latitude interval and nucleotide substitution using the following formula (equation 4) [35, 36], where C% is contribution in percentage, R is residual value and χ^2indicates the chi-square statistic. In other words, the C value represents which variable levels contributes the most to the association.

$$C\% = (R^2 / \chi^2) \times 100 \quad (4)$$

The distribution of accessions that conveyed high-impact SNPs was broad, although some latitude intervals had very few individuals. Therefore, the chi-square test was not performed on those intervals. The data was reported as frequency of accession in each latitude interval (Table 3). Finally, the results were visualized by corrplot [36], sf [37], rgeos [38], rnaturalearth [39], and ggplot2 [40] packages in R.

Table 3. Count of accessions that integrated with high-impact SNPs occurrence in coding and non-coding regions of the genes encoding PM H^+-ATPase pumps in *A. thaliana* through the different latitude intervals

A)- Coding region							**B)- Non-coding region**				
	A>C	A>T	C>A	C>T	G>A	G>T		A>T	C>T	G>A	G>T
Lat60-65	2	4	0	0	4	0	Lat60-65	2	0	0	0
Lat55-60	4	0	0	4	5	0	Lat55-60	0	0	2	0
Lat50-55	0	0	0	1	2	1	Lat50-55	0	0	0	0
Lat45-50	1	0	1	0	3	1	Lat45-50	0	0	1	0
Lat40-45	2	0	1	7	8	28	Lat40-45	0	0	5	2
Lat35-40	0	0	0	1	2	1	Lat35-40	1	1	1	0

3. Results and Discussion

The present study was conducted to evaluate whether nucleotide substitution within the H^+-ATPase gene family was associated with latitude intervals in the Northern Hemisphere. For this purpose, 1119 of 1135 accessions from 1001 Genomes project were considered in six latitude intervals from 35 to 65° N, which were divided by 5 degrees. The reason for the elimination of 16 accessions from 1135 was the lack of geographical information for five, and location inside latitude intervals with less than 30 individuals (incongruent sample size) for the remaining eleven accessions. Then, a set of 3657 SNP sites was identified within total length of 12 H^+-ATPase genes in comparison to *A. thaliana* Col-0 reference genome, which classified 1506 sites in coding regions and 2151 in non-coding regions. From the window of SNPs percentage in total length of the gene, *AHA12*, *AHA6* and *AHA10* genes with 13.9, 11.4 and 7.8% were ranked first to third, respectively (Table 4). The lowest frequencies belonged to *AHA9* (3.5%), *AHA2* (2.8%) and *AHA7* (2.7%) (Table 4). Moreover, the nucleotide substitutions C>T and C>G had the highest and lowest frequencies in the total length of H^+-ATPase genes, respectively (Figure 2).

Geographical distribution of 1001 Genomes accessions integrated with nucleotide substitution occurrence within total length of the genes encoding H^+-ATPase pumps is illustrated in Figure 1. Europe was the distribution centre of the accessions and lower levels of distribution were assigned to Central Asia and North America. It should not be forgotten that the most accessions of 1001 Genome project were collected from Europe and consequently, this part is recognized as the central zone. To determine the relationships between the geographical distributions of accessions (latitude intervals) and nucleotide substitutions within total length, coding and non-coding regions of the PM H^+-ATPase gene family, three Pearson's chi-square tests were performed.

The results of the total chi-square tests were 168.64, 114.52, 206.35 in total gene length, coding region and non-coding region, respectively (Table 5). Highly significant chi-square p-values of all tests demonstrated strong signals of association between latitude intervals and nucleotide substitutions. In order to elucidate the association between 6 latitude intervals and 12 nucleotide substitutions, the residuals of Pearson's chi-square tests were calculated. Based on the total length of the gene test, A>T nucleotide substitution showed the strongest both positive and negative association with 55-60° and 35-40° latitude intervals, respectively. Similarly, T>A substitutions had a positive association with higher latitudes and negative association with lower latitudes. Moreover, A>G showed positive and negative associations with higher and lower latitudes, respectively. The lowest association of substitution with all latitude intervals was observed in A>C. Also, the substitutions C>T and T>C showed positive associations with central latitude intervals of 45-50° and 50-55° (Figure 3A).

In respect of coding region, the association for higher intervals is more impressive than of both middle and lower latitude intervals. Substitutions C>T and A>T within 60-65° latitude interval, A>T, G>A and C>G in 55-60° latitude interval as well as G>T and C>A inside the lower interval latitudes had a positive strong association with latitude. Nonetheless, the highest negative associations were observed in G>A nucleotide substitution with higher interval. Actually, A>T showed significant positive association with higher latitudes and negative association with central and lower latitudes (Figure 3B). In the non-coding region category, A>T and T>A nucleotide substitutions had positive and negative associations with higher and lower latitudes, respectively. By comparison, A>G showed negative and positive associations with higher and lower latitudes, respectively (Figure 3C). Thus, nucleotide substitutions within middle latitude intervals showed less

Table 4. Araport11 annotation of PM H^+-ATPase genes in *A. thaliana* Col-0

Symbol	Gene ID	Gene position in Chr.	Gene Length (bp)	CDS position in Chr.	Chr.	Str-and	No. of SNP site in total length of the gene	No. of SNP site in coding region	No. of SNP site in non-coding region	SNP percent-age in total length of the gene (%)
AHA1	AT2G18960	8221514: 8227622	6109	8221858: 8227268	Chr2	1	276	67	209	4.5
AHA2	AT4G30190	14770395:14776247	5853	14770820:14775920	Chr4	-1	165	44	121	2.8
AHA3	AT5G57350	23230874:23236682	5809	23231208:23236381	Chr5	-1	298	73	225	5.1
AHA4	AT3G47950	17692666:17697999	5334	17693015:17697801	Chr3	1	348	134	214	6.5
AHA5	AT2G24520	10414963:10420029	5067	10415522:10419730	Chr2	1	304	129	175	6.0
AHA6	AT2G07560	3169908:3174102	4195	3170394:3173952	Chr2	-1	480	267	213	11.4
AHA7	AT3G60330	22298373:22303718	5346	22298763:22303509	Chr3	1	141	41	100	2.6
AHA8	AT3G42640	14724103:14728327	4225	14724309:14728062	Chr3	1	310	172	138	7.3
AHA9	AT1G80660	30315964:30320197	4234	30316227:30319948	Chr1	-1	150	73	77	3.5
AHA10	AT1G17260	5904058:5908898	4841	5904058:5908898	Chr1	1	379	149	230	7.8
AHA11	AT5G62670	25159208:25165209	6002	25159495:25164957	Chr5	1	260	71	189	4.3
AHA12	AT4G11730	7067035:7070968	3934	7067035:7070968	Chr4	1	546	286	260	13.9
Total	-	-	-	-	-	-	3657	1506	2151	-

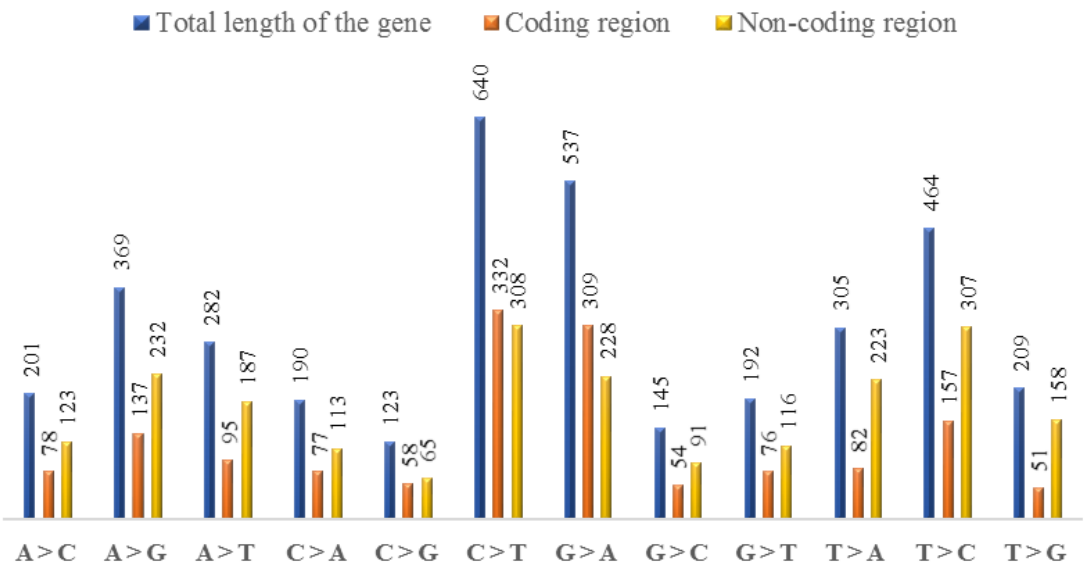

Figure 2. Number of different nucleotide substitution sites within total length of the H^+-ATPase gene family (*AHA1-AHA12*), and 3657 SNP sites within total length of the genes classified into 1506 and 2151 sites within coding and non-coding regions, respectively

Table 5. Chi-square statistic of three tests; one for total length of the H^+-ATPase gene family, second for coding region of this family and the last for non-coding region

	Chi-squared	**df**	**P-value**
Total length of gene	168.64	55	1.81E-13 **
Coding region	114.52	55	4.52E-06 **
Non-coding region	206.35	55	< 2.2e-16 **

Note: ** indicates significant level at 1% of probability

power of association than both lower and higher latitude intervals, probably due to latitude effect. Furthermore, A>G, A>T, G>A and T>A had more obvious association with interval latitudes than the others. In the context of geospatial pattern, unlike the A>G substitution, A>T and T>A exhibited a close positive association with higher and negative association with lower latitudes.

Accordingly, the nature of dependency between latitude interval and nucleotide substitution was determined using their contribution to the value of chi-square. The most contributing cell to the chi-square value was 55-60°/ A>T with 21.16% and then, 35-40°/A>T, 40-45°/ T>A, 50-55°/A>T and 60-65°/A>G ranked second to fifth with 8.74%, 8.0%, 5.69% and 5.51%, respectively in total length of the genes. In the coding regions, latitude 60-65°/G>A (8.66%), 55-60°/A>T (7.24%), 60-65°/C>T (6.11%), 40-45°/G>T (5.58%), 55-60°/G>A (5.12%), 55-60°/C>T (5.02%) and 35-40°/C>G (3.87%), gave large contributions to the chi-square test. In the non-coding regions, the nucleotide substitutions A>T, T>A, A>G and G>A gave the greatest contribution in both higher and lower latitudes as follows; 55-60°/A>T (10.58%), 35-40°/A>T (7.71%), 55-60°/T>A (4.14%), 40-45°/T>A (12.32%), 60-65°/A>G (8.41%), 40-45°/A>G (5.25%), 60-65°G>A (4.0%) and 55-60°/G>A (5.35%) (Table 6).

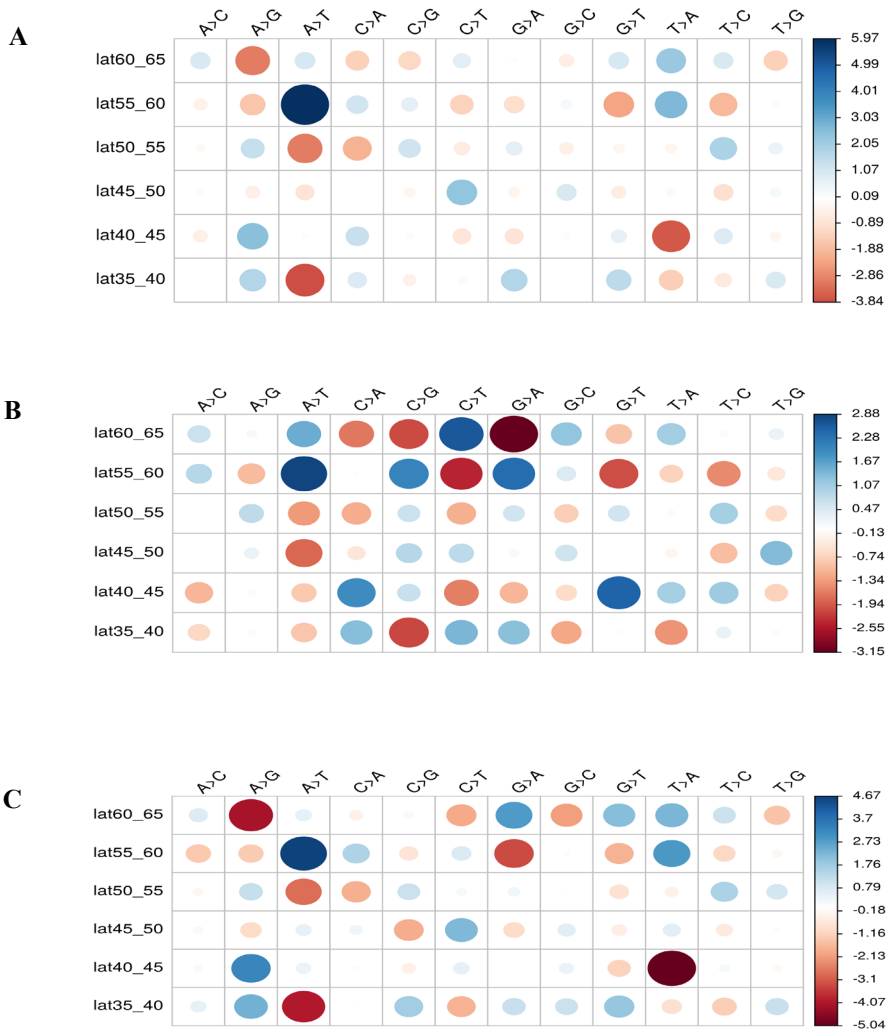

Figure 3. The pattern of Pearson's chi-square residuals for 12 nucleotide substitutions within 6 latitude intervals in **A**: Total length, **B**: Coding region and **C**: Non-coding region of the PM H^+-ATPase gene family in *A. thaliana*. For a given cell, the size of the circle is proportional to the amount of the cell contribution. Positive residuals are in blue, which present positive association between the corresponding genetic variants and latitudes. Negative residuals are in red, which present negative association between the corresponding genetic variants and latitudes.

Table 6. Contribution in percentage considering the total length, coding and non-coding region of the genes encoding the PM H^+ ATPase pumps in *A. thaliana*

A) Total length of the genes

	A>C	A>G	A>T	C>A	C>G	C>T	G>A	G>C	G>T	T>A	T>C	T>G	Sum
Lat60_65	0.60	5.51	0.64	1.13	0.90	0.36	0.01	0.18	0.64	2.93	0.55	1.21	14.66
Lat55_60	0.12	1.56	21.16	0.85	0.29	1.11	0.64	0.04	3.24	4.19	2.26	0.01	**35.48**
Lat50_55	0.02	1.18	5.69	2.58	0.90	0.24	0.27	0.14	0.06	0.06	2.07	0.14	13.33
Lat45_50	0.01	0.17	0.45	0.00	0.05	3.26	0.07	0.57	0.18	0.02	0.55	0.05	5.39
Lat40_45	0.17	3.62	0.01	1.07	0.01	0.41	0.44	0.01	0.23	8.00	0.43	0.06	14.46
Lat35_40	0.00	1.84	8.74	0.49	0.11	0.02	1.85	0.00	1.47	1.31	0.29	0.57	16.69
Sum	0.93	**13.88**	**36.68**	6.12	2.27	5.40	3.28	0.95	5.80	**16.50**	6.16	2.03	100.00

B) Coding region of the genes

	A>C	A>G	A>T	C>A	C>G	C>T	G>A	G>C	G>T	T>A	T>C	T>G	Sum
Lat60_65	0.42	0.01	2.15	2.37	3.69	6.11	8.66	1.38	0.73	1.04	0.01	0.07	26.63
Lat55_60	0.69	0.86	7.24	0.00	3.83	5.02	5.12	0.21	3.59	0.45	1.95	0.15	**29.10**
Lat50_55	0.00	0.57	1.59	1.13	0.40	1.11	0.32	0.51	0.32	0.01	0.95	0.30	7.20
Lat45_50	0.00	0.07	2.90	0.15	0.69	0.59	0.01	0.35	0.00	0.03	0.83	1.62	7.25
Lat40_45	0.95	0.01	0.59	3.38	0.43	2.25	0.96	0.30	5.58	0.97	1.12	0.44	16.95
Lat35_40	0.36	0.01	0.66	1.60	3.87	1.79	1.51	1.26	0.00	1.74	0.07	0.00	12.88
Sum	2.41	1.53	**15.13**	8.63	**12.92**	**16.86**	**16.57**	4.01	**10.22**	4.24	4.92	2.58	100.00

C) Non-coding region of the genes

	A>C	A>G	A>T	C>A	C>G	C>T	G>A	G>C	G>T	T>A	T>C	T>G	Sum
Lat60_65	0.27	8.41	0.15	0.07	0.02	1.76	4.00	2.17	2.27	2.56	0.57	0.97	23.20
Lat55_60	0.85	0.83	10.58	1.15	0.27	0.31	5.35	0.01	1.47	4.14	0.50	0.02	**25.48**
Lat50_55	0.02	0.66	3.75	1.54	0.55	0.02	0.05	0.00	0.30	0.07	1.19	0.40	8.55
Lat45_50	0.02	0.44	0.12	0.06	1.67	2.48	0.43	0.20	0.12	0.20	0.15	0.00	5.90
Lat40_45	0.02	5.25	0.10	0.00	0.08	0.15	0.00	0.10	0.61	12.32	0.02	0.01	18.65
Lat35_40	0.12	2.78	7.71	0.00	1.47	1.43	0.63	0.57	1.80	0.35	0.76	0.62	18.22
Sum	1.29	**18.36**	**22.41**	2.83	4.06	6.15	**10.46**	3.05	6.55	**19.63**	3.19	2.02	100.00

Totally, the remarkable association with latitude intervals belonged to A>T, T>A and A>G in total length of the genes (Figure 4B). Furthermore, C>T, G>A, A>T, C>G and G>T were the most dependent substitutions to the latitude intervals in the coding region (Figure 4D), while A>T, T>A, A>G and G>A were substantially associated with latitude intervals in the non-coding region (Figure 4F). Also, the results showed that the latitude interval 55-60°N had the most certain association with the nucleotide substitution (Figures 4A, 4C and 4E).

Additionally, high-impact SNPs in the H^+-ATPase gene family were identified, and the results of coding region showed that 2 missense substitutions occurred at the stop codon of *AHA12* (Table 7). Consequently, the stop codon was lost within coding region of *AHA12* with only 2 accessions that were located in the 60-65° latitude interval. Furthermore, 14 nonsense substitutions were determined in the coding region of *AHA4*, *9*, and *11* resulted in premature stop codons (Table 7). The nonsense substitutions were found in 81 accessions that were scattered in all latitude intervals and the highest number of accessions appeared in the 40-45° latitude interval. In the non-coding region, nucleotide substitutions impacted the 8 acceptor splicing sites as well as 3 donor splicing sites of the 6 genes *AHA3*, *4*, *8*, *9*, *11* and *12* (Table 7). Nucleotide substitutions in acceptor and donor sites were identified in 15 accessions which were spread over all the latitude intervals where the greatest number of accessions was found in the 40-45° latitude interval.

SNPs are the most abundant source of genetic variation as a useful tool to study association, especially in available fully sequenced genomes [26]. Owing to the SNPs importance, we defined a geospatial paradigm of nucleotide substitution within the PM H^+-ATPase gene family in *A. thaliana*. In the present study, a paradigm emerged based on chi-square method as two main factors. The first factor was six latitude intervals from 35 to 65° in the Northern Hemisphere representing thermal gradient in large scale (5 degrees of latitude) and second factor was genetic variants as single nucleotide substitutions from the 1001 Genomes project. The results displayed a significant association and consequently, the null hypothesis was rejected. Therefore, this meant that nucleotide substitution within the H^+-ATPase gene family was dependent on different latitude intervals. Actually, 3657 SNPs were identified inside the window of 60.949 Kb of *A. thaliana* genome including 12 H^+-ATPase genes of which 41.2% occurred in coding regions and 58.8% occurred in non-coding regions.

Considering the total length of the genes, the most recognizable patterns were identified in A>T, T>A and A>G. Both A>T and T>A elucidated a positive association with accessions that were located in higher latitudes (55-65°N) and negative association with lower latitude (35-40°N). Particularly, A>T and T>A were abundant in the higher latitudes in spite of being rare in lower latitudes. The behaviour of A>G was vice versa in comparison to both A>T and T>A and frequently presented in lower latitudes but was rare in higher latitudes. The explained patterns depicted that these nucleotide substitutions were specifically related to large-scale latitude intervals.

Mean annual surface temperature is generally warmer in lower latitudes and cooler in higher latitudes because the radiation angle varies with latitudes [41]. Based on latitudinal gradient of temperature, the six studied latitude intervals were categorized into three groups. The first group was considered as cold climate including the higher latitudes 55-60° and 60-65°, the second one was classified as moderate climate containing middle latitudes; 45-50° and 50-55°, and the third class with both 35-40° and 40-45° lower latitude intervals was classified as warm climate. Considering the geospatial pattern of SNP types, the most significant substitutions were identified in cold, then warm and moderate climates, respectively. In particular, extreme climates, such as cold and warm climates, shared the greatest association with nucleotide substitutions within H^+-ATPase genes. PM H^+-ATPase genes play an important role in the dynamics of plants during thermal changes like high and low temperatures in order to improve the adaptation mechanisms [22, 23, 25].

It can be demonstrated that the high frequency of nucleotide substitutions within H^+-ATPase genes in cold and warm climates is related to plant adaptation strategies to various thermal

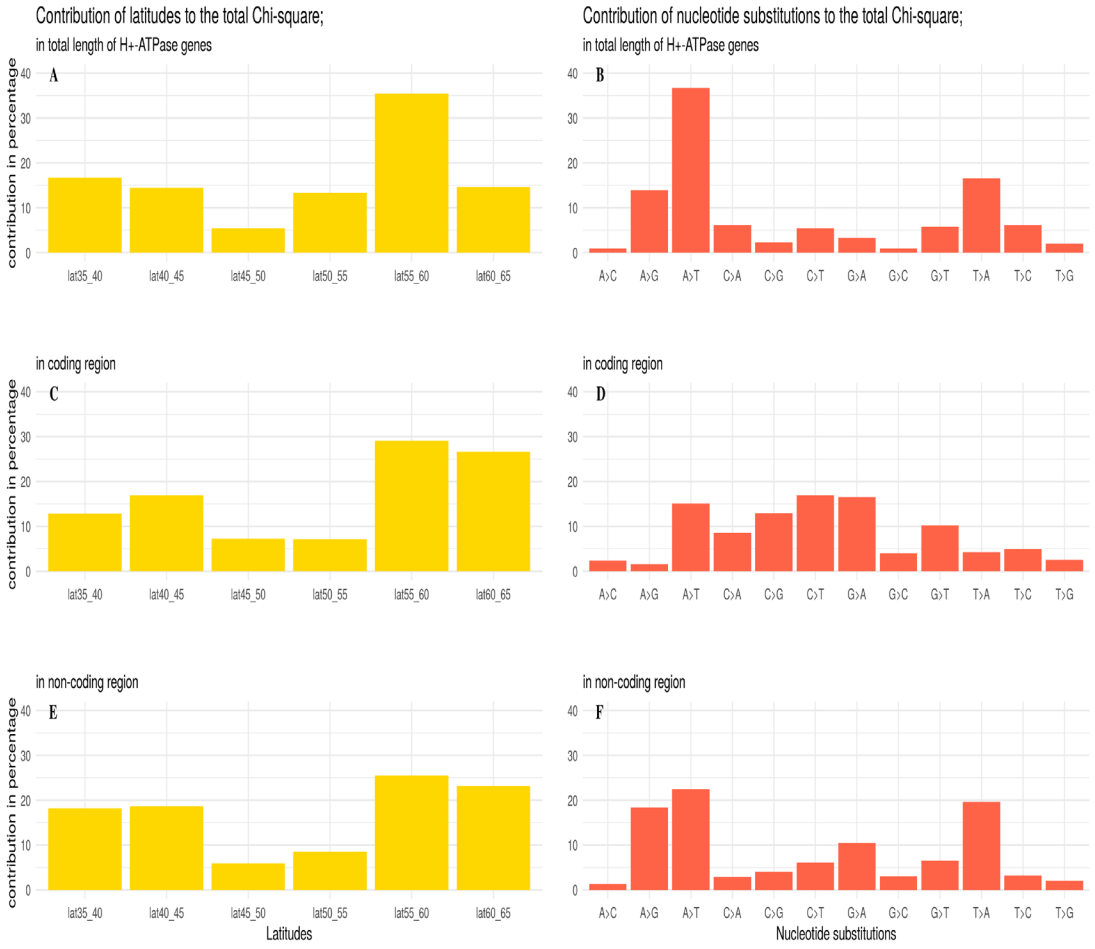

Figure 4. The cumulative contribution in percentage to the total value of chi-square, where A, C and E indicate contribution of latitudes and B, D and F denote the contribution of nucleotide substitutions in the total length, coding region and non-coding region of the PM H^+-ATPase gene family in *A. thaliana*, respectively. The bar-plots represent the marginal values of Table 6.

conditions. More specifically, both A>T and T>A substitutions are dominant in cold climate and these substitutions are vital to cold adaptation, while A>G substitutions are abundant in warm climate and they are the major candidate for driving strategies for warm climate adaptation. It should be noted that, as demonstrated by research in *Homo sapiens*, SNPs significantly correlated with winter adaptation and share a plausible contribution to adaptation [42].

An interesting result that was highlighted in the coding regions was that nucleotide substitutions such as C>T had a strong positive association with 60-65° latitude whilst showing a strong negative association with 55-60°. Both mentioned latitude intervals were recognized as higher latitudes and had a difference of only 5°. Also, such patterns were observed for C>G and G>A with strong positive and negative associations in higher latitudes (Figure 3B). Such

Table 7. A report of the number of high-impact SNPs in the H^+-ATPase gene family. Missense substitutions denoted as MS and Nonsenses as NS. SA is Splice-Acceptor which substitution occur at 3′- AG splice site while, SD is Splice-Donor which substitution occur at 5′- GT splice site

Region	Gene	A>C	A>T	C>A	C>T	G>A	G>T	Sum
Coding	*AHA12*	2 (MS)	1 (NS)	1 (NS)	4 (NS)	1 (NS)	1 (NS)	10
	AHA4	0	0	0	1 (NS)	1 (NS)	1 (NS)	3
	AHA9	0	0	1 (NS)	1 (NS)	1 (NS)	0	3
Non-coding	*AHA11*	0	0	0	0	1 (SA)	0	1
	AHA12	0	1 (SA)	0	0	2 (SA)	1 (SA)	4
	AHA3	0	0	0	0	1 (SA)	0	1
	AHA4	0	0	0	0	2 (SD)	0	2
	AHA8	0	1 (SA)	0	0	0	0	1
	AHA9	0	0	0	1 (SD)	1 (SA)	0	2
Sum	-	2	3	2	7	10	3	27

observations can prove the hypothesis of the present study, which is that nucleotide substitutions are not only associated with latitude but are also highly sensitive to intervals and temperature range. The changes affect PM H^+-ATPase as the part of membrane proteins family which is involved in functions that are crucial to the appropriate growth of plant and adaption mechanisms under various environmental conditions [43]. This is consistent with the results studied by Kim *et al.* [25] that the H^+-ATPase action dynamically changes during thermal intervals in order to enhance adaptation of plant strategies. Geospatial patterns in the non-coding regions are very similar to the patterns for the total length of the gene. This possibly related to the fact that non-coding regions share a higher number (2151) in comparison to coding regions (1506) of the total SNPs (3657).

From 3657 SNPs in total, 16 and 11 high-impact SNPs were identified in coding and non-coding regions, respectively. In coding regions, 14 nonsense nucleotide substitutions resulted in the presence of premature stop codons. A premature stop codon occurring within the normal coding sequence leads to the creation of an incomplete protein and consequently, truncated proteins are formed. Moreover, 2 missense substitutions including tAa>tCa that stop codon altered to serine amino acid as well as taA>taC substitutions that stop codon altered to tyrosine were identified within *AHA12* coding region in only 2 accessions. *AHA12* is considered as a truncated gene in *A. thaliana* Col-0 reference genome while these 2 types of missense substitutions make it elongated. Introns are removed during the splicing process by cleavage at splice sites. Splice sites are conserved sequences identified as 5′- GT (Splice donor site) and 3′- AG (Splice acceptor site) occurring at the both ends of introns [44]. Nonetheless, different splice sites could be used for the splicing process apart from GT or AG [45]. Nucleotide substitution in non-coding regions within splice sites of introns results in alternative splicing through exon skipping or inclusion of the introns and alters gene coding sequence [46]. Thus, the final return of high-impact SNPs within H^+-ATPase genes can be a truncated, elongated or alternative spliced gene which was observed in 98 accessions of the 1001 Genomes project. As reported in Table 3, the highest frequency of accessions that conveyed splice-variants and nonsense substitutions were found in warm climate areas and only 2 accessions bearing missense substitutions in H^+-ATPase genes appeared in cold climate areas.

Generally, substitutions such as A>T and T>A, and higher latitude intervals such as 55-60° and 60-65° that were cold climate areas made a large contribution to the chi-square statistics. In fact, 55-60°/ A>T was the greatest contributing cell to the value of chi-square. Therefore, 55-60° latitude interval and A>T substitution were the most influential variables in this study.

4. Conclusions

The explicit findings of the study showed T>A substitution contributed to the value of chi-square more powerful than others, and as well it had a robust significant positive association with cold climate and a negative association with warm climate. Actually, it can be concluded that PM H^+-ATPase gene family utilize the T>A substitution more than other substitution types as a general strategy through the adaptation to cold climate. High-impact SNPs resulted in truncated, elongated and alternative splicing proteins, and it can be concluded that such high-impact SNPs act as regulators in the dynamics of the H^+-ATPase gene family in order to facilitate adaption mechanisms under various environmental conditions. In spite the fact that this study found a reasonable signal of adaptation into thermal changes in the level of nucleotide substitution, further investigation in the area of gene functional annotation is needed to better explain the role of the PM H^+-ATPase family in adaptation mechanisms.

References

[1] Axelsen, K.B. and Palmgren, M.G., 1998. Evolution of substrate specificities in the P-type ATPase superfamily. *Journal of Molecular Evolution*, 46(1), 84-101.

[2] Axelsen, K.B. and Palmgren, M.G., 2001. Inventory of the superfamily of P-type ion pumps in *Arabidopsis*. *Plant Physiology*, 126(2), 696-706.

[3] Serrano, R. 1989. Structure and function of plasma membrane ATPase. *Annual Review of Plant Physiology and Plant Molecular Biology*, 40, 61-94.

[4] Sondergaard, T.E., Schulz, A. and Palmgren, M.G., 2004. Energization of transport processes in plants. Roles of the plasma membrane H^+-ATPase. *Plant Physiology*, 136(1), 2475-2482.

[5] Nordborg, M. and Weigel, D., 2008. Next-generation genetics in plants. *Nature*, 456(7223), 720-723.

[6] Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D., 2017. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal*, 89(4), 789-804.

[7] Michelet, B. and Boutry, M., 1995. The plasma membrane H^+-ATPase a highly regulated enzyme with multiple physiological functions. *Plant Physiology*, 108,1-6.

[8] Ballesteros, E., Kerkeb, B., Donaire, J.P. and Belver, A., 1998. Effects of salt stress on H^+-ATPase activity of plasma membrane-enriched vesicles isolated from sunflower roots. *Plant Science*, 134,181-190.

[9] Hamilton, C.A., Good, A.G. and Taylor, G.J., 2001. Induction of vacuolar ATPase and mitochondrial ATP synthase by aluminum in aluminum-resistant cultivar of wheat. *Plant Physiology*, 125, 2068-2077.

[10] Liang, Y.C., Zhang, W.H., Chen, Q. and Ding, R.X., 2005. Effects of silicon on H^+- ATPase and H^+-PPase activity, fatty acid composition and fluidity of tonoplast vesicles from roots of salt-stressed barley (*Hordeum vulgare L.*). *Environmental and Experimental Botany*, 53, 29-37.

[11] Palmgren, M.G., Bækgaard, L., López-Marqués, R.L. and Fuglsang, A.T., 2011. Plasma Membrane ATPases. In: A.S. Murphy, B. Schulz and W.A. Peer, eds. *The Plant Plasma Membrane*. Berlin: Springer, pp. 177-192.

[12] Haruta, M., Gray, W.M. and Sussman, M.R., 2015. Regulation of the plasma membrane proton pump (H+-ATPase) by phosphorylation. *Current Opinion in Plant Biology*, 28(1), 68-75.

[13] Palmgren, M.G., 2001. Plant plasma membrane H^+-ATPases: powerhouses for nutrient uptake. *Annual Review of Plant Biology*, 52(1), 817-845.

[14] Ueno, K., Kinoshita, T., Inoue, S.I., Emi, T. and Shimazaki, K.I., 2005. Biochemical characterization of plasma membrane H+-ATPase activation in guard cell protoplasts of *Arabidopsis thaliana* in response to blue light. *Plant and Cell Physiology*, 46(6), 955-963.

[15] Baxter, I.R., Young, J.C., Armstrong, G., Foster, N., Bogenschutz, N., Cordova, T. and Harper, J.F., 2005. A plasma membrane H+-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7), 2649-2654.

[16] Santi, S. and Schmidt, W., 2009. Dissecting iron deficiency-induced proton extrusion in *Arabidopsis* roots. *The New Phytologist*, 183, 1072-1084.

[17] Caesar, K., Elgass, K., Chen, Z., Huppenberger, P., Witthöft, J., Schleifenbaum, F. and Harter, K., 2011. A fast brassinolide-regulated response pathway in the plasma membrane of *Arabidopsis thaliana*. *The Plant Journal*, 66(3), 528-540.

[18] Janicka-Russak, M., 2011. Plant plasma membrane H^+-ATPase in adaptation of plants to abiotic stresses. Abiotic stress response in plants-physiological, biochemical and genetic perspectives, 1, 197-218.

[19] White, F., Cooke, D., Earnshaw, M., Clarkson, D. and Burden, R., 1990. Does plant growth temperature modulate the membrane composition and ATPase activities of tonoplast and plasma membranes fractions from rye roots? *Phytochemistry*, 29, 3385-3393.

[20] Kasamo, K., 1988. Response of tonoplast and plasma membrane ATPase in chilling sensitive and insensitive rice (*Oryza sativa* L.) culture cells to low temperature. *Plant and Cell Physiology*, 29(7), 1085-1094.

[21] Martz, F., Sutinen, M., Kiviniemi, S. and Palta, J., 2006. Changes in freezing tolerance, plasma membrane H+-ATPase activity and fatty acid composition in *Pinus resinosa* needles during cold acclimation and de-acclimation. *Tree Physiology*, 26, 783-790

[22] Muzi, C., Camoni, L., Visconti, S. and Aducci, P., 2016. Cold stress affects H^+-ATPase and phospholipase D activity in *Arabidopsis*. *Plant Physiology and Biochemistry*, 108, 328-336.

[23] Janicka-Russak, M., Kabala, K., Wdowikowska, A. and Klobus, G., 2012. Response of plasma membrane H^+-ATPase to low temperature in cucumber roots. *Journal of Plant Research*, 125(2), 291-300

[24] Liu, Y., Liu, H., Pan, Q., Yang, H., Zhan, J. and Huang, W., 2009. The plasma membrane H +-ATPase is related to the development of salicylic acid-induced thermo-tolerance in pea leaves. *Planta*, 229, 1087-1098.

[25] Kim, H.S., Oh, J.M., Luan, S., Carlson, J.E. and Ahn, S.J., 2013. Cold stress causes rapid but differential changes in properties of plasma membrane H^+-ATPase of camelina and rapeseed. *Journal of Plant Physiology*, 170(9), 828-837.

[26] Leaché, A.D. and Oaks, J.R., 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 48, 69-84.

[27] Zhang, P., Tillmans, L.S., Thibodeau, S.N. and Wang, L., 2019. Single-nucleotide polymorphisms sequencing identifies candidate functional variants at prostate cancer risk loci. *Genes*, 10(7), 547, https://doi.org/10.3390/genes10070547.

[28] Konôpková, A., Krajmerová, D., Kurjak, D., Kmeť, J., Pšidová, E., Kučerová, J., Hrivnák, M., Longauer, R., Ditmarová, Ľ. and Gömöry, D., 2019. Nucleotide polymorphisms associated with climate and physiological traits in silver fir (*Abies alba* Mill.) provenances. *Flora*, 250, 37-43.

[29] Krajmerová, D., Hrivnák, M., Ditmarová, Ľ., Jamnická, G., Kmeť, J., Kurjak, D. and Gömöry, D., 2017 Nucleotide polymorphisms associated with climate, phenology and physiological traits in European beech (*Fagus sylvatica* L.) *New Forests*, 48(3), 463-477.

[30] Bali, S., Robinson, B.R., Sathuvalli, V., Bamberg, J. and Goyer, A., 2018. Single nucleotide polymorphism (SNP) markers associated with high folate content in wild potato species. *PloS One*, 13(2), e0193415, https://doi.org/10.1371/journal.pone.0193415.

[31] Brookes, A.J., 1999. The essence of SNPs. *Gene*, 234(2), 177-186.
[32] Meinke, D.W., Cherry, J. M., Dean, C., Rounsley, S.D., and Koornneef, M., 1998. *Arabidopsis thaliana*: a model plant for genome analysis. *Science*, 282(5389), 662-682.
[33] Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K.M., Cao, J., Chae, E., Dezwaan, T.M., Ding, W. and Ecker, J.R., 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, 166(2), 481-491.
[34] McHugh, M.L., 2013. The chi-square test of independence. *Biochemia Medica*, 23(2), 143-149.
[35] Sharpe, D., 2015. Chi-square test is statistically significant: Now what? *Practical Assessment, Research, and Evaluation*, 20(1), 8, https://doi.org/10.7275/tbfa-x148.
[36] Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y. and Zemla, J., 2017. Package 'corrplot'. *Statistician*, 56, 316-324.
[37] Pebesma, E. 2018. Simple features for R: Standardized support for spatial vector data. *The R Journal*, 10(1), 439-446.
[38] Bivand, R., Rundel, C., Pebesma, E., Stuetz, R., Hufthammer, K.O. and Bivand, M.R., 2017. *Package 'rgeos'*. The Comprehensive R Archive Network (CRAN).
[39] South, A., 2017. *rnaturalearth: World Map Data from Natural Earth*. R package version 0.1. 0. The R Foundation.
[40] Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
[41] Legates, D.R. and Willmott, C.J., 1990. Mean seasonal and spatial variability in global surface air temperature. *Theoretical and Applied Climatology*, 41(1), 11-21.
[42] Igoshin, A.V., Gunbin, K.V., Yudin, N.S. and Voevoda, M.I., 2019. Searching for signatures of cold climate adaptation in TRPM8 gene in populations of East Asian ancestry. *Frontiers in Genetics*, 10, 759, https://doi.org/10.3389/fgene.2019.00759.
[43] Morsomme, P. and Boutry, M., 2000. The plant plasma membrane H+-ATPase: structure, function and regulation. *Biochimica et Biophysica Acta*, 1465(1-2), 1-16.
[44] Herzel, L., Ottoz, D.S.M., Alpert, T. and Neugebauer, K.M., 2017. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nature Reviews Molecular Cell Biology*, 18(10), 637-650.
[45] Sibley, C.R., Blazquez, L. and Ule, J., 2016. Lessons from non-canonical splicing. *Nature Reviews Genetics*, 17(7), 407-421.
[46] Anna, A. and Monika, G., 2018. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics*, 59(3), 253-268.