



Article

# How Machine Learning Methods Helped Find Putative Rye Wax Genes Among GBS Data

Magdalena Górska <sup>1</sup>, Jan Bińkowski <sup>1</sup>, Natalia Lenarczyk <sup>1</sup>, Anna Bienias <sup>1</sup>,  
Agnieszka Grądzielewska <sup>2</sup>, Ilona Czyczyło-Mysza <sup>3</sup> , Kamila Kapłoniak <sup>3</sup> ,  
Stefan Stojalowski <sup>1</sup> and Beata Myśków <sup>1,\*</sup>

<sup>1</sup> Department of Plant Genetics, Breeding and Biotechnology, West-Pomeranian University of Technology, Szczecin, ul. Słowackiego 17, 71–434 Szczecin, Poland; Magdalena.Goralska@zut.edu.pl (M.G.); jan.binkowski96@gmail.com (J.B.); ln21018@zut.edu.pl (N.L.); anna.bienias@zut.edu.pl (A.B.); stefan.stojalowski@zut.edu.pl (S.S.)

<sup>2</sup> Institute of Plant Genetics, Breeding and Biotechnology, University of Life Sciences in Lublin, ul. Akademicka, 20–950 Lublin, Poland; agnieszka.gradziewska@up.lublin.pl

<sup>3</sup> Polish Academy of Sciences, The Franciszek Górski Institute of Plant Physiology, Niezapominajek 21, 30–239 Kraków, Poland; i.czyczyl@ifr-pan.edu.pl (I.C.-M.); kaploniak@ifr-pan.edu.pl (K.K.)

\* Correspondence: bmyskow@zut.edu.pl

Received: 28 July 2020; Accepted: 7 October 2020; Published: 12 October 2020



**Abstract:** The standard approach to genetic mapping was supplemented by machine learning (ML) to establish the location of the rye gene associated with epicuticular wax formation (glaucous phenotype). Over 180 plants of the biparental F<sub>2</sub> population were genotyped with the DArTseq (sequencing-based diversity array technology). A maximum likelihood (MLH) algorithm (JoinMap 5.0) and three ML algorithms: logistic regression (LR), random forest and extreme gradient boosted trees (XGBoost), were used to select markers closely linked to the gene encoding wax layer. The allele conditioning the nonglaucous appearance of plants, derived from the cultivar Karlikovaja Zelenostebelnaja, was mapped at the chromosome 2R, which is the first report on this localization. The DNA sequence of DArT-Silico 3585843, closely linked to wax segregation detected by using ML methods, was indicated as one of the candidates controlling the studied trait. The putative gene encodes the ABCG11 transporter.

**Keywords:** ATP-binding cassette (ABC) transporters; fatty acid desaturase (FAD), genetic map; glaucousness; large-scale sequence-based markers; *Secale cereale* L.

## 1. Introduction

Objective of the Study The main aim of the study was to establish the localization of an unknown rye gene causing the waxless plant character (Figure 1) on a novel, high density genetic map constructed with DArTseq (sequencing-based diversity array technology) used for genotyping the biparental F<sub>2</sub> population. The research was based on the standard approach to linkage genetic mapping methods in combination with ML (three algorithms) to find markers closely linked to the gene of interest (putative candidate genes).



**Figure 1.** Stems and leaves of parental components of rye mapping population BK2: glaucous inbred line AK1 (A) and nonglaucous AKZ (B).

**Plant Waxes** The surface of primary above-ground plant organs is covered with a cuticle, a protective lipid structure sealing the tissues and insulating them from the effects of various unfavorable environmental factors. The cuticle serves as the major barrier preventing nonstomatal water loss and helps to protect plant surfaces from pathogens and ultraviolet radiation [1]. Plant cuticles consist of cutin, an insoluble polyester polymer lattice, and of soluble cuticular waxes. The waxes are either deposited within the cutin matrix (intracuticular wax) or accumulate on its surface as epicuticular wax [2].

The external appearance of plants is largely dependent on epicuticular wax structures, which are usually deposited on the plant's surfaces as crystalloid forms or films. The bluish color of plant organs, caused by the scattering of light by wax crystals, is referred to as glaucousness, and the opposite form is referred to as nonglaucous, waxless or glossy [3]. The thickness of epicuticular wax layers, their composition and the shape of wax crystals/microstructures depend on the species [4], developmental stage and anatomical location [5,6]. In addition, the presence of wax is modified by environmental conditions [7].

The chemical composition of wax is very complex. In most cases, the majority of compounds comprising the cuticular wax are derived from very long-chain fatty acids (VLCFAs; C20–C34), including alkanes, aldehydes, primary and secondary alcohols, ketones and esters [2]. Waxy epicuticular layers in cereals also contain others various hydrophobic substances including alkyl and alkenyl resorcinols [8].

Moreover, since the lipid precursors of the cuticle are synthesized within epidermal cells and must pass through the hydrophilic cell wall to the developing cuticle, molecules carrying these compounds play an important role in cuticle/wax secretion. The group of proteins responsible for the transport of both wax and cutin precursors across the plasma membrane has been shown to depend on ATP-binding cassette (ABC) transporters [2].

**Wax Genes** The presence of epicuticular wax crystals, which give the plant a glaucous appearance, has provided an easy way to detect mutants with wax crystal formation disorders. Such mutants, called eceriferum (*cer*), bloomless (*bm*) and glossy (*gl*), were the material for numerous molecular analyses, which led to an increasingly complete view of the wax biosynthesis pathway [2].

Using genetic methods, many wax genes, such as *CER1* [9], *CER2* [10], *CER3* [11], *CER4* [12], *CER5* [13], *CER6* [14,15], *CER7* [16], *CER8* [17], *CER9* [18,19] and *CER10* [20,21] have been identified in *Arabidopsis* and well characterized at the molecular level. Nonglaucous mutants of other species have also been tested quite intensively. The *BrWax1* and *BrCER4* genes in *Brassica rapa*

and *BoWax1* in *B. oleracea*, that control the glossy phenotype, have been mapped [22–24]; the maize (*Zea mays*) GLOSSY1 (GL1) gene was molecularly cloned and characterized [25] and several rice genes encoding for the wax layer have also been studied [26]. Among the studies associated with wax synthesis in wheats, those regarding genes from chromosome group 2, controlling  $\beta$ -diketone production, predominate [3,27–31]. The barley *Cer-cqu* loci that determines  $\beta$ -diketone biosynthesis and glaucousness have also been studied [32,33].

Among the genes conditioning wax formation, those associated with the transport of wax precursors belonging to the ABC family play an important role. Recent studies showed that members of the G subfamily are critical for FAs (fatty acids) export. *ABCG11/WBC11/DSO*, *ABCG12/CER5*, *ABCG13*, *ABCG32/PEC1* and *OsABCG31/HvABCG31/gl13* [13,34–41] were reported to be required for lipid export.

Rye WaxesRye (*Secale cereale* L.), belonging to the *Poaceae* family, is one of the important cereal crops, grown mainly in the north-east part of Europe. Its grain is used for animal feed, alcohol production and bread flour. Rye has a number of positive attributes, such as outstanding cold hardiness, excellent drought tolerance and strong disease resistance. Some of these attributes may be due to the intense waxy bloom on the stem and leaf sheath. Few studies of the rye cuticle have concentrated on the chemical composition of cuticular waxes, [1,42]. The study on the total wax mixture from both sides of the leaves indicated the presence of primary alcohols (71%), alkyl esters (11%), aldehydes (5%) and small amounts (<3%) of alkanes, steroids, secondary alcohols, FAs and unknowns [1].

The genetic basis for creating the waxy epidermal coating is much less known. Only two genes responsible for the waxy film on straw, leaves and spikes have been localized on the genetic map of rye [43,44]. The recessive mutations determining the nonglaucous plant character have been mapped on chromosome 4R (*w*) and on the distal part of chromosome 7RL (*wa1*), which is translocated with respect to wheat and barley and is homologous to a distal region of the group 2 chromosomes (Rognli et al., 1992 in [43]). Both localizations are not very precise due to the low density of the applied genetic map.

Genetic MappingGenetic linkage mapping in segregating populations is one of the classic approaches to gain insights into the genetic control of key characteristics in a species. Continuous progress in the area of DNA technology, and powerful genotyping systems, have allowed the development of genomics, including genetic linkage mapping ([www.maizegdb.org](http://www.maizegdb.org)), [45–51]. Genetic maps are a source of valuable information for both cognitive genetics and practical applications. Markers linked to desirable genes or QTLs are important tools for improving plants in marker assisted breeding (MAB), via marker-assisted selection (MAS) and marker-assisted backcrossing (MABC). Map design is the first step towards positional cloning of genes responsible for a particular phenotype, with ultra-dense genetic maps required for this type of research [52,53].

#### Rye Genetic Maps—Genes and QTL

The first genetic map of all seven rye chromosomes was created for over a quarter of a century ago [54] with RFLP (restriction fragment length polymorphism) and isozyme markers. Subsequent maps were constructed using the emerging marker techniques: PCR (polymerase chain reaction), microarrays and sequencing technologies, but their number is still smaller than in other important cereals. The first map created using high-throughput technology was constructed in 2011 [46]. The development of the transcript map was the next, important step in expanding the knowledge of the rye genome [55]. A consensus map of four populations included 2886 genes out of 3588 markers (80%). The map presented here is one of the few created with the use of GBS (genotyping by sequencing) markers [49,56,57]. DArTseq, like DArT, is a method in which genome complexity reduction is obtained using a combination of two restriction enzymes, at least one of which is methylation-sensitive, directing the analysis to the hypomethylated, gene-rich genome regions [49]. Compared to DArT, the new platform gives an increase in the number of available markers and provides codominant markers SNPs (single nucleotide polymorphisms), which improve the informativeness of the map. The abundance of markers, and the high probability of their localization in gene regions, make it

possible to find markers closely related to an unknown gene as a good starting point for the search for potential candidate genes.

Previous rye maps have contributed significantly to the advancement of comparative and evolutionary research [54,55,58]. They were also used to find out the localization of genes responsible for, amongst other traits plant height/dwarfism [54,55,58–61], fertility restoration [62–66], rust resistance [67] and heading date [68], and were the basis to detect QTLs (quantitative trait loci) for different morphology/agronomic traits [69–72], alpha-amylase activity and preharvest sprouting [73–75].

Phenotypic glaucousness markers have so far been located only on older generation maps, poorly saturated with markers [43,44]. The recessive mutations determining the nonglauous plant character was mapped on 4R (*w*) and 7RL (*wa1*).

The procedure used in this work consisted of constructing a highly saturated map using a moderate size mapping population, and selecting markers to create, in the future, a precise, fine map of the region with the gene responsible for a studied feature, using a larger mapping population. With such an approach it is important to select the right markers for further analysis, which can be problematic due to insufficient map precision. Genetic map design depends on many factors such as type and size of population, marker system and statistical method. It has been found that more accurate maps are obtained with F<sub>2</sub>-codominant and recombinant inbred lines than with backcrosses, double haploids and F<sub>2</sub>-dominant populations, and that a sample size of 200 individuals is sufficient for the construction of reasonably accurate maps [52].

Our experience with a population of less than 200 genotypes, and thousands of both dominant and codominant markers, shows that with such a large amount of data it is difficult to select the right set of loci to create proper linkage groups and to get the correct order of markers. To overcome this difficulty and improve marker selection, we decided to apply and test ML as a support method.

Machine Learning (ML) Applications Collecting large amounts of data, as well as hardware and software progress in recent years, have contributed significantly to the increased use of ML in various fields of science. ML is becoming a powerful approach to data analysis, especially in human diagnostics and medicine. The early application of ML to population genetics demonstrated that it outperformed traditional approaches [76]. ML is becoming more and more common in plant genetics and breeding, mainly for genotype-phenotype relationship predictions and genomic selection. Researchers proposed ML algorithms as valuable alternatives to traditional statistical techniques applied in the breeding of maize [77], wheat [78] and soybean [79]. Due to advances in genomics, leading to the identification of numerous DNA markers, ML can be a complement to classical methods of genomic analysis.

## 2. Results

### 2.1. Effect of Lack of Wax Cover on Phenotypic Features in the Mapping Population

Decreased plant height and spikelet number per spike in two generations of the rye mapping population BK2 were observed in waxy plants (Table 1). Although there was no significant difference in length of the ear, the smaller number of spikelets resulted in a smaller density of ear (lower compactness) in this group of plants. Similar relationships were observed for two parental lines (data not shown). The difference in flowering date between glaucous and nonglauous plants was also statistically significant, although small (Table 1).

**Table 1.** The significance of differences between the group of waxy (1) and waxless (2) plants of mapping population BK2, demonstrated using the *t*-test.

Trait	Generation	Mean		Standard Deviation		Sample Number		<i>t</i> -Statistic	Degrees of Freedom (df)	Probability Level ( <i>p</i> -Value)
		1	2	1	2	1	2			
PH *	F <sub>2</sub>	116.10	127.07	29.10	26.34	200	68	-2.75	266	0.006
	F <sub>3</sub>	96.40	102.25	26.19	28.44	248	128	-1.99	374	0.047
TN	F <sub>2</sub>	3.83	3.38	2.11	1.60	202	69	1.62	269	0.106
	F <sub>3</sub>	3.53	3.36	1.88	1.65	248	128	0.88	374	0.380
SL	F <sub>2</sub>	10.45	10.21	1.74	1.72	196	67	0.97	261	0.333
	F <sub>3</sub>	9.27	9.23	1.40	1.29	245	126	0.30	369	0.768
SNPS *	F <sub>2</sub>	32.84	34.27	4.84	4.97	196	67	-2.07	261	0.040
	F <sub>3</sub>	30.75	32.27	4.32	3.79	245	126	-3.34	369	0.001
CT *	F <sub>2</sub>	31.84	33.93	4.54	4.46	196	67	-3.27	261	0.001
	F <sub>3</sub>	33.41	35.34	3.73	4.38	245	126	-4.43	369	0.000
GNPS	F <sub>2</sub>	39.14	39.90	25.68	27.20	196	67	-0.20	261	0.839
	F <sub>3</sub>	21.21	21.68	17.60	17.32	245	126	-0.25	369	0.807
GWPS	F <sub>2</sub>	1.36	1.36	0.60	0.69	159	55	-0.04	212	0.965
	F <sub>3</sub>	0.64	0.57	0.39	0.41	182	106	1.41	286	0.160
TGW	F <sub>2</sub>	28.16	27.41	6.43	7.07	159	55	0.74	212	0.463
	F <sub>3</sub>	22.21	21.80	6.72	6.21	182	106	0.51	286	0.611
FD *	F <sub>3</sub>	28.61	29.01	0.87	0.89	133	50	-2.73	181	0.007

PH—plant height (cm), TN—tiller number, SL—spike length, SNPS—spikelet number per spike, CT—spike compactness (spikelet number per 10 cm), GNPS—grain number per spike, GWPS—grain weight per spike (g), TGW—thousand grain weight (g), FD—flowering data (days from May 1st), \* statistically significant differences.

## 2.2. Genotyping by Sequencing the BK2 Mapping Population with DArTseq Technology

The used GBS platform supplied 39,615 DArTseqs (28,889 Silico and 10,726 SNP). Among them 7258 were rejected from further analysis because of missing data about the parental alleles or because of monomorphic character (Table 2). Most of the markers differentiating parental lines had an unknown genomic location.

**Table 2.** DArTseq statistic in rye mapping population BK2-F<sub>2</sub>.

	Silico	SNP	Total
incomplete data	447	2914	3361
monomorphic	2630	1267	3897
polymorphic	25812	6545	32357
	including		
1R *	917	191	1108
2R *	927	208	1135
3R *	977	152	1129
4R *	873	188	1061
5R *	1086	169	1255
6R *	1165	231	1396
7R *	789	183	972
mean per chromosome	962	189	1151
unassigned **	19152	5070	24222

\*/\*\* DArTseq assigned/unassigned to chromosomes based on literature data [56,57].

### 2.3. Position of the Wax Locus in the Rye Genome Determined Using Mapping Procedure

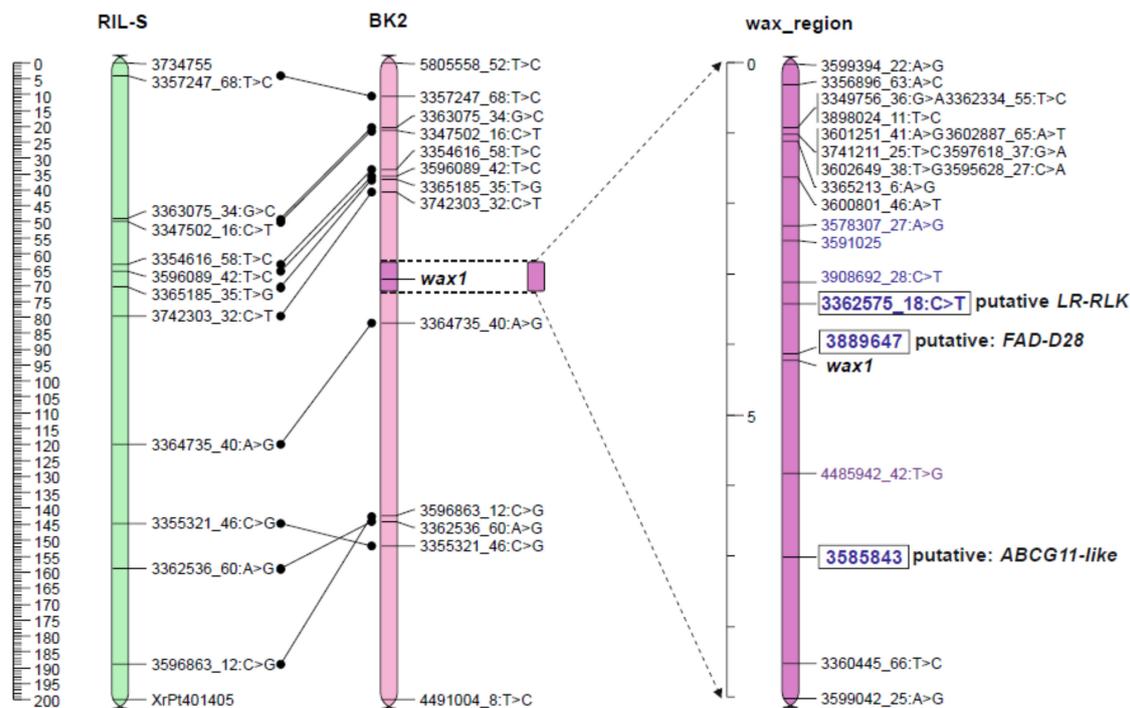
Using the information from the GBS platform, based on localization of DNA sequences reported by Bauer et al. [57] and the publication on the dense genetic map of rye containing GBS markers [56], 8,135 (6,660 Silico and 1,475 SNP) were assigned to the seven respective chromosomes. Preliminary linkage analysis (grouping command) with these seven groups suggested localization of the *wax* locus on chromosome 2R (Table S1). From the set of loci unassigned to chromosomes, listed in Table S2 (24,222: 19,152 Silico and 5,070 SNP), those with which *wax* was linked were selected. As a selection criterion, the linkage at LOD (logarithm of odds) 50 was applied in the JoinMap 5.0 software (Kyazma B.V., Wageningen, The Netherlands). Further linkage analyses was performed on the group of DArTseqs assigned to chromosome 2R, showed in Table S1 (1135: 927 Silico and 208 SNP) combined with 698 DArTseqs selected from a group of loci with a previously unknown location.

The set of 1833 DArTseqs, and one *wax* segregation recorded in two ways (dominant and codominant), were subjected to grouping analysis (Table S3). Both the dominant type of *wax* layer segregation noted in F<sub>2</sub> generation, and the codominant one (based on F<sub>3</sub> generation), were used to control the mapping results correctness.

Grouping parameters were established at the LOD range from 10 to 50. At the lowest assumed level, 1756 of 1835 loci remained linked. LOD 36 was the maximum value at which both records of *wax* segregation remain associated in one group. The trial maps were created at LOD 36 (1670 loci), LOD 45 (982 loci) and LOD 50 (882 loci). As *wax* dominant segregation fell out of the group at highly restrictive values exceeding LOD 36, it was manually attached to the linkage groups created at LOD 45 and 50.

The JoinMap 5.0 software gives two options for mapping procedure. The regression mapping algorithm (RM) simultaneously analyzes the relationships between several closely related loci, so this method is more accurate but more complex. When the number of loci is high, it requires very efficient computer equipment and is extremely time consuming. The alternative method, based on the maximum likelihood (MLH) procedure, considers relationships only between two neighbor markers in one step. This method does not allow removal of single errors in genotyping and usually gives over-estimated distances between loci but needs less effective computers and can be performed within hours, not weeks. Due to the large number of markers in linkage groups, the maps were constructed using the MLH algorithm. The maps created at LOD 36 and LOD 45 were not comparable to the

reference map (Table S4). Common loci positions were strongly displaced, which made it impossible to determine the position of the *wax* locus. The map at LOD 50 had eleven common loci with a reference map of the RIL-S population [56], and their arrangement was mostly consistent (Figure 2). At this level, both *wax* segregations remained unseparated by other markers, as was the case with maps formed from loci grouped at lower LODs. This map (Table S5) was used to select loci nearest to the *wax* locus to map the *wax* region more precisely with these markers, and those detected by ML methods. The *wax* locus was located in the middle part of chromosome 2R (Figure 2).



**Figure 2.** Position of the *wax* locus on the rye chromosome 2R of the BK2 mapping population with reference to the RIL-S map [56]. DARtseqs indicated in the ML analysis are blue.

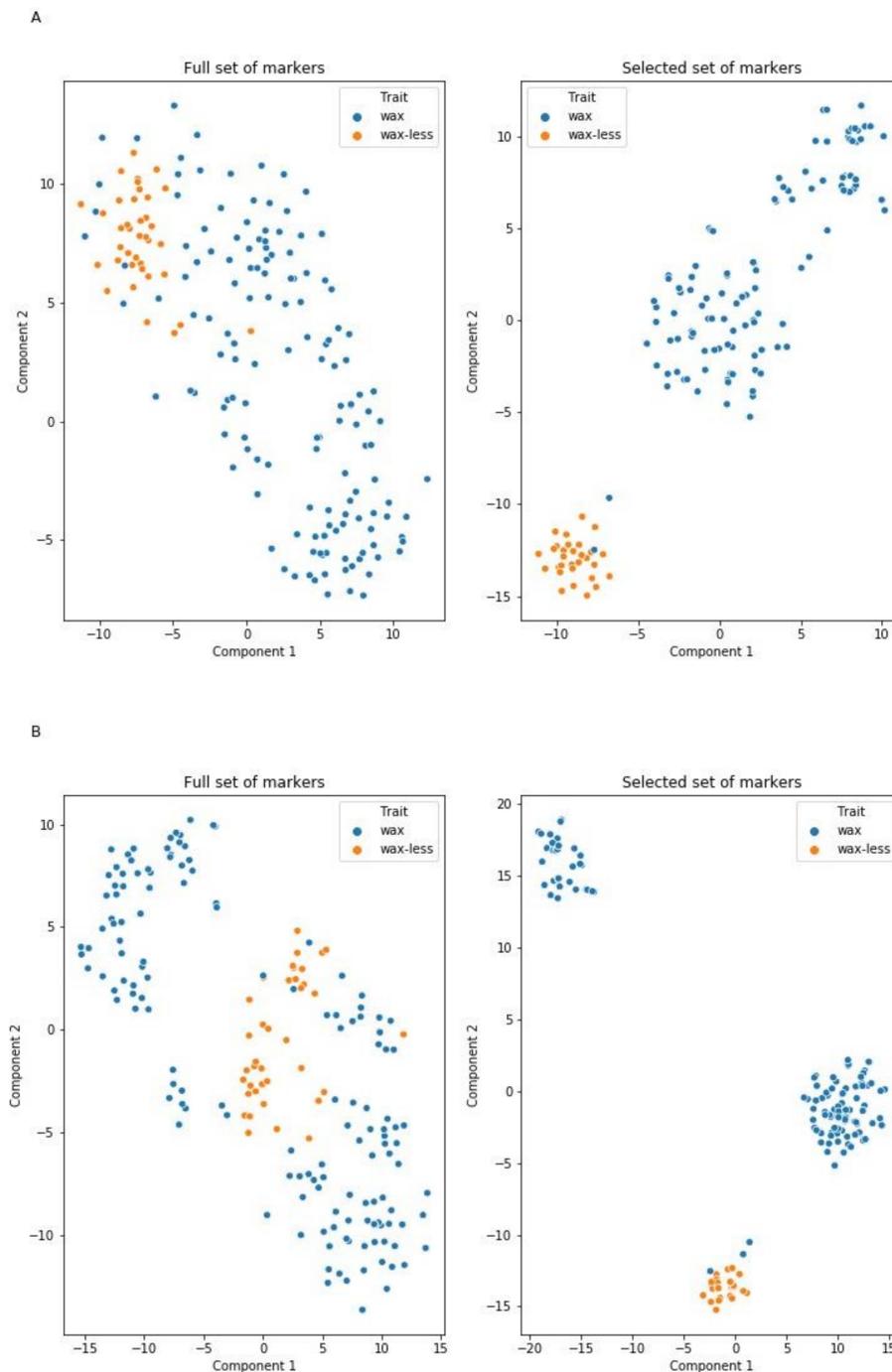
#### 2.4. Markers Distinguishing Between Waxy and Waxless Plants Detected Using Machine Learning (ML) Methods

To select markers that differentiate between waxy and waxless plants, three classifiers (LR, random forests and XGBClassifier) were trained. Each of the used algorithms assesses the features' significance based on their impact on the classification process.

For training purpose, the dominant wax dominant segregation was used, dividing the plants into two classes. Two groups of DARtseqs were used independently, 1,135 assigned to chromosome 2R (group A) based on literature data [56,57], and 1067 linked to *wax* at LOD 45 selected from the group of markers previously not qualified for any chromosome (group B). Models prepared on the training dataset had strong predictive skills measured with classification metrics. Trained models had an F1 metric at the levels 97–98% for the glaucous class and 88–94% for the nonglucous class.

Marker selection to subset considered by models as informative allowed reduction of the number of markers from 1135 to 20 and from 1067 to 35 (Table S6). New models trained on the selected features matrix with the same hyperparameters had significantly higher metric values (F1-score at levels 98–99% for the glaucous class and 91–96% for the nonglucous class). Furthermore, two-dimensional visualization of raw and selected DARtseqs performed using the T-distributed stochastic neighbor embedding (t-SNE) method shows that markers indicated by the models as informative form linearly separable groups representing opposing states of the feature (Figure 3). This observation, as well as increasing the models' metrics, indicate that the ML-based selection process reduced noise resulting from the presence of irrelevant markers, and retained the most informative features from the original

dataset allowing prediction of the type of phenotype (glaucous or nonglaucous) using 20 or 35 markers (models trained on dataset A and B, respectively) with 97–99% general prediction accuracy.



**Figure 3.** Projection of two classes of plants (wax and waxless) in a two-dimensional space characterized on the basis of DArTseqs assigned to chromosome 2R (A) and not assigned (B). The figures on the right show the ordering result using markers selected based on ML algorithms.

The most important markers appointed in the classical mapping procedure (MLH, LOD 50), and chosen by three ML models, were used to map the *wax* region. A group of 17 loci flanking *wax* on the map of the whole chromosome 2R (Table S5), 13 markers detected by all three ML methods (Table 3), and two records of *wax* segregation, were directed to the mapping procedure. A total of 31 loci were used because one DArTseq-SNP, 3362575\_18:C>T, was common to both modes of

selection. Using JoinMap 5.0 and the RM algorithm, a 30 cM map (Table S7) was created from which approximately a 10 cM interval was shown (Figure 2).

**Table 3.** DArTseqs indicated by three machine learning (ML) algorithms as important for distinguishing between waxy and waxless plants. Coefficient values are drawn out directly from models (in the case of logistic regression (LR) absolute values of coefficients were used). Impact value is the sum of coefficients.

Marker	Assignment to 2R	LR Coefficient	Random Forest Coefficient	XGBoost Coefficient	Impact	Distance from Wax Locus [cM]	Annotation
3591025	A	0.636	0.017	0.825	1.478	1.691	-
3593882	B	0.553	0.022	0.024	0.599	unmapped	+
3578307_27:A>G	B	0.413	0.013	0.013	0.439	1.944	-
3889647	A	0.285	0.025	0.035	0.345	0.088	+
3908692_28:C>T	B	0.315	0.014	0.008	0.337	1.097	-
3362575_18:C>T	B	0.192	0.014	0.086	0.293	0.81	+
4485942_42:T>G	B	0.192	0.002	0.069	0.263	0.86	-
3597393_10:T>G	B	0.062	0.005	0.162	0.228	unmapped	-
3358122	A	0.162	0.014	0.019	0.195	7.162	-
3585843	A	0.169	0.015	0.01	0.193	2.056	+
3341848	A	0.064	0.002	0.002	0.067	6.917	+
4092788_55:G>A	B	0.044	0.004	0.009	0.058	22.593	+
3750485	B	0.022	0.01	0.002	0.034	unmapped	-

## 2.5. Genetic Map of Wax Region

### 2.5.1. Candidate Genes for Wax Layer Creation

Sequences of DArTseqs (Table S3) linked to *wax* were directed to an NCBI-BLAST (The National Center for Biotechnology Information—Basic Local Alignment Tool) algorithm to find homologues. The marker most strongly linked to *wax*, DArT-Silico 3889647, was homologous to mRNA of the gene involved in long chain FA metabolism (Table 4). The second locus (3362575\_18:C>T) nearest to *wax* was probably a gene for inactive leucine-rich repeat receptor-like protein kinase (LRR-RLK). The next DArT-Silico 3585843 strongly linked to *wax* was homologous to mRNA for the ABC transporter G family (*ABCG11-like*). Both DArT-Silico markers were selected using three ML methods, and DArT-SNP was appointed by both ML and classical mapping procedure (MLH). The sequence of DArTseq 4485942\_42:T>G located between *wax* and *ABCG11-like* had no annotation found.

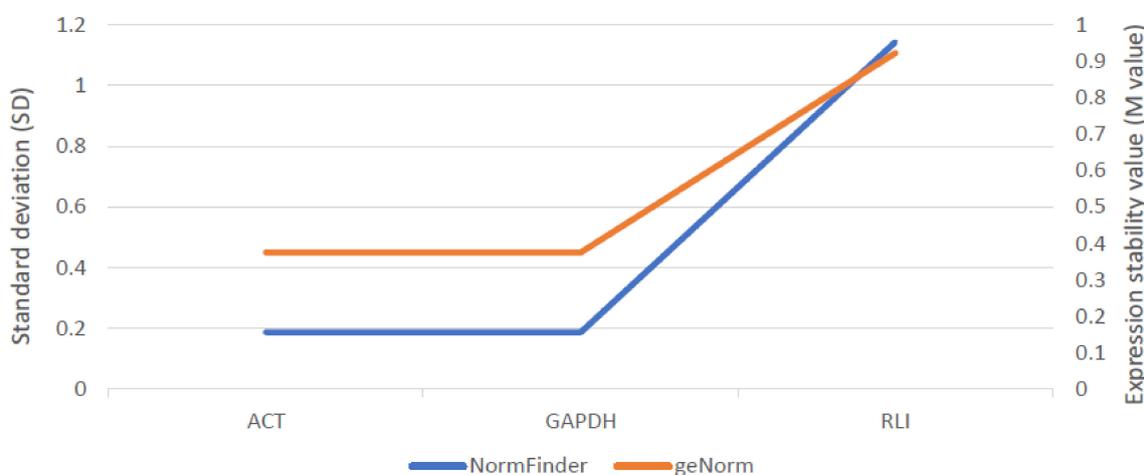
**Table 4.** Annotations of markers most strongly linked to *wax* locus (<3 cM) found in the NCBI database.

Marker	Description	Total Score	E Value	Identity	Accession
3889647 <sup>1</sup>	Predicted: glycine soja long-chain-fatty-acid-AMP ligase FadD28-like (LOC114378589), mRNA	71	1E-09	90%	XM_028337226.1
3362575_18:C>T <sup>1,2</sup>	Predicted: <i>Aegilops tauschii</i> subsp. <i>tauschii</i> probably inactive leucine-rich repeat receptor-like protein kinase At5g06940 (LOC109756806), mRNA	112	4E-22	96%	XM_020315644.1
3585843 <sup>1</sup>	Predicted: <i>Oryza brachyantha</i> ABC transporter G family member 11-like (LOC102717335), mRNA	128	2E-26	100%	XM_006652486.2

<sup>1</sup> marker pointed with use ML; <sup>2</sup> marker pointed with use MLH (JoinMap 5.0).

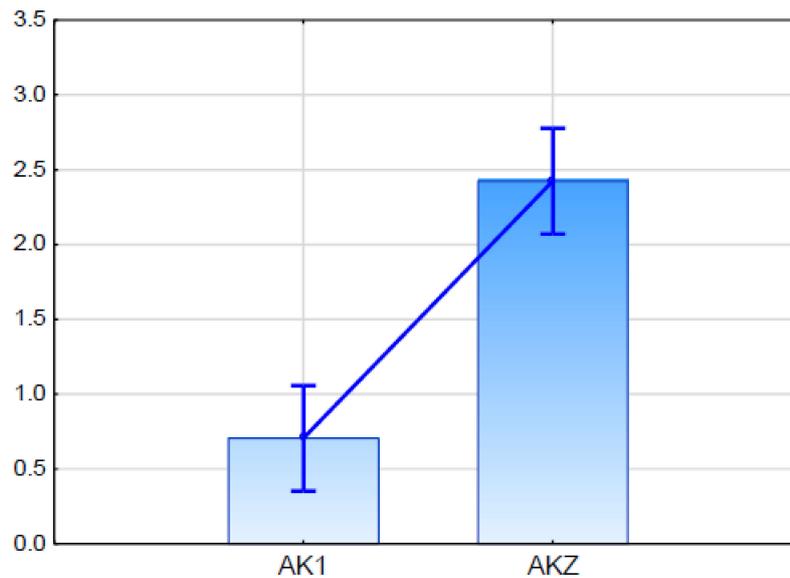
### 2.5.2. Expression Analysis of Candidate Gene for ABCG11 Transporter

Among three examined reference genes, actin (ACT), glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and RNase L inhibitor (RLI), expression of ACT and GAPDH were the lowest and most stable across all samples. Therefore, both these genes were used as internal controls subsequently (Figure 4).



**Figure 4.** Stability assessment of three reference genes across all rye samples using two different algorithms: NormFinder (SD) and geNorm (M-Value) provided in GeneEx 7.0 software (bioMCC, Freising-Weihenstephan, Germany).

Real Time PCR analysis demonstrated that the putative *ScABCG* was expressed in both studied rye lines. The expression level in glaucousness line AKZ was higher than in the glaucous line AK1, and the difference was statistically significant (Figure 5).



**Figure 5.** Relative expression (fold change in relation to the ACT and GAPDH) of putative *ScABCG11* established in qPCR (quantitative PCR) for glaucous (AK1) and nonglauous (AKZ) rye inbred line. Vertical bars indicate +/- standard error. Differences are statistically significant in Kruskal-Wallis test.

### 3. Discussion

Molecular analysis of wax-deficient mutants has led to the identification of a large number of genes encoding functional proteins in the biosynthesis, transport and regulation of cuticular wax in different plant species. Many elements of the cuticle and wax biosynthesis pathways, including genes encoding key enzymes and regulatory transcription factors, were uncovered primarily by characterizing cuticle mutants in arabidopsis, rice, maize, barley and wheat [9–21]. As many as 47 cuticle-associated genes were discussed in the review of Yeats and Rose [2]. At least 21 genes of arabidopsis were predicted to encode b-ketoacyl-CoA synthase [2] that catalyzes the initial condensation step in the elongation pathway of VLCFA biosynthesis. In maize (*Zea mays*), at least 18 loci (the GLOSSY or GL loci) have been found to affect the quantity and/or the composition of cuticular waxes on the surface of seedling leaves (in: [25]).

Despite considerable knowledge about the genetic basis of wax formation in various species, rye remains almost unrecognized in this area of knowledge. With regard to the literature on rye wax genes, references to the location of only two genes on low density genetic maps can be found [43,44]. Our study, aimed at mapping the allele conditioning the nonglauous appearance derived from the rye cultivar Karlikovaja Zelenostebelnaja, showed that the studied gene was mapped for the first time because it was linked to group 2R, while the previously tested genes were located on chromosome 4R and 7R [43,44].

#### 3.1. Mapping Procedure with ML Methods Support

Molecular markers are assigned to linkage groups on the basis of a logarithm of odds (LOD) value, which refers to the ratio of the probability that two loci are linked with a given recombination value over a probability that the two are not linked. The first genetic maps composed of a relatively small number of markers were constructed with use a critical LOD (linklod) value of three as the minimum threshold value in order to decide whether or not loci were linked, which indicates that linkage is 1000 times more likely than no linkage. Modern, high-density maps, based on thousands of marker data, require a significant increase in the level of linklod. Establishing this value, and determining number of linkage groups, is not a straightforward task. Loci on different chromosomes may appear to be linked by chance and, on the other hand, two or more linkage groups can be obtained for each chromosome, which results in a total number of linkage groups higher than the haploid chromosome

numbers [53]. In practice, different LOD values for each experiment have to be tested until in order to establish satisfactory linkage groups equal to haploid chromosome number. The correctness of marker grouping depends on many factors, including the type and size of population and the number and type of marker data (dominant or codominant).

Increasing the size of the mapping population is a good approach to improve the reliability of linkage group creation, but it must also be economically feasible. The main purpose of our research was to map specific gene encoding wax and to preselect markers closest to the studied gene. That is why about 180 genotypes were targeted for genotyping, which was a balance between map quality and the economic aspect of research.

Adopting the right LOD level in a mapping procedure is frequently problematic. In our study, although the linklod at which the two records of wax segregation data (co- and dominant) remained coupled was high (36), it resulted in a map design with an unsettled order relative to the reference map [56]. In addition, mapping of the *wax* region with several sets with different numbers of wax-flanking loci resulted in instability of the marker order. A way to increase the reliability of the map was to reduce the linkage group by further increasing the LOD up to 50. This, in turn, separated the two wax segregations and caused the loss of all DArTseq-Silico loci, suggesting that this criterion was too restricted. To improve the mapping quality, an additional selection of markers was made using three ML methods. Both DArTseq, SNP and Silico were identified in this procedure. Loci selected as flanking in a classical way and indicated by three ML methods, were used for final mapping of the wax region.

This combined mapping strategy led to the construction of a stable map and detected three potential candidate genes for forming a cuticular wax in the 2 cM region from the phenotypic marker. Two of them (DArTseq Silico) were detected only as a result of the use of ML methods, and the third (3362575\_18:C>T) was selected in both approaches, which indicates the usefulness of ML models in choosing a group of markers linked to a specific gene, during the mapping procedure.

One of DArTseq Silico (3585843) seems to be a strong functional marker candidate for the ABC transporter gene. The next two annotated markers were even more closely linked to glaucousness but their homology to known genes and their role was not so obvious. They can be considered as putative genes coding proteins of enzymatic character.

### 3.2. Candidate Gen for ABC Transporter G Family Member

DArTseq 3585843 was considered a functional candidate for the ABC transporter gene marker because it showed 100% sequence similarity to the sequence of predicted rice ABC transporter G family member 11-like (LOC102717335) mRNA (Table 4). The expression of this candidate gene in qPCR (Figure 5) showed differences between the glaucous and the nonglauous parent of mapping population BK2.

The ATP-binding cassette (ABC) transporter's role in transporting wax precursors has already been reported in the literature. ABC transporter proteins belong to a large, diverse and ubiquitous superfamily involved in the active transport of a wide range of molecules across membranes. Arabidopsis and rice genomes encode 130 and 132 members, respectively, and their encoded proteins are categorized into different subfamilies [80]. Recently released plant genome sequences allowed the identification of 803 ABC transporters in four vascular plants and 76 transporters in a green alga, by comparing them with those reannotated in *Arabidopsis thaliana* and the yeast *Saccharomyces cerevisiae* [81]. The ABCG is the largest one among eight (A-I) plant subfamilies [40,81]. Particular classes of multiple ABC transporters participate in the transport of different lipids, including FAs, waxes, and sterols [82].

By analyzing mutants in arabidopsis, ABCG half-transporters required to accumulate both cutin and wax at the cell surface were identified. AtABCG11 is responsible for the formation of cutin, through transport of lipid precursors from epidermis to cuticula, and the *abcg11/wbc11/desperado/cof1* mutants displays reduced cutin and wax levels [34,35,83–85]. ABCG11 has a broad substrate

specificity for a variety of structurally diverse cuticular lipids, and can directly transport cuticular lipids [34,35,83–85].

ABCG were also considered to be essential components of the plant immune system [86]. This is consistent with the observation that epicuticular waxes play roles in plant-insect interactions [87,88]. This aspect makes our putative gene an interesting research object in the context of rye resistance to pathogens.

### 3.3. Putative Candidate Gene for Fatty Acid Desaturase (FAD)

DArTseq 3889647 was the most closely linked to the wax locus. Its sequence showed homology mostly to uncharacterized, unannotated sequences of rye-related species. Among them was the complete sequence of *Oryza sativa* genomic DNA, chromosome 4, BAC clone (Table 4), which was consistent with the collinearity rule. Comparative genomic studies indicated that rye chromosome 2R is a counterpart of, among others, rice chromosome Os04 [55].

The blasting procedure suggested that the DArTseq 3889647 sequence was 90% similar to predicted the glycine soja long chain-fatty acid- AMP ligase FadD28-like (LOC114378589) mRNA sequence (Table 4). FADs are widely distributed throughout the biological world. They are present in all groups of organisms, namely bacteria, fungi, plants and animals, and play a key role in the maintenance of the proper structure and functioning of biological membranes [89]. Their main function is to remove hydrogen from hydrocarbon chains of FAs in the biosynthesis of unsaturated FAs to synthesize C-C double bonds at specified positions. Although the reports of FadD28 were not found in the plant literature, the other FAD family enzymes are common among plants and are involved in the metabolism of FAs. Long-chain-FA-AMP ligase FadD28 occurs in mycobacterium and catalyzes the activation of long-chain FAs (C22–24) as acyl-adenylates (acyl-AMP), which are then transferred to the multifunctional polyketide synthase Mas for further chain extension (UniProtKB-9WQ59: FAA28\_MYCTU).

In current genomic studies of plants, 23 full-length FAD genes have been identified in cucumber (*Cucumis sativus* L.) through database searches. They were distributed on all seven chromosomes and two additional scaffolds [90]. In an upland cotton (*Gossypium hirsutum* L.) study, 39 full-length FAD genes, based on database searches, were identified in tetraploid and were phylogenetically clustered into four subfamilies. Genomic localization revealed that 34 genes were mapped on 22 chromosomes, and five genes were positioned on the scaffold sequences [91].

Recently, a total of 20 full-length desaturase genes were identified from the rice genome. The rice desaturase genes were phylogenetically classified into six subfamilies with the arabidopsis counterparts FAB2, FAD2, FAD3/7/8, FAD6, DES1 and SLD1, and distributed on ten of 12 chromosomes. According to genome collinearity order, rye chromosome 2R is a counterpart of rice chromosomes 4, 6 and 7 [55]. Rice chromosome Os07 contains four desaturase genes; chromosome Os04 has one member [92]. Sequences of these five desaturases have a low level of similarity to DArTseq 3889647 (data not shown), which undermines our hypothesis. However, among twenty rice desaturases, nine members were expanded via chromosomal tandem or segmental duplications [92]. Since rye is separated from rice by a long evolutionary path (probably over 60 million years) and the much larger rye genome may have many genomic sequences with no homology to rice, it is possible that the rye gene is the result of duplication and changes of one of the desaturases. The second possible explanation is that the rye putative FadD28 gene is derived from bacterial DNA built into the rye genome. Another possibility is that we are dealing with an unknown gene unrelated to FAD.

### 3.4. Putative Candidate Gene for Leucine-Rich Repeat Receptor-Like Protein Kinase (LRR RLK)

The marker 3362575\_18:C>T is at a distance of less than 1 cM from the wax locus. Its sequence shows high similarity (96%) to predicted *Aegilops tauschii* subsp. *tauschii* probably inactive LRR-RLK At5g06940 (LOC109756806) mRNA.

Receptor-like kinases (RLKs) containing leucine-rich repeats (LRRs) are members of the family of membrane-localized receptor kinases containing extracellular leucine rich repeat regions. LRR RLK acts as both signal receptor and signal transducer in ligand-mediated communication between cells. The function of many LRR RLKs is still unknown. However, there are some papers describing the relationship between LRR RLK and brassinosteroids (BRs). BRs are perceived by the LRR RLK, binding to a subdomain of these repeats, thereby initiating intracellular signal transduction via activating a kinase cascade beginning with receptor autophosphorylation and culminating in altered gene expression [93]. BRs belonging to lipidic plant hormones may affect the FA composition of cell membranes [94]. Various possible directions of BRs' action on a membrane can be noted. The first is the influence of BRs on the FA composition towards an increased proportion of unsaturated FAs. This suggests the involvement of BRs in the biosynthesis of FAs, or FA transport, and incorporation into cell membranes. BRs, as some other sterols, may enter the cell membrane directly and modify its properties through, for example, increasing the distance between FA chains, which also may improve the functioning of the membrane under unfavorable temperature conditions [94].

The RLK-BRs-FA relationship indicated here suggests a possible gene candidate effect on the formation of a normal wax coat.

The close proximity of three discussed loci linked to wax on the genetic map may suggest that these genes, selected as candidates for participation in wax metabolism, have a similar function and, therefore, are grouped in a metabolic gene cluster. Recent progress in research on the wheat wax inhibitor has shown that the *W1* locus constitutes a  $\beta$ -diketone biosynthetic gene cluster including diketone metabolism-polyketide synthase (DMP), diketone metabolism-P450 CYP709J4 (DMC) and diketone metabolism-hydrolase/carboxylesterase (DMH) [33].

### 3.5. Summary

Rye is a small grain cereal closely related to wheat, barley and triticale. All these crops contain very large genomes (5–17 GB). Sequencing of so abundant DNA sets has been a great challenge. The first significant approach done for *S. cereale* in 2017 [57] released 1.3 million scaffolds covering approximately 30–35% of the rye genome. This database was applied in our study, allowing indication of putative genes for wax synthesis in rye. Unfortunately, we have not succeeded in the indisputable targeting of the gene of interest. Recently released new data on the wheat genome sequence [95], and the more advanced version of rye genome sequence [96], opened new opportunities in candidate gene identification. Indication of the *ScGA2ox12* gene as a candidate for determinant of the dominant type of dwarfing named *Ddw1* in rye [97] can be showed as a case in point. Thus, the further analyses should tend to combine the current DNA-sequence databases with novel methods of computational analyses based on machine learning.

Although knowledge about the genetic basis of wax formation is still expanding, understanding cuticle biosynthesis at the molecular level remains incomplete. Progress in identifying these pathways is necessary to complete the knowledge and to enable selective modification of cuticle properties to improve agricultural productivity. Even for intensively studied plant species, there is a need to continue and extend current knowledge about cuticular wax metabolism processes. To better understand the molecular mechanism of cuticular wax biosynthesis, more putative genes that are thought to participate in the process need to be characterized.

Scientific progress depends on interdisciplinary analyses and requires the integration of bioinformatic analyses of DNA sequences with our knowledge at the biochemical and molecular level. Machine learning is one of the options that can help humans to integrate these areas of research activity.

## 4. Material and Methods

The plant material for this study was a rye mapping population named BK2, which was the F<sub>2</sub> generation of an interline hybrid between inbred lines AK1 and AKZ. The line AK1 (S<sub>10</sub>) was obtained by selfing one dwarf plant found among offspring of the intergeneric hybrid *S. cereale* cv

Amilo  $\times$  *Dasypyrum villosum*. The second parent line AKZ, originates from *S. cereale* cv Karlikovaja Zelenostebel'naja, the old Russian cultivar carrying the unknown gene responsible for wax formation abnormalities. Unlike typical rye plants, individuals with this gene look like they have no blue coating on their ears, leaves and stems (Figure 1). This kind of phenotype is called nonglaucous or waxless.

Experimental trials were conducted in the field of the West Pomeranian University of Technology in Szczecin (53.45° N, 14.53° E) in two vegetation seasons (2016–2017 and 2017–2018). Meteorological data for the years of conducting experiments are shown in the Table S8. All applicable ethical standards were followed. On 20 September 2016, 300 grains of BK2-F<sub>2</sub> were sown at a spacing of 20  $\times$  20 cm. In the spring, before flowering, two ears of each of 271 surviving plants were covered with cellophane casings for self-pollination. After harvesting, the sheltered ears were used for measurements and the grains collected from them to prepare a field experiment the following season. The studied generation BK2-F<sub>3</sub>, was the offspring of F<sub>2</sub> individuals, each represented by 8–16 plants grown in one to two rows.

Phenotyping of BK2 was done on plants of the F<sub>2</sub> and F<sub>3</sub> generations. The assessment of the presence of a wax coating was done visually on the whole plants in the stem elongation phase. Obtaining codominant segregation of glaucousness was possible by assessing the F<sub>3</sub> progeny of each individual F<sub>2</sub> plant. Glaucous plant of F<sub>2</sub> was considered the dominant homozygote if at least 12 of its F<sub>3</sub> offspring were glaucous in the absence of any nonglaucous plant.

In addition to wax, the following traits were studied on mature plants: number of tillers (TN), plant height (PH), length of main spike (SL), number of spikelets per main spike (SNPS), number of grains per main spike under isolation (GNPS) and weight of grains per main spike (GWPS). Compactness of the main spike (CT), determined as the number of spikelets per 10 cm, and average thousand grain weight (TGW), were calculated based on counted/measured parameters. In the case of the F<sub>3</sub> generation, the height of all offspring plants was measured (max. 16 per genotype) and the ear / grain parameters were assessed for three isolated ears from different plants of a given genotype (the progeny of each F<sub>2</sub> plant). Flowering date (FD) was given as the number of days from May 1st. It was evaluated for the F<sub>3</sub> generation as the day on which half of the plants in a single row bloomed (visible anthers).

Genotyping was conducted on the DNA extracted from the young leaves of F<sub>2</sub> plants using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. A set of 184 individuals, randomly chosen from 271, and two parental lines, were sent to the Diversity Arrays Technology Pty Ltd. (Canberra, Australia) for genotyping by sequencing (<http://www.diversityarrays.com/>). Received segregations of DArTseq were used to localize the *wax* locus on the newly developed genetic map of population BK2.

Total RNA for expression analyses was isolated from leaves in the stem elongation phase (BBCH 34–37) using the Qiagen RNeasy Plant Miniprep Kit (Qiagen, Hilden, Germany) according to the manufacturer's procedure. The leaves were frozen in liquid nitrogen and homogenized using TissueLyser II (Qiagen, Hilden, Germany). The quantity and quality of RNA was assessed using an Epoch Microplates Spectrophotometer (BioTek Instruments, Inc., Winooski, VT, USA). For all samples, the 260/280 ratio fluctuated around 2.0. 2  $\mu$ g of total RNA was used for reverse transcription which was carried out immediately after RNA extraction using the QuantiTect ReverseTranscription Kit (Qiagen, Hilden, Germany).

Real-time PCR was performed in a RealPlex4 Mastercycler (Eppendorf, Hamburg, Germany) using GoTaq<sup>®</sup> qPCR Master Mix (Promega, Madison, WI, USA), 0.3  $\mu$ M of each primer (Table 5) and 1  $\mu$ L cDNA templates in a final volume of 20  $\mu$ L. The reaction was carried out starting from 120 s of activation at 95 °C followed by denaturation for 15 s at 95 °C and annealing/elongation at 60 °C for 60 s. The reactions were carried out in 40 cycles. Melting curve analysis (60–95 °C) was performed at the end of each PCR thermal profile. A negative control for each set of primers was performed to ensure amplification specificity.

**Table 5.** Primers used in qPCR for amplification of reference genes and studied gene *ScABCG11*.

Gene	Primer Pair	Sequence 5'–3'
Actin (ACT)	ACT Fw	AAGATGGGACGTCCTTGATGG
	ACT Rev	GGATCTTCATCGGCATCACT
Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)	GAPDH Fw	AGATGCCCTATGTTTGTGG
	GAPDH Rev	GTGGTGCAGCTAGCATTGA
RNase L inhibitor (RLI)	RLI Fw	TTGAGCAACTCATGGACCAG
	RLI Rev	TGCTTTCCAAGGCACAAACAT
ATP binding cassette transporter, subfamily G ( <i>ABCG11</i> -like)	ABCG_F_1297	GGTGATGGATTCAAGGGGCA
	ABCG_R_1382	CGCGCGACATGTTGATGAAT

*ABCG11* primers (Table 5) were developed based on a consensus sequence created using the DArTseq 3585843 sequence and homologous sequences of wheat (>AK331244.1 *Triticum aestivum* cDNA, clone: WT007\_A05, cultivar: Chinese Spring) and barley (>AK355515.1 *Hordeum vulgare* subsp. *vulgare* mRNA for predicted protein, complete cds, coding sequences, clone: NIAShv1021001), which were selected indirectly as homologous to >XM\_006652486.2 predicted: *Oryza brachyantha* ABC transporter G family member 11-like (LOC102717335) mRNA. A Bio-Edit 7.2 computer program [98] was used to establish consensus sequence and Primer 3 [99] to create primers.

Stability of three reference genes: actin (ACT), glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and RNase L inhibitor (RLI), was evaluated using geNorm and NormFinder algorithms. *ScABCG11* activity was assessed relative to the ACT and GAPDH. The relative gene expression was calculated using GeneEx 7.0 (MultiD Analysis) software. The qPCR reaction was performed for three biological and three technical replicates.

Linkage analysis was made using the JoinMap 5.0 package ([www.kyazma.nl](http://www.kyazma.nl)). MLH mapping was used to construct a map of the whole chromosome 3R, and the RM procedure (Kosambi function) was used for mapping a narrowed region with markers linked to *wax* locus.

Statistical analysis. The statistical relationship between the segregation of phenotypic traits of the mapping population was analyzed with the Student's *t*-test using STATISTICA v. 12.0 software (<http://www.statsoft.com>). Significance of differences between parental lines (phenotypic traits and expression level in qPCR) was established by employing the nonparametric Kruskal-Wallis test.

The ML models and selection process were realized in a Python environment using inter alia Scikit-learn and XGBoost libraries [100]. Data were randomly split into a 60% set to train models and a 40% test-set used to evaluate classification metrics. Three independent classifiers were trained: LR, and two ensembled models, random forest and extra boosted trees (XGBClassifier). The crucial models' hyperparameters (number of trees in ensemble classifiers and penalty strength in LR and XGBClassifier) were set using the RandomizedSearchCV method using 100 iterations on specific uniform distributions. The impact of each of the randomly selected parameters on classification accuracy was measured. Parameters that assured the highest accuracy of the models were used as final ones. Feature (marker) selection was performed using the SelectFromModel method which removes features considered as less or noninformative when the importance of the feature is below a specific threshold (mean of all feature importances). Each marker indicated as informative by all of the models was considered as significant.

Models were evaluated using classification report including accuracy (fraction of correctly predicted labels), precision (ratio of true positives divided by sum of true positive and false positive predictions), recall (ratio of true positive divided by sum of true positive and false negative predictions) and f1 metrics (harmonic mean of precision and recall), for each class [100]. A t-distributed stochastic neighbor embedding (t-SNE) algorithm was used to visualize high-dimensional data in a two-dimensional space before visualization data was standardized according to rules described

in the scikit-learn documentation [100]. The analysis described above is available on GitHub (<https://bit.ly/3b4sBNI>).

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1422-0067/21/20/7501/s1>, Table S1: Segregations of DARTseqs in rye mapping population BK2, assigned to 2R based on literature; Table S2: Segregations of DARTseqs in rye mapping population BK2, not assigned to chromosomes based on literature; Table S3: Segregations of DARTseqs in rye mapping population BK2, selected to map 2R; Table S4: The reference map of chromosome 2R of rye RIL-S population [56]; Table S5: Map of 2R of rye BK2 population constructed of 883 loci linked to wax at LOD 50 using maximum likelihood algorithm; Table S6: Loci connected with wax segregation, selected based on three machine learning methods; Table S7: Wax locus region map constructed using 31 loci (17 nearest flanking loci from whole BK2 map and selected 13 loci pointed out in three machine learning methods) using a regression mapping algorithm; Table S8: Meteorological data for the years of conducting experiments at the field of the West Pomeranian University of Technology in Szczecin (Poland).

**Author Contributions:** Conceptualization, M.G. and B.M.; methodology, M.G., J.B. and B.M.; software, J.B.; validation, M.G.; formal analysis, M.G., J.B., N.L., I.C.-M. and B.M.; investigation, M.G., N.L., A.B. and K.K.; resources, A.G., M.G. and B.M.; data curation, M.G., J.B. and B.M.; writing—original draft preparation, B.M.; writing—review and editing, S.S. and B.M.; visualization, M.G., J.B. and B.M.; supervision, B.M.; project administration, B.M.; funding acquisition, B.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The study was financially supported by the National Science Centre, Poland, under grant No 2015/17/B/NZ9/01694.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data, in the writing of the manuscript or in the decision to publish the results.

## Abbreviations

ABC	ATP-binding cassette
BRs	brassinosteroids
CER	eceriferum
cM	centimorgan
CT	compactness of the spike
cv	cultivar
DARTseq	sequencing-based diversity array technology
FA	fatty acid
FAD	fatty acid desaturase
FD	flowering date
GBS	genotyping by sequencing
GNPS	number of grains per spike
GWPS	weight of grains per spike
LOD	logarithm of odds
linklod	critical LOD
LR	logistic regression
LRR	leucine-rich repeat
ML	machine learning
MLH	maximum likelihood
NCBI	National Center of Biotechnology Information
PH	plant height
RIL	recombinant inbred line
RLK	receptor-like protein kinase
RM	regression mapping
SL	spike length
SNPS	number of spikelets per spike
TGW	thousand grain weight
TN	tiller number
VLCFA	very long-chain fatty acid

## References

1. Ji, X.; Jetter, R. Very long chain alkylresorcinols accumulate in the intracuticular wax of rye (*Secale cereale* L.) leaves near the tissue surface. *Phytochemistry* **2008**, *69*, 1197–1207. [[CrossRef](#)] [[PubMed](#)]
2. Yeats, T.H.; Rose, J.K. The Formation and Function of Plant Cuticles1. *Plant. Physiol.* **2013**, *163*, 5–20. [[CrossRef](#)] [[PubMed](#)]
3. Würschum, T.; Langer, S.M.; Longin, C.F.H.; Tucker, M.R.; Leiser, W.L. Refining the genetic architecture of flag leaf glaucousness in wheat. *Theor. Appl. Genet.* **2020**, *133*, 981–991. [[CrossRef](#)] [[PubMed](#)]
4. Koch, K.; Dommissie, A.; Niemietz, A.; Barthlott, W.; Wandelt, K. Nanostructure of epicuticular plant waxes: Self-assembly of wax tubules. *Surf. Sci.* **2009**, *603*, 1961–1968. [[CrossRef](#)]
5. Zhang, J.-Y.; Broeckling, C.D.; Blancaflor, E.B.; Sledge, M.K.; Sumner, L.W.; Wang, Z.-Y. Overexpression of WXP1, a putative *Medicago truncatula* AP2 domain-containing transcription factor gene, increases cuticular wax accumulation and enhances drought tolerance in transgenic alfalfa (*Medicago sativa*). *Plant J.* **2005**, *42*, 689–707. [[CrossRef](#)]
6. Kim, K.W.; Ahn, J.J.; Lee, J.-H. Micromorphology of epicuticular wax structures of the garden strawberry leaves by electron microscopy: Syntopism and polymorphism. *Micron* **2009**, *40*, 327–334. [[CrossRef](#)]
7. Maier, C.G.-A.; Post-Beittenmiller, D. Epicuticular wax on leek in vitro developmental stages and seedlings under varied growth conditions. *Plant Sci.* **1998**, *134*, 53–67. [[CrossRef](#)]
8. Magnucka, E.G.; Suzuki, Y.; Pietr, S.J.; Kozubek, A.; Zarnowski, R. Cycloate, an inhibitor of fatty acid elongase, modulates the metabolism of very-long-side-chain alkylresorcinols in rye seedlings. *Pest. Manag. Sci.* **2009**, *65*, 1065–1070. [[CrossRef](#)]
9. Aarts, M.G.; Keijzer, C.J.; Stiekema, W.J.; Pereira, A. Molecular characterization of the CER1 gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* **1995**, *7*, 2115–2127.
10. Haslam, T.M.; Mañas-Fernández, A.; Zhao, L.; Kunst, L. *Arabidopsis* ECERIFERUM2 Is a Component of the Fatty Acid Elongation Machinery Required for Fatty Acid Extension to Exceptional Lengths. *Plant Physiol.* **2012**, *160*, 1164–1174. [[CrossRef](#)]
11. Hannoufa, A.; Negruk, V.; Eisner, G.; Lemieux, B. The CER3 gene of *Arabidopsis thaliana* is expressed in leaves, stems, roots, flowers and apical meristems. *Plant J.* **1996**, *10*, 459–467. [[CrossRef](#)] [[PubMed](#)]
12. Rowland, O.; Zheng, H.; Hepworth, S.R.; Lam, P.; Jetter, R.; Kunst, L. CER4 Encodes an Alcohol-Forming Fatty Acyl-Coenzyme A Reductase Involved in Cuticular Wax Production in *Arabidopsis*. *Plant. Physiol.* **2006**, *142*, 866–877. [[CrossRef](#)] [[PubMed](#)]
13. Pighin, J.A. Plant Cuticular Lipid Export Requires an ABC Transporter. *Science* **2004**, *306*, 702–704. [[CrossRef](#)] [[PubMed](#)]
14. Fiebig, A.; Mayfield, J.A.; Miley, N.L.; Chau, S.; Fischer, R.L.; Preuss, D. Alterations in CER6, a gene identical to CUT1, differentially affect long-chain lipid content on the surface of pollen and stems. *Plant Cell* **2000**, *12*, 2001–2008. [[CrossRef](#)]
15. Hooker, T.S.; Millar, A.A.; Kunst, L. Significance of the Expression of the CER6 Condensing Enzyme for Cuticular Wax Production in *Arabidopsis*. *Plant Physiol.* **2002**, *129*, 1568–1580. [[CrossRef](#)]
16. Hooker, T.S.; Lam, P.; Zheng, H.; Kunst, L. A Core Subunit of the RNA-Processing/Degrading Exosome Specifically Influences Cuticular Wax Biosynthesis in *Arabidopsis*. *Plant Cell* **2007**, *19*, 904–913. [[CrossRef](#)]
17. Lü, S.; Song, T.; Kosma, D.K.; Parsons, E.P.; Rowland, O.; Jenks, M.A. *Arabidopsis* CER8 encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1) that has overlapping functions with LACS2 in plant wax and cutin synthesis. *Plant J.* **2009**, *59*, 553–564. [[CrossRef](#)]
18. Lü, S.; Zhao, H.; Marais, D.L.D.; Parsons, E.P.; Wen, X.; Xu, X.; Bangarusamy, D.K.; Wang, G.; Rowland, O.; Juenger, T.; et al. *Arabidopsis* ECERIFERUM9 Involvement in Cuticle Formation and Maintenance of Plant Water Status. *Plant Physiol.* **2012**, *159*, 930–944. [[CrossRef](#)]
19. Zhao, H.; Zhang, H.; Cui, P.; Ding, F.; Wang, G.; Li, R.; Jenks, M.A.; Lu, S.; Xiong, L. The Putative E3 Ubiquitin Ligase ECERIFERUM9 Regulates Abscisic Acid Biosynthesis and Response during Seed Germination and Postgermination Growth in *Arabidopsis*. *Plant Physiol.* **2014**, *165*, 1255–1268. [[CrossRef](#)]
20. Zheng, H.; Rowland, O.; Kunst, L. Disruptions of the *Arabidopsis* Enoyl-CoA Reductase Gene Reveal an Essential Role for Very-Long-Chain Fatty Acid Synthesis in Cell Expansion during Plant Morphogenesis. *Plant Cell* **2005**, *17*, 1467–1481. [[CrossRef](#)]

21. Suo, B.; Seifert, S.; Kirik, V. Arabidopsis GLASSY HAIR genes promote trichome papillae development. *J. Exp. Bot.* **2013**, *64*, 4981–4991. [[CrossRef](#)] [[PubMed](#)]
22. Zhang, X.; Liu, Z.; Wang, P.; Wang, Q.; Yang, S.; Feng, H. Fine mapping of BrWax1, a gene controlling cuticular wax biosynthesis in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Mol. Breed.* **2013**, *32*, 867–874. [[CrossRef](#)]
23. Wang, C.; Li, H.; Li, Y.; Meng, Q.; Xie, F.; Xu, Y.; Wan, Z. Genetic characterization and fine mapping BrCER4 in involved cuticular wax formation in purple cai-tai (*Brassica rapa* L. var. *purpurea*). *Mol. Breed.* **2019**, *39*, 12. [[CrossRef](#)]
24. Liu, D.; Dong, X.; Liu, Z.; Tang, J.; Zhuang, M.; Zhang, Y.; Lv, H.; Liu, Y.; Li, Z.; Fang, Z.; et al. Fine Mapping and Candidate Gene Identification for Wax Biosynthesis Locus, BoWax1 in *Brassica oleracea* L. var. *capitata*. *Front. Plant. Sci.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
25. Sturaro, M.; Hartings, H.; Schmelzer, E.; Velasco, R.; Salamini, F.; Motto, M. Cloning and Characterization of GLOSSY1, a Maize Gene Involved in Cuticle Membrane and Wax Production1[w]. *Plant Physiol.* **2005**, *138*, 478–489. [[CrossRef](#)]
26. Zhou, L.; Ni, E.; Yang, J.; Zhou, H.; Liang, H.; Li, J.; Jiang, D.; Wang, Z.; Liu, Z.; Zhuang, C. Rice OsGL1-6 Is Involved in Leaf Cuticular Wax Accumulation and Drought Resistance. *PLoS ONE* **2013**, *8*, e65139. [[CrossRef](#)]
27. Adamski, N.M.; Bush, M.S.; Simmonds, J.; Turner, A.S.; Mugford, S.G.; Jones, A.; Findlay, K.; Pedentchouk, N.; Von Wettstein-Knowles, P.; Uauy, C.; et al. The Inhibitor of wax 1 locus (Iw1) prevents formation of  $\beta$ - and OH- $\beta$ -diketones in wheat cuticular waxes and maps to a sub-cM interval on chromosome arm 2BS. *Plant J.* **2013**, *74*, 989–1002. [[CrossRef](#)]
28. Wu, H.; Qin, J.; Han, J.; Zhao, X.; Ouyang, S.; Liang, Y.; Zhang, N.; Wang, Z.; Wu, Q.; Xie, J.; et al. Comparative High-Resolution Mapping of the Wax Inhibitors Iw1 and Iw2 in Hexaploid Wheat. *PLoS ONE* **2013**, *8*, e84691. [[CrossRef](#)]
29. Zhang, Z.; Wang, W.; Li, W. Genetic Interactions Underlying the Biosynthesis and Inhibition of  $\beta$ -Diketones in Wheat and Their Impact on Glauousness and Cuticle Permeability. *PLoS ONE* **2013**, *8*, e54129. [[CrossRef](#)]
30. Nishijima, R.; Iehisa, J.C.M.; Matsuoka, Y.; Takumi, S. The cuticular wax inhibitor locus Iw2 in wild diploid wheat *Aegilops tauschii*: Phenotypic survey, genetic analysis, and implications for the evolution of common wheat. *BMC Plant. Biol.* **2014**, *14*, 246. [[CrossRef](#)]
31. Xu, Z.; Yuan, C.; Wang, J.; Fu, D.; Wu, J. Mapping the glauousness suppressor Iw1 from wild emmer wheat “PI 481521.”. *Crop J.* **2015**, *3*, 37–45. [[CrossRef](#)]
32. Von Wettstein-Knowles, P.; Wettstein-Knowles, P. Biosynthetic relationships between  $\beta$ -diketones and esterified alkan-2-ols deduced from epicuticular wax of barley mutants. *Mol. Genet. Genom.* **1976**, *144*, 43–48. [[CrossRef](#)]
33. Hen-Avivi, S.; Savin, O.; Racovita, R.C.; Lee, W.-S.; Adamski, N.M.; Malitsky, S.; Almekias-Siegl, E.; Levy, M.; Vautrin, S.; Bergès, H.; et al. A Metabolic Gene Cluster in the Wheat W1 and the Barley Cer-cqu Loci Determines  $\beta$ -Diketone Biosynthesis and Glauousness. *Plant Cell* **2016**, *28*, 1440–1460. [[CrossRef](#)] [[PubMed](#)]
34. Bird, D.A.; Beisson, F.; Brigham, A.; Shin, J.; Greer, S.; Jetter, R.; Kunst, L.; Wu, X.; Yephremov, A.; Samuels, L.; et al. Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. *Plant J.* **2007**, *52*, 485–498. [[CrossRef](#)] [[PubMed](#)]
35. Panikashvili, D.; Savaldi-Goldstein, S.; Mandel, T.; Yifhar, T.; Franke, R.B.; Höfer, R.; Schreiber, L.; Chory, J.; Aharoni, A. The Arabidopsis DESPERADO/AtWBC11 Transporter Is Required for Cutin and Wax Secretion. *Plant Physiol.* **2007**, *145*, 1345–1360. [[CrossRef](#)]
36. Panikashvili, D.; Shi, J.; Bocobza, S.; Franke, R.B.; Schreiber, L.; Aharoni, A. The Arabidopsis DSO/ABCG11 Transporter Affects Cutin Metabolism in Reproductive Organs and Suberin in Roots. *Mol. Plant.* **2010**, *3*, 563–575. [[CrossRef](#)] [[PubMed](#)]
37. Bessire, M.; Borel, S.; Fabre, G.; Carraça, L.; Efremova, N.; Yephremov, A.; Cao, Y.; Jetter, R.; Jacquat, A.-C.; Métraux, J.-P.; et al. A Member of the PLEIOTROPIC DRUG RESISTANCE Family of ATP Binding Cassette Transporters Is Required for the Formation of a Functional Cuticle in Arabidopsis. *Plant Cell* **2011**, *23*, 1958–1970. [[CrossRef](#)]
38. Chen, G.; Komatsuda, T.; Ma, J.F.; Nawrath, C.; Pourkheirandish, M.; Tagiri, A.; Hu, Y.-G.; Sameri, M.; Li, X.; Zhao, X.; et al. An ATP-binding cassette subfamily G full transporter is essential for the retention of leaf water in both wild barley and rice. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12354–12359. [[CrossRef](#)]

39. Panikashvili, D.; Shi, J.; Schreiber, L.; Aharoni, A. The Arabidopsis ABCG13 transporter is required for flower cuticle secretion and patterning of the petal epidermis. *New Phytol.* **2011**, *190*, 113–124. [[CrossRef](#)]
40. Qin, P.; Tu, B.; Wang, Y.; Deng, L.; Quilichini, T.D.; Li, T.; Wang, H.; Ma, B.; Li, S. ABCG15 Encodes an ABC Transporter Protein, and is Essential for Post-Meiotic Anther and Pollen Exine Development in Rice. *Plant Cell Physiol.* **2012**, *54*, 138–154. [[CrossRef](#)]
41. Li, L.; Li, D.; Liu, S.; Ma, X.; Dietrich, C.R.; Hu, H.-C.; Zhang, G.; Liu, Z.; Zheng, J.; Wang, G.; et al. The Maize glossy13 Gene, cloned via BSR-Seq and Seq-Walking Encodes a Putative ABC Transporter Required for the Normal Accumulation of Epicuticular Waxes. *PLoS ONE* **2013**, *8*, e82333. [[CrossRef](#)] [[PubMed](#)]
42. Tulloch, A.P.; Hoffman, L.L. Epicuticular waxes of *Secale cereale* and *Triticale* hexaploide leaves. *Phytochemistry* **1974**, *13*, 2535–2540. [[CrossRef](#)]
43. Korzun, V.; Malyshev, S.; Voylokov, A.; Börner, A. RFLP-based mapping of three mutant loci in rye (*Secale cereale* L.) and their relation to homoeologous loci within the Gramineae. *Theor. Appl. Genet.* **1997**, *95*, 468–473. [[CrossRef](#)]
44. Malyshev, S.; Korzun, V.; Voylokov, A.; Smirnov, V.; Börner, A. Linkage mapping of mutant loci in rye (*Secale cereale* L.). *Theor. Appl. Genet.* **2001**, *103*, 70–74. [[CrossRef](#)]
45. Lawrence, C.J.; Dong, Q.; Polacco, M.L.; Seigfried, T.E.; Brendel, V. MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* **2004**, *32*, 393D–397. [[CrossRef](#)] [[PubMed](#)]
46. Milczarski, P.; Bolibok-Bragoszewska, H.; Myśków, B.; Stojalowski, S.; Heller-Uszyńska, K.; Góralska, M.; Bragoszewski, P.; Uszyński, G.; Kilian, A.; Rakoczy-Trojanowska, M. A High-Density Consensus Map of Rye (*Secale cereale* L.) Based on DArT Markers. *PLoS ONE* **2011**, *6*, e28495. [[CrossRef](#)]
47. Gawroński, P.; Pawelkiewicz, M.; Tofil, K.; Uszyński, G.; Sharifova, S.; Ahluwalia, S.; Tyrka, M.; Wędzony, M.; Kilian, A.; Bolibok-Bragoszewska, H. DArT Markers Effectively Target Gene Space in the Rye Genome. *Front. Plant Sci.* **2016**, *7*. [[CrossRef](#)]
48. Hussain, W.; Baenziger, P.S.; Belamkar, V.; Guttieri, M.J.; Venegas, J.P.; Easterly, A.; Sallam, A.; Poland, J. Genotyping-by-Sequencing Derived High-Density Linkage Map and its Application to QTL Mapping of Flag Leaf Traits in Bread Wheat. *Sci. Rep.* **2017**, *7*, 16394. [[CrossRef](#)]
49. Borzęcka, E.; Hawliczek, A.; Bolibok, L.; Gawroński, P.; Tofil, K.; Milczarski, P.; Stojalowski, S.; Myśków, B.; Targońska-Karasek, M.; Grądzielewska, A.; et al. Effective BAC clone anchoring with genotyping-by-sequencing and Diversity Arrays Technology in a large genome cereal rye. *Sci. Rep.* **2018**, *8*, 8428. [[CrossRef](#)]
50. Yao, X.; Wu, K.; Yao, Y.; Bai, Y.; Ye, J.; Chi, D. Construction of a high-density genetic map: Genotyping by sequencing (GBS) to map purple seed coat color (Psc) in hulless barley. *Heredity* **2018**, *155*, 37. [[CrossRef](#)]
51. Gutierrez-Gonzalez, J.J.; Mascher, M.; Poland, J.; Muehlbauer, G.J. Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. *Sci. Rep.* **2019**, *9*, 1–15. [[CrossRef](#)]
52. Ferreira, A.; Da Silva, M.F.; Silva, L.D.C.E.; Cruz, C.D. Estimating the effects of population size and type on the accuracy of genetic maps. *Genet. Mol. Biol.* **2006**, *29*, 187–192. [[CrossRef](#)]
53. Kumar, J.; Pratap, A.; Solanki, R.K. Basics of Molecular Genetic Mapping and QTL Analysis in Plants. *Adv. Biotechnol.* **2010**, 35–52. [[CrossRef](#)]
54. Devos, K.M.; Atkinson, M.D.; Chinoy, C.N.; Francis, H.A.; Harcourt, R.L.; Koebner, R.M.D.; Liu, C.J.; Masojć, P.; Xie, D.X.; Gale, M.D. Chromosomal rearrangements in the rye genome relative to that of wheat. *Theor. Appl. Genet.* **1993**, *85*, 673–680. [[CrossRef](#)] [[PubMed](#)]
55. Martis, M.M.; Zhou, R.; Haseneyer, G.; Schmutzer, T.; Vrána, J.; Kubaláková, M.; König, S.; Kugler, K.G.; Scholz, U.; Hackauf, B.; et al. Reticulate Evolution of the Rye Genome[W][OPEN]. *Plant Cell* **2013**, *25*, 3685–3698. [[CrossRef](#)]
56. Milczarski, P.; Hanek, M.; Tyrka, M.; Stojalowski, S. The application of GBS markers for extending the dense genetic map of rye (*Secale cereale* L.) and the localization of the Rfc1 gene restoring male fertility in plants with the C source of sterility-inducing cytoplasm. *J. Appl. Genet.* **2016**, *57*, 439–451. [[CrossRef](#)]
57. Bauer, E.; Schmutzer, T.; Barilar, I.; Mascher, M.; Gundlach, H.; Martis, M.M.; Twardziok, S.O.; Hackauf, B.; Gordillo, A.; Wilde, P.; et al. Towards a whole-genome sequence for rye (*Secale cereale* L.). *Plant J.* **2017**, *89*, 853–869. [[CrossRef](#)]
58. Hackauf, B.; Rudd, S.; Van Der Voort, J.R.; Miedaner, T.; Wehling, P. Comparative mapping of DNA sequences in rye (*Secale cereale* L.) in relation to the rice genome. *Theor. Appl. Genet.* **2008**, *118*, 371–384. [[CrossRef](#)]

59. Korzun, V.; Melz, G.; Börner, A. RFLP mapping of the dwarfing (Ddw1) and hairy peduncle (Hp) genes on chromosome 5 of rye (*Secale cereale* L.). *Theor. Appl. Genet.* **1996**, *92*, 1073–1077. [[CrossRef](#)]
60. Stojalowski, S.; Myśków, B.; Hanek, M. Phenotypic effect and chromosomal localization of Ddw3, the dominant dwarfing gene in rye (*Secale cereale* L.). *Euphytica* **2014**, *201*, 43–52. [[CrossRef](#)]
61. Bienias, A.; Góralaska, M.; Masojć, P.; Milczarski, P.; Myśków, B. The GAMYB gene in rye: Sequence, polymorphisms, map location, allele-specific markers, and relationship with  $\alpha$ -amylase activity. *BMC Genom.* **2020**, *21*, 1–15. [[CrossRef](#)] [[PubMed](#)]
62. Miedaner, T.; Glass, C.; Dreyer, F.; Wilde, P.; Wortmann, H.; Geiger, H.H. Mapping of genes for male fertility restoration in ‘Pampa’ CMS winter rye (*Secale cereale* L.). *Theor. Appl. Genet.* **2000**, *101*, 1226–1233. [[CrossRef](#)]
63. Bednarek, P.T.; Lewandowska, R.; Kubicka, H.; Masojć, P. Linkage groups and the indirect chromosome location of cms-P-linked AFLPs. *Cell Mol. Biol. Lett.* **2002**, *7*, 721–736. [[PubMed](#)]
64. Stojalowski, S.; Jaciubek, M.; Masojć, P. Rye SCAR markers for male fertility restoration in the P cytoplasm are also applicable to marker-assisted selection in the C cytoplasm. *J. Appl. Genet.* **2005**, *46*, 371–373. [[PubMed](#)]
65. Stojalowski, S.; Milczarski, P.; Hanek, M.; Bolibok-Bragoszewska, H.; Myśków, B.; Kilian, A.; Rakoczy-Trojanowska, M. DArT markers tightly linked with the Rfc1 gene controlling restoration of male fertility in the CMS-C system in cultivated rye (*Secale cereale* L.). *J. Appl. Genet.* **2011**, *52*, 313–318. [[CrossRef](#)] [[PubMed](#)]
66. Hackauf, B.; Bauer, E.; Korzun, V.; Miedaner, T. Fine mapping of the restorer gene Rfp3 from an Iranian primitive rye (*Secale cereale* L.). *Theor. Appl. Genet.* **2017**, *108*, 1149–1189. [[CrossRef](#)] [[PubMed](#)]
67. Mago, R.; Miah, H.; Lawrence, G.; Wellings, C.; Spielmeyer, W.; Bariana, H.; McIntosh, R.; Pryor, A.; Ellis, J. High-resolution mapping and mutation analysis separate the rust resistance genes Sr31, Lr26 and Yr9 on the short arm of rye chromosome 1. *Theor. Appl. Genet.* **2005**, *112*, 41–50. [[CrossRef](#)]
68. Święcka, S.; Berdzik, M.; Myśków, B. Genetic mapping of the ScHd1 gene in rye and an assessment of its relationship with earliness per se and plant morphology. *J. Appl. Genet.* **2014**, *55*, 469–473. [[CrossRef](#)]
69. Börner, A.; Korzun, V.; Voylokov, A.; Worland, A.; Weber, W. Genetic mapping of quantitative trait loci in rye (*Secale cereale* L.). *Euphytica* **2000**, *116*, 203–209. [[CrossRef](#)]
70. Miedaner, T.; Koch, S.; Seggl, A.; Schmiedchen, B.; Wilde, P. Quantitative genetic parameters for selection of biomass yield in hybrid rye. *Plant. Breed.* **2011**, *131*, 100–103. [[CrossRef](#)]
71. Myśków, B.; Hanek, M.; Banek-Tabor, A.; Maciorowski, R.; Stojalowski, S. The application of high-density genetic maps of rye for the detection of QTLs controlling morphological traits. *J. Appl. Genet.* **2013**, *55*, 15–26. [[CrossRef](#)]
72. Hackauf, B.; Haffke, S.; Fromme, F.J.; Roux, S.R.; Kusterer, B.; Musmann, D.; Kilian, A.; Miedaner, T. QTL mapping and comparative genome analysis of agronomic traits including grain yield in winter rye. *Theor. Appl. Genet.* **2017**, *130*, 1801–1817. [[CrossRef](#)] [[PubMed](#)]
73. Myśków, B.; Stojalowski, S.; Milczarski, P.; Masojć, P. Mapping of sequence-specific markers and loci controlling preharvest sprouting and alpha-amylase activity in rye (*Secale cereale* L.) on the genetic map of an F2(S120×S76) population. *J. Appl. Genet.* **2010**, *51*, 283–287. [[CrossRef](#)] [[PubMed](#)]
74. Myśków, B.; Stojalowski, S.; Łań, A.; Bolibok-Bragoszewska, H.; Rakoczy-Trojanowska, M.; Kilian, A. Detection of the quantitative trait loci for  $\alpha$ -amylase activity on a high-density genetic map of rye and comparison of their localization to loci controlling preharvest sprouting and earliness. *Mol. Breed.* **2011**, *30*, 367–376. [[CrossRef](#)] [[PubMed](#)]
75. Milczarski, P.; Masojć, P.; Krajewski, P.; Stochmal, A.; Kowalczyk, M.; Angelov, M.; Ivanova, V.; Schollenberger, M.; Wakuliński, W.; Banaszak, Z.; et al. QTL mapping for benzoxazinoid content, preharvest sprouting,  $\alpha$ -amylase activity, and leaf rust resistance in rye (*Secale cereale* L.). *PLoS ONE* **2017**, *12*, e0189912. [[CrossRef](#)]
76. Schrider, D.R.; Kern, A.D. Supervised Machine Learning for Population Genetics: A New Paradigm. *Trends Genet.* **2018**, *34*, 301–312. [[CrossRef](#)]
77. Ornella, L.; Cervigni, G.; Tapia, E. Applications of Machine Learning for Maize Breeding. In *Crop Stress and its Management: Perspectives and Strategies*; Venkateswarlu, B., Shanker, A.K., Shanker, C., Eds.; Springer: New York, NY, USA, 2013; pp. 163–192.
78. González-Camacho, J.M.; Ornella, L.; Pérez-Rodríguez, P.; Gianola, D.; Dreisigacker, S.; Crossa, J. Applications of Machine Learning Methods to Genomic Selection in Breeding Wheat for Rust Resistance. *Plant Genome* **2018**, *11*, 170104. [[CrossRef](#)]

79. Parmley, K.A.; Higgins, R.H.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A. Machine Learning Approach for Prescriptive Plant Breeding. *Sci. Rep.* **2019**, *9*, 1–12. [[CrossRef](#)]
80. Verrier, P.J.; Bird, D.A.; Burla, B.; Dassa, E.; Forestier, C.; Geißler, M.; Klein, M.; Kolukisaoglu, Ü.; Lee, Y.; Martinoia, E.; et al. Plant ABC proteins—a unified nomenclature and updated inventory. *Trends Plant. Sci.* **2008**, *13*, 151–159. [[CrossRef](#)]
81. Andolfo, G.; Ruocco, M.; Di Donato, A.; Frusciante, L.; Lorito, M.; Scala, F.; Ercolano, M.R. Genetic variability and evolutionary diversification of membrane ABC transporters in plants. *BMC Plant. Biol.* **2015**, *15*, 51. [[CrossRef](#)]
82. Borghi, L.; Kang, J.; Francisco, R.D.B. Filling the Gap: Functional Clustering of ABC Proteins for the Investigation of Hormonal Transport in planta. *Front. Plant. Sci.* **2019**, *10*. [[CrossRef](#)]
83. Luo, B.; Xue, X.-Y.; Hu, W.-L.; Wang, L.-J.; Chen, X.-Y. An ABC Transporter Gene of Arabidopsis thaliana, AtWBC11, is Involved in Cuticle Development and Prevention of Organ Fusion. *Plant. Cell Physiol.* **2007**, *48*, 1790–1802. [[CrossRef](#)] [[PubMed](#)]
84. Ukitsu, H.; Kuromori, T.; Toyooka, K.; Goto, Y.; Matsuoka, K.; Sakuradani, E.; Shimizu, S.; Kamiya, A.; Imura, Y.; Yuguchi, M.; et al. Cytological and Biochemical Analysis of COF1, an Arabidopsis Mutant of an ABC Transporter Gene. *Plant. Cell Physiol.* **2007**, *48*, 1524–1533. [[CrossRef](#)] [[PubMed](#)]
85. McFarlane, H.E.; Shin, J.J.; Bird, D.A.; Samuels, A.L. Arabidopsis ABCG Transporters, Which Are Required for Export of Diverse Cuticular Lipids, Dimerize in Different Combinations. *Plant. Cell* **2010**, *22*, 3066–3075. [[CrossRef](#)] [[PubMed](#)]
86. Banasiak, J.; Biała, W.; Staszko, A.; Swarczewicz, B.; Kępczyńska, E.; Figlerowicz, M.; Jasiński, M. A Medicago truncatula ABC transporter belonging to subfamily G modulates the level of isoflavonoids. *J. Exp. Bot.* **2013**, *64*, 1005–1015. [[CrossRef](#)]
87. Wójcicka, A. Surface Waxes as a Plant Defense Barrier Towards Grain Aphid. *Acta Biol. Cracoviensia S. Bot.* **2015**, *57*, 95–103. [[CrossRef](#)]
88. Dhanyalakshmi, K.H.; Soolanayakanahally, R.Y.; Rahman, T.; Tanino, K.T.; Nataraja, K.N. Leaf Cuticular Wax, a Trait for Multiple Stress Resistance in Crop Plants. In *Abiotic and Biotic Stress in Plants*; de Oliveira, B.A., Ed.; IntechOpen: London, UK, 2019. [[CrossRef](#)]
89. Los, D.A.; Murata, N. Structure and expression of fatty acid desaturases. *Biochim. et Biophys. Acta (BBA) Lipids Lipid Metab.* **1998**, *1394*, 3–15. [[CrossRef](#)]
90. Dong, C.-J.; Cao, N.; Zhang, Z.-G.; Shang, Q. Characterization of the Fatty Acid Desaturase Genes in Cucumber: Structure, Phylogeny, and Expression Patterns. *PLoS ONE* **2016**, *11*, e0149917. [[CrossRef](#)]
91. Feng, J.; Dong, Y.; Liu, W.; He, Q.; Daud, M.K.; Chen, J.; Zhu, S. Genome-wide identification of membrane-bound fatty acid desaturase genes in Gossypium hirsutum and their expressions during abiotic stress. *Sci. Rep.* **2017**, *7*, srep45711. [[CrossRef](#)]
92. Zhiguo, E.; Chen, C.; Yang, J.; Tong, H.; Li, T.; Wang, L.; Chen, H. Genome-wide analysis of fatty acid desaturase genes in rice (*Oryza sativa* L.). *Sci. Rep.* **2019**, *9*, 1–11. [[CrossRef](#)]
93. Wang, Y.H.; Irving, H.R. Developing a model of plant hormone interactions. *Plant. Signal. Behav.* **2011**, *6*, 494–500. [[CrossRef](#)] [[PubMed](#)]
94. Sadura, I.; Janeczko, A. Physiological and molecular mechanisms of brassinosteroid-induced tolerance to high and low temperature in plants. *Biol. Plant* **2018**, *62*, 601–616. [[CrossRef](#)]
95. The International Wheat Genome Sequencing Consortium (IWGSC); Appels, R.; Eversole, K.; Stein, N.; Feuillet, C.; Keller, B.; Rogers, J.; Pozniak, C.J.; Choulet, F.; Distelfel, D.A.; et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [[CrossRef](#)] [[PubMed](#)]
96. Rabanus-Wallace, M.T.; Hackauf, B.; Mascher, M.; Lux, T.; Wicker, T.; Gundlach, H.; Báez, M.; Houben, A.; Mayer, K.F.X.; Guo, L.; et al. Chromosome-scale genome assembly provides insights into rye biology, evolution, and agronomic potential. *bioRxiv* **2019**. [[CrossRef](#)]
97. Braun, E.-M.; Tsvetkova, N.; Rotter, B.; Siekmann, D.; Schwefel, K.; Krezdorn, N.; Plieske, J.; Winter, P.; Melz, G.; Voylokov, A.V.; et al. Gene Expression Profiling and Fine Mapping Identifies a Gibberellin 2-Oxidase Gene Co-segregating With the Dominant Dwarfing Gene Ddw1 in Rye (*Secale cereale* L.). *Front. Plant. Sci.* **2019**, *10*. [[CrossRef](#)]
98. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp. Ser.* **1999**, *41*, 95–98.

99. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)]
100. Varoquaux, G.; Buitinck, L.; Louppe, G.; Grisel, O.; Pedregosa, F.; Mueller, A. Scikit-learn. *GetMobile Mob. Comput. Commun.* **2015**, *19*, 29–33. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).