EPIGENOMIC POLYMORPHISMS: THE DRIVERS OF DIVERSITY AND HETEROGENEITY

EDITED BY: Sajad Ahmad Dar, Tanvir Ul Hassan Dar and Naseem Akhter

PUBLISHED IN: Frontiers in Genetics and Frontiers in Ecology and Evolution

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-644-8
DOI 10.3389/978-2-83250-644-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

EPIGENOMIC POLYMORPHISMS: THE DRIVERS OF DIVERSITY AND HETEROGENEITY

Topic Editors:
Sajad Ahmad Dar, Jazan University, Saudi Arabia
Tanvir Ul Hassan Dar, Baba Ghulam Shah Badshah University, India
Naseem Akhter, Albaha University, Saudi Arabia

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Citation: Dar, S. A., Dar, T. U. H., Akhter, N., eds. (2022). Epigenomic Polymorphisms: the Drivers of Diversity and Heterogeneity. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-644-8

Table of Contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 17 October 2022
DOI 10.3389/fgene.2022.1008178

OPEN ACCESS

EDITED AND REVIEWED BY
Fei Gao,
Minzu University of China, China

*CORRESPONDENCE
Sajad Ahmad Dar,
sdar@jazanu.edu.sa,
darsajad990@gmail.com

SPECIALTY SECTION
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

RECEIVED 31 July 2022
ACCEPTED 07 October 2022
PUBLISHED 17 October 2022

CITATION
Dar T-U-H, Akhter N and Dar SA (2022), Editorial: Epigenomic polymorphisms: The drivers of diversity and heterogeneity.
Front. Genet. 13:1008178.
doi: 10.3389/fgene.2022.1008178

Editorial: Epigenomic polymorphisms: The drivers of diversity and heterogeneity

Tanvir-Ul-Hassan Dar[1], Naseem Akhter[2] and Sajad Ahmad Dar[3]*

[1]Department of Biotechnology, Baba Ghulam Shah Badshah University, Rajouri, India, [2]Department of Neurology, Henry Ford Health System, Detroit, MI, United States, [3]Research and Scientific Studies Unit, College of Nursing, Jazan University, Jizan, Saudi Arabia

KEYWORDS

epigenomic diversity, genomic variation, DNA methylation, phenotypic characters, epigenome

Editorial on the Research Topic
Epigenomic polymorphisms: The drivers of diversity and heterogeneity

Evolutionary potential of a species is primarily driven by its genetic diversity, however accumulating evidence underscores the important role of epigenomic diversity (Agarwal et al., 2020; Neinavaie et al., 2021). Genomic diversity resulting from changes in DNA nucleotide sequences is not the only heritable information influencing population survival, evolution and ecology in plant and animal species; epigenomic variations, such as DNA methylation or chromatin states, percolate from generation to generation influencing phenotypic characteristics (Flatscher et al., 2012; Miryeganeh and Saze, 2019). Recent studies have found that epigenomic diversity substantially compensates for the loss of genomic variation(s) in small wild populations of genetically homogeneous colonies, thereby demonstrating an additional component of genomic variation(s) (Jueterbock et al., 2020; Mounger et al., 2021). This suggests that both genomic and epigenomic changes in plants and animals affect species and population diversity.

Even though research on inter-and intra-species heterogeneity has progressed significantly, the interplay between genomic (genetic) and epigenomic changes in the wild populations remains to be elucidated. The idea that epigenomic diversity can compensate for genomic diversity loss is relatively new and the extent and patterns of genomic and epigenomic diversity in eukaryotic species, especially those that are closely related but ecologically distinct, warrants comprehensive investigation.

In the present Research Topic, we have collated ten articles (eight original research and two reviews), illustrating the patterns of genomic and epigenomic diversity in eukaryotic species. Barrera-Redondo et al., in their review, have summarized the theoretical and technical bases for conducting domestication genomics, from acquiring a reference genome and genome assembly to population genomics, transcriptomics, epigenomics, and experimental validation of domestication-related genes. The mechanism of epigenetic changes and their dynamic role in maintaining genomic integrity during plant growth and reproduction have been reviewed here by Kumari et al. in an elaborated way.

In addition to genetic variation patterns linked to various environmental challenges (Hodgins et al., 2015; Neinavaie et al., 2021), mounting evidence suggests that epigenetic variation plays a role in ecology and this variation can be both environmentally induced and contribute to phenotypic plasticity (Ashe et al., 2021; Stajic and Jansen, 2021). In their research, Mounger et al. have dealt with differences between genetic and epigenetic parameters in *Spartina alterniflora* (*Sporobolus alterniflorus*, foundation plant), across intertidal gradients. They used epigenotype-by-sequencing (epiGBS), in combination with environmental factors and plant phenotypic variation, in wild *S. alterniflora* populations to connect patterns of genomic and epigenomic diversity with environmental and phenotypic variations. A small but considerable amount of genetic and epigenetic diversity is accounted for by the habitat within populations. While differences in ABC transporter methylation patterns under various environmental conditions and their role in plant growth, development and response to biotic and abiotic stresses have been well documented (Tani et al., 2016; Moretti et al., 2017), little is known about variation in ABC transporter methylation patterns in native and non-native plant species. The results of changes in the methylation of ABC transporters of *Conyza canadensis* (*Erigeron canadensis*) in its native (North America) and non-native (Kashmir Himalayas) ranges are presented in this Research Topic by Shah et al. The DNA methylation of ABC transporter genes has been found to be lower in Kashmir Himalayas than in North America.

The B chromosome has recently been discovered to affect the cell's DNA methylation status, thereby impacting the global gene expression profile (Mendioroz et al., 2015). In this Research Topic, Cardoso et al. have used immunocytogenetics to analyse the epigenetic DNA modification status of B chromosomes in cichlid fish (*Astatotilapia latifasciata*), and the effect of B chromosome presence on the global contents of 5-methylcytosine (5 mC) and 5-hydroxymethylcytosine (5hmC). They found that *A. latifasciata's* B chromosome has an energetic pattern of DNA epimarks, and that its presence promotes the loss of 5 mC in females with the B chromosome's gonads and the loss of 5hmC in males with the B element's muscle.

FOXP3 (Forkhead box P3) is a member of the Forkhead/winged-helix family of transcription factors that causes X-linked autoimmune disorders in mice and humans (Schubert et al., 2001). In their Research Topic, Sadaf et al. have focussed on epigenetic changes and their link to the downregulated FOXP3 gene in female breast cancer patients. FOXO1 promoter methylation and expression at the mRNA and protein levels in different stages of breast cancer, as well as their relationship with various clinical indicators, are still to be investigated (Jiang et al., 2018; Xing et al., 2018). Using methylation-specific PCR, mRNA expression, and immunohistochemistry, Khan et al. have examined FOXO1 mRNA and protein expression in breast cancer. The downregulated protein expression and promoter hypermethylation of the FOXO1 gene are found to have a significant relationship.

Using whole-exome sequencing analysis, Alharazy et al. have explored genetic variations in genes related to vitamin D metabolism in Saudi Arabian families with vitamin D deficiency. Their study has revealed relevant and novel exonic missense variants in both DHCR7 and LRP2 genes, stressing the need to find their association with vitamin-D deficiency. Ochwedo et al. have assessed the genetic polymorphism and temporal stability of Pfs230 domain one and Pfs48/45 domain three in *Plasmodium falciparum* parasites from western Kenya. They found that the domains of the Pfs230 and Pfs48/45 from various malaria-prone regions, including areas where clinical trials are undertaken, should be followed indefinitely subsequent to the discovery of novel polymorphic sites. Srivastava et al. used an integrated transcriptomic method and bioinformatic analysis to uncover altered molecular processes that explain the underlying aetiology of Kawasaki disease (KD). Their approach revealed deregulated molecular mechanisms explaining the underlying etiology of KD which could aid in identifying therapeutic targets and a biomarker panel for early diagnosis and severity of this disease.

In summary, the Research Topic offers an updated assembly of articles assessing the genomic and epigenomic variations in health and diseases of different species of plants and animals. The recent development in methodologies and techniques, such as epigenotyping-by-sequencing, whole-exome sequencing, methylated DNA immunoprecipitation, chromatin immunoprecipitation assay, DNA epi-marks, immunocytogenetics and methylation-specific PCR will help researchers in investigating population genetics and trajectories of the evolution of mutations causing DNA methylation changes. This, combined with genome editing, could unravel the evolutionary significance of epigenome variations.

Author contributions

SD and T-U-HD conceptualized the Research Topic; NA edited the Research Topic; T-U-HD drafted the first version of the manuscript; SD and NA reviewed and edited the first manuscript draft. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Agarwal, G., Kudapa, H., Ramalingam, A., Choudhary, D., Sinha, P., Garg, V., et al. (2020). Epigenetics and epigenomics: Underlying mechanisms, relevance, and implications in crop improvement. *Funct. Integr. Genomics* 20, 739–761. doi:10.1007/s10142-020-00756-7

Ashe, A., Colot, V., and Oldroyd, B. P. (2021). How does epigenetics influence the course of evolution? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20200111. doi:10.1098/rstb.2020.0111

Flatscher, R., Frajman, B., Schönswetter, P., and Paun, O. (2012). Environmental heterogeneity and phenotypic divergence: Can heritable epigenetic variation aid speciation? *Genet. Res. Int.* 2012, 698421–698429. doi:10.1155/2012/698421

Hodgins, K. A., Bock, D. G., Hahn, M. A., Heredia, S. M., Turner, K. G., and Rieseberg, L. H. (2015). Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa. *Mol. Ecol.* 24, 2226–2240. doi:10.1111/mec.13026

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Jueterbock, A., Boström, C., Coyer, J. A., Olsen, J. L., Kopp, M., Dhanasiri, A. K. S., et al. (2020). The seagrass methylome is associated with variation in photosynthetic performance among clonal shoots. *Front. Plant Sci.* 11, 571646. doi:10.3389/fpls.2020.571646

Mendioroz, M., Do, C., Jiang, X., Liu, C., Darbary, H. K., Lang, C. F., et al. (2015). Trans effects of chromosome aneuploidies on DNA methylation patterns in human Down syndrome and mouse models. *Genome Biol.* 16, 263. doi:10.1186/s13059-015-0827-6

Miryeganeh, M., and Saze, H. (2019). Epigenetic inheritance and plant evolution. *Popul. Ecol.* 62, 17–27. doi:10.1002/1438-390X.12018

Moretti, M. L., Ala, R., Pearce, S., Morran, S., and Hanson, D. (2017). Transcription of putative tonoplast transporters in response to glyphosate and paraquat stress in Conyza bonariensis and Conyza canadensis and selection of reference genes for qRT-PCR. *PLoS One* 12, e0180794. doi:10.1371/journal.pone.0180794

Mounger, J., Ainouche, M. L., Bossdorf, O., Cavé-Radet, A., Li, B., Parepa, M., et al. (2021). Epigenetics and the success of invasive plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20200117. doi:10.1098/rstb.2020.0117

Neinavaie, F., Ibrahim Hashim, A., Kramer, A. M., Brown, J. S., and Richards, C. L. (2021). The genomic processes of biological invasions: From invasive species to cancer metastases and back again. *Front. Ecol. Evol.* 2021, 681100. doi:10.3389/fevo.2021.681100

Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F., and Ziegler, S. F. (2001). Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. *J. Biol. Chem.* 276 (40), 37672–37679. doi:10.1074/jbc.M104521200

Stajic, D., and Jansen, L. (2021). Empirical evidence for epigenetic inheritance driving evolutionary adaptation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20200121. doi:10.1098/rstb.2020.0121

Tani, E., Chachalis, D., Travlos, I. S., and Bilalis, D. (2016). Environmental conditions influence induction of key ABC-transporter genes affecting glyphosate resistance mechanism in Conyza canadensis. *Int. J. Mol. Sci.* 17, 342. doi:10.3390/ijms17040342

Xing, Y.-q., Li, A., Yang, Y., Li, X.-x., Zhang, L.-n., and Guo, H.-c. (2018). The regulation of FOXO1 and its role in disease progression. *Life Sci.* 193, 124–131. doi:10.1016/j.lfs.2017.11.030

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 April 2019
doi: 10.3389/fgene.2019.00324

Epigenetic DNA Modifications Are Correlated With B Chromosomes and Sex in the Cichlid *Astatotilapia latifasciata*

*Adauto Lima Cardoso[1], Bruno Evaristo de Almeida Fantinatti[1], Natália Bortholazzi Venturelli[1], Bianca de Oliveira Carmello[1], Rogério Antonio de Oliveira[2] and Cesar Martins[1]**

[1] *Integrative Genomics Laboratory, Department of Morphology, Institute of Biosciences, São Paulo State University – Universidade Estadual Paulista, Botucatu, Brazil,* [2] *Department of Biostatistics, Institute of Biosciences, São Paulo State University – Universidade Estadual Paulista, Botucatu, Brazil*

Supernumerary B chromosomes are dispensable elements found in several groups of eukaryotes, and their impacts in host organisms are not clear. The cichlid fish *Astatotilapia latifasciata* presents one or two large metacentric B chromosomes. These elements affect the transcription of several classes of RNAs. Here, we evaluated the epigenetic DNA modification status of B chromosomes using immunocytogenetics and assessed the impact of B chromosome presence on the global contents of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) and the molecular mechanisms underlying these variations. We found that the B chromosome of *A. latifasciata* has an active pattern of DNA epimarks, and its presence promotes the loss of 5mC in gonads of females with B chromosome (FB+) and promotes the loss of 5hmC in the muscle of males with the B element (MB+). Based on the transcriptional quantification of DNA modification genes (*dnmt*, *tet*, and *tdg*) and their candidate regulators (*idh* genes, microRNAs, and long non-coding RNAs) and on RNA-protein interaction prediction, we suggest the occurrence of passive demethylation in gonads of FB+ and 5hmC loss by Tet inhibition or by 5hmC oxidation in MB+ muscle. We suggest that these results can also explain the previously reported variations in the transcription levels of several classes of RNA depending on B chromosome presence. The DNA modifications detected here are also influenced by sex. Although the correlation between B chromosomes and sex has been previously reported, it remains unexplained. The B chromosome of *A. latifasciata* seems to be active and impacts cell physiology in a very complex way, including at the epigenetic level.

Keywords: supernumerary chromosome, 5-methylcytosine, 5-hydroxymethylcytosine, DNA methylation, DNA demethylation, microRNA

OPEN ACCESS

Edited by:
Jose M. Eirin-Lopez, Florida International University, United States

Reviewed by:
Astrid Böhne, Universität Basel, Switzerland
Ciro Rivera-Casas, Florida International University, United States
Celine Cosseau, Université de Perpignan Via Domitia, France

***Correspondence:**
Cesar Martins
cesar.martins@unesp.br

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *20 November 2018*
Accepted: *22 March 2019*
Published: *12 April 2019*

Citation:
Cardoso AL, Fantinatti BEA, Venturelli NB, Carmello BO, Oliveira RA and Martins C (2019) Epigenetic DNA Modifications Are Correlated With B Chromosomes and Sex in the Cichlid Astatotilapia latifasciata. Front. Genet. 10:324. doi: 10.3389/fgene.2019.00324

INTRODUCTION

Supernumerary B chromosomes (B) are numerical chromosome polymorphisms reported in several groups of eukaryotes, including plants, fungi and animals (Jones, 2017). B chromosomes are dispensable, exhibit non-Mendelian patterns of inheritance and were traditionally seen as inert elements (Beukeboom, 1994; Camacho et al., 2000). However, the identification of functional

sequences in B chromosomes and the effects of these elements in hosts changed the view of these chromosomes as non-functional units (Camacho et al., 2000; Banaei-Moghaddam et al., 2015). Moreover, recent advances based on large-scale DNA/RNA analyses have allowed an understanding of B chromosome biology at a level never considered before (Martis et al., 2012; Valente et al., 2014, 2017; Huang et al., 2016; Li et al., 2017; Ruban et al., 2017; Navarro-Domínguez et al., 2019). This new scenario has provided evidence of complex B chromosome effects in the cells and organisms. In several species B chromosomes have sex-associated differences in frequency, effects and transmission, and they can even generate or were derived from sex chromosomes (Camacho et al., 2011). The influence of Bs over sex seems to be one of the most fascinating and still not understood phenotypic effects of the presence of extra chromosomal elements.

The African cichlid *Astatotilapia latifasciata* has a standard karyotype with 44 chromosomes and supernumerary chromosomes that can be found in several individuals, ranging from 0 to 2 (Poletto et al., 2010; Fantinatti et al., 2011). This B chromosome is large, metacentric, fully heterochromatic and rich in repetitive sequences (Poletto et al., 2010; Fantinatti et al., 2011; Valente et al., 2014). Among the repetitive DNA classes, a sequence that is enriched in the B chromosome and called BncDNA was characterized as a potentially long non-coding RNA and identified as differentially processed and differentially expressed by the effect of supernumerary chromosomes (Ramos et al., 2017). Expansion of several transposable elements in the B chromosome was found (Coan and Martins, 2018). The B chromosome carries potentially duplicated protein-coding genes (Valente et al., 2014) and retro-inserted hnRNP Q-like genes (Carmello et al., 2017). To the best of our knowledge, the B chromosome presence seems to impact the transcription levels of several classes of RNA, thereby influencing cellular function (Carmello et al., 2017; Ramos et al., 2017; Valente et al., 2017; Coan and Martins, 2018).

Despite the reported functional effects of the B chromosome in *A. latifasciata* cells, it is still not clear how the B element has such an influence. Moreover, total or partial inactivation of this element should be necessary to avoid gene dosage effects, such as those observed in aneuploidies (Han et al., 2007). A wide range of regulatory mechanisms may be involved in these processes, and the epigenetic regulation of chromatin seems to be one of possible pathways that B chromosome uses to impact cell physiology, since chromatin modifications directly impact the regulation of gene expression (Cedar and Bergman, 2009). Among these modifications, DNA methylation is described as inactivating complete B chromosomes (Bugno-Poniewierska et al., 2014) or specific B-sequences (López-León et al., 1991; Leach et al., 1995; Langdon et al., 2000; Stitou et al., 2000; Koo et al., 2011) and can be a mechanism that silences transposable elements in the B chromosome of *A. latifasciata* (Coan and Martins, 2018). In turn, the presence of a B chromosome can affect the DNA methylation status of the cell and thereby impact the global profile of gene expression, as observed in aneuploidies (Mendioroz et al., 2015), another type of numerical variation.

DNA methylation (5mC formation) is carried out by the enzyme class DNA methyltransferases (Dnmt1, Dnmt3a, and Dnmt3b) (Plongthongkum et al., 2014). These enzymes catalyze the addition of methyl radicals on cytosines. Dnmt3a and Dnmt3b have *de novo* methylation function, and Dnmt1 is the maintainer Dnmt (Goll and Bestor, 2005). Although DNA methylation is well studied, active and passive DNA demethylation (5mC removal) is poorly explored. Active DNA demethylation is conducted by dioxygenases of the ten-eleven translocation family (Tet1, Tet2, and Tet3) and thymine DNA glycosylase (Tdg), which promote oxidation and base excision, respectively (Kohli and Zhang, 2013). This process involves the transient formation of the variants 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), from 5mC until the recovery of the unmethylated cytosine status (Ito et al., 2011; Kohli and Zhang, 2013; Shen et al., 2014). In turn, passive DNA demethylation occurs during DNA replication in the absence of the maintainer Dnmt activity (Kohli and Zhang, 2013). In fish, the percentage of DNA methylation in CpG contexts is usually higher than in mammal and bird species (Varriale and Bernardi, 2006; Goll and Halpern, 2011; Head, 2014), which suggests the relevance of this epigenetic change to genome function. Epigenetic DNA modifications can be assessed by several methods reaching from cytological approaches, as chromosome immunostaining (Rens et al., 2010; Costa et al., 2016), to strategies that provide the epigenetic status of cytosines (Harrison and Parle-McDermott, 2011; Kurdyukov and Bullock, 2016).

Deregulation of Dnmt, Tet, and Tdg activity can change the pattern of DNA epimarks, and this can result in disturbances of the transcriptional or posttranscriptional mechanisms that regulate the expression of the *dnmt*, *tet*, and *tdg* genes (Denis et al., 2011; Arvinden et al., 2017). Among these mechanisms, the action of microRNAs (miRNA) is widely recognized (Kuc et al., 2017). miRNAs are a class of short (∼20–22 nucleotides), non-coding RNAs that regulate gene expression by target 3′UTR regions of mRNAs (Bartel, 2004) in normal and pathological conditions (Alvarez-Garcia and Miska, 2005). Protein–RNA interactions are a regulatory mechanism of DNA modification enzymes (Di Ruscio et al., 2013). Deregulated activity of isocitrate dehydrogenase enzymes (Idh1 and Idh2), which catalyze the conversion of isocitrate to α-ketoglutarate, a cofactor to Tet, can affect the functions of these dioxygenases and DNA demethylation (Figueroa et al., 2010; Turcan et al., 2012).

Considering that B chromosomes of *A. latifasciata* are the source of several classes of RNAs (Valente et al., 2014, 2017; Carmello et al., 2017; Ramos et al., 2017; Coan and Martins, 2018), it is possible these chromosomal elements regulate DNA modification genes in several ways and promote epigenetic effects on a DNA level. Therefore, to understand if the B chromosome promotes epigenetic DNA modifications in the host cells, we analyzed the DNA modification status of this chromosome and its effects on the routes of DNA methylation and demethylation. In terms of DNA epimarks, the B chromosome of *A. latifasciata* seems to be active and impacts heterogeneously the levels of 5mC and 5hmC, which can in part explain the differential expression of several classes of RNAs induced by B chromosomes.

MATERIALS AND METHODS

Samples

Twenty-nine mature *A. latifasciata* individuals [5 males without B (MB−), 10 males with B (MB+), 4 females without B (FB−), and 10 females with (FB+)], which constituted the offspring of two crosses between MB+ and FB+ individuals maintained in standardized environmental conditions in the fish facility of Integrative Genomics Laboratory at São Paulo State University (UNESP), Botucatu, Brazil, were used in this study. The first cohort was 180 days old and was composed of 3 MB−, 6 MB+, 3 FB−, and 5 FB+. The second cohort was 210 days old and was composed of 2 MB−, 4 MB+, 1 FB−, and 5 FB+. All the procedures were in accordance with the ethical principles of the Brazilian College of Animal Experimentation and were approved by the Institute of Biosciences (UNESP) ethics committee on the use of animals (Protocol No. 486-2013; 769-2015). The individuals were genotyped for B chromosome presence by PCR with a set of primers developed by Fantinatti and Martins (2016). The PCR protocol only detected the B chromosome presence and did not allow discrimination of 1 or 2 B chromosomes. Therefore, B+ samples included 1 or 2 B chromosomes.

DNA and RNA Extraction

Total DNA was obtained from the caudal fin for genotyping (B chromosome absence or presence) and from the encephalon, muscle and gonad for 5mC and 5hmC global quantification. Previous gene ontology analysis of B chromosome gene copies (Valente et al., 2014) revealed enrichment of terms related to these tissues and, therefore, they were selected for the present study. DNA was extracted with phenol-chloroform (Sambrook and Russel, 2001), and its integrity was assessed using agarose gel electrophoresis. Total RNA was extracted from 10 mg of encephalon, muscle and gonads with TRIzol® Reagent (Thermo Fisher Scientific, cat. 15596026) following the manufacturer's instructions. RNA integrity number (RIN) was assessed in a Bioanalyzer 2011, and only samples with RIN greater than or equal to 7 and 8 were used in the Quantitative Reverse Transcription PCR (RT-qPCR) experiments and small RNA sequencing, respectively. Nucleic acid concentrations were determined with a Nanodrop 2000 spectrophotometer.

Chromosome Immunostaining

Metaphasic chromosomes used in the immunostaining of DNA modifications were obtained from kidney cells, which constitute the hematopoietic tissue of fish, following the protocol of Bertollo et al. (1978). Chromosomes were fixed in Carnoy solution (3:1 methanol:acetic acid). Immunofluorescence with 5mC (Abcam, cat. Ab-124936) and 5hmC antibodies (Active Motif, cat. 39770) was performed following the protocol of Rens et al. (2010) with modifications. Cell suspensions containing the metaphasic chromosomes were dropped on slides, followed by dehydration in an ethanol series (70, 85, and 100%). The slides were denatured in 2 N HCl for 15 min at 37°C, incubated in 0.1 M borax (pH 8.4) for 1 min, incubated in PBS for 1 min, dehydrated in an ethanol series and rehydrated in PBS for 3 min. The slides were covered with 10% bovine serum for 7 min, followed by treatment with a primary antibody overnight at 4°C. The slides were washed three times for 5 min each in PBS-Tween 0.01%. The slides were then covered with a 1:100 ratio of secondary antibody anti-rabbit IgG-FITC (Sigma-Aldrich, cat. F0382) in PBS-Tween 0.01% for 40 min and washed as before. Finally, the slides were mounted in Vectashield with DAPI (Vector Laboratories, cat. H-1200). Line scans of the chromosomes were obtained using ImageJ software to determine the levels of fluorescence along each chromosome. Thirty cells of each three individuals (one male and two female) with one B chromosome were analyzed to detect 5mC and 5hmC signal distribution and levels of fluorescence. One metaphase was selected to represent the results of immunostaining and line scan.

Global DNA Methylation and Hydroxymethylation

The global DNA methylation (5mC) and hydroxymethylation (5hmC) were quantified in three individuals of each group (MB−, MB+, FB−, and FB+) with the colorimetric MethylFlash Methylated DNA Quantification and MethylFlash Hydroxymethylated DNA Quantification kits (Epigentek, cat. P-1034, P-1036), respectively, following the manufacturer's instructions. The experiments were performed in technical duplicates with an initial amount of DNA of 100 ng per sample. The relative percentages of methylated and hydroxymethylated DNA in relation to positive control were detected by measuring the absorbance at 450 nm and using the formula described by the manufacturer: 5mC% or 5hmC% = ((Sample OD – NC OD)/S × 100%)/((PC OD – NC OD) × F*)/R, where OD is optical density; NC is negative control (a unmethylated polynucleotide containing 50% of cytosine used for 5mC quantification or a methylated polynucleotide containing 20% of 5-methylcytosine used for 5hmC quantification) supplied by the manufacturer; PC is positive control (a methylated polynucleotide containing 50% of 5-methylcytosine used for 5mC quantification or a hydroxymethylated polynucleotide containing 20% of hydroxymethylcytosine used for 5hm quantification) supplied by the manufacturer; S is amount of input sample DNA in nanograms; F is a factor to normalize 5mC or 5hmC in the positive control to 100%, as the positive control contained only 50% of 5mC or 20% of 5hmC, where $F = 2$ for 5mC and $F = 5$ for 5hmC quantification; and R, amount of input positive control in nanograms.

miRNA Target Prediction

To determine if miRNAs regulate DNA modification genes, the 3′UTRs of these genes were obtained in the gene annotations available on SaciBase[1] and BouillaBase[2]. Sequences of the expressed miRNAs were acquired by alignments of the reads of the small RNA sequenced libraries (Fantinatti, 2015) against the fish mature miRNA sequences downloaded from miRBase v.21[3]. MicroRNA target predictions were performed using PITA (Kertesz et al., 2007), miRanda (Enright et al., 2004)

[1]http://sacibase.ibb.unesp.br/

[2]http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html

[3]http://www.mirbase.org/

and RNAhybrid (Krüger and Rehmsmeier, 2006). Overlapping interactions in the three predictors with p-value less than 0.05 and free energy less than −10 kcal/mol for PITA and −18 kcal/mol for miRanda and RNAhybrid were considered significant.

miRNA Identification and Expression

Small RNA sequencing of two libraries of muscle and three of encephalon and gonads of both groups were performed using single-end Illumina HiSeq 2000 platform. The quality of sequencing data was analyzed with FastQC software, and reads with low quality were eliminated with FastX-toolkit following quality thresholds: 90% of read extension with a phred score of at least 30. Mature miRNA was identified by alignment of filtered reads against a mature miRNA sequence dataset of fish downloaded from miRBase v.21 using Bowtie2 software. The alignments did not accept mismatches and all the parameters were kept on default. Only the miRNAs that were found as candidates to regulate DNA modification genes had their expression determined. Count data normalization and differential expression (DE) of mature miRNAs were carried out using R/Bioconductor DESeq1 package (Anders and Huber, 2010) that is based on the negative binomial distribution with variance and mean linked by local regression. No filtering after read count obtention was carried out. We considered DE miRNAs those that showed fold-change $\geq$2 and p-value $\leq$ 0.05.

mRNA and miRNA Expression Validation

Relative expression of mRNA and miRNA was assessed in 5 MB−, 9 MB+, 4 FB−, and 8 FB+. RNA was treated with DNase I, RNase-free Kit 1 U/μL (Thermo Fisher Scientific, cat. EN0521). Reverse transcription (RT) of mRNA was performed using a High Capacity kit (Thermo Fisher Scientific, cat. 4368814) with an initial 1,000 ng of RNA. RT of miRNA was conducted in agreement with Mei et al. (2012), which included a first step of miRNA polyuridylation with the initial 1000 ng of RNA and cDNA synthesis with a poly(A)-stem-loop primer. For quantification based on qPCR, we used 4 ng of cDNA amplified with GoTaq qPCR Master Mix kit (Promega, cat. A6001) and 400 nM of each primer in a final volume of 20 μL. U6 snRNA and *ubce* genes were used as a reference for normalization of the miRNA and mRNA expression, respectively. Both targets and reference genes were analyzed in duplicate. Cycling conditions were as follows: 95°C for 5 min and 40 cycles at 95°C for 15 s and 60°C for 1 min. The thermal

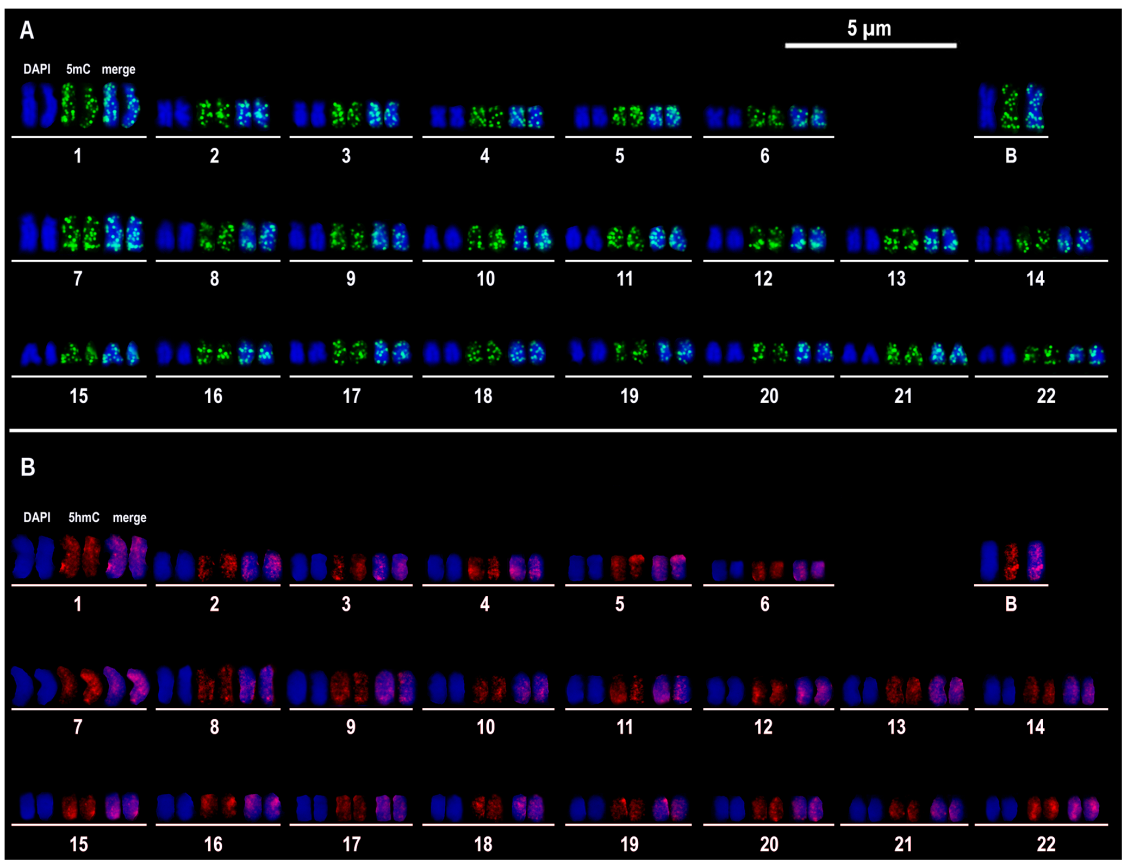

FIGURE 1 | Immunocytogenetics in *Astatotilapia latifasciata*. Representative image of immunostaining of 5mC **(A)** and 5hmC **(B)** in the metaphasic chromosomes of one cell. 5mC and 5hmC signals are observed in green and red, respectively. The chromosomes are counterstained in blue with DAPI. For each karyotype chromosome pair, the fluorescent signals of DAPI and 5mC or 5hmC are shown separated and merged.

cycler was a StepOne Plus Real-Time PCR System (Thermo Fisher Scientific). The $\Delta\Delta$Ct method was used to compare the relative gene expression. The data were normalized using Q-Gene software (Muller et al., 2002; Simon, 2003). The sequences of the primers (specific for cDNA amplification) used are available in **Supplementary Table S1**.

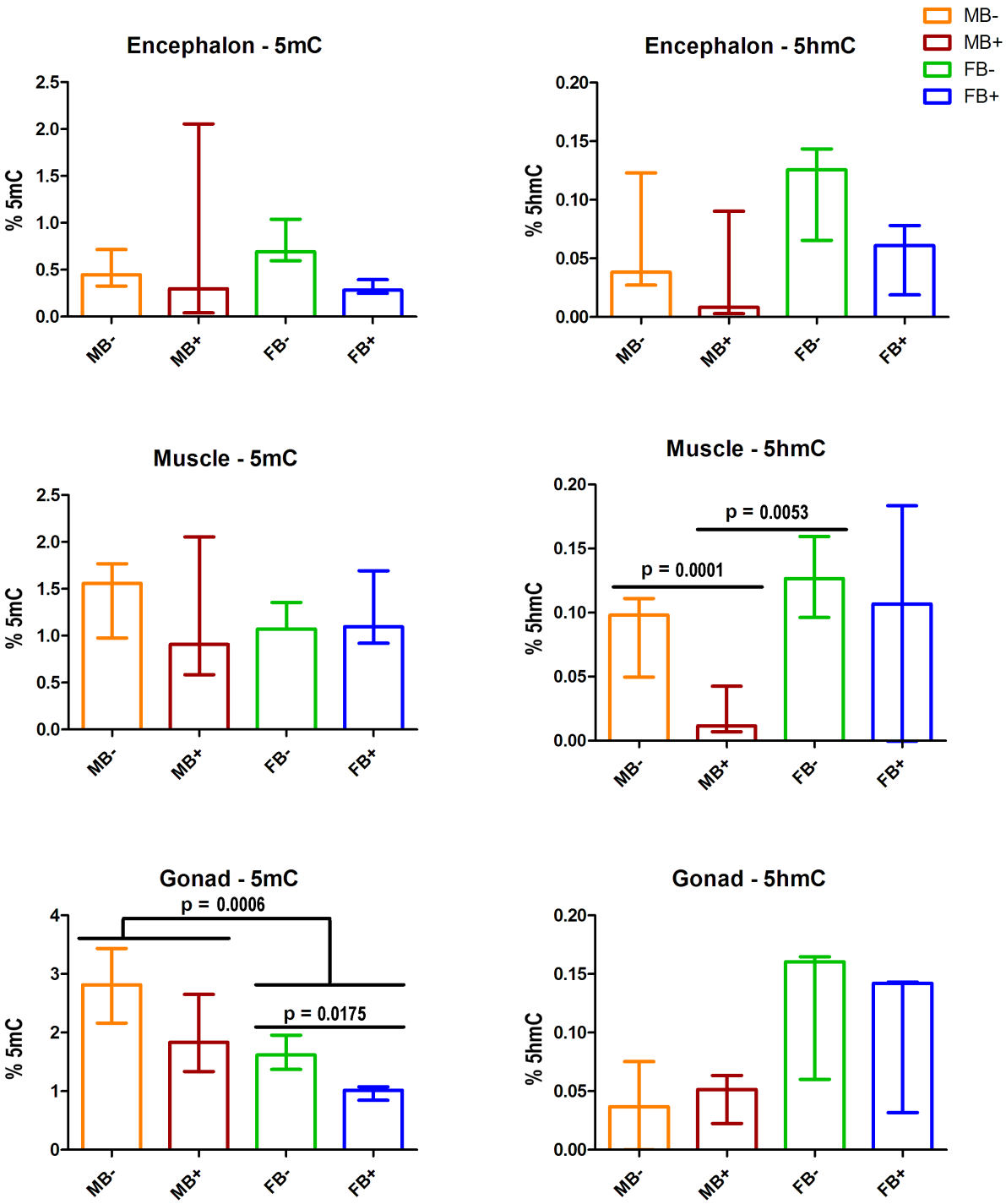

FIGURE 2 | DNA modifications in the presence of B chromosome in *A. latifasciata*. Global quantification of 5mC and 5hmC in the encephalon, muscle and gonads. Statistically significant *p*-values determined by *t*-test ($\alpha = 0.05$) are indicated by horizontal lines. Bar colors represent the different samples analyzed, which included males without B chromosome (orange, MB–), males with B chromosome (brown, MB+), females without B chromosome (green, FB–), and females with B chromosome (blue, FB+).

RNA-Protein Interaction Prediction

The potential non-coding BncRNA is differentially expressed in B chromosome samples, mainly the BncRNA region 2, which is upregulated in B+ samples (Ramos et al., 2017). Moreover, non-coding RNAs were described as inhibitors of Dnmt1 (Di Ruscio et al., 2013). Thus, to explore the possible association of BncRNA and Dnmt, we performed interaction predictions of the BncRNA complete sequence and BncRNA region 2 with the Dnmts amino acid sequences using the tools RPIseq (Muppirala et al., 2011) and RPI-pred (Suresh et al., 2015).

Statistics

The results of the global quantification of 5mC and 5hmC, mRNA and miRNA transcription were analyzed as in Carmello et al. (2017) and Ramos et al. (2017) using a generalized linear model (GLM) statistical approach with averages adjusted by the asymmetric gamma distribution of all the variables, implemented in SAS version 9.3, procedure GENMOD, followed by Bonferroni's *post hoc* test. Associations of 5mC level and BncRNA expression were determined by Pearson's correlation. In all the cases, significant results were accepted at a maximum *p*-value of 0.05.

RESULTS

Chromosomal Immunolocalization of 5mC and 5hmC Marks

The chromosomal immunolocalization of 5mC showed marks scattered over all A chromosomes and on the B chromosome, without accumulation in any specific region and without differences between A and B chromosomes (**Figure 1A**). Similarly, the 5hmC marks were dispersed over the A and B chromosomes without accumulation in any specific region and without differences between A and B chromosomes (**Figure 1B**). Line scan provided a more detailed signal distribution of marks and showed peaks of 5mC marks along all the chromosomes and regions of depletion of these marks, as the centromeric regions of most chromosomes, including the centromeric region of the B chromosome (**Supplementary Figure S1**). In turn, line scan showed that 5hmC marks

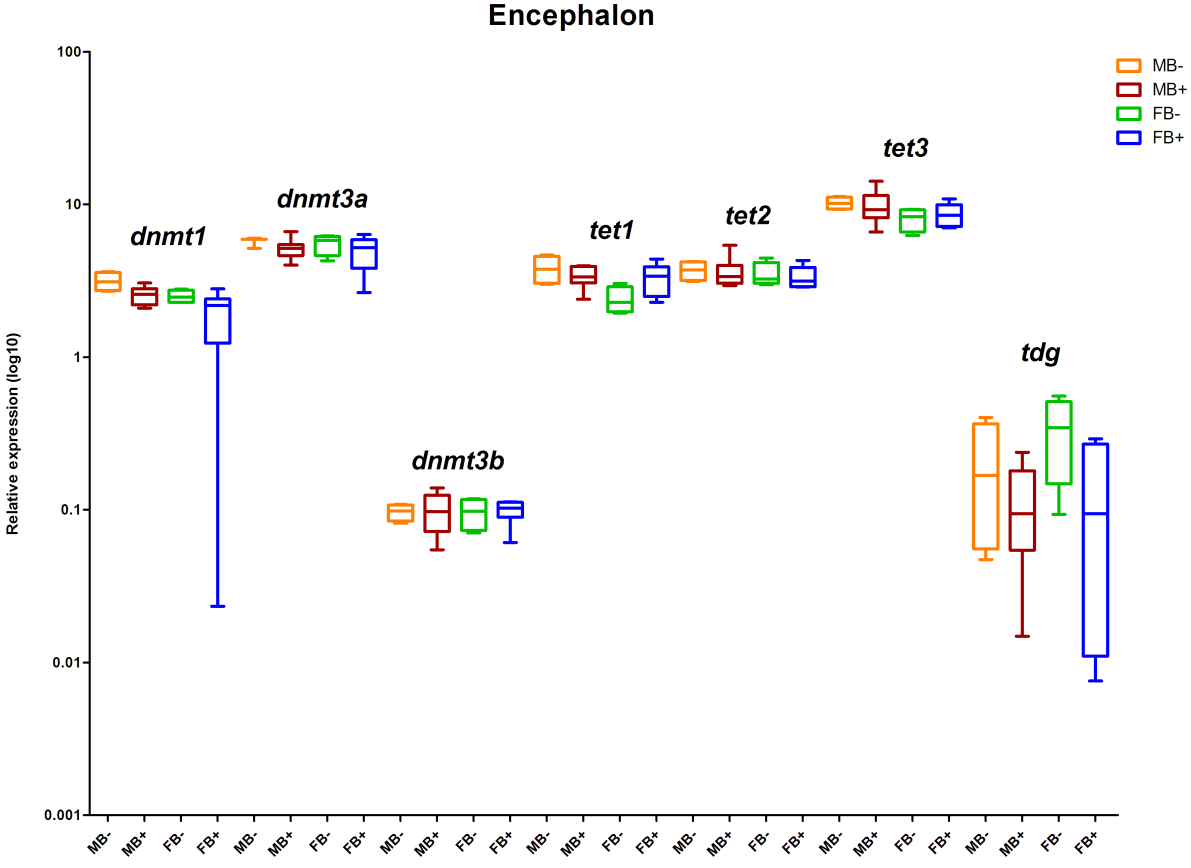

FIGURE 3 | Relative expression of DNA modification genes in *A. latifasciata*. Transcription levels of the genes *dnmt1*, *dnmt3a*, *dnmt3b*, *tet1*, *tet2*, *tet3*, and *tdg* in encephalon measured by RT-qPCR. *p*, *p*-values determined by generalized linear model ($\alpha = 0.05$) followed by Bonferroni's *post hoc* test. Bar colors represent the different samples analyzed, which included males without B chromosome (orange, MB−), males with B chromosome (brown, MB+), females without B chromosome (green, FB−), and females with B chromosome (blue, FB+). The box define median and the lines the minimum and maximum values.

were distributed similarly along A and B chromosomes (**Supplementary Figure S2**). Since we studied chromosomes of one MB+ and two FB+ samples, we did not observe differences among the three individuals.

Effects of B Chromosomes and Sex on the Global Levels of 5mC and 5hmC

We assessed the contents of 5mC and 5hmC in encephalon, muscle and gonads of MB−, MB+, FB−, and FB+ samples (**Figure 2**). In encephalon, we did not find statistically significant differences in the levels of 5mC and 5hmC among the groups, although we observed a tendency of reduction of 5mC in FB+ compared to FB−. In muscle, we did not observed variations in the level of 5mC among the groups, but the level of 5hmC was different between MB− and MB+ ($p = 0.0001$; $\alpha = 0.05$) and between MB+ and FB− ($p = 0.0053$; $\alpha = 0.05$). In gonads, we identified statistical differences in the level of 5mC between males and females ($p = 0.0006$; $\alpha = 0.05$) and between FB− and FB+ ($p = 0.0175$; $\alpha = 0.05$). In addition, we observed a tendency of increase of the level of 5hmC in females, but that was not statistically significant.

Transcription of Genes of DNA Modification Cycle

We determined the relative expression of DNA modification genes in encephalon, muscle and gonads of MB−, MB+, FB−, and FB+ samples. In encephalon, we did not detect statistically significant differences in the transcription of *dnmt1*, *dnmt3a*, *dnmt3b*, *tet1*, *tet2*, *tet3*, or *tdg* among the groups (**Figure 3**). In addition, we observed a tendency of reduction in the transcription of *tdg* gene in encephalon of MB+ and FB+, although these variations have not been statistically significant. In muscle, we identified significant differences only in the level of expression of the gene *tet3* between FB− and FB+ ($p = 0.0143$; $\alpha = 0.05$) (**Figure 4**). Moreover, we found a tendency of increase of *dnmt3b* transcription in muscle of MB+ and FB+, and reduction of *tdg* trancription in MB+ and increase of *tdg* transcription in FB+, although these variations have not been statistically proven. In gonads, we detected differences in the transcription of *dnmt1*, *dnmt3b*, and *tdg* between males and females independent of B chromosome presence ($p < 0.0001$; $\alpha = 0.05$) (**Figure 5**). We also found a tendency of reduction of *tet3* expression in ovaries of FB+, although this has not

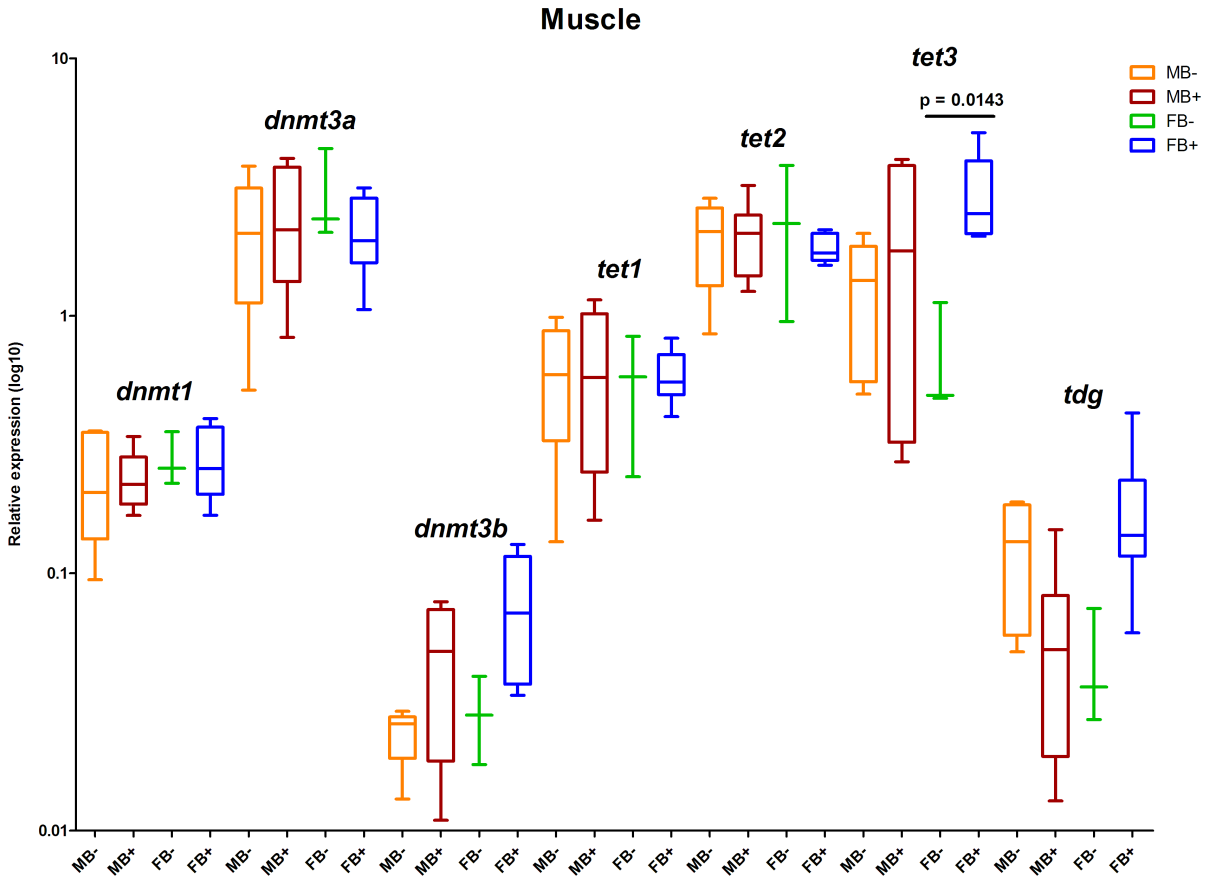

FIGURE 4 | Relative expression of DNA modification genes in *A. latifasciata*. Transcription levels of the genes *dnmt1*, *dnmt3a*, *dnmt3b*, *tet1*, *tet2*, *tet3*, and *tdg* in muscle measured by RT-qPCR. *p*-values determined by generalized linear model ($\alpha = 0.05$) followed by Bonferroni's *post hoc* test. Bar colors represent the different samples analyzed, which included males without B chromosome (orange, MB−), males with B chromosome (brown, MB+), females without B chromosome (green, FB−), and females with B chromosome (blue, FB+). The box define median and the lines the minimum and maximum values.

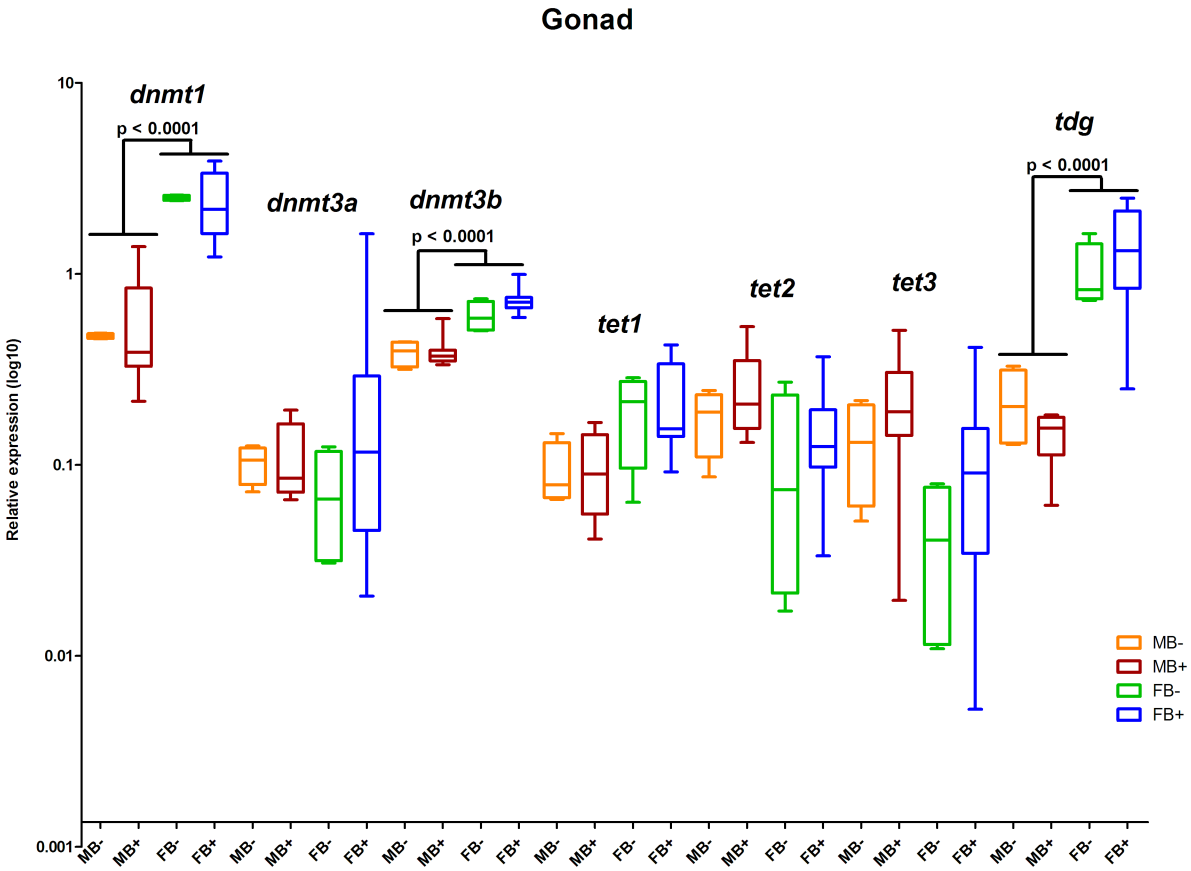

FIGURE 5 | Relative expression of DNA modification genes in *A. latifasciata*. Transcription levels of the genes *dnmt1*, *dnmt3a*, *dnmt3b*, *tet1*, *tet2*, *tet3*, and *tdg* in gonads measured by RT-qPCR. *p*, *p*-values determined by generalized linear model (α = 0.05) followed by Bonferroni's *post hoc* test. Bar colors represent the different samples analyzed, which included males without B chromosome (orange, MB−), males with B chromosome (brown, MB+), females without B chromosome (green, FB−), and females with B chromosome (blue, FB+). The box define median and the lines the minimum and maximum values.

been statistically significant. Additionally, in muscle, we observed differences in the transcription of *idh1* between MB+ and FB− ($p = 0.0008$; $\alpha = 0.05$) and between FB− and FB+ ($p = 0.0085$; $\alpha = 0.05$) (**Figure 6A**).

Identification and Expression of epi-miRNAs

Using alignments of mature miRNA sequences against the reads of small RNA-seq libraries (Fantinatti, 2015), we identified the set of miRNAs expressed in *A. latifasciata* (data not shown). Then, we used these identified miRNAs to predict miRNA/3′UTR interactions and found candidate miRNAs to regulate DNA modification genes in this species (**Table 1**). Next, we explored their expression profile in small RNA-seq libraries (Fantinatti, 2015), and in agreement with the cutoff used here, we did not found different expressed miRNAs among the groups (**Supplementary Table S2**). We selected some miRNAs to validate their expression by RT-qPCR and found no differences in the expression of the miR-29b in the encephalon among the groups (**Figure 6B**). We did not observe changes in the expression of miRNAs in muscle (**Figure 6C**). We found variations in the expression of miR-132b ($p = 0.0002$; $\alpha = 0.05$) and miR-199a-5p ($p = 0.0002$; $\alpha = 0.05$) between testicles and ovaries independent of B chromosome presence (**Figure 6D**). In gonads, we also identified differences in the expression of miR-17a-2-3p between MB− and FB− ($p = 0.0286$; $\alpha = 0.05$) and between MB− and FB+ ($p = 0.0286$; $\alpha = 0.05$).

BncRNA Is a Candidate for Dnmt Regulation

We performed RNA-protein interaction predictions based on software that considers only the sequence of the RNA and protein (RPIseq) and detected scores that indicated a positive interaction of both the complete BncRNA and region 2 of the BncRNA (Ramos et al., 2017) with Dnmt proteins (**Supplementary Dataset 1**). Similarly, we conducted predictions that considered sequence and structural information of RNA and protein (RPI-Pred) and again identified positive interactions. We also assessed the transcriptional status of BncRNA available in Ramos et al. (2017), which studied the same samples of the present study, and we correlated this data with the level of 5mC. This analysis revealed negative correlation in encephalon ($R = -0.3336$;

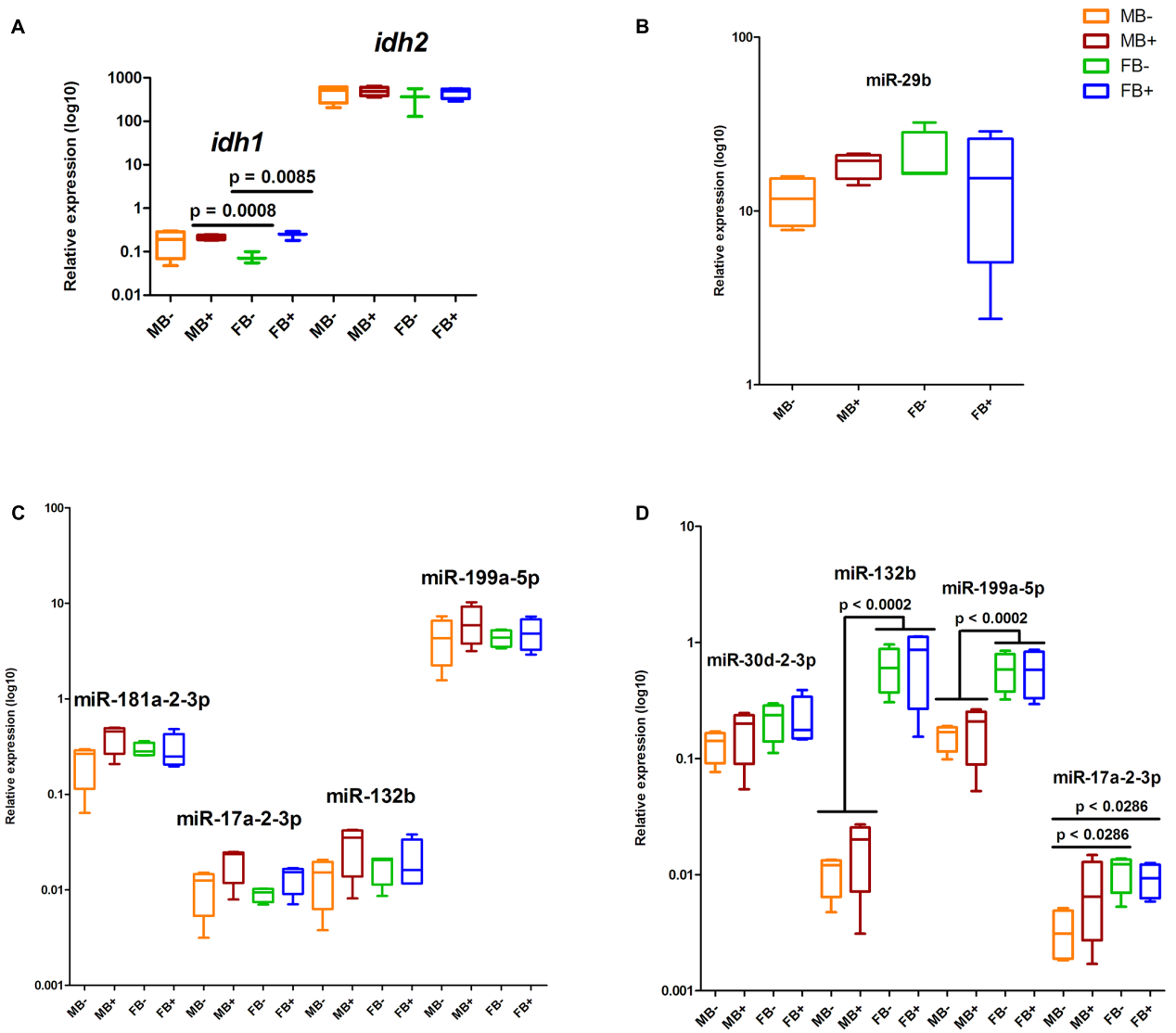

FIGURE 6 | Relative expression of the regulators of DNA modification genes in *A. latifasciata*. Transcription levels of *idh* genes in the muscle measured by RT-qPCR **(A)**. Transcription levels of selected candidate miRNAs regulating the DNA modification genes in encephalon **(B)**, muscle **(C)**, and gonads **(D)**, as revealed by RT-qPCR. *p*, *p*-values determined by generalized linear model (α = 0.05) followed by Bonferroni's *post hoc* test. Bar colors represent the different samples analyzed, which included males without B chromosome (orange, MB−), males with B chromosome (brown, MB+), females without B chromosome (green, FB−), and females with B chromosome (blue, FB+). The box define median and the lines the minimum and maximum values.

$p < 0.05$) and ovaries (R = -0.4689; $p < 0.05$), suggesting that B chromosome promotes reduction of 5mC level by increase of BncRNA expression (**Figure 7**). Together, these data could indicate a positive interaction between the BncRNA and Dnmts.

DISCUSSION

5mC and 5hmC Localization on the B Chromosome

For decades, B chromosomes were thought of as non-functional units, since they are dispensable and heterochromatic (Camacho, 2005). However, evidence of transcriptionally active B-sequences has been emerging from several organisms (Miao et al., 1991; Carchilan et al., 2007; Trifonov et al., 2013; Li et al., 2017; Ma et al., 2017). Despite the discovery of these expressed sequences in B chromosomes, the inactivation of these elements should be necessary to avoid dosage effects, like those observed in aneuploidies (Han et al., 2007). Thus, DNA methylation could be a mechanism that acts in the silencing of specific genic regions or at the chromosomal level.

The B chromosome is completely enriched by 5mC marks in the canid *Nyctereutes procyonoides*, which is believed to be a possible mechanism of silencing (Bugno-Poniewierska et al., 2014). However, our chromosomal immunolocalization of 5mC in *A. latifasciata* shows marks scattered over all A and B

TABLE 1 | Candidate miRNAs in the regulation of DNA modification genes predicted by the software PITA, miRanda, and RNAhybrid.

Target gene	miRNA
dnmt1	ssa-miR-143-5p, ssa-miR-30d-2-3p
dnmt3a	dre-miR-29b
dnmt3b	ccr-miR-132b/dre-miR-29b
tet1	unidentified
tet2	ipu-miR-212
tet3	ccr-miR-199-5p/ola-miR-199a-5p
tdg	dre-miR-181a-2-3p, dre-miR-17a-2-3p

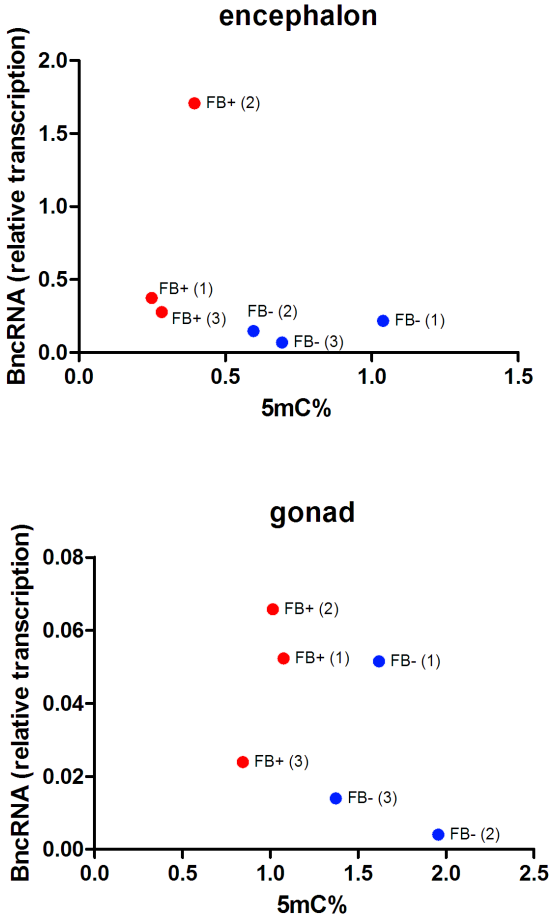

FIGURE 7 | Comparison of the percentages of 5mC and transcriptional levels of BncRNA per sample in FB− females without B chromosome (blue, FB−) and females with B chromosome (red, FB+). Numbers inside parentheses indicate sample ID. Pearson's correlation indicates inverse correlation between 5mC levels and BncRNA expression in encephalons ($R = -0.3336$ and $p < 0.05$) and ovaries ($R = -0.4689$ and $p < 0.05$) of FB+.

chromosomes. Likewise, a scattered and uniform pattern of 5mC marks in A and B chromosomes was also observed in *Secale cereale* (Carchilan et al., 2007; Pereira et al., 2016). In these cases, high DNA methylation level likely is not involved in chromosome inactivation. In turn, our line scan showed reduced 5mC marks in centromeric regions of most chromosomes, including the B, which is in agreement with previous results that revealed hypomethylation of the active centromeres, even those located in B chromosomes (Koo et al., 2011). This active epigenetic status of the B chromosome centromere can represent an important feature for the maintenance and transmission of this element.

At the sequence level, in the leaf tissue of *S. cereale*, the E3900 subtelomeric sequence repeat of the B chromosome is methylated (Langdon et al., 2000), as is the inactive centromeric tandem repeat Bd49 in *Brachycome dichromosomatica* (Leach et al., 1995). Moreover, in *Zea mays* the B-specific satellite ZmB is hypomethylated in the active centromere, while its inactive version is hypermethylated (Koo et al., 2011). In *Eyprepocnemis plorans* (López-León et al., 1991) and *Rattus rattus* (Stitou et al., 2000), ribosomal RNA genes localized in the B chromosome are silenced by 5mC marks. Therefore, although DNA methylation seems not to act at the chromosomal level of *A. latifasciata*, this modification could be important to the inactivation at the sequence level.

Similar patterns of 5mC observed among A and B chromosomes of *A. latifasciata* could indicate that this epimark has the same role in the regulation of both types of chromosomes. Therefore, since A chromosomes are active, this could also indicate functional activity of B chromosomes. This hypothesis can be supported by evidence of active B gene copies in this species (Valente et al., 2014). However, although DNA methylation may not act in B chromosome inactivation, other mechanisms, such as histone modifications and late replication, can be acting, as suggested for the supernumerary chromosome of *B. dichromosomatica* (Houben et al., 1997). In mouse, for example, Rens et al. (2010) observed hypomethylation both of active and inactive X chromosomes in females, which is not common in eutherian mammals, and the authors attributed the inactivation of one X chromosome to histone modifications. Therefore, other possible mechanisms of B chromosome inactivation in *A. latifasciata* need better investigation.

Chromosomal distribution of 5hmC is still poorly explored, with studies focused only on mammals (Szulwach et al., 2011; Kubiura et al., 2012; Li et al., 2013; Yamaguchi et al., 2013; Bogomazova et al., 2014; Efimova et al., 2015, 2017), which reported localization of this mark mostly in active regions, such as R-bands (Kubiura et al., 2012; Efimova et al., 2015) and the reactivated X chromosome (Bogomazova et al., 2014). Here, we conducted the first localization of 5hmC in chromosome spreads of a non-mammalian species and the first analysis focused on supernumerary B chromosomes. Our immunostaining of 5hmC showed marks scattered over all A chromosomes and on the B chromosome, without accumulation of this mark in any region, which can indicate that 5hmC does not act at the chromosome level. However, studies at the sequence level need to be performed to advance our understanding of 5hmC control of B chromosomes. Moreover, similar patterns of 5hmC distribution between A and B chromosomes can be further evidence of supernumerary chromosome activity, similarly to that discussed for 5mC marks.

The chromosomal profiles of 5mC and 5hmC of the B chromosome indicate that this element can be functional and impact the global epigenetic DNA modification status of the

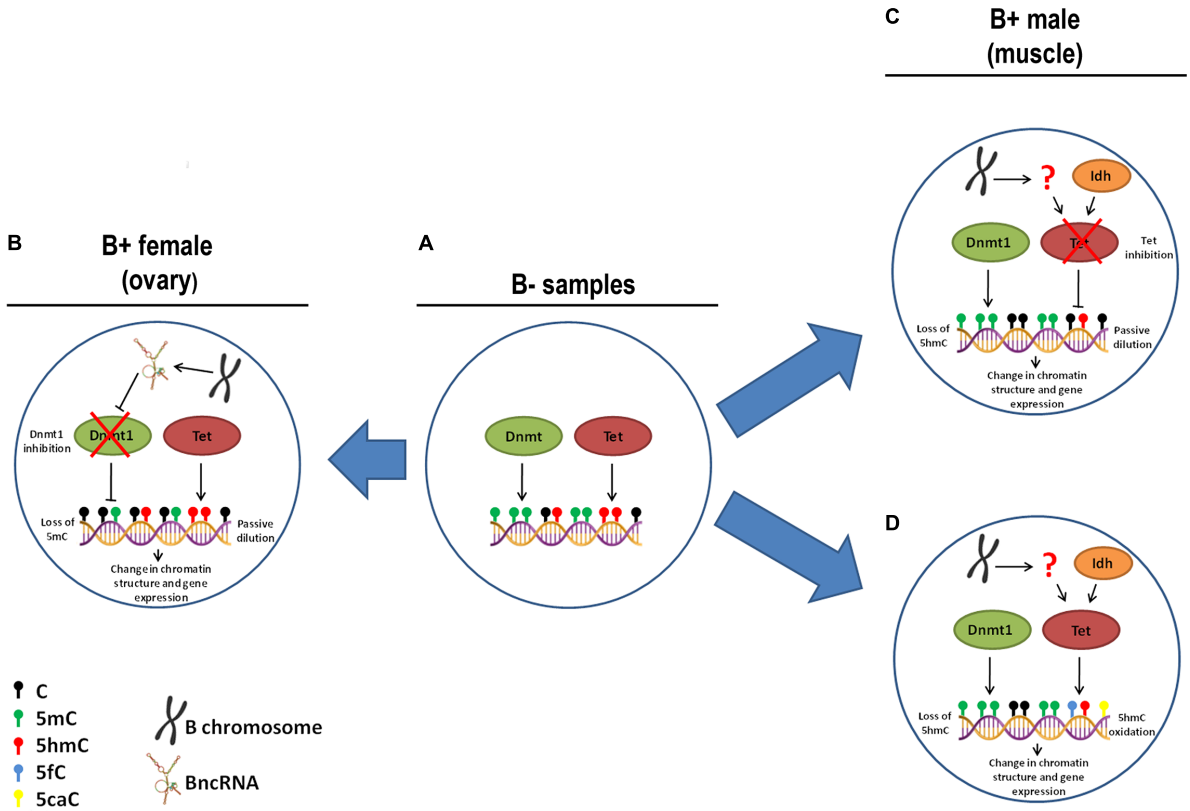

FIGURE 8 | Hypothesis of causes of 5mC and 5hmC variations with B chromosome presence in *A. latifasciata*. **(A)** In individuals without B chromosomes, regular function of Dnmt and Tet is observed. **(B)** In ovary of FB+, the chromosome transcribes BncRNA, which can inhibit Dnmt1 and promote loss of 5mC due to passive dilution. **(C)** In muscle of MB+, the B chromosome reduces the inhibition of Tet function by an unidentified mechanism and promotes loss of 5mhC due to passive dilution. **(D)** Alternatively, in muscle of MB+, the B chromosome can regulate Tet by an unidentified mechanism, which promotes 5hmC oxidation.

cell, which was confirmed in the quantification of 5mC and 5hmC. Therefore, the exploration of these effects is relevant to elucidate the mechanisms of transmission and maintenance of the B chromosome and to understand the consequences of epigenetic modifications.

Impact of B Chromosomes in the Global Levels of 5mC and 5hmC

To explore the molecular mechanisms underlying the loss of 5mC and 5hmC in some tissues of B+ individuals, we quantified the transcription levels of the genes related to epigenetic DNA modification (*dnmt1*, *dnmt3a*, *dnmt3b*, *tet1*, *tet2*, *tet3*, and *tdg*) and detected upregulation of *tet3* in muscle of FB+. However, we did not observe variations in the level of 5mC or 5hmC, indicating that these changes are not enough to impact epigenetic DNA modifications. Moreover, we explored the expression of candidate miRNAs targeting DNA modification genes and the transcription of the *idh1* and *idh2* genes (regulators of the DNA modification genes) in the presence of supernumerary chromosome, and no variation was observed. In ovaries of FB+, we did not observe 5mC reduction followed by 5hmC accumulation. Moreover, we did not observe substantial alterations in the transcription of *tet* genes, which can indicate that active modifications are not the mechanisms that promoted reduction of 5mC to 5hmC, so passive mechanisms might be responsible for this. Passive demethylation can be achieved by reduced activity of Dnmt1 during replication (Kohli and Zhang, 2013). However, we did not observe changes in the transcription of the *dnmt1* gene in these samples, indicating the possible occurrence of a posttranscriptional mechanism of regulation of Dnmt1, such as the action of miRNAs. In turn, we did not observe alteration in any candidate miRNA that would inhibit *dnmt1* RNA. In addition, another posttranscriptional mechanism of Dnmt1 inhibition can explain the interaction of Dnmt1 enzyme with RNAs. The association of the non-coding RNA ecCEBPA and Dnmt1 is involved in the reduction of Dnmt1 function in human (Di Ruscio et al., 2013). Interestingly, we observed positive interaction of BncRNA and Dnmt1 protein, which indicated that the BncRNA element, which is upregulated in B+ samples (Ramos et al., 2017), can act as an inhibitor of this enzyme, promoting passive DNA demethylation in ovaries of FB+ (**Figures 8A,B**). It is important to highlight that we found a tendency of reduction or increase in the level of 5mC and 5hmC and expression of some genes among the groups, but that was not statistically significant. This can be a consequence of the sample size.

With regard to loss of 5hmC in muscle of B+ males, this variation could be explained by reduction in the activity of Tet enzymes, but we did not observe differences in the transcription of *tet* genes or their candidate miRNA regulators. Moreover, we did not identify altered transcription of *idh1* and *idh2* genes, which encode cofactors of Tet enzymes. Therefore, two possible scenarios can explain the loss of 5hmC: passive dilution and 5hmC oxidation. In the first case, the B chromosome could express any factor that can inhibit Tet enzymes and avoid 5mC oxidation, while remaining 5hmC marks are lost during DNA replication (**Figures 8A,C**). In the second situation, the supernumerary chromosome could express any factor that stimulates Tet to convert 5hmC to 5fC and 5caC forms without 5mC conversion to 5hmC, which could explain the stable levels of 5mC between B− and B+ samples (**Figures 8A,D**).

Sex Effects in DNA Modifications Routes

We observed reduced level of 5mC in ovaries compared to testes independent of B chromosome presence, similar to observed in *Danio rerio* (Laing et al., 2018). These differences can result of high proportion of maturing gametes in testes, which are known to be hypermethylated among vertebrates (Jiang et al., 2013; Potok et al., 2013; Laing et al., 2018). Although the reduced 5mC level in ovaries, we found opposite upregulation of *dnmt1* and *dnmt3b* in this organ. This can be partially explained by miRNA repression of *dnmt3b* since the miR-132b is upregulated in ovaries. Moreover, we observed evidences of increased level of 5hmC (although not statistically significant) and *tdg* transcription in ovaries, which can indicate active DNA demethylation, although increased transcription of *tet* genes has not been observed.

Sex-Biased Effects of B Chromosomes

The epigenetic pattern of B chromosome presence observed in *A. latifasciata* was sex-associated, although B chromosomes occur in both sexes in this species. The B chromosome presence was correlated with global reduction of 5mC in ovaries of FB+ and reduction of 5hmC in muscle of MB+. Association of sex and B chromosomes has already been reported for *A. latifasciata*, as in the transcriptional variation of the non-coding BncRNA (Ramos et al., 2017) and the hnRNP Q-like gene (Carmello et al., 2017). Furthermore, the sex–B chromosome association has long been described in many species as diverse as fish and invertebrates (Camacho et al., 2011). Cichlid fish represent a promising model to investigate B chromosomes. The element has been described in 21 species of the group, and in eight species, all of the B-carrying individuals were female (Yoshida et al., 2011; Clark et al., 2017). Although the association of B chromosomes and sex is still enigmatic, we could hypothesize that in cichlids, the B chromosome bias for females could favor their drive during female meiosis. Considering that disturbances in the transmission of B chromosome during mitosis have been reported under DNA demethylation in rye (Neves et al., 1992), the sex variation detected for the *A. latifasciata* female B carriers could indicate that B chromosomes and sex physiology are in some way connected in this species.

CONCLUSION

The B chromosome of *A. latifasciata* has a pattern of 5mC and 5hmC epimarks that can suggest its active status or that DNA methylation, at least, is not involved in B-silencing. Moreover, our data correlate B chromosome presence with passive DNA demethylation associated with sex, and the epigenetic effects of the B chromosome presence can also explain the previously reported variations in the transcription levels of several classes of RNA. B chromosomes represent additional chromatin in the nucleus, and their presence seems to have an extensive impact on several cellular processes, including epigenetic modification. The state of the art of B chromosome science suggests that besides B chromosomes favoring their own drive during cell division, these accessory elements seem to cause major impacts in the cell and in the organism.

ETHICS STATEMENT

This study was carried out in accordance with the Brazilian College of Animal Experimentation and was approved by the Institute of Biosciences (UNESP) ethics committee on the use of animals (Protocol No. 486-2013 and 769-2015).

AUTHOR CONTRIBUTIONS

AC, BF, NV, BC, RO, and CM provided the substantial contributions to the conception of the work, wrote, read, and approved the manuscript. AC and NV performed the acquisition, analyses, and interpretation of cytogenetic data. AC and BC performed the nucleic acid acquisition and RNA expression analyses. AC and CM performed the epigenetic analyses and critically edited the final manuscript. AC and BF performed the bioinformatic analyses. AC and RO performed the statistical analyses.

FUNDING

This work was financially supported through grants from the São Paulo Research Foundation (FAPESP) (2012/21546-9, 2013/04533-3, and 2015/16661-1) and the National Council for Scientific and Technological Development (CNPq) (474684/2013-0 and 305321/2015-3).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00324/full#supplementary-material

REFERENCES

Alvarez-Garcia, I., and Miska, E. A. (2005). MicroRNA functions in animal development and human disease. *Development* 132, 4653–4662. doi: 10.1242/dev.02073

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Arvinden, V. R., Rao, A. K. D. M., Rajkumar, T., and Mani, S. (2017). Regulation and functional significance of 5-hydroxymethylcytosine in cancer. *Epigenomes* 1:19. doi: 10.3390/epigenomes1030019

Banaei-Moghaddam, A. M., Martis, M. M., Macas, J., Gundlach, H., Himmelbach, A., Altschmied, L., et al. (2015). Genes on B chromosomes: old questions revisited with new tools. *Biochim. Biophys. Acta* 1849, 64–70. doi: 10.1016/j.bbagrm.2014.11.007

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanisms, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5

Bertollo, L. A. C., Takahashi, C. S., and Moreira-Filho, O. (1978). Cytotaxonomic consideration on Hoplias lacerdae (Pisces, Erythrinidae). *Braz. J. Genet.* 1, 103–120.

Beukeboom, L. W. (1994). Bewildering Bs: an impression of the 1st B-Chromosome Conference. *Heredity* 73, 328–336. doi: 10.1038/hdy.1994.140

Bogomazova, A. N., Lagarkova, M. A., Panova, A. V., Nekrasov, E. D., and Kiselev, S. L. (2014). Reactivation of X chromosome upon reprogramming leads to changes in the replication pattern and 5hmC accumulation. *Chromosoma* 123, 117–128. doi: 10.1007/s00412-013-0433-x

Bugno-Poniewierska, M., Solek, P., Wronski, M., Potocki, L., Jezewska-Witkowska, G., and Wnuk, M. (2014). Genome organization and DNA methylation patterns of B chromosomes in the red fox and Chinese raccoon dogs. *Hereditas* 151, 169–176. doi: 10.1111/hrd2.00049

Camacho, J. P. M. (2005). "B chromosomes," in *The Evolution of the Genome*, ed. T. R. Gregory (San Diego, CA: Elsevier), 223–286. doi: 10.1016/B978-012301463-4/50006-1

Camacho, J. P. M., Schmid, M., and Cabrero, J. (2011). B chromosomes and sex in animals. *Sex. Dev.* 5, 155–166. doi: 10.1159/000324930

Camacho, J. P. M., Sharbel, T. F., and Beukeboom, L. W. (2000). B-chromosome evolution. *Philos. T. R. Soc. B* 355, 163–178. doi: 10.1098/rstb.2000.0556

Carchilan, M., Delgado, M., Ribeiro, T., Costa-Nunes, P., Caperta, A., Morais-Cecílio, L., et al. (2007). Transcriptionally active heterochromatin in rye B chromosomes. *Plant Cell* 19, 1738–1749. doi: 10.1105/tpc.106.046946

Carmello, B. O., Coan, R. L. B., Cardoso, A. L., Ramos, E., Fantinatti, B. E. A., Marques, D., et al. (2017). The hnRNP Q-like gene is retroinserted into the B chromosomes of the cichlid fish *Astatotilapia latifasciata*. *Chromosome Res.* 25, 277–290. doi: 10.1007/s10577-017-9561-0

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* 10, 295–304. doi: 10.1038/nrg2540

Clark, F. E., Conte, M. A., Ferreira-Bravo, I. A., Poletto, A. B., Martins, C., and Kocher, T. (2017). Dynamic sequence evolution of a sex-associated B chromosome in Lake Malawi cichlid fish. *J. Hered.* 108, 53–62. doi: 10.1093/jhered/esw059

Coan, R. L. B., and Martins, C. (2018). Landscape of transposable elements focusing on the B chromosome of the cichlid fish *Astatotilapia latifasciata*. *Genes* 9:269. doi: 10.3390/genes9060269

Costa, G., Barra, V., Lentini, L., Cillufo, D., and Di Leonardo, A. (2016). DNA demethylation caused by 5-aza-2′-deoxycytidine induces mitotic alterations and aneuploidy. *Oncotarget* 7, 3726–3739. doi: 10.18632/oncotarget.6897

Denis, H., Ndlovu, M. N., and Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Rep.* 12, 647–656. doi: 10.1038/embor.2011.110

Di Ruscio, A., Ebralidze, A. K., Benoukraf, T., Amabile, G., Goff, L. A., Terragni, J., et al. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503, 371–376. doi: 10.1038/nature12598

Efimova, O. A., Pendina, A. A., Tikhonov, A. V., Fedorova, I. D., Krapivin, M. I., Chiryaeva, O. G., et al. (2015). Chromosome hydroxymethylation patterns in human zygotes and cleavage-stage embryos. *Reproduction* 149, 223–233. doi: 10.1530/REP-14-0343

Efimova, O. A., Pendina, A. A., Tikhonov, A. V., Parfenyev, S. E., Mekina, I. D., Komarova, E. M., et al. (2017). Genome-wide 5-hydroxymethylcytosine patterns in human spermatogenesis are associated with semen quality. *Oncotarget* 8, 88294–88307. doi: 10.18632/oncotarget.18331

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, M. S. (2004). MicroRNAs targets in Drosophila. *Genome Biol.* 5:R1. doi: 10.1186/gb-2003-5-1-r1

Fantinatti, B. E. A. (2015). *Elucidação Dos Efeitos De Cromossomos Supernumeráveis Na Espécie De Ciclídeo Africano Astatotilapia latifasciata, Com Base Na Análise De Expressão De Micro-RNAs*. Doctoral Thesis (In Portuguese). Universidade Estadual Paulista, Instituto de Biociências de Botucatu.

Fantinatti, B. E. A., and Martins, C. (2016). Development of chromosomal markers based on next-generation sequencing: the B chromosome of the cichlid fish *Astatotilapia latifasciata* as a model. *BMC Genet.* 17:119. doi: 10.1186/s12863-016-0427-9

Fantinatti, B. E. A., Mazzuchelli, J., Valente, G. T., Cabral-de-Mello, D. C., and Martins, C. (2011). Genomic content and new insights on the origin of the B chromosome of the cichlid fish Astatotilapia latifasciata. *Genetica* 139, 1273–1282. doi: 10.1007/s10709-012-9629-x

Figueroa, M. E., Abdel-Wahab, O., Lu, C., Ward, P. S., Patel, J., Shih, A., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567. doi: 10.1016/j.ccr.2010.11.015

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Goll, M. G., and Halpern, M. E. (2011). DNA methylation in zebrafish. *Prog. Mol. Biol. Transl. Sci.* 101, 193–218. doi: 10.1016/B978-0-12-387685-0.00005-6

Han, F., Lamb, J. C., Yu, W., Gao, Z., and Birchler, J. A. (2007). Centromere function and nondisjunction are independent components of the maize B chromosome accumulation mechanism. *Plant Cell* 19, 524–533. doi: 10.1105/tpc.106.049577

Harrison, A., and Parle-McDermott, A. (2011). DNA methylation: a timeline of methods and applications. *Front. Genet.* 2:74. doi: 10.3389/fgene.2011.00074

Head, J. A. (2014). Patterns of DNA methylation in animals: an ecotoxicological perspective. *Integr. Comp. Biol.* 54, 77–86. doi: 10.1093/icb/icu025

Houben, A., Belyaev, N. D., Leach, C. R., and Timmis, J. N. (1997). Differences of histone H4 acetylation and replication timing between A and B chromosomes of Brachycome dichromosomatica. *Chromosome Res.* 5, 233–237. doi: 10.1023/B:CHRO.0000032297.10876.86

Huang, W., Du, Y., Zhao, X., and Jin, W. (2016). B chromosome contains active genes and impacts the transcription of A chromosomes in maize (*Zea mays* L.). *BMC Plant Biol.* 18:88. doi: 10.1186/s12870-016-0775-7

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303. doi: 10.1126/science.1210597

Jiang, L., Zhang, J., Wang, J. J., Wang, L., Zhang, L., Li, G., et al. (2013). Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell* 153, 773–784. doi: 10.1016/j.cell.2013.04.041

Jones, N. (2017). New species with B chromosomes discovered since 1980. *Nucleus* 60, 263–281. doi: 10.1007/s13237-017-0215-6

Kertesz, M., Iovino, M., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39, 1278–1284. doi: 10.1038/ng2135

Kohli, R. M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472–479. doi: 10.1038/nature12750

Koo, D. H., Han, F., Birchler, J. A., and Jiang, J. (2011). Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. *Genome Res.* 21, 908–914. doi: 10.1101/gr.116202.110

Krüger, J., and Rehmsmeier, M. (2006). RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* 34, W451–W454. doi: 10.1093/nar/gkl243

Kubiura, M., Okano, M., Kimura, H., Kawamura, F., and Tada, M. (2012). Chromosome-wide regulation of euchromatin-specific 5mC to 5hmC conversion in mouse ES cells and female human somatic cells. *Chromosome Res.* 20, 837–848. doi: 10.1007/s10577-012-9317-9

Kuc, C., Richard, D. J., Johnson, S., Bragg, L., Servos, M. R., Doxey, A. C., et al. (2017). Rainbow trout exposed to benzo[a]pyrene yields conserved microRNA binding sites in DNA methyltranferases across 500 milion years of evolution. *Sci. Rep.* 7:16843. doi: 10.1038/s41598-017-17236-x

Kurdyukov, S., and Bullock, M. (2016). DNA methylation analyses: choosing the right method. *Biology* 5:3. doi: 10.3390/biology5010003

Laing, L. V., Viana, J., Dempster, E. L., Uren-Webster, T. M., van-Aerle, R., Mill, J., et al. (2018). Sex-specific transcription and DNA methylation profiles of reproductive and epigenetic associated genes in the gonads and livers of breeding zebrafish. *Comp. Biochem. Physiol.* 222, 16–25. doi: 10.1016/j.cbpa.2018.04.004

Langdon, T., Seago, C., Jones, R. N., Ougham, H., Thomas, H., Forster, J. W., et al. (2000). De novo evolution of satellite DNA on the rye B Chromosome. *Genetics* 154, 869–884.

Leach, C. R., Tamzin, M. D., Franks, T. K., Spiniello, S. S., Hanrahan, C. F., and Timmis, J. N. (1995). Organisation and origin of a B chromosome centromeric sequence from Brachycome dichromosomatica. *Chromosoma* 103, 708–714. doi: 10.1007/BF00344232

Li, Y., Jing, X. A., Aldrich, J. C., Clifford, C., Chen, J., Akbari, O. S., et al. (2017). Unique sequence organization and small RNA expression of a "selfish" B chromosome. *Chromosoma* 126, 753–768. doi: 10.1007/s00412-017-0641-x

Li, Y., Miyanari, Y., Shirane, K., Nitta, H., Kunota, T., Ohashi, H., et al. (2013). Sequence-specific microscopic visualization of DNA methylation status at satellite repeats in individual cell nuclei and chromosomes. *Nucleic Acids Res.* 41:e186. doi: 10.1093/nar/gkt766

López-León, M. D., Cabrero, J., and Camacho, J. P. M. (1991). A nucleolus organizer region in a B chromosome inactivated by DNA methylation. *Chromosoma* 100, 134–138. doi: 10.1007/BF00418247

Ma, W., Gabriel, T. S., Martis, M. M., Gursinky, T., Schubert, V., Vrána, J., et al. (2017). Rye B chromosomes encode a functional Argonaute-like protein with in vitro slicer activities similar to its A chromosome paralog. *New Phytol.* 213, 916–928. doi: 10.1111/nph.14110

Martis, M. M., Klemme, S., Banaei-Moghaddam, A. M., Blattner, F. R., Macas, J., Schmutzer, T., et al. (2012). Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13343–13346. doi: 10.1073/pnas.1204237109

Mei, Q., Li, X., Meng, Y., Wu, Z., Guo, M., Zhao, Y., et al. (2012). A facile and specific assay for quantifying microRNA by an optimized RT-qPCR approach. *PLoS One* 7:e46890. doi: 10.1371/journal.pone.0046890

Mendioroz, M., Do, C., Jiang, X., Liu, C., Darbary, H. K., Lang, C. F., et al. (2015). Trans effects of chromosome aneuploidies on DNA methylation patterns in human Down syndrome and mouse models. *Genome Biol.* 16:263. doi: 10.1186/s13059-015-0827-6

Miao, V. P., Covert, S. F., and VanEtten, H. D. (1991). A fungal gene for antibiotic resistance on a dispensable ("B") chromosome. *Science* 254, 1773–1776. doi: 10.1126/science.1763326

Muller, P. Y., Janovjak, H., Miserez, A. R., and Dobbie, Z. (2002). Processing of gene expression data generated by quantitative real-time RT-PCR. *Biotechniques* 32, 1372–1374.

Muppirala, U. K., Honavar, V. G., and Dobbs, D. (2011). Predicting RNA-protein interactions using only sequence information. *BMC Bioinformatics* 12:489. doi: 10.1371/journal.pone.0046890

Navarro-Domínguez, B., Martín-Peciña, M., Ruiz-Ruano, F. J., Cabrero, J., Corral, J. M., López-León, M. D., et al. (2019). Gene expression changes elicited by a parasitic B chromosome in the grasshopper *Eyprepocnemis plorans* are consistent with its phenotypic effects. *Chromosoma* 128, 53–67. doi: 10.1007/s00412-018-00689-y

Neves, N., Barão, A., Castilho, A., Silva, M., Morais, L., Carvalho, V., et al. (1992). Influence of DNA methylation on rye B-chromosome non disjunction. *Genome* 35, 650–652. doi: 10.1139/g92-098

Pereira, H. S., Delgado, M., Viegas, W., Rato, J. M., Barão, A., and Caperta, A. D. (2016). Rye (Secale cereale) supernumerary (B) chromosomes associated with heat tolerance during early stages of male sporogenesis. *Ann. Bot.* 119, 325–337. doi: 10.1093/aob/mcw206

Plongthongkum, N., Diep, D. H., and Zhang, K. (2014). Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat. Rev. Gen.* 15, 647–661. doi: 10.1038/nrg3772

Poletto, A. B., Ferreia, I. A., and Martins, C. (2010). The B chromosomes of the African cichlid fish Haplochromis obliquidens harbour 18S rRNA gene copies. *BMC Genet.* 11:1. doi: 10.1186/1471-2156-11-1

Potok, M. E., Nix, D. A., Parnell, T. J., and Cairns, B. R. (2013). Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* 153, 759–772. doi: 10.1016/j.cell.2013.04.030

Ramos, E., Cardoso, A. L., Brown, J., Marques, D. F., Fantinatti, B. E. A., Cabral-de-Mello, D. C., et al. (2017). The repetitive DNA element BncDNA, enriched in the B chromosome of the cichlid fish Astatotilapia latifasciata, transcribes a potentially noncoding RNA. *Chromosoma* 126, 313–323. doi: 10.1007/s00412-016-0601-x

Rens, W., Wallduck, M. S., Lovell, F. L., Ferguson-Smith, M. A., and Ferguson-Smith, A. (2010). Epigenetic modifications on X chromosomes in marsupial and monotreme mammals and implications for evolution of dosage compensation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17657–17662. doi: 10.1073/pnas.0910322107

Ruban, A., Schmutzer, T., Scholz, U., and Houben, A. (2017). How next-generation sequencing has aided our understanding of the sequence composition and origin of b chromosomes. *Genes* 8:294. doi: 10.3390/genes8110294

Sambrook, J., and Russel, D. W. (2001). *Molecular Cloning. A Laboratory Manual, 3rd Edn*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Shen, L., Song, C. X., He, C., and Zuang, Y. (2014). Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev. Biochem.* 83, 585–614. doi: 10.1146/annurev-biochem-060713-035513

Simon, P. (2003). Q-Gene: processing quantitative real-time RT-PCR data. *Bioinformatics* 19, 1439–1440. doi: 10.1093/bioinformatics/btg157

Stitou, S., Díaz-de-la-Guardia, R., Jiménez, R., and Burgos, M. (2000). Inactive ribosomal cistrons are spread throughout the B chromosomes of Rattus rattus (Rodentia, Muridae). Implications for their origin and evolution. *Chromosome Res.* 8, 305–311. doi: 10.1023/A:1009227427575

Suresh, V., Liu, L., Adjeroh, D., and Zhou, X. (2015). RPI-Pred: predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res.* 43, 1370–1379. doi: 10.1093/nar/gkv020

Szulwach, K. E., Li, X., Li, Y., Song, C. X., Han, J. W., Kim, S. S., et al. (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* 7:e1002154. doi: 10.1371/journal.pgen.1002154

Trifonov, V. A., Dementyeva, P. V., Larkinm, D. M., O'Brien, P. C. M., Perelman, P. L., Yang, F., et al. (2013). Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer (Capreolus pygargus). *BMC Biology* 11:90. doi: 10.1186/1741-7007-11-90

Turcan, S., Rohle, D., Goenka, A., Walsh, L. A., Fang, F., Yilmaz, E., et al. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* 483, 479–483. doi: 10.1038/nature10866

Valente, G., Nakajima, R., Fantinatti, B. E. A., Marques, D. F., Almeida, R. O., Simões, R. P., et al. (2017). B chromosomes: from cytogenetics to system biology. *Chromosoma* 126, 73–81. doi: 10.1007/s00412-016-0613-6

Valente, G. T., Conte, M. A., Fantinatti, B. E. A., Cabral-de-Mello, D. C., Carvalho, R. F., Vicari, M. R., et al. (2014). Origin and evolution of B chromosomes in the cichlid fish Astatotilapia latifasciata based on integrated genomic analyses. *Mol. Biol. Evol.* 31, 2061–2072. doi: 10.1093/molbev/msu148

Varriale, A., and Bernardi, G. (2006). DNA methylation and body temperature in fishes. *Gene* 385, 111–121. doi: 10.1016/j.gene.2006.05.031

Yamaguchi, S., Hong, K., Liu, R., Inoue, A., Shen, L., Zhang, K., et al. (2013). Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res.* 23, 329–339. doi: 10.1038/cr.2013.22

Yoshida, K., Terai, Y., Mizoiri, S., Aibara, M., Nishihara, H., Watanabe, M., et al. (2011). B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. *PLoS Genet.* 7:e1002203. doi: 10.1371/journal.pgen.1002203

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

REVIEW
published: 15 July 2020
doi: 10.3389/fgene.2020.00742

Genomic, Transcriptomic and Epigenomic Tools to Study the Domestication of Plants and Animals: A Field Guide for Beginners

*Josué Barrera-Redondo, Daniel Piñero and Luis E. Eguiarte**

Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico

In the last decade, genomics and the related fields of transcriptomics and epigenomics have revolutionized the study of the domestication process in plants and animals, leading to new discoveries and new unresolved questions. Given that some domesticated taxa have been more studied than others, the extent of genomic data can range from vast to nonexistent, depending on the domesticated taxon of interest. This review is meant as a rough guide for students and academics that want to start a domestication research project using modern genomic tools, as well as for researchers already conducting domestication studies that are interested in following a genomic approach and looking for alternate strategies (cheaper or more efficient) and future directions. We summarize the theoretical and technical background needed to carry out domestication genomics, starting from the acquisition of a reference genome and genome assembly, to the sampling design for population genomics, paleogenomics, transcriptomics, epigenomics and experimental validation of domestication-related genes. We also describe some examples of the aforementioned approaches and the relevant discoveries they made to understand the domestication of the studied taxa.

Keywords: population genomics, pangenomics, ancient DNA, differential expression analysis, epialleles, genome editing

OPEN ACCESS

Edited by:
TingFung Chan,
The Chinese University of Hong Kong, China

Reviewed by:
Eric Von Wettberg,
The University of Vermont, United States
Martin Mascher,
Leibniz Institute for Plant Genetics and Cultural Plant Research (IPK), Germany
David Irwin,
University of Toronto, Canada

***Correspondence:**
Luis E. Eguiarte
fruns@unam.mx

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *25 April 2020*
Accepted: *22 June 2020*
Published: *15 July 2020*

Citation:
Barrera-Redondo J, Piñero D and Eguiarte LE (2020) Genomic, Transcriptomic and Epigenomic Tools to Study the Domestication of Plants and Animals: A Field Guide for Beginners. Front. Genet. 11:742. doi: 10.3389/fgene.2020.00742

INTRODUCTION

The modern study of domestication of plants and animals is multidisciplinary, and relevant contributions come from botany, zoology, archeology, genetics, ethnobiology, biogeography, and linguistics (Larson et al., 2014). Modern domestication studies seek to understand the dates of domestication, the places where domestication started and number of times that domestication took place, as well as the details of the evolutionary and ecological forces that led to the divergence between the domesticated taxa and their wild relatives and ancestors (Zeder, 2006; Larson et al., 2014).

Given that domestication is an evolutionary process, genetics emerged as a powerful tool to understand the domestication of plants and animals, revealing the demographic history of the domesticated taxa and the genetic variants that underlie their domesticated phenotypes (Zeder et al., 2006; Gepts, 2014). The advent of high-throughput sequencing technologies sparked the use of genomic studies to understand the domestication of crops and animals in a much deeper

level than previously imagined, as researchers can now pinpoint the genetic changes that allowed domestication to happen (Ross-Ibarra et al., 2007; Gepts, 2014).

WHY AND HOW TO USE A GENOMIC APPROACH IN DOMESTICATION STUDIES? TOP-DOWN AND BOTTOM-UP APPROACHES FOR THE STUDY OF DOMESTICATION

In genetics, we refer to *top* or *up* when referring to a specific phenotype, while we refer to *bottom* or *down* when referring to the underlying genotype responsible for that trait. Thus, top-down approaches start by studying a particular phenotype and searching for its genetic basis. Huge advances in the genetic study of domestication traits have been made using classic top-down approaches (*e.g.*, Sax, 1923; Paterson et al., 1988; Doebley and Stec, 1991; Doebley et al., 1995), which are performed by analyzing the phenotypic traits of interest between wild and domesticated taxa, and then finding the genetic variant or variants that correlate with the phenotypic traits through the mapping of quantitative trait loci and linkage disequilibrium (Ross-Ibarra et al., 2007; Kantar et al., 2017). These top-down approaches are precise in finding causal variants involved in the evolution of specific traits, but usually they are very labor-intensive and are biased towards *a priori* selected phenotypes to be compared between wild and domesticated taxa (Ross-Ibarra et al., 2007; Kantar et al., 2017).

In contrast to top-down approaches, bottom-up approaches start by analyzing the genetic variation within genomes in order to detect potential signals of selection related to the domestication process and finally associate such evolutionary signals to important loci and domestication phenotypes (Ross-Ibarra et al., 2007; Kantar et al., 2017). In the last decade, high-throughput sequencing technologies allowed us to analyze entire genomes of one or several individuals of domesticated taxa, and to compare them to different varieties or to their wild relatives (*e.g.*, Hufford et al., 2012; Yang et al., 2012; Li et al., 2013; Wang et al., 2019; Zeng et al., 2019).

Bottom-up approaches do not need an *a priori* phenotypic target, enabling a genome-wide search of domestication-related loci without previous background of possible candidates, revealing important traits that can hardly be studied using a top-down approach (Ross-Ibarra et al., 2007; Kantar et al., 2017). Nevertheless, the results of bottom-up approaches can be limited by the sampling scheme, the density of genetic markers, and the detection of false positives (Tiffin and Ross-Ibarra, 2014), so these genomic approaches have to be properly and carefully designed in order to obtain satisfying results (De Mita et al., 2013).

Genomic data facilitated the widespread and reliable use of bottom-up approaches to study plant and animal domestication, but top-down strategies were also aided by genomics, allowing a more efficient search of genotype-phenotype correlations through genome-wide association studies (GWAS; Wang G.-D. et al., 2014), which can be defined as experimental designs that are used to detect the association between genetic variation in a population and phenotypical traits of interest (Visscher et al., 2017).

Genome-wide genetic markers allows to differentiate between global and local evolutionary signals occurring throughout the genome (Diao and Chen, 2012), discerning the signals of selection during domestication (Vitti et al., 2013) from other fine-scale signals of demographic events that occurred during the domestication process (Meyer and Purugganan, 2013; Guerra García and Piñero, 2017).

The use of modern genomic tools is not limited to population genetics, as other interesting approaches can reveal important aspects of the domestication process. For instance, one can analyze changes in the transcriptional activity of genes related to domestication (Hekman et al., 2015), demonstrate the phenotypic effects of certain alleles through the use of genomic editing tools (Zhou J. et al., 2019), search for epigenetic patterns that changed between domesticated and wild taxa (Janowitz Koch et al., 2016) or analyze the genetic makeup of archeological samples (Irving-Pease et al., 2019).

This review describes the necessary steps and data to start a genomic research project towards understanding domestication, the questions that can be approached using genomic data and the main results obtained from previous studies using these methods (**Figure 1**).

WHOLE-GENOME ASSEMBLY AND REFERENCE GENOMES

Whole-genome assembly is one of the first steps in modern domestication studies, since it generates a reference genome that is useful for downstream analyses. Whole-genome assembly projects require the use of high-throughput sequencing technologies such as Illumina (*e.g.,* Sun et al., 2017), PacBio (*e.g.,* Badouin et al., 2017; VanBuren et al., 2018), Oxford Nanopore (*e.g.,* Belser et al., 2018) or a combination of these (*e.g.,* Bickhart et al., 2017; Zhou Y. et al., 2019) to sequence the genome of interest of a single individual. Before starting a genome assembly project, a rough estimate of the haploid genome size must be known as well as the ploidy of the organism, since the assembly difficulty and sequencing cost are determined by both factors (Sims et al., 2014). In order to successfully assemble eukaryotic genomes, where repetitive elements usually comprise a significant portion of its content [ranging from 3% in tiny genomes such as *Utricularia gibba* (Ibarra-Laclette et al., 2013) up to 65.5% in huge genomes such as *Ambystoma mexicanum* (Nowoshilow et al., 2018)], it is necessary to generate sequencing libraries with large insert sizes – called mate-pair libraries – or use long-read sequencing technologies such as PacBio or Oxford Nanopore (Levy and Myers, 2016; Sohn and Nam, 2016). Additionally, the use of chromosome conformation capture (Mascher et al., 2017), optical mapping (Dong et al., 2013) or linkage maps obtained from crosses (Fierst, 2015) will help achieve chromosome-level assemblies that are highly desirable to adequately assess haplotypes, linkage disequilibrium, putative genomic rearrangements and the

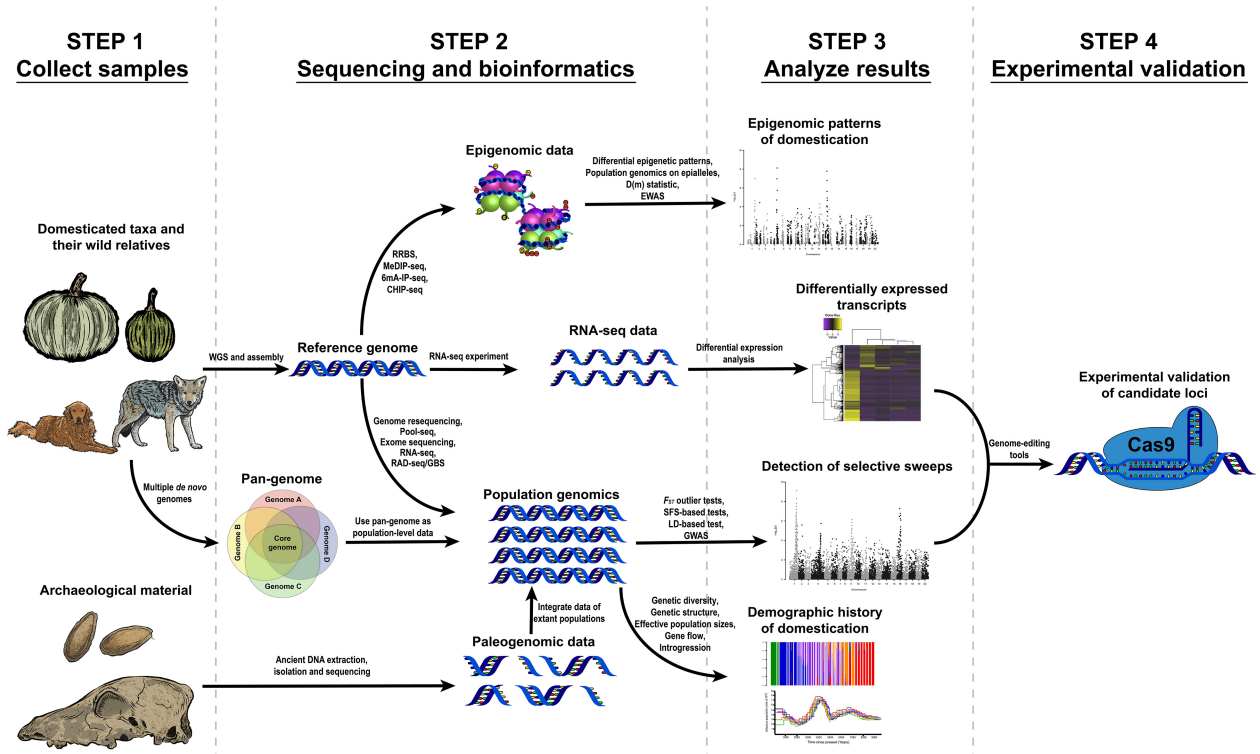

FIGURE 1 | Proposed workflows to study different problems related to the domestication of plants and animals through genomic, transcriptomic and epigenomic tools.

genomic location of candidate loci (Sohn and Nam, 2016; see **Table 1**).

After sequencing and assembling the genome of at least one individual, it must be properly annotated before it can be of any use. Since eukaryotic genes are structurally complex, genome assemblies require the additional sequencing of RNA data from the same species to be used as transcriptomic evidence, alongside homology evidence from other curated genomes and *ab initio* predictions based on the underlying structure of genes, in order to be successfully annotated (Yandell and Ence, 2012; see **Table 1**). Even though whole-genome assembly projects were previously restricted to large research groups (*e.g.*, Schnable et al., 2009; Tomato Genome Consortium, 2012), the sequencing cost per nucleotide is declining constantly in all the aforementioned technologies, making genome analyses accessible for a large part of the research community (Muir et al., 2016). The current bottleneck for small research groups is usually not the cost of sequencing itself, but rather the availability of computational resources capable of storing and analyzing huge amounts of data (Muir et al., 2016).

The main purpose of assembling a genome in a domestication study is to use it as a reference for high-quality population data to infer the selection, introgression and recombination processes, and to design posterior studies for experimental validation of candidate loci. Even though several population-level analyses based on reduced-representation genome sequencing can be performed in the absence of a reference genome (De Wit et al., 2012; Mastretta-Yanes et al., 2015), the use of a reference genome alongside population data enables the correct identification of otherwise anonymous loci into specific genes or regions within the genome and it makes possible the identification and the proper handling of linkage between loci (Fitz-Gibbon et al., 2017). Also, it can help to discriminate between orthologous and paralogous loci, which is critical given the large size of many genomes and the frequent genome duplication processes experienced during the evolution of plant and animal lineages (Clark and Donoghue, 2018; Zadesenets and Rubtsov, 2018).

Thus, the availability of a reference genome is desired for genomic analyses concerning domestication. Luckily, domesticated taxa are usually economically relevant, drawing the attention of several research groups worldwide and in some cases helping to fund the projects. Therefore, reference genomes are usually available for domesticated species, since such data is also relevant for other research areas, such as crop improvement and breeding programs (Ellegren, 2014). However, it should be noted that using a single reference genome can lead to reference bias, where sequenced individuals that are more distantly related to the reference will tend to have fewer predicted variants due to mismatches while mapping the reads (Günther and Nettelblad, 2019).

Besides its use as a reference genome for population-level data, the analysis of several whole-genome assemblies between

TABLE 1 | List of key publications with other reviews that are focused on specific topics, as well as some notable examples of research articles using some of the methods described in this review with reliable results.

Topic	Citation	Usefulness/importance
Genome assembly and reference genomes	Sohn and Nam, 2016	In-depth review on genome assembly. Includes compelling explanations behind the genome assembly algorithms and an extensive list of genome assembly strategies.
	Yandell and Ence, 2012	In-depth review on eukaryotic genome annotation, a description of the available tools to predict genes and best practices when predicting genes.
Sequencing strategies	Meirmans, 2015	Classic review concerning common pitfalls that should be avoided in a population genomic study. A compulsory review for any newcomer to population genomics.
	Dorant et al., 2019	Empirical study that compares the efficiency of Pool-seq, RADseq and Rapture to detect weak signals of genetic structure in lobsters.
	Inbar et al., 2020	Empirical study that compares the efficiency of whole-genome sequencing, Pool-seq and RADseq for GWAS in ants.
Pan-genomics	Golicz et al., 2016a	In-depth review about pan-genomics in plant species, its advantages over the use of reference genomes, a guide on how to generate pan-genomes and the importance of studying structural variants. The article is dedicated to plants, but the rationale and methods can also be applied to other eukaryotes.
	Khan et al., 2020	Opinion article detailing the relevance of pan-genomes as a necessary next step from reference genomes. The authors also highlight the importance of including wild taxa into pan-genomics and propose the idea of genus-level super-pan-genomes.
	Gao et al., 2019	Landmark study of the tomato pan-genome. The authors sequenced 725 accessions from the domesticated tomato and its wild relatives. They found 4,873 additional genes, including several well-characterized genes that were absent from the reference-genome. They also evaluated the presence-absence variants between the wild and domesticated tomatoes, which were enriched in disease-resistance genes.
Demographic analyses	Linck and Battey, 2019	Research study focused on the effects of minor allele-frequency filters to detect genetic structure in populations. Gives a good explanation on the rationale behind the clustering-based methods to detect structure.
	Mather et al., 2020	Review dedicated to the theoretical background and technical requirements of PSMS and MSMC to infer changes in effective population sizes and coalescent times.
	Gerbault et al., 2014	Excellent review on how to use Bayesian approaches to test different demographic models of domestication.
	Frantz et al., 2015	Landmark study on pig domestication. The authors make use of Approximate Bayesian computation to compare domestication scenarios, they use clustering-based methods to detect genetic structure and used a graph-based method to infer the genetic relationship between pig and wild boar populations.
Selection scans	Vitti et al., 2013	Good review focused on explaining the rationale behind many of the bottom-up tests to detect selection and the genomic signals they are sensitive to.
	De Mita et al., 2013; Lotterhos and Whitlock, 2015	Classic simulation-based studies that compare different scenarios to evaluate the best sampling strategies and the most powerful methods to detect selection throughout the genome, according to the reproductive nature of the organism under study.
	Gibson, 2018	A primer dedicated to understanding the principles behind GWAS and its ability to detect polygenic effects on quantitative traits.
	Hufford et al., 2012	A landmark paper that illustrates how to perform genome scans to detect domestication-related loci in domesticated taxa, and the importance of these loci for crop improvement. The paper studies the domestication of maize, but a similar study design can be applied to domesticated animals.
Paleogenomics	Irving-Pease et al., 2019	Exhaustive book chapter dedicated to the study of ancient DNA to understand domestication.
	Allaby et al., 2019	Research study that casts into doubt the long-lasting idea that domestication processes lead to strong population bottlenecks by re-analyzing data based on ancient DNA samples.
	Daly et al., 2018	Remarkable study that sequenced and analyzed 83 mitochondrial genomes and 51 nuclear genomes from ancient goat samples. The authors found signals of ancient introgression events, as well as ancient selective signals related to several traits that are shared with modern goats.
Transcriptomics	Fang and Cui, 2011	General guideline on how to adequately design an RNA-seq experiment to avoid technical mistakes and generate meaningful results.
	Yang and Kim, 2015	General guideline on how to analyze RNA-seq data to assess differential expression.
	Hekman et al., 2015	In-depth review dedicated to study the domestication process through transcriptomics, including methodological strategies and challenges.
	Hradilová et al., 2017	An excellent study that combines transcriptomic data with metabolomic data and morphological data between domesticated and wild peas. The analysis of multi-omic data allowed them to get a better understanding behind seed dormancy and pod dehiscence in domesticated peas.
Epigenomics	Guerrero-Bosagna, 2012; Heard and Martienssen, 2014; Burggren, 2016	Contrasting views on the role of transgenerational epigenetic inheritance in evolution. The topic is still debated and should be viewed critically.
	Jensen, 2015	In-depth review on the rationale and advances of epigenetic studies to understand domestication. The manuscript is focused on animal behavior, but many of the ideas can also be applied to domesticated plants.

(Continued)

TABLE 1 | Continued

Topic	Citation	Usefulness/importance
	Janowitz Koch et al., 2016	A landmark paper showing the importance of epigenetic marks on dog domestication and its association with behavioral traits. The study doesn't just compare the methylation marks between wolves and dogs, but also assess the heritability of the methylation marks and proposes a formal test to detect selection on epialleles.
Genome-editing tools	Boettcher and McManus, 2015).	Review on novel genome-editing techniques and RNA interference. Useful to compare and choose the best tool to validate candidate loci.
	Shan et al., 2020	A general guide on how to develop a CRISPR/Cas9 system on a non-model plant species.
	Soyk et al., 2017	Landmark paper that uses genome-editing to validate two candidate genes related to fruit size and reduced fruit dropping in tomato. The authors also detect the emergence of undesirable traits in domesticated tomatoes due to an epistatic effect between both domesticated loci and introduce wild alleles to generate new tomato phenotypes with reduced degrees of the undesirable traits.
Perspectives	Piperno, 2017	Review centered on the potential application of an extended synthesis framework to understand domestication. Centered around the concepts of niche construction, transgenerational epigenetic inheritance and developmental plasticity.

domesticated and wild taxa will help us reveal structural differences between the genome of a domesticated taxon and its closest wild relatives, such as duplications, chromosome rearrangements or presence/absence of entire genes and genomic regions (Yang et al., 2012; Wang W. et al., 2014; Xie et al., 2019). Since selection and bottlenecks during domestication often leads to the fixation of mutations that involve a loss of function (Renaut and Rieseberg, 2015; Moyers et al., 2018), comparative analyses using genome assemblies of wild ancestors may also reveal these changes in genes that could not be properly predicted within the domesticated genome (Moyers et al., 2018). In this sense, further efforts should be made to assemble high-quality genomes of wild relatives alongside the domesticated taxon of interest (Brozynska et al., 2016; Xie et al., 2019).

STRATEGIES TO GATHER ADEQUATE POPULATION GENOMICS DATA

Genome assemblies alone give us a limited view on domestication, unless several genomes of wild relatives (if known and available) and domesticated individuals are sequenced, because evolution is a population-level process, and in consequence population data is necessary to address most of the evolutionary questions in domestication (Wang G.-D. et al., 2014; Guerra García and Piñero, 2017). Population genomics examines the genetic variation within and between populations that is scattered across the entire genome to assess the demographic history, phylogenetic relations and selective pressures of a species (Jorde, 2001). Several types of genomic data can be evaluated at the population level, including single nucleotide polymorphisms (SNPs), indels and copy number variations; but SNPs are the most commonly analyzed of the three (Seal et al., 2014).

All population-level sequencing techniques share common pitfalls that should be known and avoided before investing any money on sequencing. Population sampling should be planned carefully, as the sampling scheme has a stronger impact over sequencing to obtain reliable results in any analysis (Meirmans, 2015). Also, different populations should be mixed, rather than being sequenced on separate libraries or sequencing lanes, as failing to do so will generate sequencing biases that can be confused with biological patterns (Meirmans, 2015; see **Table 1**).

Once adequate genomic population data is gathered, we need to analyze the demographic processes that shaped the genetic variation and the population structure of contemporary populations during the domestication process. This data is necessary to perform tests to detect natural and artificial selection, which are required to understand the genetic base of domestication syndromes (Ross-Ibarra et al., 2007). There are several approaches to obtain population data at a genomic scale, which differ in the fraction of the genome that is sequenced, therefore determining the sequencing cost of each sample (Schreiber et al., 2018).

Whole-Genome Sequencing of Populations

After assembling a reference genome, one of the next possible strategies to understand domestication is to sequence the complete genome of several individuals. This approach requires the alignment of the sequencing reads back to a reference genome, in order to infer the variable sites between individuals and know the genetic elements (e.g., genes, upstream regulators, repetitive elements, non-coding RNAs) associated to those sites. The main benefit of this approach is its potential to retrieve all the variant sites within an individual's genome that are structurally represented in the reference genome. Whole-genome sequencing can be used in almost any population-level test of interest (Schreiber et al., 2018). Common practices recommend a sequencing depth around 30× per individual, but empirical studies in pigs suggests that even 10x is enough to cover up to 99% of a genome with accurate detection of variant sites (Jiang L.G. et al., 2019). The main drawback of this approach is the sequencing cost of each sample, which is significantly higher compared with other approaches, especially for organisms with large genomes such as polyploid crops or mammals (Schreiber et al., 2018). This can lead researchers to evaluate a trade-off between sequencing depth and number of sampled individuals to optimize their resources. Simulation studies suggest that sequencing more individuals is more convenient to obtain

reliable results, even at the expense of lower sequencing depths per individual (Fumagalli, 2013).

Alternatives to Whole-Genome Sequencing

Other approaches aim to reduce the sequencing cost per samples by pooling the DNA of several individuals into a single sequencing library (Futschik and Schlötterer, 2010) or by reducing the portion of the genome that is sequenced (often named as reduced-representation sequencing), either by sequencing arbitrary defined segments scattered across the genome, by targeting the desired portions of the genome or by sequencing the transcriptionally active portions of the genome (Schreiber et al., 2018). These techniques are especially helpful for organisms with very large genomes, and some of these methods can even be used in the absence of a reference genome (Mastretta-Yanes et al., 2015; Schreiber et al., 2018). Furthermore, the reduced representation of the genome means that those fewer regions that are targeted can have a high sequencing depth, leading to higher accuracy of the observed genetic variation and better heterozygosity estimations (Schreiber et al., 2018). Additionally, the reduced sequencing cost per sample allows for a large number of sequenced individuals and populations that, with a proper sampling strategy, can lead to robust results (De Mita et al., 2013; Lotterhos and Whitlock, 2015). Due to the fragmented nature of these sequencing techniques, reduced representation data alone may be insufficient to pinpoint all or even the most important possible causal genetic variants associated to the domestication syndromes (Lowry et al., 2017), but they are still useful to infer basic genetic statistics, infer demographic properties and past demographic scenarios, detect some signatures of selective sweeps across the genome and even perform GWAS for domestication traits of interest (Andrews et al., 2016; Schreiber et al., 2018).

Pool Sequencing

Pool sequencing (Pool-seq) is a promising alternative to whole-genome sequencing with a much lower cost (Futschik and Schlötterer, 2010). As the name suggests, Pool-seq consists of sequencing a large pool of individuals for a given population into a single high-throughput sequencing library, instead of sequencing each individual separately, allowing an accurate estimation of allele frequencies and other parameters of population genetics at the expense of losing individual-level information (Futschik and Schlötterer, 2010). This method requires to map the reads against a reference genome of the same species in order to work (Schlötterer et al., 2014). It is intended for sequencing large pools of individuals (>40 individuals per population is recommended, but >100 is optimal), otherwise the allele frequencies will not be estimated accurately (Schlötterer et al., 2014). The relative amount of pooled DNA of each individual in a Pool-seq study should be similar in order to avoid overrepresentation of individual alleles, a task that is often challenging (Schlötterer et al., 2014).

Pool-seq has several limitations that should be considered based on the objectives of the research project. It is difficult to discard a low-frequency allele from a sequencing error, but this problem is potentially fixed by either establishing a minor allele frequency threshold for SNP calling or by using pool replicates (Schlötterer et al., 2014). One important limitation is the inability of Pool-seq data to estimate linkage disequilibrium and haplotype phasing, which is particularly important to evaluate the non-independence of genetic signals in demographic studies and selective scans (Schlötterer et al., 2014). Finally, assessing genetic structure can be difficult and sometimes misleading when using Pool-seq, due to potential biases in individual allele representations within the pool (Dorant et al., 2019). This makes Pool-seq an adequate method for GWAS, selective sweeps and some methods based on allele frequencies when resources are limited (Luu et al., 2017; Inbar et al., 2020), but the loss of individual-level information makes many of the demographic inferences difficult, as populations need to be predefined before sequencing (Dorant et al., 2019), and the bioinformatic tools that handle Pool-seq data are scarce.

Exome Capture and Sequencing

Exome sequencing is another lower-cost alternative to whole-genome sequencing which targets the protein-coding regions of the genome (Warr et al., 2015; Kaur and Gaikwad, 2017). Protein-coding genes represent a small fraction of eukaryotic genomes, which is particularly useful for most population genomic studies, since it represents mostly functional elements within genomes (Kaur and Gaikwad, 2017). This technique is usually performed using hybridization probes, which requires previous knowledge of the genome content as well as a *priori* selection of regions of interest in order to design probes (Kaur and Gaikwad, 2017). Fortunately, hybridization probes are already available for several domesticated plants and animals (Warr et al., 2015; Kaur and Gaikwad, 2017).

Despite its advantages, exome sequencing can generate an uneven sequencing depth in certain genomic positions, unlike whole-genome sequencing that shows a uniform distribution of reads throughout the genome (Lelieveld et al., 2015). Another important limitation of exome sequencing is its bias towards the protein-coding portion of the genome, since increasing evidence shows that many of the genetic changes that have been directly associated to domestication traits are located within *cis*-regulatory elements, noncoding RNAs and other *trans*-regulatory elements, rather than within the open reading frame of the genes (Swinnen et al., 2016). Despite its limitations, demographic history and selective sweeps can still be detected using this sequencing method (Pankin et al., 2018).

RNA Sequencing of Populations

Transcriptome sequencing (also known as RNA-seq) is another useful approach to obtain population-level data from the transcriptionally active elements within genomes (De Wit et al., 2012). RNA-seq can be mapped against a reference genome to detect genetic variants and determine the genomic regions of interest, but it can also be analyzed in the absence of a reference genome (De Wit et al., 2012), since transcriptomes can be assembled *de novo* (Haas et al., 2013) and the functional

annotation of the assembled transcripts is relatively easy (Bryant et al., 2017).

However, transcription profiles are dependent on the sequenced tissues and organs, the development stage of the organism, and the influence of external stimuli, capturing just the transcripts that are active at the moment of RNA extraction (Hekman et al., 2015). This complexity can generate important biases in the relative abundance of certain transcripts over others and overlook potential adaptative genes whose expression are context dependent (Hekman et al., 2015; Kaur and Gaikwad, 2017). Nonetheless, RNA-seq is still a good option for species with large genomes that are hard to assemble (De Wit et al., 2012). Similarly to exome-sequencing, RNA-seq data can be used to evaluate demographic history and selective sweeps, but the selective signals are restricted to the transcriptionally active part of the genome, and cannot be used to evaluate structural variants (Schreiber et al., 2018).

Restriction Site-Associated DNA Sequencing

Restriction site-associated DNA sequencing (RAD-seq), which may also be referred to as genotyping by sequencing (GBS), has been one of the most popular options for cost-affordable population genomics in the last decade (Davey and Blaxter, 2010). The technique consists in using restriction enzymes to digest the DNA and sequence the regions adjacent to the restriction sites that are scattered across the genome (Davey and Blaxter, 2010). It can also be combined with sequence capture techniques to target specific loci of interest (Ali et al., 2016). RADseq data can either be mapped against a reference genome or it can be assembled *de novo* (Catchen et al., 2013; Mastretta-Yanes et al., 2015), making it a versatile technique for species with scant genomic resources.

However, empirical studies show that using certain *de novo* approaches for RAD-seq data can lead to fewer predicted SNPs due to errors in the definition of loci and treatment of sequencing errors (Shafer et al., 2017), all which may subsequently alter downstream analyses, especially those based on the distribution of allele frequencies within the genome of a population, also known as the site frequency spectrum (SFS) (Shafer et al., 2017). For this reason, a reference-based approach is highly recommended as long as the reference genome is closely related to the population dataset (Shafer et al., 2017). Furthermore, RADseq data could involve errors when a polymorphism resides within a restriction site, which prevents the enzyme to cut in individuals carrying such polymorphism, leading to failures in sequencing that region in homozygous individuals (null alleles) and makes heterozygous individuals to look like homozygotes (allele dropout) (Andrews et al., 2016). Finally, the capacity of RADseq libraries to adequately perform selective scans has been casted into serious question (Lowry et al., 2017). Its potential capacity to detect selective sweeps is dependent on the genome size, the density of variants detected for a given genomic region and specially the length of the extent of linkage disequilibrium in the genome (Lowry et al., 2017). Thus, when a species genome has short regions in linkage disequilibrium (due to high recombination rates) and the SNP density is low (particularly in large genomes), odds are that the selective scans will likely miss a significant portion of selective sweeps associated to domestication (Lowry et al., 2017).

PAN-GENOME ANALYSES IN DOMESTICATED AND WILD TAXA

An increasing number of studies are revealing that structural variants (copy-number variation, presence/absence of genomic regions, inversions, transversions, translocations) are common within plant and animal populations (Khan et al., 2020). Thus, the use of a single reference genome hampers our ability to study the full repertoire of genetic variation within a species (Golicz et al., 2016a; Zhao et al., 2018). Structural variants such copy-number variation can contain functional genomic elements that are usually under relaxed selective pressures and can serve as the basis of adaptation given specific environments and selective regimes (Lye and Purugganan, 2019). Coincidentally, copy-number variation and other structural variants play an important role in the emergence of domestication traits, as well as diversification traits in landrace varieties (Lye and Purugganan, 2019). Some studies estimate that at least one third of the known domestication loci are structural variants, and up to one in seven genes can be hemizygous (*i.e.*, with one copy) in grapevine individuals (Zhou Y. et al., 2019). Despite its importance, structural variants cannot be properly analyzed using any of the aforementioned techniques. This led the research community to adopt the concept of the pan-genome, an idea that first appeared in microbiology (Tettelin et al., 2005), into the study of plant and animal genomes (Golicz et al., 2016a).

The concept of pan-genome rests on the idea that the genomes of individuals within a population or species share a core set of genes that unifies them (i.e., the core genome), but also contain a fraction of genes that are absent from one or more individuals (i.e., the accessory or dispensable genome), which altogether give rise to the pan-genome of such population or species (Tettelin et al., 2005).

There are three main methods to generate a pan-genome: the alignment and comparison of multiple *de novo* genome assemblies, the iterative assembly of several genomes from an initial reference or the use of *de Bruijn* graph assemblers to jointly assemble several genomes (Golicz et al., 2016a; see **Table 1**). Since domestication reduces the genetic diversity of a taxon, often eliminating portions of the dispensable genome that contain genes involved in local adaptation, the use of wild relatives is crucial to generate a representative pan-genome for a species (Khan et al., 2020). Once a pan-genome is generated, it can be used alongside whole-genome sequencing data to analyze the structural variants between and within populations, revealing novel loci involved in the development of domestication-related traits that would have stayed hidden when using a single reference genome (Li et al., 2014; Zhao et al., 2018). Besides, the use of a pan-genome alleviates the inherent reference biases of a single reference genome (Günther and Nettelblad, 2019).

Pan-genome studies have revealed additional selective sweeps and structural variants associated to the domestication process,

which were not identified using sequencing data with a single reference genome (Li et al., 2014; Zhao et al., 2018). Pan-genomes are already available for several species (**Figure 2**) such as maize (Brohammer et al., 2018), wheat (Montenegro et al., 2017), *Brassica oleracea* (Golicz et al., 2016b) or *Brassica napus* (Hurgobin et al., 2018); and pan-genome analyses to study domestication have already been performed in soybean (Li et al., 2014), rice (Zhao et al., 2018), sunflower (Hübner et al., 2019) and tomato (Gao et al., 2019). While current eukaryote pan-genome analyses are focused on plant species (Golicz et al., 2016a, see **Table 1**) and goats (Li et al., 2019), other livestock researchers may soon venture into this field. As sequencing technologies become cheaper, multiple pan-genomes from different species of the same genus should eventually be combined to create a super-pan-genome that represents the entire genetic content available in a genus with one or more domesticated taxa, as it would include the diversity of all their wild relatives (Khan et al., 2020).

POPULATION GENETICS AND DEMOGRAPHIC ANALYSES OF THE DOMESTICATION PROCESS

Demography and population size changes during the domestication process is tightly related to unraveling some of the most fundamental questions of the domestication process. These analyses can help answer questions such as possible centers of origin and diversification, patterns of migration and expansion throughout these centers, gene flow between domesticated and wild taxa, number of domestication events, the extent of genetic erosion in the domesticated taxon, levels of global genetic differentiation between wild and domesticated taxa, the patterns of adaptive and neutral introgression among them, and in some cases even the number of generations that have elapsed since domestication and other processes such as differentiation and local adaptation of domesticated taxa (Meyer and Purugganan, 2013; Guerra García and Piñero, 2017).

Genetic Diversity in Populations

A first necessary step for the SNP data is to extract and compare the summary statistics of population genetics within and between populations (Andrews et al., 2016). This information describes the genetic diversity in populations, including the estimate of allele frequencies (usually denoted as p or the frequency of the most abundant allele), observed heterozygosity (H_O), expected heterozygosity (H_E), nucleotide diversity (π), number of segregating sites (S) and number of private alleles (*i.e.*, alleles only found in one population). These summary statistics can reveal the level of genetic erosion in domesticated plants and animals when compared to the ancestral wild population, which is expected due to severe bottlenecks, selective sweeps and inbreeding (Groeneveld et al., 2010; Gepts, 2014). One should be aware that reference bias can influence the relative genetic variation observed between the wild and domesticated populations, which could be alleviated using more than one reference or using a pan-genome (Günther and Nettelblad, 2019).

Population Structure

It is also important to describe the population structure (*i.e.*, the genetic differentiation among populations) of domesticated taxa and of their wild relatives, as it can reveal the influence of historical events that shaped the genetic diversity of the organisms (Linck and Battey, 2019). The level of population structure between wild and domesticated taxa can be determined by several factors, such as the number of generations since domestication started, the intensity of the selective pressures imposed to the domesticated taxon, the intensity of the bottlenecks suffered though the domestication process, and the frequency of gene flow between the domesticated taxon and its wild relative (Meyer and Purugganan, 2013).

The F-statistics are classic estimates of population genetics that are based on the heterozygosity values within and among populations, which can reveal patterns of inbreeding, gene flow and differentiation between and within populations (Andrews et al., 2016). Of these, the F_{ST} statistic is of particular interest, since it can be used to detect population structure between wild and domesticated populations, or between different domesticated varieties (Andrews et al., 2016). These estimates are relatively simple to calculate, but they require *a priori* assignment of individuals to discrete populations, which may be wrongly assigned, may not reflect natural populations or may simply be unknown (Linck and Battey, 2019).

Methods based on population clustering have become popular for describing genetic structure, as they do not require *a priori* population assignment. These clustering methods can be classified into parametric and non-parametric methods (Linck and Battey, 2019). Parametric methods, also known as model-based methods, assign individuals into a predefined number of K populations based on their genotypes and the allele frequency of each locus (Pritchard et al., 2000). Several parametric methods have been described that successfully analyze genomic datasets to infer population structure (*e.g.*, Tang et al., 2006; Alexander et al., 2009; Raj et al., 2014), but one has to be careful when using them, as they assume linkage equilibrium and Hardy-Weinberg equilibrium in the dataset (Linck and Battey, 2019), so SNPs should be filtered accordingly before these methods can be confidently used (Wigginton et al., 2005; Mathew et al., 2018). Furthermore, parametric methods have been found to be susceptible to changes in the SFS generated by minor allele frequency thresholds that are commonly used to filter population genomics data because low-frequency polymorphisms are expected to contain information about recent events, which adds uncertainty to the assignation of individuals in populations that reflect ancient demographic events (Linck and Battey, 2019).

Non-parametric methods include principal component analyses, discriminant analyses of principal components and K-means clustering. These methods define populations and genetic structure by transforming the genetic data into uncorrelated variables – named eigenvectors or principal components – to identify groups within the dataset (Patterson et al., 2006; Jombart et al., 2010; Linck and Battey, 2019). Non-parametric methods were designed to work with large amounts

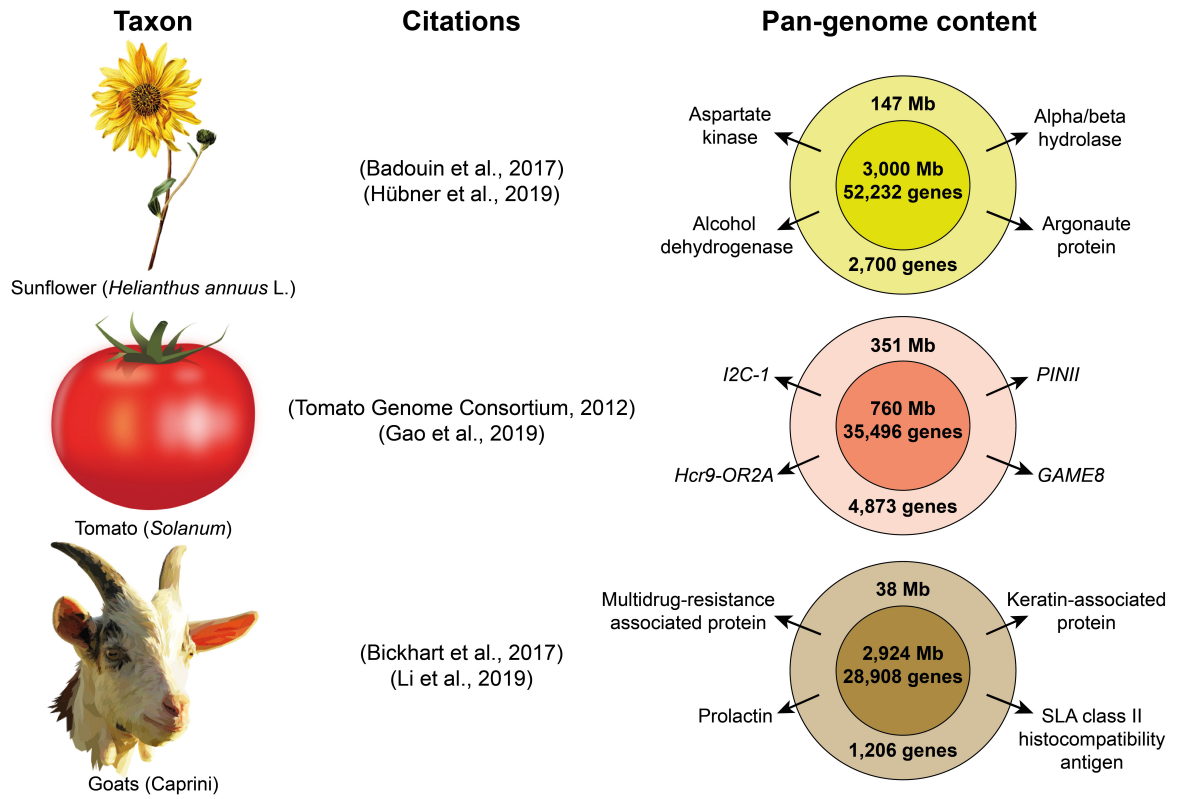

FIGURE 2 | Examples of three pan-genomes of domesticated taxa. The citations included correspond to the publications of the original reference genomes and the subsequent pan-genome assemblies. The inner circles on the right represent the content of the reference genome, while the outer circles represent the additional nonreference content retrieved with the pan-genome. Some examples of important genes are show for the nonreference part of each pan-genome. (images obtained from Openclipart).

of genomic data (Patterson et al., 2006; Jombart et al., 2010) and they are more robust to changes in the SFS than the parametric methods, so it is recommended to run both types of methods and compare their results before making further inferences (Linck and Battey, 2019).

Inferences in Changes of Population Sizes Throughout Time

An important aspect of the demographic history of domesticated taxa is the analysis of the change in the effective population size (N_e) in the populations throughout time (Chen J. et al., 2018). The concept of N_e reflects the estimated populations size in a Wright-Fisher model given an observed genetic variation, so these estimations hardly reflect the census population size of real populations (Charlesworth, 2009), and can also be affected by reference biases and allele dropouts. Changes in N_e can reveal or at least hint on the demographic history of taxa throughout the domestication process, such as expansions or bottlenecks. These changes can help to understand other evolutionary aspects of domestication concerning natural and artificial selection, such as the efficiency of selection and the accumulation of deleterious mutations in domesticated taxa (Chen J. et al., 2018; Allaby et al., 2019).

The domestication process is expected to include a bottleneck as a consequence of subsampling the genetic diversity in the wild ancestor, followed by a population expansion as domesticated taxa diversify (Meyer and Purugganan, 2013), although this idea has been recently challenged by paleogenomic studies (Allaby et al., 2019). Many methods exist to explore the changes in N_e throughout time, whose approach sometimes depends on the type of data available. It should be noted that all the methods to infer historical changes in N_e are susceptible to predicting false bottlenecks when populations are structured, so as indicated above, genetic structure should be evaluated and properly accounted for (Nielsen and Beaumont, 2009).

Studies with few individuals and high sequencing depth may use the Pairwise Sequentially Markovian Coalescent model (PSMC; Li and Durbin, 2011) or the Multiple Sequential Markovian Coalescent model (MSMC; Schiffels and Durbin, 2014) to analyze the demographic history of domesticated and wild taxa. The PSMC and MSMC models can infer changes in N_e throughout time (bottlenecks and expansions) by calculating the distribution of the time of coalescence between all the heterozygous loci in complete diploid genomes (Li and Durbin, 2011; Schiffels and Durbin, 2014). These models can also calculate the time of coalescence (*i.e.*, separation, and in some cases the domestication time) between two genomes given a

specified mutation rate, recombination rate and generation time (Li and Durbin, 2011).

However, the genomes used in PSMC or MSMC must be of very good quality, having an average sequencing depth of the very least 18x, at least 10 reads per site, and less than 25% of missing data (Nadachowska-Brzyska et al., 2016). Besides, PSMC has several limitations compared to other estimators of N_e and is particularly susceptible to predicting false bottlenecks when populations are structured (Mazet et al., 2015). Nevertheless, this can be properly handled by comparing models of instantaneous N_e size change against models of classical symmetric islands using a maximum-likelihood approach (Mazet et al., 2015).

Multiple Sequential Markovian Coalescent can infer more recent changes in N_e compared to PSMC (Schiffels and Durbin, 2014), so it may be convenient to explore recent demographic expansions in diversified domesticated taxa (Allaby et al., 2019). For example, MSMC was used to infer population bottlenecks in East Asian and Western Eurasian dogs, as well as divergence times between wolves and dogs around 60,000–20,000 years ago (Frantz et al., 2016), while PSMC was used to determine a severe bottleneck in African rice around 15,000–13,000 years ago (Meyer et al., 2016).

Other methods rely on population data at a genomic scale from many (sometimes hundreds) individuals (as obtained from exome sequencing or RAD-seq), namely the extended Bayesian skyline plots (Heled and Drummond, 2008; Trucchi et al., 2014) and the stairway plots (Liu and Fu, 2015). Since N_e is a crucial concept in coalescent theory, extended Bayesian skyline plots and stairway plots rely on the SFS calculated from the population data to estimate N_e (Heled and Drummond, 2008; Liu and Fu, 2015). The inferences made from these two methods are comparable to those obtained from PSMC and MSMC, although they rely on different kinds of datasets (Liu and Fu, 2015). Furthermore, stairway plots are more efficient in inferring recent demographic history, whereas PSMC is more reliable for ancient demographic events (Liu and Fu, 2015).

Estimating Gene Flow and Introgression Between Populations

Ancient gene flow and local ancestry (*i.e.*, the genetic ancestry of an individual for an specific chromosomal position; Thornton and Bermejo, 2014) are also important aspects of plant and animal domestication that need to be addressed, since they can describe the genetic contribution of different ancestral populations in the genomic architecture of extant populations, such as wild and domesticated taxa (Price et al., 2009; Pickrell and Pritchard, 2012).

One approach to assess ancient gene flow are graph-based methods that incorporate the possibility of ancient gene flow between distantly related populations (Pickrell and Pritchard, 2012). This type of methods represents the relationships between populations as a bifurcating tree, where internal nodes can also be interconnected forming a graph that represents ancient gene flow that contributed to modern genetic variation (Pickrell and Pritchard, 2012). For example, graph-based analyses have revealed constant gene flow between sympatric populations of domesticated and wild pearl millet (Burgarella et al., 2018), constant gene flow between domesticated and wild pigs (Frantz et al., 2015) but lack of hybridization events between wild and domesticated populations of goats and sheep (Alberto et al., 2018).

Another popular test to infer ancient admixture is the ABBA-BABA test, also known as the *D*-statistic, which evaluates the allelic patterns of three taxa and compares them to an outgroup to identify genomic regions with an excess of shared derived variants that are not concordant to the species tree (i.e., ABBA-BABA patterns), which suggest introgression events (Durand et al., 2011). The $f_d^{\wedge}$ test, which is derived from the *D*-statistic, can help discriminate between introgression events and nonrandom mating in ancestral structured populations (Martin et al., 2015). The *D*-statistic is sensitive to both introgression and incomplete lineage sorting, so both signals can be separated by testing deviations in the symmetry of branch lengths between the gene trees and the species tree (Edelman et al., 2019). By the same logic, the D_3 test can also infer introgression events by analyzing the symmetry in branch lengths, without the need for an outgroup (Hahn and Hibbins, 2019). The *D*-statistic has been used to infer several introgression events between species of the *Bos* genus during domestication (Wu et al., 2018).

On the other hand, local ancestry methods can reveal which chromosomal segments in the genome were inherited from different ancestral source populations (Price et al., 2009). These methods use the data obtained from linkage disequilibrium between loci to assign ancestry in each portion of the genome in comparison to reference populations that depict ancestral source populations, requiring an *a priori* assignation of unadmixed reference populations in order to assign local ancestry to the populations of interest (Price et al., 2009). The analysis reveals chromosomic blocks that can be assigned to either a wild or a domesticated ancestry in hybrid populations, which may reveal historical processes of introgression and local adaptation in modern domesticated populations, as well as potential targets for selective breeding (Janzen et al., 2019).

Many methods exist that can infer local ancestry using genome-wide population data, and all of them require a high-quality reference genome (preferably assembled at a chromosome-level) in order to detect the ancestry of chromosomal segments (*e.g.*, Price et al., 2009; Baran et al., 2012; Maples et al., 2013; Dias-Alves et al., 2018). For example, a local ancestry analysis of East Asian domestic cattle revealed introgressed blocks inherited from ancient banteng and yak populations that contained genes enriched in sensory perception of smell, transmembrane transport and antigen processing (Chen N. et al., 2018).

Using Demographic Simulations to Infer Domestication Scenarios

The previous descriptive tools can help us explore possible evolutionary and demographic scenarios in the absence of *a priori* hypotheses (Liu and Fu, 2015). However, for domesticated taxa we usually have additional classic botanical, zoological,

morphological, paleoclimatic, archeological, ethnobiological and biogeographical data that may suggest some likely scenarios (Gerbault et al., 2014). Thus, demographic modeling can be used to test explicit demographic scenarios by comparing simulations of SFS in such scenarios to the observed data (Gerbault et al., 2014; Liu and Fu, 2015). There are many methods available for demographic modeling, which can be more suitable depending on the type of scenarios that need to be tested (Anderson et al., 2005; Gutenkunst et al., 2009; Excoffier and Foll, 2011; Cornuet et al., 2014). All these methods rely on some basic tenets of coalescent theory (Liu and Fu, 2015), so they are also susceptible to possible biases in the observed genetic variation in the populations.

For example, the approximate Bayesian computation (ABC) method compares the summary statistics of several simulated scenarios against the observed data to accept or reject certain demographic hypotheses (Cornuet et al., 2014; Gerbault et al., 2014). This method can help us determine certain parameters of our models and can be used with genome-wide datasets (Cornuet et al., 2014).

Other methods based on diffusion approximation can help us infer the demographic history of multiple populations and their interaction through migration and admixture using biallelic SNP data (Gutenkunst et al., 2009). Demographic modeling has helped test the number of domestication events as well as intercontinental migratory events in cattle (Pitt et al., 2019). Coalescent simulations have supported a common origin for all the domesticated varieties of pearl millet (Burgarella et al., 2018), while the ABC method has revealed that the most likely scenario in the domestication of the scarlet runner bean consists of a single domestication event around 21,000 years ago with a mild bottleneck effect (Guerra-García et al., 2017).

IDENTIFYING GENES UNDER SELECTION DURING DOMESTICATION

Demographic processes are important to understand the general history that led to the domestication of plant and animal taxa, but many studies are specially interested in finding the selected genes that explain the phenotypic differences between domesticated taxa and their wild counterparts (Wang G.-D. et al., 2014; Kantar et al., 2017). Indeed, the detection of these genes under selection during domestication is critical to understand the genetic basis of domestication syndromes, especially for detecting genetic variation relevant for future improvement and selective breeding (Hufford et al., 2012).

When a genetic variant increases its frequency due to positive selection (*i.e.*, selection favoring the fixation of a new allele), the adjacent alleles (*i.e.*, physically connected in the same chromosomal region) also increase their frequency in a process known as hitchhiking (Smith and Haigh, 2007). Once the genetic variant under selection reaches a high frequency or fixation, the hitchhiking effect reduces or even eliminates the genetic variation around the selected locus, producing what is known as a selective sweep (Vitti et al., 2013; Pavlidis and Alachiotis, 2017). The size and intensity of a selective sweep depends on the rate of recombination in the genome, and on the intensity of the selective pressure (Smith and Haigh, 2007), which may be weaker in conscious selection compared to some cases of natural selection (Fugère and Hendry, 2018; Yang et al., 2019). Luckily, the signals of a selective sweep can be detected when the selection event occurred "recently" in an evolutionary timescale, as it is the case for domestication (Vitti et al., 2013).

Different bottom-up methods using population genomics data have been developed to detect the regions in the genome that were selected for during domestication, which we will refer to as candidate loci. We can mention methods for detecting regions with higher population differentiation compared to the rest of the genome, methods for detecting local changes in the SFS throughout the genome, and methods that detect extended regions with strong linkage disequilibrium compared to other haplotypes in the genome (see **Supplementary Table S1** for a summary of methods to detect selective sweeps). Alternatively, a GWAS can be performed to detect the association of a genetic variant to a specific phenotype of interest (Wang G.-D. et al., 2014).

F_{ST} Outlier Tests to Detect Candidate Genes

Besides the standard use of F_{ST} to detect global population structure, the F_{ST} statistic can also be used to detect signals of selective sweeps between populations, namely between wild and domesticated taxa (Gepts, 2014). While a global F_{ST} statistic (involving all the analyzed loci or SNPs) can reveal the overall genetic structure between populations, a local F_{ST} statistic calculated for each locus or SNP along the genome can evaluate whether particular regions of the genome are more differentiated from what is expected due to demographic processes, which can be interpreted as signals of a selective sweep (Nei and Maruyama, 1975). Many different methods exist that are based on the F_{ST} statistic, which are collectively known as F_{ST} outlier tests (Foll and Gaggiotti, 2008; Excoffier et al., 2009; Bonhomme et al., 2010; de Villemereuil and Gaggiotti, 2015; Lotterhos and Whitlock, 2015), that differ mainly on the underlying model used to calculate the null distribution of the F_{ST} values, and thus its ability to detect outliers (**Supplementary Table S1**).

F_{ST} outlier tests are able to detect selective pressures following a bottom-up approach, but their efficiency is determined by a multitude of factors that should be carefully accounted for before using them, such as the sampling scheme used to obtain the population data, the total size of the dataset (*i.e.*, number of populations, of individual per population and of SNPs analyzed), the intensity of the selective pressure, the selfing or allogamous nature of its sexual reproduction, and the migration patterns and genetic structure among populations (De Mita et al., 2013; Lotterhos and Whitlock, 2014, 2015).

Some successful examples in the use of F_{ST} outlier tests include the detection of domestication candidate genes in apple involved in fruit development, size, acidity and sugar metabolism (Khan et al., 2014), the finding of candidate domestication genes involved in metabolism and oil biosynthesis in sunflower

(Baute et al., 2015), the description of candidate diversification genes between pig breeds associated to the shape of the skull (Wilkinson et al., 2013), and the identification of candidate loci between wild and domesticated salmon strains involved in body weight, condition factor, male maturation and a brain related protein (Vasemägi et al., 2012).

Site Frequency Spectrum Based Tests to Detect Selective Sweeps

Selective sweeps alter the SFS that would be expected under neutral evolution processes because of the reduction in the genetic diversity around the loci under selection (Vitti et al., 2013). The genomic region under selection skews the SFS into an excess of high frequency derived alleles when the selective sweep was recent, since the alleles that were linked to the favored selected locus also reach high frequencies (Fay and Wu, 2000). However, after all the high-frequency alleles reached fixation, the genomic region under the selective sweep will have little to no variation, while mutations will slowly generate new allelic variants, skewing the SFS into an excess of low frequency variants (Zeng et al., 2006). Several tests have been developed to detect skews in the SFS, each of them capable of detecting changes in different parts of the SFS (**Supplementary Table S1**), making them complementary to one another (Zeng et al., 2006; Vitti et al., 2013).

Even though SFS based tests are powerful tools to detect selection, it is important to remember that the SFS at the global genomic scale is also altered by demographic events such as bottlenecks that produces an excess of low frequency variants, and expansions that generates an excess of intermediate frequency variants (Vitti et al., 2013). Thus, it is mandatory to have a previous prediction of the demographic history of the populations in order to properly adjust the null hypothesis in each test (Ross-Ibarra et al., 2007).

The well-known summary statistic called Tajima's *D* is sensitive to changes in low-frequency variants, making it particularly useful to detect selective sweeps before and after the selected locus reaches fixation, although low-frequency variants can also be observed in loci under purifying selection (Tajima, 1989; Zeng et al., 2006). Tajima's *D* is also sensitive to intermediate-frequency alleles, making it useful to detect balancing selection (Tajima, 1989) or even some forms of soft selective sweeps generated by standing genetic variation (Prezeworski et al., 2005).

Conversely, Fay and Wu's *H* is sensitive to changes in high-frequency variants, which are only altered by positive selection, making it very useful when used alongside Tajima's *D* (Fay and Wu, 2000). Unlike Tajima's *D*, Fay and Wu's *H* needs an outgroup species in order to differentiate ancestral alleles from derived alleles and thereby to know whether the derived alleles are at high or low frequencies (Fay and Wu, 2000).

Zeng et al. (2006)'s *E* is sensitive to both low and high frequency variants, making it particularly powerful to detect selective sweeps before or after the selected locus reached fixation, also needing an outgroup in order to differentiate derived alleles from ancestral alleles).

There are some tools available to implement SFS based tests using genome-wide data, that can perform all the above tests (*i.e.,* Korneliussen et al., 2013, 2014; Rozas et al., 2017). For example, Tajima's *D* test was used alongside other methods to detect selective sweeps associated to the domestication of yaks (Qiu et al., 2015), Zeng's *E* test helped discover 125 selective sweeps associated to the domestication of horses (Librado et al., 2016), and the complementary implementation of Tajima's *D*, Fay and Wu's *H* and Zeng's *E* revealed several candidate genes that share similar functions between peach and almond (Velasco et al., 2016).

The reduction of diversity (ROD) test is another popular SFS-based method to detect selective sweeps that has been particularly useful for the study of domestication (**Supplementary Table S1**). ROD compares local π values of domesticated taxa against the local π values of its wild relatives, using sliding windows alongside the genome (Guo et al., 2012; Huang et al., 2012; Qi et al., 2013; Schmutz et al., 2014). The ROD method has been used to successfully detect candidate domestication genes in rice (Huang et al., 2012), watermelon (Guo et al., 2012), cucumber (Qi et al., 2013), common bean (Schmutz et al., 2014), and chickpea (Varshney et al., 2019), to name a few.

Linkage Disequilibrium (LD) Based Methods to Detect Selection

Given that selective sweeps remove the variation in regions adjacent to the locus under selection, they can form haplotype blocks that extend in strong LD compared to other haplotypes in the same locus because they reached a medium-to-high frequency in the population swift enough so they are not yet disrupted by recombination (Sabeti et al., 2002; Vitti et al., 2013). This pattern has been exploited to develop several methods based on LD to detect selective sweeps of recent origin (Vitti et al., 2013). Interestingly, LD-based methods are sensitive enough to detect both strong and soft selective sweeps (Garud et al., 2015), as well as partial or incomplete selective sweeps (Vitti et al., 2013), making them excellent tools to study recent and ongoing selection events, such as those occurring during domestication and the subsequent diversification of landraces (**Supplementary Table S1**).

Since the above rationale relies on LD decay due to recombination, any method based on LD requires to control for local variation in recombination rates in order to reduce false positives (Sabeti et al., 2002). The extended haplotype homozygosity (EHH) is a widely used statistic in LD-based methods that is defined as the probability that two orthologous genomic regions carrying a "core" haplotype of interest (*i.e.*, the part of the haplotype that is shared by all the individuals carrying it, such as the allele under positive selection) in the population are identical by descent (*i.e.,* they were inherited by the same ancestor), as one looks to a specified distance farther away from the core region (Sabeti et al., 2002).

Among the LD based methods that uses the EHH, we can mention the long-range haplotype (LRH) test, sometimes named the relative EHH (rEHH) test, which controls for local

recombination rates by comparing the EHH of several haplotypes localized within the same locus (Sabeti et al., 2002). Other EHH based methods include the whole-genome long-range haplotype (WGLRH) test that uses sliding windows to perform the LRH test (Zhang et al., 2006), the long-range haplotype similarity (LRHs) test (Hanchard et al., 2006), the integrated haplotype score (iHS) which is particularly sensitive to incomplete selective sweeps and soft sweeps (Voight et al., 2006) and the cross-population extended haplotype homozygosity (XP-EHH) statistic that is able to detect selective sweeps after the selected allele reached fixation (Sabeti et al., 2007). The iHS and the XP-EHH statistics can be regarded as complementary to each other, enabling the detection of incomplete and complete selective sweeps in the target population (Vitti et al., 2013).

All the LD-based tests that make use of the EHH statistic require the previous phasing of the chromosomes in order to work (*i.e.*, assignation of alleles in an individual to their corresponding maternal and paternal haplotypes), which may or may not be possible depending on the sequencing depth and type of data available for the analysis (Delaneau et al., 2013). For instance, a reference genome is usually needed in order to phase genotypes, since most methods rely on the information of proximity between alleles and their distribution within individuals in a population to assign haplotypes (Delaneau et al., 2013) although new methods are emerging that can phase genotypes without a reference genome (Money et al., 2017).

There are other LD-based methods that do not make use of the EHH statistic, such as the LD decay (LDD) test, which rely on individuals that are homozygous for any given SNPs to look for LD differences between alleles in a population (Wang et al., 2006) or the ω statistic that scans for high SNP correlation coefficients around a site under selection (Kim and Nielsen, 2004; Alachiotis et al., 2012). Another method that do not require chromosome phasing is the regression-based test, which relies on the reduction of heterozygosity as one approaches the locus under selection in a genome to infer selective sweeps (Wiener and Pong-Wong, 2011). Other LD-based methods exploit the estimation of identity-by-descent using genome-wide data to detect haplotypes that are shared between several unrelated individual (> 10 generations) to infer selective sweeps without previous knowledge of the pedigree of individual (Han and Abney, 2013), so they might prove useful to study recent domestication processes.

Some examples of LD-based methods used to explore the domestication process includes an analysis using LRH to detect signatures of selection associated to dairy and beef cattle breeds (Bomba et al., 2015), a study using the XP-EHH statistic to find signals of selective sweeps in Jinhua pigs (Li et al., 2016), and a paper focused on the diversification of goat landraces that calculated the iHS and the XP-EHH statistics alongside other tests to detect selective sweeps between goat breeds (Bertolini et al., 2018).

Other important tests include the XP-CLR test (Chen et al., 2010) and the μ statistic (Alachiotis and Pavlidis, 2018) which implement multiple signatures to detect selective sweeps (**Supplementary Table S1**) and have been used to detect candidate loci in maize and African rice, respectively (Hufford et al., 2012; Ndjiondjop et al., 2019).

Using GWAS to Detect Domestication-Associated Loci

Genome-wide association studies have been used extensively to uncover the genetic variants that underlie domestication traits (Shi and Lai, 2015). The domestication traits that can be analyzed through a GWAS can encompass any biological characteristic from simple morphological traits (Jiao et al., 2012) to the production of certain metabolites (Shang et al., 2014), tame behavior in animals (Ilska et al., 2017), resistance or susceptibility to certain diseases (Wang et al., 2012), or adaptation to certain environmental conditions (Song et al., 2018).

An important advantage of the GWAS over the bottom-up approaches is its ability to detect polygenic effects on single traits of interest, which is commonplace considering that genes interact between them and the environment to generate phenotypes (Gibson, 2018).

A prerequisite before preforming a GWAS is to have large sample sizes in both the number of sequenced genetic variants and the number of individuals included in the study, as they are necessary to obtain the statistical power to detect variants with small effects and to reduce the risk of false positives (Wang G.-D. et al., 2014).

Some recent examples include the use of a GWAS to identify candidate genes with unknown functions involved in several agronomic traits, including drought and heat tolerance in chickpea (Varshney et al., 2019); a GWAS that revealed loci associated to fruit size and quality in peach (Cao et al., 2019); and a GWAS that uncovered the genetic variants involved in the absence of anthocyanin in domesticated rice compared to its wild relative (Zheng et al., 2019).

ANCIENT DNA AND PALEOGENOMICS OF DOMESTICATED TAXA

Extant domesticated taxa lack the information of ancient genetic diversity that was lost through bottlenecks, selection and genetic drift (Ramos-Madrigal et al., 2016). However, the analysis of ancient DNA can allow the research community to overcome some of these limitations (Irving-Pease et al., 2019). Ancient DNA retrieved from archeological sites allows the study of the rate at which domestication happened, as well as revealing which genes were important at the beginning of this process (Vallebueno-Estrada et al., 2016; Irving-Pease et al., 2019). Thus, paleogenomics is becoming a novel research area for understanding the process of plant and animal domestication (Irving-Pease et al., 2019).

Extraction and Sequencing of Ancient DNA

An important limitation of paleogenomic analyses is the level of preservation of the ancient DNA itself, as well as the total yield of extracted DNA (Sawyer et al., 2012). The DNA molecules that are extracted from tissues that are not conserved on permafrost and are older than 100 years are usually shorter than 100 bp (Sawyer et al., 2012). The strand breaks of

such fragments are also non-random, as purines are enriched before the strand breaks (Sawyer et al., 2012). Additionally, these fragments incorporate cytosine-to-uracil mutations on their ends, further hindering the analysis of the sequenced fragments (Sawyer et al., 2012). Even though these characteristics hamper the sequencing and analysis of ancient DNA, they are also useful to differentiate between real ancient DNA and extant DNA contamination (Sawyer et al., 2012). Furthermore, due to the scarce ancient material located throughout few archeological sites worldwide, sample sizes in paleogenomic studies are very small, usually one or few individuals per location and sometimes only one locality (*e.g.,* Wales et al., 2016; Ramos-Madrigal et al., 2016).

Given the above difficulties and the uniqueness of the biological material retrieved from archeological sites, it is crucial to extract and sequence as much ancient DNA as possible while avoiding DNA contamination (Gamba et al., 2016). Major efforts have been made to develop efficient protocols for ancient DNA extraction (Gamba et al., 2016) and single-strand library preparation for high-throughput sequencing (*e.g.*, Gansauge et al., 2017). Organelle genomes were usually the target for ancient DNA sequencing because multiple copies of these can be found within each plant and animal cell and can reveal several demographic processes (Wales et al., 2016; Irving-Pease et al., 2019). Nonetheless, more evolutionary information can be retrieved from nuclear DNA, which is the main target for modern paleogenomic studies (Wales et al., 2016; Irving-Pease et al., 2019).

Insights of Paleogenomic Data in Domestication

Paleogenomic studies are challenging some of our previous ideas of the domestication process, such as the occurrence of ancient domestication bottlenecks, which appear to be absent in several archeological plant genomes, suggesting that the reduced diversity in domesticated taxa may be a more gradual process from what was expected using DNA of extant populations (Allaby et al., 2019). For example, several archeological samples of *Sorghum bicolor* from different time periods (ranging from 1800 to 100 years ago) were compared to extant individuals of the species, revealing that this crop did not suffered an initial domestication bottleneck, but rather that the reduction in genetic diversity, and its associated mutational load, occurred gradually throughout time (Smith et al., 2019).

Paleogenomics is also revealing important aspects of plant and animal domestication, such as the first genetic steps towards domestication syndromes as well as the overall graduality of the process (Ramos-Madrigal et al., 2016; Vallebueno-Estrada et al., 2016; Daly et al., 2018). For example, archeological remains of goat populations have revealed multiple domestication processes in ancient wild goats, possible dispersal routes of ancient goat populations and signs of early selective pressures towards candidate genes involved in pigmentation, milk production, size, reproduction and changes in diet (Daly et al., 2018). Likewise, several archeological maize samples retrieved from the Tehuacán Valley in Mexico have revealed that early domesticates already presented signals of selective sweeps on important candidate genes, such as *teosinte branched1* and *brittle endosperm2*, but lacked selective sweep signals in other important candidate genes present on modern maize populations, even though these ancient maize populations were already endogamous and more closely related to modern maize than to wild teosinte, revealing that maize domestication was a gradual process ranging thousands of years (Ramos-Madrigal et al., 2016; Vallebueno-Estrada et al., 2016).

Other examples demonstrate the importance of paleogenomic studies in domesticated taxa, including grapevine (Wales et al., 2016), barley (Mascher et al., 2016), sunflower (Wales et al., 2019), horses (Schubert et al., 2014), dogs (Frantz et al., 2016) and cats (Ottoni et al., 2017).

RNA SEQUENCING TO DETECT DIFFERENTIALLY EXPRESSED GENES ASSOCIATED TO DOMESTICATION

Besides the use of RNA-seq to obtain population-level data, comparative transcriptomics is a good way to find or support the validity of candidate genes (Hekman et al., 2015). Transcriptomic analyses between domesticated and wild taxa can reveal important changes in gene expression associated to domestication (Koenig et al., 2013; Hekman et al., 2015; Hradilová et al., 2017). Likewise, the analysis of hybrids between domesticated and wild individuals can reveal important patterns of allele-specific regulation and the role of *cis*/*trans* regulatory elements in the emergence of domestication traits (Bell et al., 2013; Lemmon et al., 2014).

The Experimental Design of Differential Expression Analyses

Transcriptomic profiles are tissue-specific and time-dependent (Hekman et al., 2015). Thus, a good experimental design can reveal important loci involved in the phenotypic differences associated to domestication syndromes, such as suppression of secondary metabolites, changes in form, size, taste, absence of defense mechanisms, seed dormancy, docile behavior, among other traits (Hekman et al., 2015). This can be done by comparing the total RNA expression of the tissue or organ of interest (Koenig et al., 2013), as well as comparing RNA expression throughout the developmental stages of such tissue or organ (Hradilová et al., 2017).

Since transcriptomic analyses are experimental by nature, experimental designs require biological replicates for each treatment, condition or organ to assess the variability in the data; as well as controlled environmental conditions to reduce possible biases and sources of error (Fang and Cui, 2011; Schurch et al., 2016). Empirical studies recommend using at least six biological replicates for each condition in the experiment, even though the use of three replicates is common, but discouraged (Burden et al., 2014; Schurch et al., 2016). Additionally, it is important to avoid committing

errors in the experimental design that can bias the results of the RNA-seq experiment, such as using different sequencing technologies for each sample, using different methods for library preparations throughout the samples, sequencing each treatment in a different sequencing flowcell or different lanes within a flowcell (Fang and Cui, 2011). Other technical biases associated to adapter ligation and within-lane variation can be properly assessed when using biological replicates (Fang and Cui, 2011; see **Table 1**).

RNA-seq data can also be complemented with metabolomic data to infer the association between the differential expression of genes and the presence/absence of metabolites between wild and domesticated taxa (Hradilová et al., 2017).

After obtaining high-quality data with an appropriate experimental design, RNA-seq analyses usually follow a similar workflow, which should culminate in the detection of differentially expressed genes between a wild plant and its domesticated counterpart (Yang and Kim, 2015; see **Table 1**). These differentially expressed genes are most likely candidates that may explain to some degree the changes associated to domestication (Koenig et al., 2013; Hradilová et al., 2017). Nonetheless, one must be careful while interpreting the results of these studies, as some differentially expressed genes between wild and domesticated taxa may be a consequence, rather than a cause, of the domestication traits under study (Albert et al., 2012).

Successful Examples of RNA-seq Experiments to Understand Domestication

RNA-seq analysis has been successfully employed to discover differentially expressed genes involved in the domestication of several plant species. For example, RNA-seq analyses between maize and teosinte found 600 differentially expressed genes and 1,100 genes with altered patterns of co-expression, mainly involved in biotic stress responses, and many of which were previously found as candidate genes using selective scans (Myers et al., 2012). Similar results have been found in tomato (Koenig et al., 2013), pea (Hradilová et al., 2017), common bean (Singh et al., 2018), and carrot (Machaj et al., 2018). This approach has also led to the discovery of differentially expressed genes between dogs and wolves associated to tameness (Li et al., 2013), as well as changes related to the immune system and aerobic capacity (Yang et al., 2018). Another study found differential isoform expression between wild and domesticated sorghum accessions, revealing that domestication can alter the patterns of alternative spicing (Ranwez et al., 2017). Hybrid studies have been performed between maize and teosinte, suggesting potential selection on *cis* regulatory elements associated with changes in ear tissue and previously reported candidate genes (Lemmon et al., 2014). Another hybrid study in *Capsicum annuum* using network analyses revealed that loss of function in *cis* regulatory sequences lead to transcriptional changes in *trans* elements that are associated with fruit morphology (Díaz-Valenzuela et al., 2020).

MODERN EPIGENOMICS AND METHODOLOGICAL STRATEGIES TO EXPLORE DOMESTICATION

Epigenetics is classically defined as the heritable mechanisms that regulate gene expression without direct modifications to the DNA sequence, namely DNA methylation, RNA methylation, covalent histone modifications and chromatin assembly states (Sakurada, 2010; Zhao et al., 2017). Epigenetic variants, sometimes called epialleles, are local differences in these epigenetic marks between individuals in a population, which can have similar dynamics to genetic variants (Weigel and Colot, 2012; Guo et al., 2015). Since epigenetic mechanisms underly the ability of organisms to respond to changing environmental conditions, some epigenetic marks associated to these responses are more susceptible to change due to environmental input, while other marks involved in cell differentiation, embryonic development and core cellular functions might be more stable (Turner, 2009).

Most of the domestication studies that explain phenotypic differences between wild and domesticated taxa focus on genetic variation. However, the study of epigenomics may explain some of the missing heritability in domestication traits (i.e., the gap between the heritability of a trait estimated by classic genetics and GWAS), the patterns of differentially expressed genes that do not have clear signs of selective sweeps, or even connect the causality between the genetic variation that was selected for during domestication and the resulting phenotypes (Schmitz et al., 2013; Trerotola et al., 2015; Janowitz Koch et al., 2016; Bélteky et al., 2018).

Epigenetic variation can be inherited from one generation to the next in a process known as trans-generational epigenetic inheritance, which has been documented in plants and animals (Heard and Martienssen, 2014), even though the overall importance of this trans-generational epigenetic inheritance in plant and animal evolution is still debated (see **Table 1**). Nevertheless, we consider that studying epigenetic patterns associated to transcriptional activity and phenotypic traits should help understand the emergence of domestication phenotypes (Bélteky et al., 2018). If epigenetic variants such as single methylation polymorphisms (SMPs) show complete transgenerational inheritance, they can even be analyzed using the theoretical tools of population genetics to detect selective sweeps (Schmitz et al., 2013; Janowitz Koch et al., 2016).

In a similar fashion to GWAS, the use of epigenome-wide association studies (EWAS) can also reveal the association of an epigenetic variant to a trait of interest in domesticated taxa (Feeney et al., 2014). The same precautions taken in transcriptomic data should also be taken for epigenomic data, since the patterns of epigenetic marks in organisms are tissue-specific, time-dependent and sensitive to environmental input, meaning that epigenomic data should be analyzed for specific organs or tissues of interest in a controlled environment (Jensen, 2015). This is particularly important for the epigenetic marks that respond to environmental input, since domesticated taxa and their wild relatives live under different environmental

conditions. Growing both taxa under controlled conditions will alter the natural state of these marks, but will also help differentiate the heritable epialleles associated to domestication traits (Turner, 2009).

Obtaining Population Data From Epigenetic Marks

The most studied epigenetic mark is DNA 5-methylcytosine, which refers to the DNA methylation in cytosines which are usually associated to transcriptional gene silencing (He et al., 2011). Cytosine methylome data can be obtained using high-throughput sequencing technologies alongside bisulfite sequencing (Meissner, 2005). Bisulfite sequencing consists in the deamination of unmethylated cytosines through a bisulfite reaction, converting them into uracil, which are encoded as thymine by sequencing technologies (Frommer et al., 1992). The comparison of sequenced DNA that was treated with bisulfite alongside sequenced DNA without treatment can discriminate between methylated and unmethylated cytosines in an organ, tissue or cell-type of interest (Frommer et al., 1992).

Reduced representation bisulfite sequencing (RRBS) is a high-throughput technique with a similar rationale to RAD-seq that enriches the sequencing of CG rich regions of the genome after the digestion of restriction enzymes (Meissner, 2005). This makes the RRBS technique a cost-effective option to analyze cytosine methylation patterns in mammals, since its cytosine DNA methylation happens at CG sites (Meissner, 2005; He et al., 2011). Plant cytosine methylomes should instead be analyzed through MethylC-seq, which consists of whole-genome sequencing and bisulfite treatment (Urich et al., 2015), as cytosine methylation can also happen in CHG and CHH sites in plant genomes (He et al., 2011). Cytosine methylation can also be detected using methylated DNA immunoprecipitation sequencing (MeDIP-seq), which consists in shearing the genomic DNA into small pieces followed by the immunoprecipitation of the methylated cytosines using antibodies that recognizes 5-methylcytosine and finally sequencing the DNA sequences with the methylated sites using standard high-throughput sequencing technologies (Weber et al., 2005).

Besides cytosine methylation, adenine has also been shown to be methylated in both plants and animals (N6-methyldeoxyadenosine), which cannot be detected using bisulfite sequencing (Luo et al., 2015). However, genomic regions with methylated adenines can be detected using N6-methyldeoxyadenosine immunoprecipitation sequencing (6mA-IP-seq), which uses the same rationale as MeDIP-seq but requires antibodies that specifically targets N6-methyldeoxyadenosine (Fu et al., 2015). PacBio and Nanopore sequencing technologies are known to be sensitive to DNA methylation, regardless of it being on a cytosine or adenine, so they are currently being used as powerful, albeit expensive tools to evaluate DNA methylation patterns in genomes (Gouil and Keniry, 2019).

Histone modifications refers to either posttranslational covalent modifications in histones (methylations, acetylations, phosphorylations, ubiquitylations, ADP-ribosylations, sumoylations, crotonylations, malonylations, succinylations) or the substitution of canonical histones by histone variants with different amino acid composition (Bowman and Poirier, 2015). These histone modifications determine the functionality of local genomic regions by changing the state of the chromatin either through its direct effects on the chemical interactions between DNA and histones or through the recruitment of chromatin remodeling complexes (Bowman and Poirier, 2015).

Chromatin immunoprecipitation sequencing (ChIP-seq) can be used to assess the genome-wide association between DNA regions and specific histone modifications (Schmidt et al., 2009). ChIP-seq consists in the initial fixation of DNA-protein interactions using formaldehyde followed by DNA fragmentation and subsequent enrichment of the target histone modification using magnetic beads coupled to antibodies in order to sequence the genomic regions where the histone modification is present (Schmidt et al., 2009). ChIP-seq can also be used to assess the interaction between any DNA-binding protein such as transcriptional factors and specific genomic regions (Schmidt et al., 2009).

Epigenomic Studies Applied to Understand Domestication

The current epigenomic analyses regarding domestication have focused on DNA methylation patterns (Jensen, 2015; Ding and Chen, 2018), but some studies have also ventured into histone modification patterns (He et al., 2014). Recent efforts are trying to connect the discoveries of genomics and epigenetics to understand the evolution of tameness in domesticated animals (Jensen, 2015). A study using RRBS that compared the DNA methylation patterns between wolves and dogs revealed signals of natural selection acting on SMPs which are enriched in transposons and genes involved in the regulation of neurotransmitters, suggesting a dog-specific silencing of genes involved in behavior (Janowitz Koch et al., 2016). Similarly, a recent study using MeDIP-seq in red junglefowl populations that were bred to have either high or low fear to humans discovered genomic region that were differentially methylated in genes that were previously related to tameness (Bélteky et al., 2018).

Other studies focused on plant domestication have found differentially methylated sites associated to domestication syndromes (Song et al., 2017; Shen et al., 2018). A study using MethylC-seq found 519 differentially methylated genes between domesticated and wild cotton from which some of them are associated with the observed differences in flowering time and seed dormancy between the wild and domesticated taxa (Song et al., 2017). Another study using MethylC-seq found 4,248 differentially methylated regions between wild and domesticated soybean and 1,164 differentially methylated regions between domesticated and improved soybean (Shen et al., 2018). As expected, the differentially methylated regions in soybean had higher genetic diversity compared to the regions with evidence of selective sweeps that were previously found, and interestingly, 22.5% of the differentially methylated sites could be associated to a causal genetic variant (suggesting that these genetic variants were responsible for the observed epigenetic

patterns), whereas the rest of the differentially methylated regions could be interpreted as genuine epialleles located within genes involved in carbohydrate metabolism (Shen et al., 2018).

EXPERIMENTAL VALIDATION OF CANDIDATE GENES

Once we have evidence of candidate genes involved in the domestication syndrome, the necessary next step to understand the genetic basis of domestication is to design *in vitro* systems, knock-out, knock-down or knock-in experiments that validate the involvement of such genes in the observed phenotypes (Zhang et al., 2017). This can be performed either by direct alteration of the genome in the organism of interest, by using RNA interference or by designing heterologous systems in a model organism (Boettcher and McManus, 2015). As an example, a knock-out experiment with backcrosses between domesticated and wild mice elucidated the role of some genes involved in behavioral changes associated to mouse domestication (Chalfin et al., 2014).

Previous knock-out and knock-in experiments were restricted to model organisms, but nowadays experimental validation of candidate genes can be supported via knock-out and knock-in experiments, using novel genome editing tools (*e.g.*, Shalem et al., 2014; Hahn et al., 2017; Ueta et al., 2017). Genome-editing tools are already available for a broad range of taxa, including dozens of crop species, but developing a working system in non-model organisms can still be a difficult task that can take several months or even years to accomplish (Shan et al., 2020), so doing collaborative studies alongside experimental researchers is recommended. In this moment, the leading toolset to perform genome editing is the Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) system alongside the CRISPR associated protein 9 (Cas9), commonly known as CRISPR/Cas9, which can be used to eliminate, introduce or replace specific segments of DNA within a targeted site in a genome (Cong et al., 2013). Another useful tool for genome editing is the Transcription Activator-Like Effector Nuclease (TALEN) technology, which has its own advantages in comparison to CRISPR/CAS9 (Zhang et al., 2017). RNA interference can also help in validating the function of candidate genes, although it is limited to knock-down experiments (Boettcher and McManus, 2015). Heterologous expression in model organisms is a cost-effective alternative to validate candidate genes (e.g., Schweiger et al., 2010), although this method overlooks the interaction networks that exist *in vivo* which are accountable for the emergence of phenotypes (Rodríguez-Mega et al., 2015).

Regardless the genome-editing tool of choice (Boettcher and McManus, 2015; Zhang et al., 2017), genome edition is proving its usefulness to validate the effect of candidate genes involved in domestication through the introduction of domesticated alleles on wild relatives and vice-versa (Zhou J. et al., 2019), which can prove that the gene is indeed involved in the appearance of the domesticated phenotype (Zhou J. et al., 2019). This can be performed in the same way as a usual knock-out or knock-in experiment, where the edited locus must be validated through PCR and Sanger sequencing, a PCR-RFLP analysis or using Western-blot in case of a protein knock-out (*e.g.,* Ueta et al., 2017). The expected result of these type of studies is to find a modified phenotype after editing a candidate locus, either a wild individual with a domesticated-like phenotypic trait or a domesticated individual with a wild-like phenotypic trait (Zhou J. et al., 2019).

Of course, the above studies will hardly reproduce a complete domesticated or wild phenotype, since genetic elements interact in complex regulatory networks, including other elements within the genome as well as epigenetic and environmental components (Rodríguez-Mega et al., 2015), but nonetheless will be useful to understand the role of those genes in the emergence of domesticated phenotypes.

Once the candidate genes are validated, genome-editing tools can also become useful to introduce desirable traits from wild relatives to its domesticated counterparts, a goal of great interest for crop improvement (Zhou J. et al., 2019) and currently used to accelerate plant breeding and to fine-tune desirable traits (Wolter et al., 2019). Furthermore, recent efforts are trying to domesticate plant crops *de novo* by inserting the desired domestication alleles into their wild relatives, generating crops with the desired domestication phenotypes but without the problems of low genetic variation and accumulation of deleterious mutations that are an inevitable consequence of regular domestication processes (Fernie and Yan, 2019).

CONCLUSION AND PERSPECTIVES

Plant and animal domestication can be studied using genomic, transcriptomic and epigenomic strategies, revealing the action of evolutionary, ecological and anthropogenic processes (Kantar et al., 2017). These tools can lead us beyond the description of the possible historical scenarios that shaped the domesticated species, since we can explore the effects of domestication on the transcriptomic activity of a species (Hekman et al., 2015), test the validity of candidate genes associated to domestication phenotypes (Zhou J. et al., 2019) and analyze epigenetic patterns associated to domestication traits (Jensen, 2015). Many domesticated taxa remain genetically unexplored, and as sequencing technologies become cheaper and more efficient, domestication genomics will soon be available for polyploids and species with huge genomes (*e.g.*, Edger et al., 2019).

Nonetheless, the modern study of domestication of plants and animals should still be multidisciplinary, since genetics only tells us part of the story (Larson et al., 2014). An extended synthesis framework should also be considered to understand domestication, as these new studies are helping us understand niche construction and the emergence of domesticated phenotypes (Piperno, 2017). Other potential lines of work remain to be addressed in domestication studies, such as the changes in the chromatin architecture (e.g., Concia et al., 2020), the use of comparative proteomic atlases (e.g., Jiang Y. et al., 2019) and the analysis of cell-type divergences during development using single-cell RNA-seq data (Arendt et al., 2016). The use of this multi-omic approaches will help us create and

compare developmental atlases (e.g., Walley et al., 2016) between wild and domesticated taxa to understand how morphology diverged during domestication.

AUTHOR CONTRIBUTIONS

JB-R, DP, and LE wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) KE004 "Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de Cucurbita" and CONABIO PE001 "Diversidad genética de las especies de Cucurbita en México. Fase II. Genómica evolutiva y de poblaciones, recursos genéticos y domesticación, both awarded to Rafael Lira-Saade and LE. DP was funded by Instituto de Ecología, UNAM". JB-R is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México and received fellowship 583146 from Consejo Nacional de Ciencia y Tecnología (CONACyT).

ACKNOWLEDGMENTS

We acknowledge the Doctorado en Ciencias Biomédicas for the academic support provided to JB-R during the development of this project. We thank Alejandra Moreno Letelier for her insights, which greatly improved the quality of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00742/full#supplementary-material

REFERENCES

Alachiotis, N., and Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun. Biol.* 1, 1–11. doi: 10.1038/s42003-018-0085-8

Alachiotis, N., Stamatakis, A., and Pavlidis, P. (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* 28, 2274–2275. doi: 10.1093/bioinformatics/bts419

Albert, F. W., Somel, M., Carneiro, M., Aximu-Petri, A., Halbwax, M., Thalmann, O., et al. (2012). A comparison of brain gene expression levels in domesticated and wild animals. *PLoS Genet.* 8:e1002962. doi: 10.1371/journal.pgen.1002962

Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., et al. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* 9:813. doi: 10.1038/s41467-018-03206-y

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., et al. (2016). RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* 202, 389–400. doi: 10.1534/genetics.115.183665

Allaby, R. G., Ware, R. L., and Kistler, L. (2019). A re−evaluation of the domestication bottleneck from archaeogenomic evidence. *Evol. Appl.* 12, 29–37. doi: 10.1111/eva.12680

Anderson, C. N. K., Ramakrishnan, U., Chan, Y. L., and Hadly, E. A. (2005). Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21, 1733–1734. doi: 10.1093/bioinformatics/bti154

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17, 81–92. doi: 10.1038/nrg.2015.28

Arendt, D., Musser, J. M., Baker, C. V., Bergman, A., Cepko, C., Erwin, D. H., et al. (2016). The origin and evolution of cell types. *Nat. Rev. Genet.* 17, 744–757. doi: 10.1038/nrg.2016.127

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., et al. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546, 148–152. doi: 10.1038/nature22380

Baran, Y., Pasaniuc, B., Sankararaman, S., Torgerson, D. G., Gignoux, C., Eng, C., et al. (2012). Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* 28, 1359–1367. doi: 10.1093/bioinformatics/bts144

Baute, G. J., Kane, N. C., Grassa, C. J., Lai, Z., and Rieseberg, L. H. (2015). Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytol.* 206, 830–838. doi: 10.1111/nph.13255

Bell, G. D., Kane, N. C., Rieseberg, L. H., and Adams, K. L. (2013). RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol. Evol.* 5, 1309–1323. doi: 10.1093/gbe/evt072

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., Falentin, C., et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* 4, 879–887. doi: 10.1038/s41477-018-0289-4

Bélteky, J., Agnvall, B., Bektic, L., Höglund, A., Jensen, P., and Guerrero-Bosagna, C. (2018). Epigenetics and early domestication: differences in hypothalamic DNA methylation between red junglefowl divergently selected for high or low fear of humans. *Genet. Sel. Evol.* 50:13. doi: 10.1186/s12711-018-0384-z

Bertolini, F., Servin, B., Talenti, A., Rochat, E., Kim, E. S., Oget, C., et al. (2018). Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* 50:57. doi: 10.1186/s12711-018-0421-y

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., et al. (2017). Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat. Genet.* 49, 643–650. doi: 10.1038/ng.3802

Boettcher, M., and McManus, M. T. (2015). Choosing the right tool for the job: RNAi, TALEN, or CRISPR. *Mol. Cell* 58, 575–585. doi: 10.1016/j.molcel.2015.04.028

Bomba, L., Nicolazzi, E. L., Milanesi, M., Negrini, R., Mancini, G., Biscarini, F., et al. (2015). Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. *Genet. Sel. Evol.* 47:25. doi: 10.1186/s12711-015-0113-9

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., et al. (2010). Detecting selection in population trees: the lewontin and krakauer test extended. *Genetics* 186, 241–262. doi: 10.1534/genetics.104.117275

Bowman, G. D., and Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. *Chem. Rev.* 115, 2274–2295. doi: 10.1021/cr500350x

Brohammer, A. B., Kono, T. J. Y., and Hirsch, C. N. (2018). "The maize pan-genome," in *The Maize Genome. Compendium of Plant Genomes*, eds J. Bennetzen, S. Flint-Garcia, C. Hirsch, and R. Tuberosa (Cham: Springer), 13–29. doi: 10.1007/978-3-319-97427-9_2

Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085. doi: 10.1111/pbi.12454

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. doi: 10.1016/j.celrep.2016.12.063

Burden, C. J., Qureshi, S. E., and Wilson, S. R. (2014). Error estimates for the analysis of differential expression from RNA-seq count data. *PeerJ* 2:e576. doi: 10.7717/peerj.576

Burgarella, C., Cubry, P., Kane, N. A., Varshney, R. K., Mariac, C., Liu, X., et al. (2018). A western Sahara centre of domestication inferred from pearl millet genomes. *Nat. Ecol. Evol.* 2, 1377–1380. doi: 10.1038/s41559-018-0643-y

Burggren, W. (2016). Epigenetic inheritance and its role in evolutionary biology: re-evaluation and new perspectives. *Biology (Basel).* 5:24. doi: 10.3390/biology5020024

Cao, K., Li, Y., Deng, C. H., Gardiner, S. E., Zhu, G., Fang, W., et al. (2019). Comparative population genomics identified genomic regions and candidate genes associated with fruit domestication traits in peach. *Plant Biotechnol. J.* 17, 1954–1970. doi: 10.1111/pbi.13112

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., and Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1016/j.biotechadv.2011.08.021.Secreted

Chalfin, L., Dayan, M., Levy, D. R., Austad, S. N., Miller, R. A., Iraqi, F. A., et al. (2014). Mapping ecologically relevant social behaviours by gene knockout in wild mice. *Nat. Commun.* 5:4569. doi: 10.1038/ncomms5569

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10, 195–205. doi: 10.1038/nrg2526

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109

Chen, J., Ni, P., Li, X., Han, J., Jakovliæ, I., Zhang, C., et al. (2018). Population size may shape the accumulation of functional mutations following domestication. *BMC Evol. Biol.* 18:4. doi: 10.1186/s12862-018-1120-6

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., et al. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* 9:2337. doi: 10.1038/s41467-018-04737-0

Clark, J. W., and Donoghue, P. C. J. (2018). Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* 23, 933–945. doi: 10.1016/j.tplants.2018.07.006

Concia, L., Veluchamy, A., Ramirez-Prado, J. S., Martin-Ramirez, A., Huang, Y., Perez, M., et al. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol.* 21, 1–20. doi: 10.1186/s13059-020-01998-1

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143

Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., et al. (2014). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30, 1187–1189. doi: 10.1093/bioinformatics/btt763

Daly, K. G., Maisano Delser, P., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., et al. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* 361, 85–88. doi: 10.1126/science.aas9411

Davey, J. W., and Blaxter, M. L. (2010). RADSeq?: next-generation population genetics. *Brief. Funct. Genomics* 9, 416–423. doi: 10.1093/bfgp/elq031

De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., et al. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol. Ecol.* 22, 1383–1399. doi: 10.1111/mec.12182

de Villemereuil, P., and Gaggiotti, O. E. (2015). A new F ST -based method to uncover local adaptation using environmental variables. *Methods Ecol. Evol.* 6, 1248–1258. doi: 10.1111/2041-210X.12418

De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., et al. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* 12, 1058–1067. doi: 10.1111/1755-0998.12003

Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F., and Marchini, J. (2013). Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* 93, 687–696. doi: 10.1016/j.ajhg.2013.09.002

Diao, L., and Chen, K. C. (2012). Local ancestry corrects for population structure in *Saccharomyces cerevisiae* genome-wide association studies. *Genetics* 192, 1503–1511. doi: 10.1534/genetics.112.144790

Dias-Alves, T., Mairal, J., and Blum, M. G. B. (2018). Loter: a software package to infer local ancestry for a wide range of species. *Mol. Biol. Evol.* 35, 2318–2326. doi: 10.1093/molbev/msy126

Díaz-Valenzuela, E., Sawers, R. H., and Cibrián-Jaramillo, A. (2020). *Cis*-and trans-regulatory variations in the domestication of the chili pepper fruit. *Mol. Biol. Evol.* 37, 1593–1603. doi: 10.1093/molbev/msaa027

Ding, M., and Chen, Z. J. (2018). Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. *Curr. Opin. Plant Biol.* 42, 37–48. doi: 10.1016/j.pbi.2018.02.003

Doebley, J., and Stec, A. (1991). Genetic analysis of the morphological differences between maize and teosinte. *Genetics* 129, 285–295.

Doebley, J., Stec, A., and Gustus, C. (1995). teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* 141, 333–346.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* 31, 135–141. doi: 10.1038/nbt.2478

Dorant, Y., Benestan, L., Rougemont, Q., Normandeau, E., Boyle, B., Rochette, R., et al. (2019). Comparing Pool−seq, Rapture, and GBS genotyping for inferring weak population structure: the American lobster (*Homarus americanus*) as a case study. *Ecol. Evol.* 9, 6606–6623. doi: 10.1002/ece3.5240

Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252. doi: 10.1093/molbev/msr048

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science* 366, 594–599. doi: 10.1126/science.aaw2090

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51, 541–547. doi: 10.1038/s41588-019-0356-4

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* 29, 51–63. doi: 10.1016/j.tree.2013.09.008

Excoffier, L., and Foll, M. (2011). fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27, 1332–1334. doi: 10.1093/bioinformatics/btr124

Excoffier, L., Hofer, T., and Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb).* 103:285. doi: 10.1038/hdy.2009.74

Fang, Z., and Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Brief. Bioinform.* 12, 280–287. doi: 10.1093/bib/bbr004

Fay, J. C., and Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413.

Feeney, A., Nilsson, E., and Skinner, M. K. (2014). Epigenetics and transgenerational inheritance in domesticated farm animals. *J. Anim. Sci. Biotechnol.* 5:48. doi: 10.1186/2049-1891-5-48

Fernie, A. R., and Yan, J. (2019). De novo domestication: an alternative route toward new crops for the future. *Mol. Plant* 12, 615–631. doi: 10.1016/j.molp.2019.03.016

Fierst, J. L. (2015). Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Front. Genet.* 6:220. doi: 10.3389/fgene.2015.00220

Fitz-Gibbon, S., Hipp, A. L., Pham, K. K., Manos, P. S., and Sork, V. L. (2017). Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (Quercus section Quercus). *Genome* 60, 743–755. doi: 10.1139/gen-2016-0202

Foll, M., and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180, 977–993. doi: 10.1534/genetics.108.092221

Frantz, L. A. F., Mullin, V. E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., et al. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 352, 1228–1231. doi: 10.1126/science.aaf3161

Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., et al. (2015). Evidence of long-term gene flow and selection during

domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* 47, 1141–1148. doi: 10.1038/ng.3394

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., et al. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1827–1831. doi: 10.1073/pnas.89.5.1827

Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., et al. (2015). N6-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. *Cell* 161, 879–892. doi: 10.1016/j.cell.2015.04.010

Fugère, V., and Hendry, A. P. (2018). Human influences on the strength of phenotypic selection. *Proc. Natl. Acad. Sci. U.S.A.* 115, 10070–10075. doi: 10.1073/pnas.1806013115

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One* 8:e79667. doi: 10.1371/journal.pone.0079667

Futschik, A., and Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186, 207–218. doi: 10.1534/genetics.110.114397

Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., et al. (2016). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* 16, 459–469. doi: 10.1111/1755-0998.12470

Gansauge, M.-T., Gerber, T., Glocke, I., Korleviæ, P., Lippik, L., Nagel, S., et al. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* 45:gkx033. doi: 10.1093/nar/gkx033

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2

Garud, N. R., Messer, P. W., Buzbas, E. O., and Petrov, D. A. (2015). Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet.* 11:e1005004. doi: 10.1371/journal.pgen.1005004

Gepts, P. (2014). The contribution of genetic and genomic approaches to plant domestication studies. *Curr. Opin. Plant Biol.* 18, 51–59. doi: 10.1016/j.pbi.2014.02.001

Gerbault, P., Allaby, R. G., Boivin, N., Rudzinski, A., Grimaldi, I. M., Pires, J. C., et al. (2014). Storytelling and story testing in domestication. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6159–6164. doi: 10.1073/pnas.1400425111

Gibson, G. (2018). Population genetics and GWAS: a primer. *PLoS Biol.* 16:e2005485. doi: 10.1371/journal.pbio.2005485

Golicz, A. A., Batley, J., and Edwards, D. (2016a). Towards plant pangenomics. *Plant Biotechnol. J.* 14, 1099–1105. doi: 10.1111/pbi.12499

Golicz, A. A., Bayer, P. E., Barker, G. C., Edger, P. P., Kim, H., Martinez, P. A., et al. (2016b). The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 7, 1–8. doi: 10.1038/ncomms13390

Gouil, Q., and Keniry, A. (2019). Latest techniques to study DNA methylation. *Essays Biochem.* 63, 639–648. doi: 10.1042/EBC20190027

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals – a review. *Anim. Genet.* 41, 6–31. doi: 10.1111/j.1365-2052.2010.02038.x

Guerra García, A., and Piñero, D. (2017). Current approaches and methods in plant domestication studies. *Bot. Sci.* 95:345. doi: 10.17129/botsci.1209

Guerra-García, A., Suárez-Atilano, M., Mastretta-Yanes, A., Delgado-Salinas, A., and Piñero, D. (2017). Domestication genomics of the open-pollinated scarlet runner bean (*Phaseolus coccineus* L.). *Front. Plant Sci.* 8:1891. doi: 10.3389/fpls.2017.01891

Guerrero-Bosagna, C. (2012). Finalism in darwinian and lamarckian evolution: lessons from epigenetics and developmental biology. *Evol. Biol.* 39, 283–300. doi: 10.1007/s11692-012-9163-x

Günther, T., and Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genet.* 15:e1008302. doi: 10.1371/journal.pgen.1008302

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2012). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45, 51–58. doi: 10.1038/ng.2470

Guo, Z., Song, G., Liu, Z., Qu, X., Chen, R., Jiang, D., et al. (2015). Global epigenomic analysis indicates that Epialleles contribute to Allele-specific expression via Allele-specific histone modifications in hybrid rice. *BMC Genomics* 16:232. doi: 10.1186/s12864-015-1454-z

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5:e1000695. doi: 10.1371/journal.pgen.1000695

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Hahn, F., Eisenhut, M., Mantegazza, O., and Weber, A. (2017). Generation of targeted knockout mutants in *Arabidopsis thaliana* using CRISPR/Cas9. *Bio Protoc.* 7, 1–20. doi: 10.21769/BioProtoc.2384

Hahn, M. W., and Hibbins, M. S. (2019). A three-sample test for introgression. *Mol. Biol. Evol.* 36, 2878–2882. doi: 10.1093/molbev/msz178

Han, L., and Abney, M. (2013). Using identity by descent estimation with dense genotype data to detect positive selection. *Eur. J. Hum. Genet.* 21, 205–211. doi: 10.1038/ejhg.2012.148

Hanchard, N. A., Rockett, K. A., Spencer, C., Coop, G., Pinder, M., Jallow, M., et al. (2006). Screening for recently selected alleles by analysis of human haplotype similarity. *Am. J. Hum. Genet.* 78, 153–159. doi: 10.1086/499252

He, S., Yan, S., Wang, P., Zhu, W., Wang, X., Shen, Y., et al. (2014). Comparative analysis of genome-wide chromosomal histone modification patterns in maize cultivars and their wild relatives. *PLoS One* 9:e97364. doi: 10.1371/journal.pone.0097364

He, X.-J., Chen, T., and Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Res.* 21, 442–465. doi: 10.1038/cr.2011.23

Heard, E., and Martienssen, R. A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157, 95–109. doi: 10.1016/j.cell.2014.02.045

Hekman, J. P., Johnson, J. L., and Kukekova, A. V. (2015). Transcriptome analysis in domesticated species: challenges and strategies. *Bioinform. Biol. Insights* 9(Suppl. 4), 21–31. doi: 10.4137/BBI.S29334

Heled, J., and Drummond, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* 8:289. doi: 10.1186/1471-2148-8-289

Hradilová, I., Trnìnı, O., Válková, M., Cechová, M., Janská, A., Prokešová, L., et al. (2017). A combined comparative transcriptomic, metabolomic, and anatomical analyses of two key domestication traits: pod dehiscence and seed dormancy in pea (*Pisum* sp.). *Front. Plant Sci.* 8:542. doi: 10.3389/fpls.2017.00542

Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532

Hübner, S., Bercovich, N., Todesco, M., Mandel, J. R., Odenheimer, J., Ziegler, E., et al. (2019). Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* 5, 54–62. doi: 10.1038/s41477-018-0329-0

Hufford, M. B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J.-M., Cartwright, R. A., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811. doi: 10.1038/ng.2309

Hurgobin, B., Golicz, A. A., Bayer, P. E., Chan, C. K. K., Tirnaz, S., Dolatabadian, A., et al. (2018). Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol. J.* 16, 1265–1274. doi: 10.1111/pbi.12867

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C. A., Carretero-Paulet, L., Chang, T. H., et al. (2013). Architecture and evolution of a minute plant genome. *Nature* 498, 94–98. doi: 10.1038/nature12132

Ilska, J., Haskell, M. J., Blott, S. C., Sánchez-Molano, E., Polgar, Z., Lofgren, S. E., et al. (2017). Genetic characterization of dog personality traits. *Genetics* 206, 1101–1111. doi: 10.1534/genetics.116.192674

Inbar, S., Cohen, P., Yahav, T., and Privman, E. (2020). Comparative study of population genomic approaches for mapping colony-level traits. *PLoS Comput. Biol.* 16:e1007653. doi: 10.1371/journal.pcbi.1007653

Irving-Pease, E. K., Ryan, H., Jamieson, A., Dimopoulos, E. A., Larson, G., and Frantz, L. A. F. (2019). "Paleogenomics of animal domestication," in *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, eds C. Lindqvist and O. P. Rajora (Cham: Springer International Publishing), 225–272. doi: 10.1007/13836_2018_55

Janowitz Koch, I., Clark, M. M., Thompson, M. J., Deere-Machemer, K. A., Wang, J., Duarte, L., et al. (2016). The concerted impact of domestication and

transposon insertions on methylation patterns between dogs and grey wolves. *Mol. Ecol.* 25, 1838–1855. doi: 10.1111/mec.13480

Janzen, G. M., Wang, L., and Hufford, M. B. (2019). The extent of adaptive wild introgression in crops. *New Phytol.* 221, 1279–1288. doi: 10.1111/nph.15457

Jensen, P. (2015). Adding "epi-" to behaviour genetics: implications for animal domestication. *J. Exp. Biol.* 218, 32–40. doi: 10.1242/jeb.106799

Jiang, L. G., Li, B., Liu, S. X., Wang, H. W., Li, C. P., Song, S. H., et al. (2019). Characterization of proteome variation during modern maize breeding. *Mol. Cell. Proteomics* 18, 263–276. doi: 10.1074/mcp.RA118.001021

Jiang, Y., Jiang, Y., Wang, S., Zhang, Q., and Ding, X. (2019). Optimal sequencing depth design for whole genome re-sequencing in pigs. *BMC Bioinform.* 20:556. doi: 10.1186/s12859-019-3164-z

Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., et al. (2012). Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* 44, 812–815. doi: 10.1038/ng.2312

Jombart, T., Devillard, S., and Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94. doi: 10.1186/1471-2156-11-94

Jorde, L. B. (2001). Population genomics: a bridge from evolutionary history to genetic medicine. *Hum. Mol. Genet.* 10, 2199–2207. doi: 10.1093/hmg/10.20.2199

Kantar, M. B., Nashoba, A. R., Anderson, J. E., Blackman, B. K., and Rieseberg, L. H. (2017). The genetics and genomics of plant domestication. *Bioscience* 67, 971–982. doi: 10.1093/biosci/bix114

Kaur, P., and Gaikwad, K. (2017). From genomes to GENE-omes: exome sequencing concept and applications in crop improvement. *Front. Plant Sci.* 8:2164. doi: 10.3389/fpls.2017.02164

Khan, A. W., Garg, V., Roorkiwal, M., Golicz, A. A., Edwards, D., and Varshney, R. K. (2020). Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends. Plant Sci.* 25, 148–158. doi: 10.1016/j.tplants.2019.10.012

Khan, M. A., Olsen, K. M., Sovero, V., Kushad, M. M., and Korban, S. S. (2014). Fruit quality traits have played critical roles in domestication of the apple. *Plant Genome* 7, 1–18. doi: 10.3835/plantgenome2014.04.0018

Kim, Y., and Nielsen, R. (2004). Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167, 1513–1524. doi: 10.1534/genetics.103.025387

Koenig, D., Jimenez-Gomez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2655–E2662. doi: 10.1073/pnas.1309606110

Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* 15:356. doi: 10.1186/s12859-014-0356-4

Korneliussen, T. S., Moltke, I., Albrechtsen, A., and Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinform.* 14:289. doi: 10.1186/1471-2105-14-289

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., et al. (2014). Current perspectives and the future of domestication studies. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6139–6146. doi: 10.1073/pnas.1323964111

Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A., and Gilissen, C. (2015). Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum. Mutat.* 36, 815–822. doi: 10.1002/humu.22813

Lemmon, Z. H., Bukowski, R., Sun, Q., and Doebley, J. F. (2014). The role of *Cis* regulatory evolution in maize domestication. *PLoS Genet.* 10:e1004745. doi: 10.1371/journal.pgen.1004745

Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115. doi: 10.1146/annurev-genom-083115-022413

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496. doi: 10.1038/nature10231

Li, R., Fu, W., Su, R., Tian, X., Du, D., Zhao, Y., et al. (2019). Towards the complete goat pan-genome by recovering missing genomic segments from the reference genome. *Front. Genet.* 10:1169. doi: 10.3389/fgene.2019.01169

Li, Y., Von Holdt, B. M., Reynolds, A., Boyko, A. R., Wayne, R. K., Wu, D. D., et al. (2013). Artificial selection on brain-expressed genes during the domestication of dog. *Mol. Biol. Evol.* 30, 1867–1876. doi: 10.1093/molbev/mst088

Li, Y., Zhou, G., Ma, J., Jiang, W., Jin, L., Zhang, Z., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979

Li, Z., Chen, J., Wang, Z., Pan, Y., Wang, Q., Xu, N., et al. (2016). Detection of selection signatures of population-specific genomic regions selected during domestication process in Jinhua pigs. *Anim. Genet.* 47, 672–681. doi: 10.1111/age.12475

Librado, P., Fages, A., Gaunitz, C., Leonardi, M., Wagner, S., Khan, N., et al. (2016). The evolutionary origin and genetic makeup of domestic horses. *Genetics* 204, 423–434. doi: 10.1534/genetics.116.194860

Linck, E., and Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Mol. Ecol. Resour.* 19, 639–647. doi: 10.1111/1755-0998.12995

Liu, X., and Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra. *Nat. Genet.* 47, 555–559. doi: 10.1038/ng.3254

Lotterhos, K. E., and Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Mol. Ecol.* 23, 2178–2192. doi: 10.1111/mec.12725

Lotterhos, K. E., and Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* 24, 1031–1046. doi: 10.1111/mec.13100

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., et al. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* 17, 142–152. doi: 10.1111/1755-0998.12635

Luo, G.-Z., Blanco, M. A., Greer, E. L., He, C., and Shi, Y. (2015). DNA N6-methyladenine: a new epigenetic mark in eukaryotes? *Nat. Rev. Mol. Cell Biol.* 16, 705–710. doi: 10.1038/nrm4076

Luu, K., Bazin, E., and Blum, M. G. B. (2017). pcadapt?: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* 17, 67–77. doi: 10.1111/1755-0998.12592

Lye, Z. N., and Purugganan, M. D. (2019). Copy number variation in domestication. *Trends Plant Sci.* 24, 352–365. doi: 10.1016/j.tplants.2019.01.003

Machaj, G., Bostan, H., Macko-Podgórni, A., Iorizzo, M., and Grzebelus, D. (2018). Comparative transcriptomics of root development in wild and cultivated carrots. *Genes (Basel).* 9:431. doi: 10.3390/genes9090431

Maples, B. K., Gravel, S., Kenny, E. E., and Bustamante, C. D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* 93, 278–288. doi: 10.1016/j.ajhg.2013.06.020

Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32, 244–257. doi: 10.1093/molbev/msu269

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433. doi: 10.1038/nature22043

Mascher, M., Schuenemann, V. J., Davidovich, U., Marom, N., Himmelbach, A., Hübner, S., et al. (2016). Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat. Genet.* 48, 1089–1093. doi: 10.1038/ng.3611

Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., and Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Mol. Ecol. Resour.* 15, 28–41. doi: 10.1111/1755-0998.12291

Mather, N., Traves, S. M., and Ho, S. Y. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecol. Evol.* 10, 579–589. doi: 10.1002/ece3.5888

Mathew, B., Léon, J., and Sillanpää, M. J. (2018). A novel linkage-disequilibrium corrected genomic relationship matrix for SNP-heritability estimation and genomic prediction. *Heredity (Edinb).* 120, 356–368. doi: 10.1038/s41437-017-0023-4

Mazet, O., Rodríguez, W., and Chikhi, L. (2015). Demographic inference using genetic data from a single individual: separating population size variation from population structure. *Theor. Popul. Biol.* 104, 46–58. doi: 10.1016/j.tpb.2015.06.003

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Mol. Ecol.* 24, 3223–3231. doi: 10.1111/mec.13243

Meissner, A. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877. doi: 10.1093/nar/gki901

Meyer, R. S., Choi, J. Y., Sanches, M., Plessis, A., Flowers, J. M., Amas, J., et al. (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* 48, 1083–1088. doi: 10.1038/ng.3633

Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605

Money, D., Migicovsky, Z., Gardner, K., and Myles, S. (2017). LinkImputeR: user-guided genotype calling and imputation for non-model organisms. *BMC Genomics* 18:523. doi: 10.1186/s12864-017-3873-5

Montenegro, J. D., Golicz, A. A., Bayer, P. E., Hurgobin, B., Lee, H., Chan, C. K. K., et al. (2017). The pangenome of hexaploid bread wheat. *Plant Journal* 90, 1007–1013. doi: 10.1111/tpj.13515

Moyers, B. T., Morrell, P. L., and McKay, J. K. (2018). Genetic costs of domestication and improvement. *J. Hered.* 109, 103–116. doi: 10.1093/jhered/esx069

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., et al. (2016). The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol.* 17:53. doi: 10.1186/s13059-016-0917-0

Myers, C. L., Springer, N. M., Schaefer, R., Ross-Ibarra, J., Swanson-Wagner, R., Tiffin, P., et al. (2012). Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11878–11883. doi: 10.1073/pnas.1201961109

Nadachowska-Brzyska, K., Burri, R., Smeds, L., and Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Mol. Ecol.* 25, 1058–1072. doi: 10.1111/mec.13540

Ndjiondjop, M. N., Alachiotis, N., Pavlidis, P., Goungoulou, A., Kpeki, S. B., Zhao, D., et al. (2019). Comparisons of molecular diversity indices, selective sweeps and population structure of African rice with its wild progenitor and Asian rice. *Theor. Appl. Genet.* 132, 1145–1158. doi: 10.1007/s00122-018-3268-2

Nei, M., and Maruyama, T. (1975). Lewontin-Krakauer test for neutral genes. *Genetics* 80:395.

Nielsen, R., and Beaumont, M. A. (2009). Statistical inferences in phylogeography. *Mol. Ecol.* 18, 1034–1047. doi: 10.1111/j.1365-294X.2008.04059.x

Nowoshilow, S., Schloissnig, S., Fei, J. F., Dahl, A., Pang, A. W., Pippel, M., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554, 50–55. doi: 10.1038/nature25458

Ottoni, C., Van Neer, W., De Cupere, B., Daligault, J., Guimaraes, S., Peters, J., et al. (2017). The palaeogenetics of cat dispersal in the ancient world. *Nat. Ecol. Evol.* 1:0139. doi: 10.1038/s41559-017-0139

Pankin, A., Altmüller, J., Becker, C., and von Korff, M. (2018). Targeted resequencing reveals genomic signatures of barley domestication. *New Phytol.* 218, 1247–1259. doi: 10.1111/nph.15077

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., and Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335, 721–726. doi: 10.1038/335721a0

Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2:e190. doi: 10.1371/journal.pgen.0020190

Pavlidis, P., and Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *J. Biol. Res.* 24:7. doi: 10.1186/s40709-017-0064-0

Pickrell, J. K., and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. doi: 10.1371/journal.pgen.1002967

Piperno, D. R. (2017). Assessing elements of an extended evolutionary synthesis for plant domestication and agricultural origin research. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6429–6437. doi: 10.1073/pnas.1703658114

Pitt, D., Sevane, N., Nicolazzi, E. L., MacHugh, D. E., Park, S. D. E., Colli, L., et al. (2019). Domestication of cattle: two or three events? *Evol. Appl.* 12, 123–136. doi: 10.1111/eva.12674

Prezeworski, M., Coop, G., and Wall, J. D. (2005). The signature of positive selection on standing genetic variation. *Evolution* 59, 2312–2323. doi: 10.1111/j.0014-3820.2005.tb00941.x

Price, A. L., Tandon, A., Patterson, N., Barnes, K. C., Rafaels, N., Ruczinski, I., et al. (2009). Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* 5:e1000519. doi: 10.1371/journal.pgen.1000519

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. doi: 10.1111/j.1471-8286.2007.01758.x

Qi, J., Liu, X., Shen, D., Miao, H., Xie, B., Li, X., et al. (2013). A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* 45, 1510–1515. doi: 10.1038/ng.2801

Qiu, Q., Wang, L., Wang, K., Yang, Y., Ma, T., Wang, Z., et al. (2015). Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat. Commun.* 6:10283. doi: 10.1038/ncomms10283

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350

Ramos-Madrigal, J., Smith, B. D., Moreno-Mayar, J. V., Gopalakrishnan, S., Ross-Ibarra, J., Gilbert, M. T. P., et al. (2016). Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Curr. Biol.* 26, 3195–3201. doi: 10.1016/j.cub.2016.09.036

Ranwez, V., Serra, A., Pot, D., and Chantret, N. (2017). Domestication reduces alternative splicing expression variations in sorghum. *PLoS One* 12:e0183454. doi: 10.1371/journal.pone.0183454

Renaut, S., and Rieseberg, L. H. (2015). The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other compositae crops. *Mol. Biol. Evol.* 32, 2273–2283. doi: 10.1093/molbev/msv106

Rodríguez-Mega, E., Piñeyro-Nelson, A., Gutierrez, C., García-Ponce, B., Sánchez, M. D. L. P., Zluhan-Martínez, E., et al. (2015). Role of transcriptional regulation in the evolution of plant phenotype: a dynamic systems approach. *Dev. Dyn.* 244, 1074–1095. doi: 10.1002/dvdy.24268

Ross-Ibarra, J., Morrell, P. L., and Gaut, B. S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8641–8648. doi: 10.1073/pnas.0700643104

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi: 10.1038/nature01140

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. doi: 10.1038/nature06250

Sakurada, K. (2010). Environmental epigenetic modifications and reprogramming-recalcitrant genes. *Stem Cell Res.* 4, 157–164. doi: 10.1016/j.scr.2010.01.001

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* 7:e34131. doi: 10.1371/journal.pone.0034131

Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8, 552–560.

Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46, 919–925. doi: 10.1038/ng.3015

Schlötterer, C., Tobler, R., Kofler, R., and Nolte, V. (2014). Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* 15, 749–763. doi: 10.1038/nrg3803

Schmidt, D., Wilson, M. D., Spyrou, C., Brown, G. D., Hadfield, J., and Odom, D. T. (2009). ChIP-seq: using high-throughput sequencing to discover protein–DNA interactions. *Methods* 48, 240–248. doi: 10.1016/j.ymeth.2009.03.001

Schmitz, R. J., Schultz, M. D., Urich, M. A., Nery, J. R., Pelizzola, M., Libiger, O., et al. (2013). Patterns of population epigenomic diversity. *Nature* 495, 193–198. doi: 10.1038/nature11968

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534

Schreiber, M., Stein, N., and Mascher, M. (2018). Genomic approaches for studying crop evolution. *Genome Biol.* 19:140. doi: 10.1186/s13059-018-1528-8

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., et al. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U.S.A.* 111, E5661–E5669. doi: 10.1073/pnas.1416991111

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., et al. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22, 839–851. doi: 10.1261/rna.053959.115

Schweiger, W., Boddu, J., Shin, S., Poppenberger, B., Berthiller, F., Lemmens, M., et al. (2010). Validation of a candidate deoxynivalenol-inactivating UDP-glucosyltransferase from barley by heterologous expression in yeast. *Mol. Plant Microbe Interact.* 23, 977–986. doi: 10.1094/MPMI-23-7-0977

Seal, A., Gupta, A., Mahalaxmi, M., Aykkal, R., Singh, T. R., and Arunachalam, V. (2014). Tools, resources and databases for SNPs and indels in sequences: a review. *Int. J. Bioinform. Res. Appl.* 10:264. doi: 10.1504/IJBRA.2014.060762

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., et al. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol. Evol.* 8, 907–917. doi: 10.1111/2041-210X.12700

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87. doi: 10.1126/science.1247005

Shan, S., Soltis, P. S., Soltis, D. E., and Yang, B. (2020). Considerations in adapting CRISPR/Cas9 in nongenetic model plant systems. *Appl. Plant Sci.* 8:e11314. doi: 10.1002/aps3.11314

Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., et al. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215

Shen, Y., Zhang, J., Liu, Y., Liu, S., Liu, Z., Duan, Z., et al. (2018). DNA methylation footprints during soybean domestication and improvement. *Genome Biol.* 19:128. doi: 10.1186/s13059-018-1516-z

Shi, J., and Lai, J. (2015). Patterns of genomic changes with crop domestication and breeding. *Curr. Opin. Plant Biol.* 24C, 47–53. doi: 10.1016/j.pbi.2015.01.008

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., and Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132. doi: 10.1038/nrg3642

Singh, J., Zhao, J., and Vallejos, C. E. (2018). Differential transcriptome patterns associated with early seedling development in a wild and a domesticated common bean (*Phaseolus vulgaris* L.) accession. *Plant Sci.* 274, 153–162. doi: 10.1016/j.plantsci.2018.05.024

Smith, J. M., and Haigh, J. (2007). The hitch-hiking effect of a favourable gene. *Genet. Res. (Camb)* 89, 391–403. doi: 10.1017/S0016672308009579

Smith, O., Nicholson, W. V., Kistler, L., Mace, E., Clapham, A., Rose, P., et al. (2019). A domestication history of dynamic adaptation and genomic deterioration in Sorghum. *Nat. Plants* 5, 369–379. doi: 10.1038/s41477-019-0397-9

Sohn, J., and Nam, J.-W. (2016). The present and future of de novo whole-genome assembly. *Brief. Bioinform.* 19:bbw096. doi: 10.1093/bib/bbw096

Song, J., Li, J., Sun, J., Hu, T., Wu, A., Liu, S., et al. (2018). Genome-wide association mapping for cold tolerance in a core collection of rice (*Oryza sativa* L.) landraces by using high-density single nucleotide polymorphism markers from specific-locus amplified fragment sequencing. *Front. Plant Sci.* 9:875. doi: 10.3389/fpls.2018.00875

Song, Q., Zhang, T., Stelly, D. M., and Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18:99. doi: 10.1186/s13059-017-1229-8

Soyk, S., Lemmon, Z. H., Oved, M., Fisher, J., Liberatore, K. L., Park, S. J., et al. (2017). Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* 169, 1142–1155. doi: 10.1016/j.cell.2017.04.032

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid cucurbita genomes. *Mol. Plant* 10, 1293–1306. doi: 10.1016/j.molp.2017.09.003

Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from domestication: targeting Cis -regulatory elements for crop improvement. *Trends Plant Sci.* 21, 506–515. doi: 10.1016/j.tplants.2016.01.014

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

Tang, H., Coram, M., Wang, P., Zhu, X., and Risch, N. (2006). Reconstructing genetic ancestry blocks in admixed individuals. *Am. J. Hum. Genet.* 79, 1–12. doi: 10.1086/504302

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome.". *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Thornton, T. A., and Bermejo, J. L. (2014). Local and global ancestry inference and applications to genetic association analysis for admixed populations. *Genet. Epidemiol.* 38, S5–S12. doi: 10.1002/gepi.21819

Tiffin, P., and Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.* 29, 673–680. doi: 10.1016/j.tree.2014.10.004

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635. doi: 10.1038/nature11119

Trerotola, M., Relli, V., Simeone, P., and Alberti, S. (2015). Epigenetic inheritance and the missing heritability. *Hum. Genomics* 9:17. doi: 10.1186/s40246-015-0041-3

Trucchi, E., Gratton, P., Whittington, J. D., Cristofari, R., Le Maho, Y., Stenseth, N. C., et al. (2014). King penguin demography since the last glaciation inferred from genome-wide data. *Proc. R. Soc. B Biol. Sci.* 281:20140528. doi: 10.1098/rspb.2014.0528

Turner, B. M. (2009). Epigenetic responses to environmental change and their evolutionary implications. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3403–3418. doi: 10.1098/rstb.2009.0125

Ueta, R., Abe, C., Watanabe, T., Sugano, S. S., Ishihara, R., Ezura, H., et al. (2017). Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Sci. Rep.* 7:507. doi: 10.1038/s41598-017-00501-4

Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J., and Ecker, J. R. (2015). MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* 10:475. doi: 10.1038/nprot.2014.114

Vallebueno-Estrada, M., Rodríguez-Arévalo, I., Rougon-Cardoso, A., Martínez González, J., García Cook, A., Montiel, R., et al. (2016). The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *Proc. Natl. Acad. Sci. U.S.A.* 113, 14151–14156. doi: 10.1073/pnas.1609701113

VanBuren, R., Wai, C. M., Colle, M., Wang, J., Sullivan, S., Bushakra, J. M., et al. (2018). A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. *Gigascience* 7, 1–9. doi: 10.1093/gigascience/giy094

Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi: 10.1038/s41588-019-0401-3

Vasemägi, A., Nilsson, J., McGinnity, P., Cross, T., O'Reilly, P., Glebe, B., et al. (2012). Screen for footprints of selection during domestication/captive breeding of atlantic salmon. *Comp. Funct. Genomics* 2012, 1–14. doi: 10.1155/2012/628204

Velasco, D., Hough, J., Aradhya, M., and Ross-Ibarra, J. (2016). Evolutionary genomics of peach and almond domestication. *G3* 6, 3985–3993. doi: 10.1534/g3.116.032672

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22. doi: 10.1016/j.ajhg.2017.06.005

Vitti, J. J., Grossman, S. R., and Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120. doi: 10.1146/annurev-genet-111212-133526

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72. doi: 10.1371/journal.pbio.0040072

Wales, N., Akman, M., Watson, R. H. B., Sánchez Barreiro, F., Smith, B. D., Gremillion, K. J., et al. (2019). Ancient DNA reveals the timing and persistence of organellar genetic bottlenecks over 3,000 years of sunflower domestication and improvement. *Evol. Appl.* 12, 38–53. doi: 10.1111/eva.12594

Wales, N., Ramos Madrigal, J., Cappellini, E., Carmona Baez, A., Samaniego Castruita, J. A., Romero-Navarro, J. A., et al. (2016). The limits and potential

of paleogenomic techniques for reconstructing grapevine domestication. *J. Archaeol. Sci.* 72, 57–70. doi: 10.1016/j.jas.2016.05.014

Walley, J. W., Sartor, R. C., Shen, Z., Schmitz, R. J., Wu, K. J., Urich, M. A., et al. (2016). Integration of omic networks in a developmental atlas of maize. *Science* 353, 814–818. doi: 10.1126/science.aag1125

Wang, E. T., Kodama, G., Baldi, P., and Moyzis, R. K. (2006). Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proc. Natl. Acad. Sci. U.S.A.* 103, 135–140. doi: 10.1073/pnas.0509691102

Wang, G.-D., Xie, H.-B., Peng, M.-S., Irwin, D., and Zhang, Y.-P. (2014). Domestication genomics: evidence from animals. *Annu. Rev. Anim. Biosci.* 2, 65–84. doi: 10.1146/annurev-animal-022513-114129

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y., et al. (2012). Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant Sci.* 196, 125–131. doi: 10.1016/j.plantsci.2012.08.004

Wang, W., Feng, B., Xiao, J., Xia, Z., Zhou, X., Li, P., et al. (2014). Cassava genome from a wild ancestor to cultivated varieties. *Nat. Commun.* 5:5110. doi: 10.1038/ncomms6110

Wang, W., Zhang, X., Zhou, X., Zhang, Y., La, Y., Zhang, Y., et al. (2019). Deep genome resequencing reveals artificial and natural selection for visual deterioration, plateau adaptability and high prolificacy in chinese domestic sheep. *Front. Genet.* 10:300. doi: 10.3389/fgene.2019.00300

Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., and Watson, M. (2015). Exome sequencing: current and future perspectives. *G3* 5, 1543–1550. doi: 10.1534/g3.115.018564

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., et al. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37, 853–862. doi: 10.1038/ng1598

Weigel, D., and Colot, V. (2012). Epialleles in plant evolution. *Genome Biol.* 13, 249. doi: 10.1186/gb-2012-13-10-249

Wiener, P., and Pong-Wong, R. (2011). A regression-based approach to selection mapping. *J. Hered.* 102, 294–305. doi: 10.1093/jhered/esr014

Wigginton, J. E., Cutler, D. J., and Abecasis, G. R. (2005). A note on exact tests of hardy-weinberg equilibrium. *Am. J. Hum. Genet.* 76, 887–893. doi: 10.1086/429864

Wilkinson, S., Lu, Z. H., Megens, H.-J., Archibald, A. L., Haley, C., Jackson, I. J., et al. (2013). Signatures of diversifying selection in european pig breeds. *PLoS Genet.* 9:e1003453. doi: 10.1371/journal.pgen.1003453

Wolter, F., Schindele, P., and Puchta, H. (2019). Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. *BMC Plant Biol.* 19:176. doi: 10.1186/s12870-019-1775-1

Wu, D.-D., Ding, X.-D., Wang, S., Wójcik, J. M., Zhang, Y., Tokarska, M., et al. (2018). Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nat. Ecol. Evol.* 2, 1139–1145. doi: 10.1038/s41559-018-0562-y

Xie, M., Chung, C. Y.-L., Li, M., Wong, F.-L., Wang, X., Liu, A., et al. (2019). A reference-grade wild soybean genome. *Nat. Commun.* 10:1216. doi: 10.1038/s41467-019-09142-9

Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13, 329–342. doi: 10.1038/nrg3174

Yang, C. J., Samayoa, L. F., Bradbury, P. J., Olukolu, B. A., Xue, W., York, A. M., et al. (2019). The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5643–5652. doi: 10.1073/pnas.1820997116

Yang, I. S., and Kim, S. (2015). Analysis of whole transcriptome sequencing data: workflow and software. *Genomics Inform.* 13:119. doi: 10.5808/GI.2015.13.4.119

Yang, L., Koo, D. H., Li, Y., Zhang, X., Luan, F., Havey, M. J., et al. (2012). Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* 71, 895–906. doi: 10.1111/j.1365-313X.2012.05017.x

Yang, X., Zhang, H., Shang, J., Liu, G., Xia, T., Zhao, C., et al. (2018). Comparative analysis of the blood transcriptomes between wolves and dogs. *Anim. Genet.* 49, 291–302. doi: 10.1111/age.12675

Zadesenets, K. S., and Rubtsov, N. B. (2018). Genome duplication in animal evolution. *Russ. J. Genet.* 54, 1125–1136. doi: 10.1134/S1022795418090168

Zeder, M. A. (2006). Central questions in the domestication of plants and animals. *Evol. Anthropol. Issues News Rev.* 15, 105–117. doi: 10.1002/evan.20101

Zeder, M. A., Emshwiller, E., Smith, B. D., and Bradley, D. G. (2006). Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* 22, 139–155. doi: 10.1016/j.tig.2006.01.007

Zeng, K., Fu, Y.-X., Shi, S., and Wu, C.-I. (2006). Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics* 174, 1431–1439. doi: 10.1534/genetics.106.061432

Zeng, L., Tu, X.-L., Dai, H., Han, F.-M., Lu, B.-S., Wang, M.-S., et al. (2019). Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. *Genome Biol.* 20:79. doi: 10.1186/s13059-019-1686-3

Zhang, C., Bailey, D. K., Awad, T., Liu, G., Xing, G., Cao, M., et al. (2006). A whole genome long-range haplotype (WGLRH) test for detecting imprints of positive selection in human populations. *Bioinformatics* 22, 2122–2128. doi: 10.1093/bioinformatics/btl365

Zhang, H., Zhang, J., Lang, Z., Botella, J. R., and Zhu, J.-K. (2017). Genome editing—principles and applications for functional genomics research and crop improvement. *CRC. Crit. Rev. Plant Sci.* 36, 291–309. doi: 10.1080/07352689.2017.1402989

Zhao, B. S., Roundtree, I. A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell. Biol.* 18, 31–42. doi: 10.1038/nrm.2016.132

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284. doi: 10.1038/s41588-018-0041-z

Zheng, J., Wu, H., Zhu, H., Huang, C., Liu, C., Chang, Y., et al. (2019). Determining factors, regulation system, and domestication of anthocyanin biosynthesis in rice leaves. *New Phytol.* 223, 705–721. doi: 10.1111/nph.15807

Zhou, J., Li, D., Wang, G., Wang, F., Kunjal, M., Joldersma, D., et al. (2019). Application and future perspective of CRISPR/Cas9 genome editing in fruit crops. *J. Integr. Plant Biol.* 62, 269–286. doi: 10.1111/jipb.12793

Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., et al. (2019). The population genetics of structural variants in grapevine domestication. *Nat. Plants* 5, 965–979. doi: 10.1038/s41477-019-0507-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 June 2021
doi: 10.3389/fgene.2021.677780

Whole-Exome Sequencing for Identification of Genetic Variants Involved in Vitamin D Metabolic Pathways in Families With Vitamin D Deficiency in Saudi Arabia

Shatha Alharazy[1]†, Muhammad Imran Naseer[2,3]*†, Eman Alissa[4], Margaret Denise Robertson[5], Susan Lanham-New[5] and Adeel G. Chaudhary[2,3,6]*

[1] Department of Physiology, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, [2] Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia, [3] Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia, [4] Department of Clinical Biochemistry, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, [5] Department of Nutritional Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom, [6] Centre for Innovation in Personalized Medicine, King Abdulaziz University, Jeddah, Saudi Arabia

OPEN ACCESS

Edited by:
Tanvir Ul Hassan Dar, Baba Ghulam Shah Badshah University, India

Reviewed by:
Raja Amir Hassan Kuchay, Baba Ghulam Shah Badshah University, India
Showket Hussain, ICMR-National Institute of Cancer Prevention and Research, India

***Correspondence:**
Muhammad Imran Naseer
mimrannaseer@yahoo.com;
minaseer@kau.edu.sa
orcid.org/0000-0002-9343-5825
Shatha Alharazy
smalharazy@kau.edu.sa

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *08 March 2021*
Accepted: *12 April 2021*
Published: *08 June 2021*

Citation:
Alharazy S, Naseer MI, Alissa E, Robertson MD, Lanham-New S and Chaudhary AG (2021) Whole-Exome Sequencing for Identification of Genetic Variants Involved in Vitamin D Metabolic Pathways in Families With Vitamin D Deficiency in Saudi Arabia. Front. Genet. 12:677780. doi: 10.3389/fgene.2021.677780

Background: Numerous research studies have found an association between vitamin D (vitD) status and single-nucleotide polymorphisms (SNPs) in genes involved in vitD metabolism. It is notable that the influence of these SNPs on 25-hydroxyvitamin D [25(OH)D] levels might vary in different populations. In this study, we aimed to explore for genetic variants in genes related to vitD metabolism in families with vitD deficiency in Saudi Arabia using whole-exome sequencing (WES).

Methods: This family-based WES study was conducted for 21 families with vitD deficiency ($n = 39$) in Saudi Arabia. WES was performed for DNA samples, then resulting WES data was filtered and a number of variants were prioritized and validated by Sanger DNA sequencing.

Results: Several missense variants in vitD-related genes were detected in families. We determined two variants in low-density lipoprotein 2 gene (LRP2) with one variant (rs2075252) observed in six individuals, while the other LRP2 variant (rs4667591) was detected in 13 subjects. Single variants in 7-dehydrocholesterol reductase (DHCR7) (rs143587828) and melanocortin-1 receptor (MC1R) (rs1805005) genes were observed in two subjects from two different families. Other variants in group-specific component (GC), cubilin (CUBN), and calcium-sensing receptor (CASR) gene were found in index cases and controls. Polymorphisms in GC (rs9016) and CASR (rs1801726) were found in the majority of family cases (94 and 88%), respectively.

Conclusion: In vitD-deficient families in Saudi Arabia, we were able to detect a number of missense exonic variants including variants in GC (rs9016), CUBN (rs1801222), CASR (rs1801726), and LRP2 (rs4667591). However, the existence of these variants was not different between affected family members and non-affected controls. Additionally, we

were able to find a mutation in DHCR7 (rs143587828) and a polymorphism in LRP2 (rs2075252), which may affect vitD levels and influence vitD status. Further studies are now required to confirm the association of these variants with vitD deficiency.

Keywords: Saudi Arabia, family study, variants, polymorphisms, vitamin D, whole-exome sequencing, vitamin D metabolism, vitamin D deficiency

INTRODUCTION

Vitamin D (vitD) plays an important role in maintaining skeletal calcium (Ca) homeostasis by stimulating intestinal absorption of Ca and phosphate (PO4), stimulating bone resorption and inducing Ca reabsorption by the kidney, thus sustaining the level of calcium and phosphate necessary for bone formation and supporting appropriate functioning of parathyroid hormone (PTH) to maintain Ca levels in serum (Holick, 2007; Holick et al., 2011).

Clinically, serum 25-hydroxyvitamin D [25(OH)D] has been identified as the most effective predictor of vitD status, to date. Levels of 25(OH)D in serum are influenced by the vitD produced dermally and consumed orally, through diet or supplementation (Hollis, 1996; Del Valle et al., 2011). In addition, there are physiological, pathological, and lifestyle factors affecting 25(OH)D levels such as aging, obesity, liver and kidney diseases, and inadequate exposure to sunlight (Holick, 2004, 2007; Tsiaras and Weinstock, 2011; Hyppönen and Boucher, 2018). Among other significant factors influencing 25(OH)D levels are the genetic factors with the heritability of circulating 25(OH)D levels predicted to be between 23 and 80% (Bahrami et al., 2018), primarily as single-nucleotide polymorphisms (SNPs) in genes involved in the vitD metabolic pathway (Ahn et al., 2010; McGrath et al., 2010; Wang et al., 2010; Jolliffe et al., 2016).

Vitamin D metabolism undergoes numerous pathways that are genetically regulated. VitD is produced mainly in the skin epidermis under the effect of ultraviolet-B radiation which activates 7-dehydrocholesterol (7-DHC) in the skin to form pre-vitD (Datta et al., 2017). 7-DHC is converted to cholesterol by the enzyme 7-dehydrocholesterol reductase (DHCR7) and the presence of the reduced coenzyme nicotinamide adenine dinucleotide phosphate (NAD-P) which is activated by nicotinamide adenine dinucleotide synthetase 1 (NADSYN1) (Ahn et al., 2010). VitD (from all sources) is hydroxylated in the liver mainly by cytochrome p450 enzyme (CYP2R1/25-hydroxylase) to form 25(OH)D, with further hydroxylation in the kidney distal tubule by cytochrome p450 enzyme (CYP27B1/1α-hydroxylase) producing 1,25-dihydroxyvitamin D [1,25$(OH)_2$D], the active form of vitD (Eyles et al., 2005). Both 25(OH)D and l,25$(OH)_2$D are eliminated under the effect of another cytochrome P450 enzyme (CYP24A1/24-hydroxylase), transforming them into inactive waste calcitroic acid products [24,25$(OH)_2$D and 1,24,25$(OH)_3$D] (McGrath et al., 2010). Metabolites of vitD including 25(OH)D and 1,25$(OH)_2$D are primarily transported by vitamin D-binding protein (VDBP) encoded by group-specific components (GCs) (Daiger et al., 1975; Speeckaert et al., 2006). Cubilin (intrinsic factor-cobalamin receptor) and Magalin (low-density lipoprotein-related receptor), which are encoded by the cubilin gene (CUBN) and low-density lipoprotein 2 gene (LRP2), respectively, assist renal cells in the uptake of the 25(OH)D VDBP complex (McGrath et al., 2010; Kaseda et al., 2011; Kohlmeier, 2015). The active form of vitD, 1,25$(OH)_2$D, initiates its genomic action by binding to the vitamin D receptor (VDR) (Pike and Meyer, 2012). In addition to these genes that can be directly implicated in vitD metabolism, additional genes such as the calcium-sensing receptor (CASR) that contributes in vitD metabolism by regulating PTH and Ca levels (Brown, 2013) may also be clinically important in determining vitD status.

Vitamin D deficiency is highly prevalent in Saudi Arabia. Although several studies have already reported an association between vitD status and SNPs in genes involved in vitD metabolism (McGrath et al., 2010; Jolliffe et al., 2016), the influence of these SNPs on 25(OH)D levels might vary in different populations. For example, an SNP in DHCR7 (rs12800438) was related to vitD deficiency in African Americans but not in European Americans (Batai et al., 2014), and another SNP in DHCR7 (rs12785878) was linked to vitD deficiency in Chinese cohorts from Kazak ethnicity but not in Uyghurs (Xu et al., 2015).

The relationship between inherited variants in vitD-related genes and vitD deficiency has not been adequately addressed in Saudi Arabia. Whole-exome sequencing (WES) analysis is designated as state-of-art, sequencing large amounts of DNA with high throughput, providing fast and broad data about known or novel mutations in candidate genes in family members with a specific disease or trait. Therefore, we aimed to investigate the presence of genetic variants in genes related to vitD metabolism among families with vitD deficiency in Saudi Arabia using WES.

MATERIALS AND METHODS

Study Design and Recruitment

Members from families with a history of vitD deficiency were recruited for this study from a single tertiary center [King Abdulaziz University Hospital (KAUH), Jeddah, Saudi Arabia] and seven primary health care centers (PHCCs) distributed in Jeddah (a PHCC from each of the seven sectors of Jeddah area). The study was undertaken at the Center of Innovation in Personalized Medicine (CIPM), King Fahd Medical Research Center (KFMRC), in Jeddah.

The favorable ethical opinion of this study was provided by the Research Ethics Committee in Unit of Biomedical Ethics, Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University (KAU) (ref no. 05-CEGMR-Bioeth-2018). All participants provided written informed consent, with both

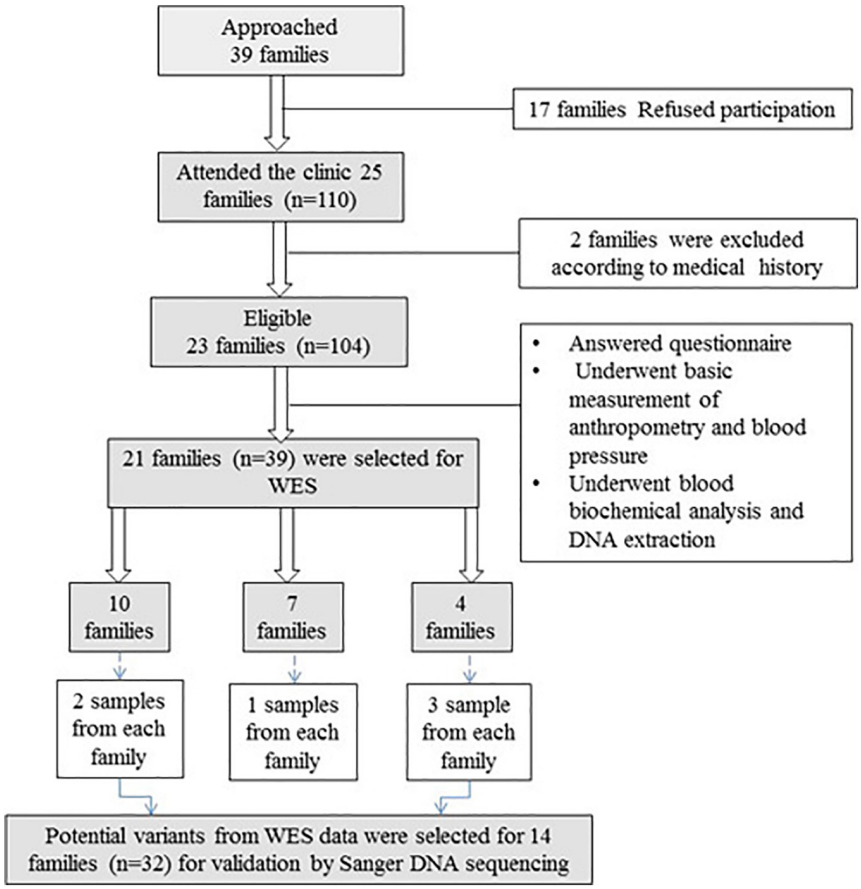

FIGURE 1 | Flowchart of recruitment of families with vitD deficiency. Thirty-nine families were approached, and 21 families ($n = 39$) were included in WES. After variant filtration and prioritization, potential variants were validated in 14 families ($n = 32$) using Sanger DNA sequencing.

parental consent and child assent obtained for participants under 16 years of age.

In total, 23 families (104 individual participants) with a history of vitD deficiency [serum 25(OH)D < 12 ng/ml] were recruited. Of these, 39 samples from 21 families were selected for WES (**Figure 1**). Exclusion criteria for inclusion in the WES analysis included history of chronic renal and liver disease, cancer, malabsorption syndrome, rheumatoid arthritis, intake of medications with possible effects on vitD (such as glucocorticoids and anticonvulsants), hyperthyroidism, hyperparathyroidism, diabetes, or any other endocrinal disorders.

Study Procedure and Blood Analysis

All participants answered a questionnaire (filled by the researcher), which requested information including socio-demographic data, medical history, drug history, and lifestyle history. Each participant underwent basic anthropometric and blood pressure measurements. Multi-generation pedigree was carefully made for each family by interviewing the family and documenting the family history of vitD deficiency. Fasting blood samples of all members of the family and from 100 unrelated controls were collected. Total serum 25(OH)D and intact PTH were measured by chemiluminescence immunoassay (CLIA), using a LIAISON auto-analyzer (DiaSorin Inc., Stillwater, MN, United States); free 25 (OH)D was directly measured by immunoassay using ELISA kit (KAPF1991, Future Diagnostics Solutions B.V., Wijchen, Netherlands); and VDBP was measured by quantitative sandwich enzyme immunoassay using Quantikine® ELISA (DVDBP0B, R&D Systems, Minneapolis, MN, United States). Serum albumin, Ca, PO_4, magnesium (Mg), lipid profile, blood glucose, and renal and liver function were all measured by the colorimetric method using a VITROS 250 Clinical Chemistry auto-analyzer (Ortho-Clinical Diagnostics Inc., Rochester, NY, United States).

Whole-Exome Sequencing

Genomic DNA was first extracted (DNA extraction kit 53104, Qiagen, Hilden, Germany), and the concentration and purity of the DNA filtrate were measured using a NanoDrop spectrophotometer (ND-1000 UV-VIS). WES with a 150-bp paired-end read length for 39 DNA samples was performed by next-generation sequencing (NGS) using the Illumina platform and Twist Human Core Exome library kit. Genomic DNA was extracted from all included blood samples, and a library was constructed by random fragmentation of DNA followed by 5′ and 3′ adapter ligation, or by "tagmentation" which coupled

the fragmentation and ligation reactions in one step, increasing the proficiency of the library preparation procedure. Afterward, adapter-ligated fragments were PCR amplified and gel purified. The library was loaded into a flow cell so that fragments get captured on a lawn of surface-bound oligos complementary to the library adapters. Next, amplification of each fragment into different clonal clusters was done by bridge amplification. Once clusters were generated completely, templates were sequenced. Illumina SBS technology which uses a reversible terminator-based approach was utilized to identify single bases integrating into DNA template strands. This technology was used due to its lower rates of raw errors compared to other technologies, as natural competition in this technology due to the presence of all four reversible terminator-bound dNTPs during each sequencing cycle reduces incorporation bias. In addition, Illumine SBS produces very precise base-by-base sequencing that practically removes sequence-context-specific errors even within repetitive sequence regions and homopolymers. Sequencing data were transformed into raw data. Raw data or images were generated by the Illumina sequencer using integrated analysis software called Real Time Analysis which is a sequencing control software for system control and base calling. The base call binaries were converted into FASTQ by using Illumina package (bcl2fastq). Reads were produced without trimming away adaptors.

Analysis of WES Data

Whole-exome sequencing data generated the raw reads in the form of FASTQ format. Insertion, deletion, and copy number variation were distinguished by utilizing SAMtools[1]. Data was aligned by using the BWA Aligner[2], after the crude information FASTQ files were adjusted. The resulting VCF files contained over 120,000 variants per samples. The variants were clarified by using different parameters, such as quality, frequency, genomic position, protein effect, and association with vitD deficiency. SNPs or variants and short indel candidates were determined at nucleotide resolution. SNPs found were compared to 1000 genomes using the international genome[3], SnpEff[4], and gnomAD databases[5]. A bioinformatics tool (laser gene Genomic Suite v. 12, DNASTAR, Madison, WI, United States) was used to look for variants involved in vitD metabolism. Variant alleles were tagged according to dbSNP142 using ArrayStar v. 12 (Rockville, MD, United States). The obtained FASTQ sequences were aligned against the human reference genome using the Borrow–Wheel arrangement tool[6] and reference genome hg19 for humans[7]. FASTQ raw data files were then transformed to BAM file format that were afterward annotated using Toolkit for Genome Analysis[8]. In this study, we targeted indels and SNPs situated in the exons and splicing junctions of the genes that caused protein-level changes, with exclusion of synonymous variants. Our selected variants were identified in around 45% of total reads.

Variant Prioritization

For variant prioritization, the coding and splicing regions of genes involved in vitD metabolic pathways were analyzed and assessed using the available online database for these variants (see text footnote 5)[9,10]. Initially, variants positioned in introns, intergenic regions, and untranslated regions were excluded, as well as synonymous variants. To comprehend potential biological functions of the variants designated, the functional influence of the selected genomic variants and pathogenicity were evaluated using prediction algorithms (Mutation Taster, PolyPhen2, SIFT, PROVEAN, and Mutation Assessor) included in ANNOVAR[11]. Lastly, candidate genes were reviewed in PubMed publications and the Online Mendelian Inheritance in human's database.

To analyze identified exonic variants related to vitD, we selected major genes involved in vitD metabolic pathways as follows: DHCR7, MC1R, GC, CYP2R1, CYP27B1, CYP24A1, VDR, RXRA, CUBN, LRP2, and CASR (Fischer, 2020).

After applying various filters, the total number of variants was reduced to 20–30 variants per sample. Finally, the variants involved in vitD metabolism were selected in the following target genes: GC, CUBN, LRP2, DHCR7, and CASR.

Validation of WES Results by Sanger Sequencing

Validation of prioritized variants in candidate genes was conducted using Sanger sequencing. Designing primers were done initially using web-based Primer3 (v. 0.4.1) software. Primers used were as follows: GC gene: Forward: 5′-TGA GACAGGCAAGTATTTCTATT-3′ Reverse: 5′-GCCAAGTTA CAATAACACCAGGA-3′, CUBN gene: Forward 1: 5′-AGAG GATAAACAGTTAGGGGCT-3′ Reverse 1: 5′-GCAGAACCA GGACACAAAACT-3′ and Forward 2: 5′-CCGAACAGAGGA GGAGACC-3′ Reverse 2: 5′-ACGTTACATTTTATAGGCGT GA-3′ and for the LRP2 gene we used Forward 1: 5′-AAAAGG GGACAGGTACATACT-3′ Rreverse1: 5′-TCCTCCTCCACTAA TGCAACA-3′ Forward 2: 5′-TGTGCTCTGCTGTAGTGGAG-3′ Rreverse 2: 5′-TCTTGAGAAAACAGGCAAAGACA-3′ and for DHCR7 Forward: 5′-CTCATGTTCCTGCTATGCGTC-3′ Reverse: 5′-GACTGGCCCCTGAGAGAAAG-3′ and for CASR Forward: 5′-ACGGTCACCTTCTCACTGAG-3′ Reverse: 5′-GACAACTCTTCAGGGTCCTCC-3′. Next, PCR amplification and purification were performed for DNA samples and Sanger sequencing conducted using a genetic analyzer (3500 genetic analyzer, Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, United States) and BigDye Terminator V3.1 Cycle Sequencing kit (cat#4337455, Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, United States).

The validated results were compared with the results of control samples ($n = 100$). Controls were matched with index samples for age, skin tone, sunlight exposure, oral vitD intake, and BMI but notably were vitD sufficient.

[1] http://samtools.Sourceforge.Net/

[2] http://bio-bwa.Sourceforge.Net/

[3] https://www.internationalgenome.org/

[4] http://snpeff.sourceforge.net/SnpEff.html

[5] https://gnomad.broadinstitute.org/

[6] http://bio-bwa.sourceforge.net/bwa.shtml

[7] http://hgdownload.cse.ucsc.edu/goldenPath

[8] http://www.broadinstitute.org/gatk

[9] http://www.1000genomes.org

[10] http://www.ncbi.nlm.nih.gov/SNP/

[11] http://www.openbioinformatics.org/annovar/

RESULTS

Results of WES Data

Various missense variants with moderate impact were determined in GC, CUBN, LRP2, DHCR7, and CASR genes (**Table 1**). The polymorphism rs9016 in GC was detected in 13 families ($n = 30$), rs1801726 in CASR was detected in 12 families ($n = 28$), while rs4667591 and rs2075252 in the LRP2 gene were observed in six families ($n = 13$) and three families ($n = 6$), respectively. In addition, rs1801222 and rs1801224 in CUBN and

TABLE 1 | Filtered WES results of vitD-related gene variants among vitD-deficient families.

Gene name	Family ID (F#)	Chromosome	HGVS.c	HGVS.p	Effect	Putative Impact	SNP
GC	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F12, F13, F14	Chr4	1391A > G	His464Arg	Missense variant	Moderate	rs9016
CUBN	F5	Chr10	758T > C	Phe253Ser	Missense variant	Moderate	rs1801222
	F6		1165C > A	Pro389Thr	Missense variant	Moderate	rs1801224
LRP2	F1, F3, F9	Chr2	12280A > G	Lys4094Glu	Missense variant	Moderate	rs2075252
	F2, F5, F7, F10, F12, F13		12628A > C	Ile4210Leu	Missense variant	Moderate	rs4667591
DHCR7	F1	Chr11	376G > A	Val126Ile	Missense variant	Moderate	rs143587828
CASR	F1, F2, F3, F4, F5, F6, F8, F9, F10, F12, F13, F14	Chr3	3061G > C	Glu1021Gln	Missense variant	Moderate	rs1801726

HGVS.c/p stands for Human Genome Variation Society nomenclature for DNA reference sequence or protein reference sequence, respectively. F1, F2, F3, F5, F7, F8, F9, F11, F12, and F13: n = 2 for each family group. F4, F6, F10, and F14: n = 3 for each family group.

TABLE 2 | General characteristics of family 1, family 3, and family 9 female participants.

Variables	*Family 1*		*Family 3*		*Family 9*	
	II-1	*III-1*	*II-1*	*III-1*	*II-1*	*III-1*
Age (years)	70	54	59	35	65	40
Age at menopause years)	45	52	47	–	55	–
Years since menopause	25	2	12	–	10	–
Ethnicity	White (Arabic)	White (Arabic)	White (Arabic)	White (Arabic)		White (Arabic)
Weight (kg)	65	110	97	72	56	97
Height (cm)	140	149	161	167	149	158
BMI (kg/m^2)	33.2	49.5	37.4	25.8	25.2	38.9
Waist circumference (cm)	90	119	100	128	90	110
Hip circumference (cm)	110	148	82	108	110	128
WHR	0.82	0.80	0.78	0.76	0.82	0.86
Hypertensive/Diabetic (according to medical records)	No	No	No	No	No	No
SBP (mmHg)	141	130	126	109	100	150
DBP (mmHg)	80	72	86	71	54	94
Marital status	Widow	Divorced	Widow	Married	Married	Married
Education	Illiterate	Secondary	University	University	Illiterate	Illiterate
Occupation	Housewife	Housewife	Self-employee	Government employee	Housewife	Housewife
Skin tone (Fitzpatrick)*	Type IV (olive and mid brown)	Type IV (olive and mid brown)	Type II (white and fair)	Type II (white and fair)	Type IV (olive and mid brown)	Type IV (olive and mid brown)
Sun exposure	<1 h/week	<1 h/week	<1 h/week	<1 h/week	1–2 h/week	1–2 h/week
Veiling type	Partially covered	Partially covered	Totally covered	Partially covered	Totally covered	Totally covered
Use of sunscreen	No	No	No	No	No	No
Dietary vitD intake (IU/day)	27	116	189	195	100	49
Use of vitD supplementation	Yes	Yes	Yes	No	No	No
Physical activity	No	No	No	Yes	No	Yes
Smoking	No	No	No	No	No	No

BMI, body mass index; WHR, waist–hip ratio; SBP, systolic blood pressure; and DBP, diastolic blood pressure.
**Fitzpatrick scale Pike and Meyer, 2012.*
EAR is the estimated average requirement.

rs143587828 in DHCR7 were each detected in a different family (each family $n = 2$).

Validation of WES Results

CUBN

The CUBN variant c.758T > C in family 5 ($n = 2$) was validated. Both family 5 samples and the controls were homozygous (CC genotype) as shown in **Figures 2A,B**.

LRP2

The general and biochemical characteristics of families (F1, F3, and F9) that exhibited the c.12280A > G (rs2075252) variant in LRP2 are shown in **Tables 2**, **3**. Validation of this SNP (rs2075252) showed that F1, F3, and F9 had this variant while the control did not. In family 1 and family 9, subject II-1 (the mother) had heterozygous AG genotype while subject III-1 (the daughter) had a homozygous GG genotype and the control samples had a homozygous AA genotype (**Figures 3A–D**). On the other hand, both subjects II-1 and III-1 in family 3 had the heterozygous AG genotype (**Figures 3E,F**).

The validation of the other polymorphism (c.12628A > C) in LRP2 that was observed with WES in F2, F5, F7, F10, F12, and F13 ($n = 13$) showed that this SNP existed in all the mentioned families and control samples ($n = 100$). All samples were homozygous CC except a single sample in F5 and four of the controls that were heterozygous AC (**Figure 3G**).

DHCR7

Whole-exome sequencing results showed variant c.376G > A in DHCR7 in Family 1 (F1). General and biochemical characteristics of F1 subjects were presented earlier in **Tables 2**, **3**, and the pedigree of this family is shown in **Figure 3A**. Validation of the observed variant c.376G > A in DHCR7 in F1 revealed that subject II-1 (mother) has a GA genotype and II-1 has an AA genotype in comparison to the controls that had a GG genotype (**Figure 3H**). When this DHCR7 c.376G > A variant (rs143587828) was evaluated, it was found to be a mutation not a polymorphism.

GC

When the WES results were validated by Sanger DNA sequencing for SNP c.1391A > G in GC in family samples (F1–F10 and F12–F14) ($n = 30$), the presence of c.1334A > G SNP as homozygous genotype (GG) was confirmed in these family samples as well as in the control healthy samples (**Figure 4A**).

CASR

Validation of the c.3061G > C variant in CASR in subjects from F1 to F6, F8 to F10, and F12 to 14 ($n = 28$) showed that this variant is present in the CC genotype in controls and in these families except F2 where the genotype was heterozygous (GC) (**Figure 4B**).

Identified Polymorphisms and Mutations

In families with vitD deficiency, all observed variants were polymorphisms with the exception of the variant in DHR7 (rs143587828) which was a mutation. We found two single variants in LRP2 with one variant (rs2075252) observed in six individuals but not in control cases, while the other LRP2 variant (rs4667591) was detected in 13 subjects and in controls. A single variant in DHCR7 (rs143587828) and one in MC1R (rs1805005) were observed in two subjects from two different families but not in controls. Other variants in GC, CUBN, and CASR were found in index cases and controls. Polymorphisms in GC (rs9016) and CASR (rs1801726) were found in the majority of family cases (94% and 88%, respectively).

TABLE 3 | Biochemical characteristics of family 1, family 3, and family 9 female participants.

Variable	*Family 1*		*Family 3*		*Family 9*	
	II-1	*III-1*	*II-1*	*III-1*	*II-1*	*III-1*
Serum total 25(OH)D (ng/ml)	7.8	10.1	16	13.4	17.2	14.3
Serum direct free 25(OH)D (pg/ml)	3.6	5.1	5.1	3.1	4.6	4.01
Percentage of free 25(OH)D out of total 25(OH)D (%)*	0.046	0.050	0.032	0.023	0.027	0.028
Serum VDBP (μg/ml)	350	402	192	332	554	664
Serum Albumin (g/L)	45	40	44	49	43	46
Serum Ca (mmol/L)	2.4	2.57	2.58	2.62	2.51	2.63
Serum PO_4 (mmol/L)	1.34	1.19	1.34	1.36	1.42	1.46
Serum Mg (mmol/L)	0.8	0.9	0.9	0.9	0.8	0.8
Fasting blood glucose (mmol/L)	6.5	6.2	5.4	4.7	4.3	4.6
Serum AST (U/L)	25	17	20	21	22	21
Serum ALT (U/L)	33	25	20	17	21	18
Serum ALP (U/L)	105	116	76	60	66	62
Serum creatinine (μmol/L)	66	50	47	38	52	48
Serum total cholesterol (mmol/L)	6.2	5.7	2.9	4.1	4.9	5.3
Serum triglyceride (mmol/L)	2.30	1.58	0.59	1.12	1.2	1.08
Serum HDL-C (mmol/L)	1.2	1.3	1.2	1.2	1.7	2.2
Serum LDL-C (mmol/L)	3.8	3.72	1.39	2.37	2.1	2.58
Serum VLDL-C (mmol/L)	1.1	0.73	0.27	0.51	0.45	0.50

**Percentage of free 25(OH)D out of the total 25(OH)D was calculated by dividing free 25(OH)D levels in ng/ml over total 25(OH)D level in ng/ml, then multiplied by 100.*

25(OH)D, 25-hydroxyvitamin D; VDBP, vitamin D-binding protein; Ca, calcium; PO4, phosphate; Mg, magnesium; AST, aspartate aminotransferase; ALT, alanine aminotransferase; ALP, alkaline phosphatase; HDL-C, high-lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; and VLDL-C, very low-density lipoprotein cholesterol.

DISCUSSION

Several studies have linked vitD deficiency with numerous variants in genes involved in vitD metabolism (McGrath et al., 2010; Jolliffe et al., 2016). Our WES study in families having vitD deficiency revealed various variants in genes related to vitD; however, the majority of these variants including the ones in GC (rs9016), CUBN (rs1801222), CASR (rs1801726), and LRP2 (rs4667591) coexisted in both the vitD-deficient families and the

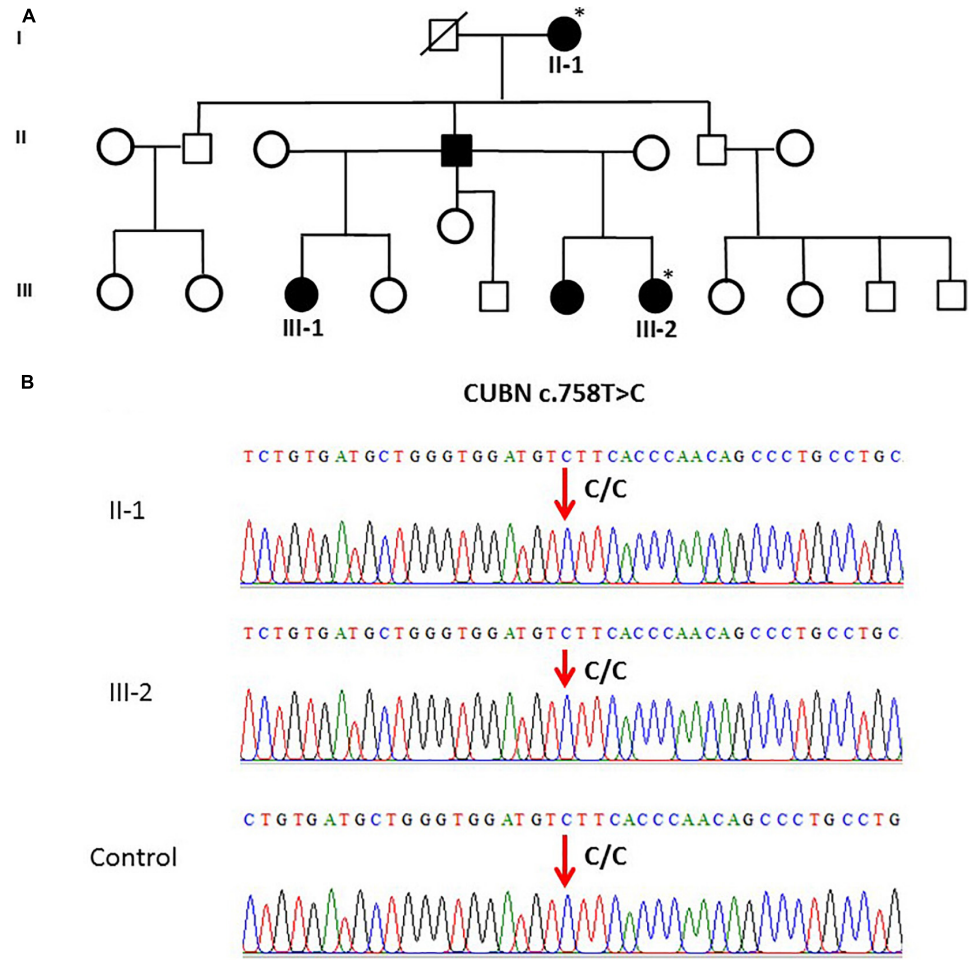

FIGURE 2 | (A) A pedigree for family 5, showing vitD-deficient II-1 and III-2 subjects. **(B)** Sanger sequencing chromatogram revealed c.758T > C polymorphism (homozygous CC genotype) in CUBN gene in members of family 5 and control. * Samples analysed.

non-affected control group (with GC and CASR SNPs having the highest frequency), suggesting no association between these SNPs and 25(OH)D levels. In agreement with our findings, a case–control study in Egyptians ($n = 328$) also found that CUBN (rs1801222) was not associated with total 25(OH)D levels Harding et al. (2006) and Elsabbagh et al. (2020) found no association between 25(OH)D and CASR (rs1801726). With regard to GC (rs9016) and LRP2 (rs4667591), no reports exist in the literature about their relationship with vitD. However, these two SNPs were reported in a family-based WES study specifically looking at SNPs in genes related to vitD metabolism in families with familial multiple sclerosis; however, no association was found with multiple sclerosis (Pytel et al., 2019).

In the present study, a mutation in DHCR7 (rs143587828) was identified in two affected subjects from one family (mother was heterozygous and daughter was homozygous for the minor allele) but not in any of the control subjects. As DHCR7 encodes for the production of the enzyme that is responsible for the conversion of 7-DHC (the precursor of vitD) to cholesterol (Berry and Hyppönen, 2011), it is suggested that this mutation in DHCR7 (rs143587828) might result in increased activity of DHCR7 leading to reduced conversion to vitD and thus vitD deficiency (Kohlmeier, 2012). Two large genome-wide association studies in subjects of European ancestry found that minor alleles of nine alternative SNPs in DHCR7/NADSYN1 were associated with vitD deficiency (Ahn et al., 2010; Wang et al., 2010). However, this may be the first report of the association of rs143587828 with 25(OH)D. This observed mutation in DHCR7 (rs143587828) now needs to be investigated in a large-scale population study, to explore further the association between this mutation and vitD status.

Cubilin and megalin, which are receptor proteins present in the proximal renal tubules encoded by CUBN and LRP2 genes, respectively, bind to the VDBP 25(OH)D complex and contribute to the process of endocytosis of the VDBP 25(OH)D complex so that 25(OH)D can be hydroxylated to $1,25(OH)_2D$, the active form of vitD (McGrath et al., 2010; Kaseda et al., 2011; Kohlmeier, 2015). Severe hypovitaminosis D was reported in LRP2 knockout mice, which suggests an important role for LRP2 (Nykjaer et al., 1999). In our study, we found an SNP (rs2075252) in LRP2 in six

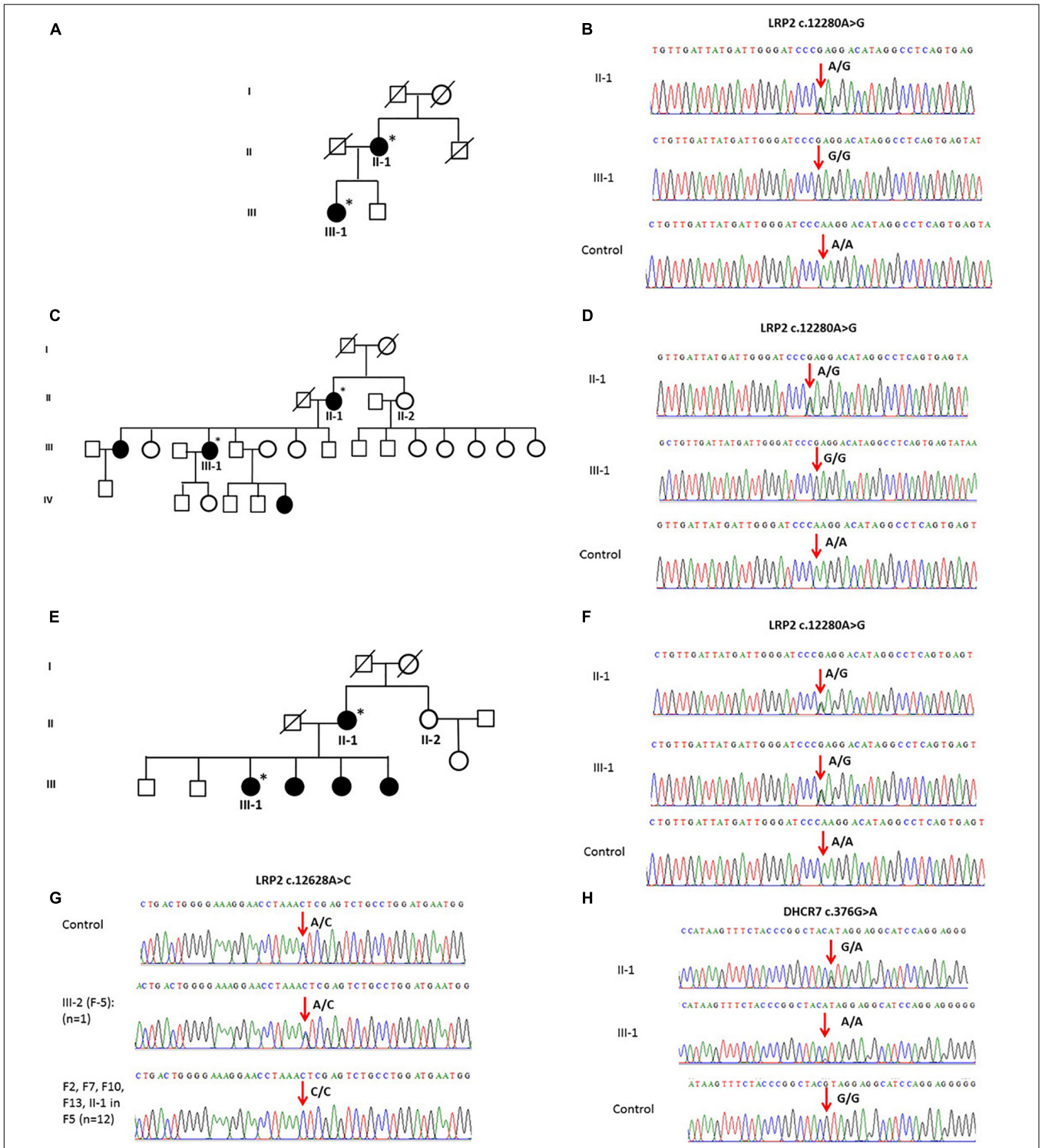

FIGURE 3 | (A) A pedigree of family 1 representing II-1 and III-1 subjects with history of vitD deficiency. **(B)** Sanger sequencing chromatogram of family 1 showing the c.12280A > G variant in LRP2 with heterozygous AG genotype in subject II-1, homozygous GG genotype in III-1 subject, and homozygous AA genotype in control. **(C)** A multi-generation pedigree for family 9 showing the analyzed samples with vitD deficiency (the mother II-1 and daughter III-1). **(D)** Sanger sequencing chromatograms for c.12280 > C LRP2 variants in control (AA), mother (AG), and daughter (GG). **(E)** A multi-generation pedigree for family 3 showing the analyzed samples with vitD deficiency (the mother II-1 and daughter III-1). **(F)** Sanger sequencing chromatograms for c.12280 > C LRP2 variants in control (AA), mother (AG), and daughter (AG). **(G)** Examples of Sanger sequencing chromatograms representing the variant c.12628A > C in the LRP2 gene. Majority of samples ($n = 12$) were homozygous (CC), while control samples ($n = 100$) and a single sample in F5 were heterozygous. **(H)** Sanger sequencing chromatogram showing c.376G > A mutation in the DHCR7 gene as (G/A) in mother and (A/A) in daughter compared to control (G/G). * Samples analysed.

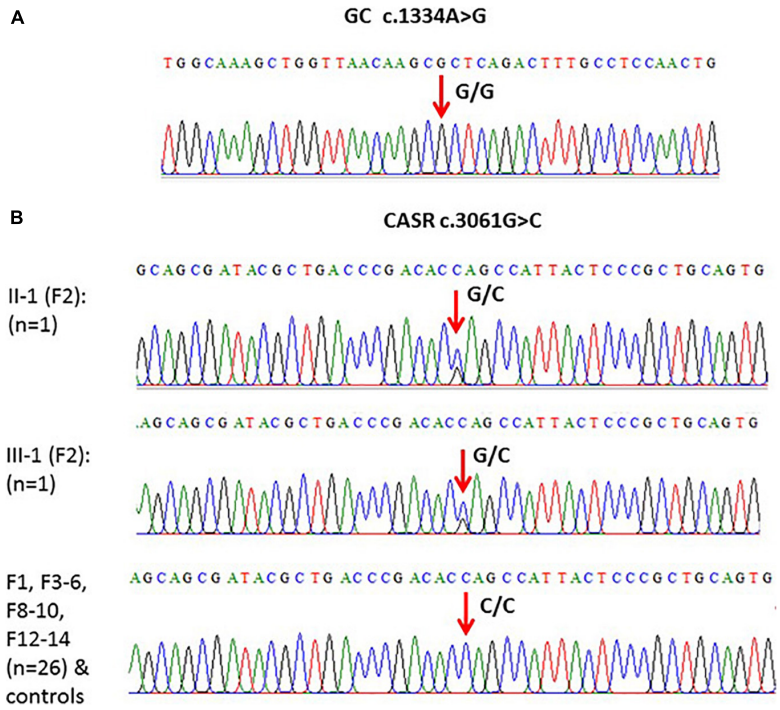

FIGURE 4 | (A) An example of one of the chromatograms that showed c.1334A > G polymorphism (G/G) in the GC gene. **(B)** Models of elicited Sanger sequencing chromatograms that showed polymorphism (c.3061G > C) in the CASR gene. Homozygous CC was the major genotype, while heterozygous GC was found in F2 only.

affected families ($n = 13$) but not in the controls. This strongly suggests that this SNP might be related to vitD deficiency and emphasizes the need for additional studies on the association between vitD status and SNPs in LRP2. To our knowledge, there is only one report in the literature and this opposes our finding, with polymorphism rs4667591 in LRP2 not found to be associated with total 25(OH)D (Elsabbagh et al., 2020).

Our study has revealed relevant and novel exonic missense variants in both DHCR7 and LRP2 in vitD-deficient families (not evident in control individuals); the association between these variants and vitD deficiency now needs to be addressed. Our results provide information on the variants related to vitD metabolism in families with vitD deficiency, thus helping researchers understand genetic factors underlying vitD deficiency in the Saudi population.

DATA AVAILABILITY STATEMENT

The datasets for this manuscript are not publicly available because family consents to share data publicly was not allowed. Requests to access the datasets should be directed to corresponding author (MN).

ETHICS STATEMENT

Ethical approval of this study was obtained from the Research Ethics Committee in Unit of Biomedical Ethics, Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University (KAU), (05-CEGMR-Bioeth-2018). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SA contributed to the study design and execution, data analysis, and manuscript drafting. MN contributed to the study design, data analysis, writing, editing, and review. EA and AC contributed to writing the review and supervision. MR contributed to the supervision and review of the manuscript. SL-N contributed to supervision. All the authors read and approved the final manuscript.

FUNDING

This work was supported by the Joint supervision program, KAU, Jeddah, Saudi Arabia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We are thankful for all families who participated in this study.

REFERENCES

Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K. C., McCullough, M. L., Gallicchio, L., et al. (2010). Genome-wide association study of circulating vitamin D levels. *Hum. Mol. Genet.* 19, 2739–2745.

Bahrami, A., Sadeghnia, H. R., Tabatabaeizadeh, S. A., Bahrami-Taghanaki, H., Behboodi, N., Esmaeili, H., et al. (2018). Genetic and epigenetic factors influencing vitamin D status. *J. Cell. Physiol.* 233, 4033–4043. doi: 10.1002/jcp.26216

Batai, K., Murphy, A. B., Shah, E., Ruden, M., Newsome, J., Agate, S., et al. (2014). Common vitamin D pathway gene variants reveal contrasting effects on serum vitamin D levels in African Americans and European Americans. *Hum. Genet.* 133, 1395–1405. doi: 10.1007/s00439-014-1472-y

Berry, D., and Hyppönen, E. (2011). Determinants of vitamin D status: focus on genetic variations. *Curr. Opin. Nephrol. Hypertens.* 20, 331–336. doi: 10.1097/mnh.0b013e328346d6ba

Brown, E. M. (2013). Role of the calcium-sensing receptor in extracellular calcium homeostasis. *Best Pract. Res. Clin. Endocrinol. Metab.* 27, 333–343. doi: 10.1016/j.beem.2013.02.006

Daiger, S. P., Schanfield, M. S., and Cavalli-Sforza, L. (1975). Group-specific component (Gc) proteins bind vitamin D and 25-hydroxyvitamin D. *Proc. Natl. Acad. Sci. U.S.A.* 72, 2076–2080. doi: 10.1073/pnas.72.6.2076

Datta, P., Philipsen, P. A., Olsen, P., Bogh, M. K., Johansen, P., Schmedes, A. V., et al. (2017). The half-life of 25 (OH) D after UVB exposure depends on gender and vitamin D receptor polymorphism but mainly on the start level. *Photochem. Photobiol. Sci.* 16, 985–995. doi: 10.1039/c6pp00258g

Del Valle, H. B., Yaktine, A. L., Taylor, C. L., and Ross, A. C. (2011). *Dietary Reference Intakes for Calcium and Vitamin D*. Washington, DC: National Academies Press.

Elsabbagh, R. A., Rahman, M. F. A., Hassanein, S. I., Hanafi, R. S., Assal, R. A., Shaban, G. M., et al. (2020). The association of megalin and cubilin genetic variants with serum levels of 25-hydroxvitamin D and the incidence of acute coronary syndrome in Egyptians: a case control study. *J. Adv. Res.* 21, 49–56. doi: 10.1016/j.jare.2019.09.006

Eyles, D. W., Smith, S., Kinobe, R., Hewison, M., and McGrath, J. J. (2005). Distribution of the vitamin D receptor and 1α-hydroxylase in human brain. *J. Chem. Neuroanat.* 29, 21–30. doi: 10.1016/j.jchemneu.2004.08.006

Fischer, K. (2020). *Vitamin D", in Principles of Nutrigenetics and Nutrigenomics*. Amsterdam: Elsevier, 245–254.

Harding, B., Curley, A. J., Hannan, F. M., Christie, P. T., Bowl, M. R., Turner, J. J., et al. (2006). Functional characterization of calcium sensing receptor polymorphisms and absence of association with indices of calcium homeostasis and bone mineral density. *Clin. Endocrinol.* 65, 598–605. doi: 10.1111/j.1365-2265.2006.02634.x

Holick, M. F. (2004). Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. *Am. J. Clin. Nutr.* 80, 1678S–1688S.

Holick, M. F. (2007). Vitamin D deficiency. *N. Engl. J. Med.* 357, 266–281.

Holick, M. F., Binkley, N. C., Bischoff-Ferrari, H. A., Gordon, C. M., Hanley, D. A., Heaney, R. P., et al. (2011). Evaluation, treatment, and prevention of vitamin D deficiency: an Endocrine Society clinical practice guideline. *J. Clin. Endocrinol. Metab.* 96, 1911–1930. doi: 10.1210/jc.2011-0385

Hollis, B. W. (1996). Assessment of vitamin D nutritional and hormonal status: what to measure and how to do it. *Calcif. Tissue Int.* 58, 4–5. doi: 10.1007/s002239900002

Hyppönen, E., and Boucher, B. J. (2018). Adiposity, vitamin D requirements, and clinical implications for obesity-related metabolic abnormalities. *Nutr. Rev.* 76, 678–692. doi: 10.1093/nutrit/nuy034

Jolliffe, D. A., Walton, R. T., Griffiths, C. J., and Martineau, A. R. (2016). Single nucleotide polymorphisms in the vitamin D pathway associating with circulating concentrations of vitamin D metabolites and non-skeletal health outcomes: review of genetic association studies. *J. Steroid Biochem. Mol. Biol.* 164, 18–29. doi: 10.1016/j.jsbmb.2015.12.007

Kaseda, R., Hosojima, M., Sato, H., and Saito, A. (2011). Role of megalin and cubilin in the metabolism of vitamin D3. *Therap. Apheresis Dial.* 15, 14–17. doi: 10.1111/j.1744-9987.2011.00920.x

Kohlmeier, M. (2012). *Nutrigenetics: Applying the Science of Personal Nutrition*. Cambridge, MA: Academic Press.

Kohlmeier, M. (2015). *Nutrient Metabolism: Structures, Functions, and Genes*. Cambridge, MA: Academic Press.

McGrath, J. J., Saha, S., Burne, T. H. J., and Eyles, D. W. (2010). A systematic review of the association between common single nucleotide polymorphisms and 25-hydroxyvitamin D concentrations. *J. Steroid Biochem. Mol. Biol.* 121, 471–477. doi: 10.1016/j.jsbmb.2010.03.073

Nykjaer, A., Dragun, D., Walther, D., Vorum, H., Jacobsen, C., Herz, J., et al. (1999). An endocytic pathway essential for renal uptake and activation of the steroid 25-(OH) vitamin D3. *Cell* 96, 507–515. doi: 10.1016/s0092-8674(00)80655-8

Pike, J. W., and Meyer, M. B. (2012). The vitamin D receptor: new paradigms for the regulation of gene expression by 1, 25-dihydroxyvitamin D3. *Rheum. Dis. Clin.* 38, 13–27. doi: 10.1016/j.rdc.2012.03.004

Pytel, V., Matías-Guiu, J. A., Torre-Fuentes, L., Montero-Escribano, P., Maietta, P., Botet, J., et al. (2019). Exonic variants of genes related to the vitamin D signaling pathway in the families of familial multiple sclerosis using whole-exome next generation sequencing. *Brain Behav.* 9:e01272. doi: 10.1002/brb3.1272

Speeckaert, M., Huang, G., Delanghe, J. R., and Taes, Y. E. (2006). Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. *Clin. Chim. Acta* 372, 33–42. doi: 10.1016/j.cca.2006.03.011

Tsiaras, W. G., and Weinstock, M. A. (2011). Factors influencing vitamin D status. *Acta Derm. Venereol.* 91, 115–124. doi: 10.2340/00015555-0980

Wang, T. J., Zhang, F., Richards, J. B., Kestenbaum, B., Van Meurs, J. B., Berry, D., et al. (2010). Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* 376, 180–188.

Xu, X., Mao, J., Zhang, M., Liu, H., Li, H., Lei, H., et al. (2015). Vitamin D deficiency in Uygurs and Kazaks is associated with polymorphisms in CYP2R1 and DHCR7/NADSYN1 genes. *Med. Sci. Monit.* 21:1960. doi: 10.12659/msm.894793

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 29 November 2021
doi: 10.3389/fgene.2021.781400

Epigenetic Alteration and its Association With Downregulated FOXP3 Gene in Indian Breast Cancer Patients

*Sadaf[1], Naseem Akhter[1], Raed A. Alharbi[2], Abdulmajeed A. A. Sindi[2], Mohammad Zeeshan Najm[3], Fahad A. Alhumaydhi[4], Mohammad Aasif Khan[1], S.V.S Deo[5] and Syed Akhtar Husain[1]**

[1]Department of Biosciences, Jamia Millia Islamia, New Delhi, India, [2]Department of Laboratory Medicine, Faculty of Applied Medical Sciences, Albaha University, Albaha, Saudi Arabia, [3]School of Biosciences, Apeejay Stya University, Gurugram, India, [4]Department of Medical Laboratories, College of Applied Medical Sciences, Qassim University, Buraydah, Saudi Arabia, [5]Department of Surgical Oncology, All India Institute of Medical Sciences, New Delhi, India

OPEN ACCESS

Edited by:
Edivaldo Herculano Correa De Oliveira, Evandro Chagas Institute, Brazil

Reviewed by:
Jogendra Pawar, Purdue University, United States
Saif Ahmad, Barrow Neurological Institute (BNI), United States

***Correspondence:**
Syed Akhtar Husain
shusain@jmi.ac.in

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *22 September 2021*
Accepted: *26 October 2021*
Published: *29 November 2021*

Citation:
Sadaf, Akhter N, Alharbi RA, Sindi AAA, Najm MZ, Alhumaydhi FA, Khan MA, Deo SVS and Husain SA (2021) Epigenetic Alteration and its Association With Downregulated FOXP3 Gene in Indian Breast Cancer Patients. Front. Genet. 12:781400. doi: 10.3389/fgene.2021.781400

Background: *FOXP3* gene, known to be a potential tumor suppressor, has been identified to interact with HER2 in mammary cancer. Moreover, the high expression of *FOXP3* serves as a good predictor of the survival of patients in breast cancer, prostate cancer, and gastric cancer. The expression and epigenetic alterations were evaluated in female breast cancer patients.

Material and Methods: Expression studies at the mRNA level and protein level were conducted in 140 breast cancer cases by real-time PCR and immunohistochemistry, respectively. Epigenetic studies were also conducted by analyzing the methylation status at the promoter region of the gene using MS-PCR.

Results: *FOXP3* mRNA expression and protein expression were downregulated in breast cancer patients. The absence of *FOXP3* protein expression is significantly associated with promoter methylation, where 70 methylated cases exhibited protein loss (70/95, 73.6%). Statistically, we also found a significant correlation between *FOXP3* protein expression and TNM stage, promoter methylation, and histological grade. The methylated *FOXP3* cases that did not express protein were also significantly associated with positive lymph node metastasis and HER-2 status.

Conclusion: The expression profile of *FOXP3* may serve as a prognostic factor. In short, *FOXP3* may stand in the most crucial list of biomarkers for breast cancer, bringing compelling results in terms of treatment and management of the disease.

Keywords: PCR-real time, methylation specific PCR (MS PCR), immunohistochemistry (IHC), biomarkers, mammary cancer, gene experession

INTRODUCTION

It is a well-known fact that females around the world are mostly affected by breast cancer (1.7 million cases, 11.9%); however, it positions fifth as the cause of death (6.4%) because of the comparatively conducive prognosis. Whereas breast cancer accounted for the most prominent cause of mortality in females in undeveloped and developing regions of the world (Ferlay et al., 2015), the development of biomarkers and their clinical use in therapy for prediction and expected response hold a remarkable potential.

FOXP3(Forkhead box P3), located on Xp11.23, is a member of the Forkhead/winged-helix family of transcription factors which is responsible for X-linked autoimmune diseases in mice as well as humans (Wildin et al., 2001; Brunkow et al., 2001; Schubert et al., 2001). Transcription factor *FOXP3* regulates the development and function of Treg cells, and Treg cells are known to regulate homeostasis (Chen et al., 2015) and immunosuppression (Müller et al., 2010) and also recognized as the most peculiar marker for Treg (Hori et al., 2003). Before the advanced research on *FOXP3* expression, it was thought to be expressed only in hematopoietic cells but now seems to be present in human tumors, particularly tumors of the breast (Merlo et al., 2009). In mammary cancers, *FOXP3* is found to regulate *HER-2* and *SKP2* by repressing their expression, and importantly these genes are linked to a poor prognosis in the cases with breast carcinoma (Martin et al., 2010). The downregulation and many functional somatic mutations in the *FOXP3* gene were usually found in human breast cancer samples. These mutations may also account for the overall down-regulation at the protein level (Karanikas et al., 2008).

FOXP3 halts the transcription of *HER-2* by attaching to the promoter region of the *ERBB2* gene (Redpath et al., 2011; Zuo et al., 2007), and it is a well-known fact that *HER-2* is a potent marker in terms of prediction and effective therapy (Presson et al., 2011; Weigelt and Reis-Filho, 2010). The study has also pointed out the high expression of *FOXP3* as a good predictor of the survival of a patient in prostate cancer, breast cancer, gastric cancer, and bladder cancer (Li et al., 2007; Wang et al., 2009; Winerdal et al., 2011; He et al., 2013; Fiori Lopes et al., 2014; Hao et al., 2014; Ma et al., 2014). Previous studies also reported that many SNPs in the *FOXP3* gene had been associated with breast cancer (Jiang and Ruan, 2014).

FOXP3 is also linked to *p21* and *LATS2*, where it is involved in transcriptional control (Li et al., 2011). Due to these captivating characteristics of *FOXP3*, the present work examines the correlation of *FOXP3* protein expression with the clinicopathological variables to strengthen its role as a putative biomarker for breast carcinoma in the Indian population.

METHODOLOGY

Ethical Statement

The University Ethical Committee of Jamia Millia Islamia (JMI), New Delhi, and the Ethical Committee for Human Study of AIIMS (All India Institute of Medical Sciences), New Delhi, have officially approved the study. The experimental work had been undertaken with written consent obtained from each subject, and the study complies with the rules and standards set by the Ethics Code of the Medical Association of the world, which have been noted as per the Declaration of Helsinki as published in British Medical Journal (1964).

Sample Collection

A total of 140 participants were included in the present case–control study. Cancer tissue from the breast and non-cancerous adjacent tissue were both obtained from the surgical oncology department of the collaborating institute (AIIMS). The samples were collected in three vials containing RNALater, phosphate-buffered saline (PBS), and formalin, respectively, for further processing.

The classification of breast cancer stages was done under the TNM staging system, and the histological grading of tumors was classified based on the Nottingham grade system. The exclusion criteria in the current study included familial cancer, any previous type of cancer, other metastasized cancer that has spread from different organs, and chemotherapy and radiotherapy exposure. In addition, included were various clinicopathological variables such as tumor distinctiveness [age, tumor size, metastasis at the lymph node level, TNM staging, grade of tumor, molecular subtype of tumor, hormonal receptor status (ER, PR, and Her2neu), and reproductive history (menopausal status parity)].

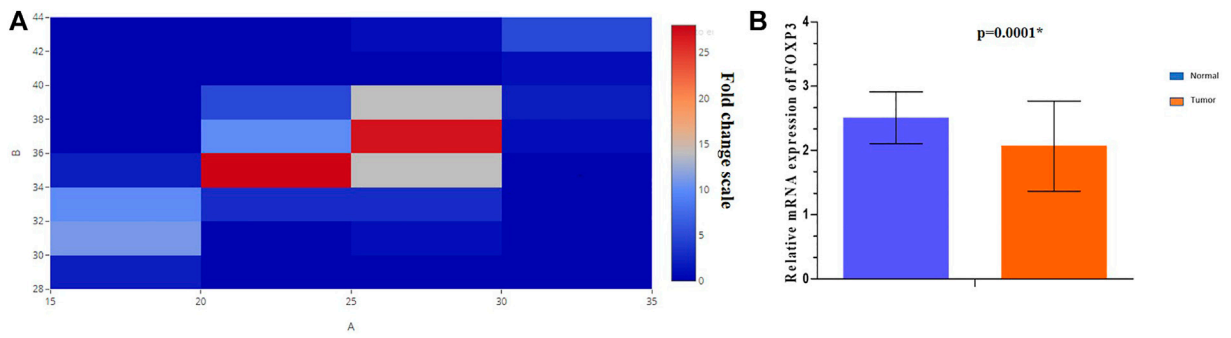

FIGURE 1 | (A) Heat map plot showing the mRNA relative expression of the *FOXP3* gene (fold change) in breast cancer patients. **(B)** Graphical representation of the relative expression of *FOXP3/GAPDH* mRNA in breast tumor and adjoining non-tumorous tissues.

TABLE 1 | Correlation study of *FOXP3* mRNA expression with clinicopathological parameters in North Indian breast cancer patients.

Characteristics	Total (N)	*FOXP3* mRNA expression relative to GAPDH (mean ± S.E)	*p*-value
Tissue[a]			
Normal	89	2.51 ± 0.4	<**0.0001***
Tumor	89	2.07 ± 0.7	
Age			
<50	50 (35.71)	2.40 ± 0.60	0.716
≥50	90 (64.29)	2.50 ± 0.40	
Menopausal status			
Premenopausal	40 (28.57)	3.07 ± 0.80	0.698
Postmenopausal	100 (71.43)	2.23 ± 0.40	
Estrogen receptor status			
Negative	37 (26.42)	2.79 ± 0.80	0.691
Positive	103 (73.58)	2.36 ± 0.42	
Progesterone receptor status			
Negative	67 (47.85)	2.36 ± 0.51	0.862
Positive	73 (52.15)	2.58 ± 0.58	
Her2 neu status			
Negative	70 (50)	2.65 ± 0.55	0.726
Positive	70 (50)	2.30 ± 0.55	
Tumor size			
<5	64 (45.71)	1.50 ± 0.35	0.482
≥5	76 (54.29)	3.29 ± 0.64	
Lymph node status			
Positive	119 (85)	2.68 ± 0.45	0.811
Negative	21 (15)	1.37 ± 0.40	
TNM staging			
Stage (I + II)	45 (32.1)	2.87 ± 0.73	0.192
Stage (III + IV)	95 (67.9)	2.28 ± 0.46	
Histological grade			
(I + II)	120 (85.7)	2.47 ± 0.42	0.803
(III)	20 (14.3)	2.51 ± 1.02	
Molecular subtypes			
Luminal A	51 (36.44)	2.13 ± 0.56	0.193
Luminal B	53 (37.86)	2.58 ± 0.64	
Her2neu enriched	18 (12.85)	1.48 ± 1.04	
TNBC	18 (12.85)	4.24 ± 1.44	

TNBC: triple negative breast cancer, FOXP3: Forkhead Box P3.
[a]*Only downregulated cases were included.*

Quantitative- PCR

RNALater (Qiagen) was used to store excised tissues from normal and breast cancer patients, and then RNA was extracted by using the TRIzol method as per the manual. cDNA was synthesized using a Thermo Fisher verso kit from the extracted RNA. The qPCR is processed using Roche LightCycler® 96 machine with SYBR Green I Master mix reagent (Roche) with the help of *FOXP3* primers (Fwd- 5′-TCCCAGAGTTCCTCCACAAC-3′ and Rev-5′ATTGAGTGTCCGCTGCTTCT-3′) that give an amplified 122-bp product. The internal control used was GAPDH gene which was also amplified in the same PCR reactions. The program used for the amplification cycles was as follows: preheating at 95°C for 1 min, 30 cycles of denaturation at 95°C for 20 s, annealing at 58°C for 15 s, extension at 72°C for 15 s, and further elongation at 72°C for 7 min. The experiments were repeated thrice.

The relative quantification of expression was calculated as the calibrator normalized ratio using LightCycler 96 (Roche) Software 1.5. The formula used, $RQ = 2^{-\Delta\Delta C}t$, was according to MIQE.

Genomic DNA Extraction

The phenol-chloroform extraction method was used to isolate high-molecular-weight total gDNA from both tumor and normal tissues stored in PBS. The genomic DNA isolated was quantified on a Nanodrop spectrophotometer, and its quality was also assessed using an A260/280 ratio. It was further visualized on the 1% agarose gel stained with ethidium bromide under a UV transilluminator.

MS-PCR for Epigenetic Analysis

Isolated gDNA from the tissues were given bisulfite treatment using Zymo research EZ DNA Methylation-Gold™ Kit per the instructions. The treated gDNA was amplified using two sets of methylated and unmethylated primers for the FOXP3 promoter. MethPrimer tool was used to design the set of primers for methylation and unmethylation. (Li and Dahiya, 2002). The methylated primer pairs for the promoter region of the *FOXP3* gene were: forward 5′- TGTAGGGGGTGT AGAATTTTTTTC-3′ and reverse 5′- AAACTAAATTCCCAA AAACCTCG-3′ and for the unmethylated were forward 5′- GTA GGGGGTGTAGAATTTTTTTTGT-3 and reverse 5′- TAAAAC TAAATTCCCAAAAACCTCA-3′. The positive controls used in the experiment were commercially available Completely methylated and unmethylated human genomic DNA.

MS-PCR was performed under the following cycles: First denaturation at 95°C for 7 min, denaturation at 95°C for 30s 52.5°C annealing for both types of primers for 30 s, 72°C for 30 s and final elongation at 72°C for 7 min which was amplified for 35 cycles. Then, 2% agarose gel stained with EtBr was used to visualize and analyze the PCR product, which was finally photographed using the Bio-Rad Gel Documentation system. All experiments were conducted in triplicates.

Immunohistochemistry

The tissue biopsies of the breast carcinoma and adjoining non-cancerous tissue were conserved in formalin-fixed blocks. The blocks were then sectioned and engraved on slides coated with poly-L-lys and exposed to deparaffinization by dipping in different concentrations of xylene and further rehydrated with grades of ethanol. By using 0.3% H_2O_2 for 30 min, the endogenous peroxidase activity was blocked, and sodium citrate buffer (pH 9.0) was used for Ag retrieval.

The sections were treated for 30 min with TENG-T (10 mM Tris; 5 mM ethylenediaminetetraacetic acid, 0.15 mol/L NaCl, 0.25% gelatin, and 0.05% (v/v) Tween20 (pH 8.0) to block the samples. Bovine serum albumin was used to limit unspecific binding to the protein. The slides were treated with the primary antibody (mAbCam#ab22510 *FOXP3* 1:50) and incubated overnight at 4°C. The slides were then treated for 30 min with a biotinylated anti-mouse secondary antibody along with streptavidin–horseradish peroxidase conjugate. The DAB chromogen was finally used as a substrate to give a brown-colored precipitate. The slides were also treated with hematoxylin as a counterstain for better contrast.

Histopathologists interpreted the slides after immunohistochemistry; the slides were photographed under a light microscope at ×400 magnification. The pathologist further graded the expression on the number scale 0–4, with 0 as no expression and 4 as highest expression; the slides with >50% protein staining were considered in the highest scale.

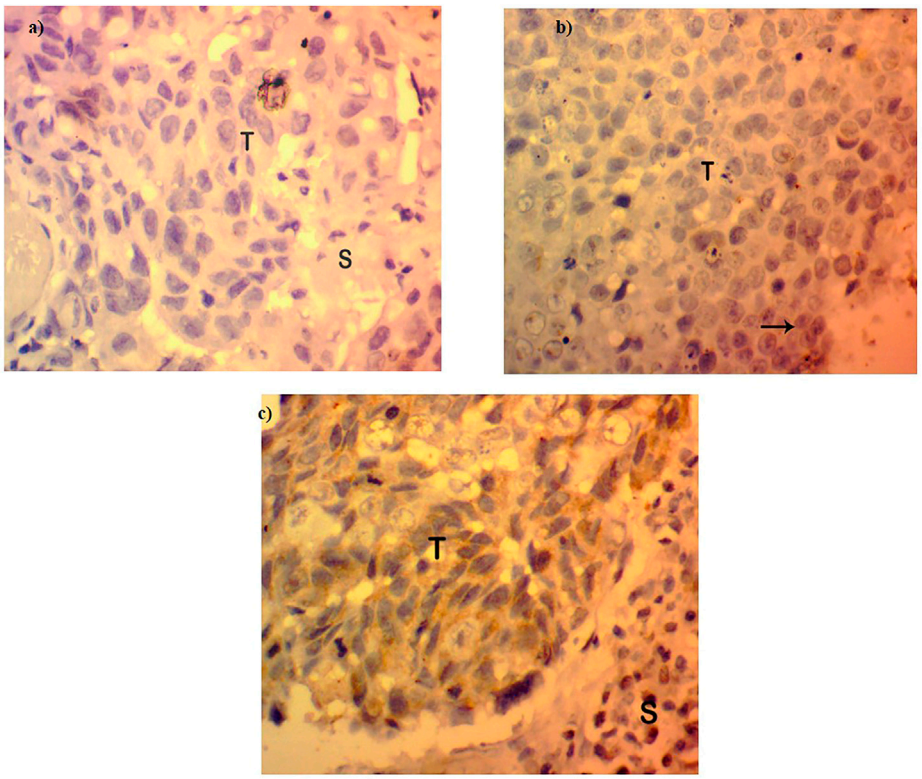

FIGURE 2 | Protein expression in *FOXP3* by immunohistochemistry. **(A)** Absence or loss of *FOXP3* expression in Breast cancer tissues. **(B)** *FOXP3* low expression in breast tumor tissues. **(C)** *FOXP3* moderate/high expression in breast tumor tissues.

Statistics

SPSS version 22.0 for Windows was used for all the statistical correlations between the outcomes and the clinicopathological parameters. All data were expressed as mean ± standard error. Fisher's exact test was used to obtain *p*-values between mRNA levels, methylation status, and protein expression with the clinicopathological parameters. Non-parametric Wilcoxon signed-rank test is used to estimate *FOXP3*/*GAPDH* mRNA expression levels significantly in both cancer and normal tissue samples. The *p*-values >0.05 were considered as significant.

RESULTS

mRNA Expression of the FOXP3 Gene Is Downregulated in the Cases of Breast Cancer

FOXP3 mRNA expression revealed downregulation in 63.5% (89/140) of cases (**Figure 1A**), out of which nearly 86.5% (77/89) of the cases registered in the study were linked to histological grade type 1 and type 2. As per the fold change analysis, 89 cases out of 140 samples seemed to be downregulated (5.09-fold), while the expression pattern of *FOXP3* at the level of mRNA, when normalized accordingly with the internal control *GAPDH* in tumor and non-tumor tissues, was 2.07 ± 0.7 (mean ± standard error) and 2.51 ± 0.4 (mean ± standard error) (**Figure 1B**), (*p*-value of <0.0001; **Table 1**) respectively. However, there was no significant association observed between *FOXP3* mRNA level and various clinical variables.

Expression of FOXP3 Protein is Either Lost or Low in Breast Cancer

The expression of *FOXP3* at the protein level was found to be either low or absent in 95 cases of the total 140 samples involved in the study (67.85%) (**Figure 2**), while in the other 45 cases, the expression pattern was either in the high or moderate range as interpreted by a histopathologist on the basis expression scale (45/140, 32.14%). The protein expression pattern was in relation to the mRNA expression. The *FOXP3* protein, as visualized by immunohistochemistry, was mainly located in the nuclear region.

FOXP3 Protein Expression and Its Significant Correlation With Clinicopathological Parameters

As revealed by immunohistochemistry staining, the majority of *FOXP3* proteins in the samples were found to be significantly

TABLE 2 | Statistical analysis of the protein expression pattern of *FOXP3* with the clinicopathological variables of patients with breast cancer.

Characteristics	Total cases (N)	*FOXP3* absent	*FOXP3* present	*p*-value
Age				
<50	50	36 (72)	14 (28)	0.457
≥50	90	59 (65.5)	31 (34.5)	
Menopausal status				
Premenopausal	40	27 (67.5)	13 (32.5)	1.0
Postmenopausal	100	68 (68)	32 (32)	
Estrogen receptor status				
Negative	37	26 (70.2)	11 (29.8)	0.838
Positive	103	69 (66.9)	34 (33.1)	
Progesterone receptor status				
Negative	67	43 (64.2)	24 (35.8)	0.469
Positive	73	52 (71.2)	21 (28.8)	
Her2 neu status				
Negative	70	44 (62.8)	26 (37.2)	0.278
Positive	70	51 (72.8)	19 (27.2)	
Tumor size				
<5	64	46 (71.8)	18 (28.1)	0.37
≥5	76	49 (64.4)	27 (35.6)	
Lymph node status				
Positive	119	80 (67.2)	39 (32.8)	0.804
Negative	21	15 (71.5)	6 (28.5)	
TNM staging				
Stage (I + II)	45	25 (55.5)	20 (44.5)	0.035*
Stage (III + IV)	95	70 (73.6)	25 (26.4)	
Histological grade				
(I + II)	120	82 (68.3)	38 (31.7)	0.70
(III)	20	13 (65)	7 (35)	
Molecular subtypes				
Luminal A	51	34 (66.6)	17 (33.4)	0.093
Luminal B	53	36 (67.9)	17 (32.1)	
Her2neu Enr	18	16 (88.8)	2 (11.2)	
TNBC	18	9 (50)	9 (50)	

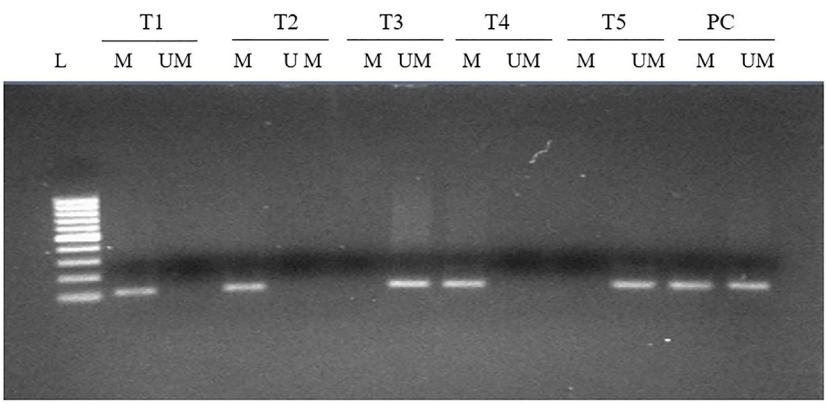

FIGURE 3 | MS-PCR analysis of the promoter region of *FOXP3* in breast carcinoma. Promoter methylation was evaluated by two types of primers to amplify either unmethylated gDNA (UM) or methylated gDNA (M). L, 100-bp DNA ladder; T, tumor tissue; PC, positive control for completely methylated and unmethylated DNA.

downregulated. Moreover, when we tried to statistically associate the protein expression with the clinical parameters of the patients, we observed a significant correlation between TNM staging and *FOXP3* protein expression ($p < 0.035$) (**Table 2**). However, with other parameters, no significant association was obtained statistically (**Table 2**), though most cases with tumor grades 1 and 2 seem to have protein loss (82/95, 86.3%).

FOXP3 Promoter Methylation and Its Association With Clinical Variables

Methylation at various CpG present in the upstream promoter area of *FOXP3* gene was observed in 73 cases (73/140, 52.14%) (**Figure 3**) and, once linked with clinical parameters, revealed a significant association with the Nottingham histological grades 1 and 2 type tumors of breast cancer patients

TABLE 3 | Statistical association study of methylated *FOXP3* gene with clinicopathological parameters of patients having breast carcinoma.

Characteristics	Total cases (N)	Methylated	Unmethylated	*p*-value
Age				
<50	50	30 (60)	20 (40)	0.217
≥50	90	43 (47.7)	47 (52.3)	
Menopausal status				
Premenopausal	40	22 (55)	18 (45)	0.711
Postmenopausal	100	51 (51)	49 (49)	
Estrogen receptor status				
Negative	37	21 (56.7)	16 (43.3)	0.568
Positive	103	52 (50.4)	51 (49.6)	
Progesterone receptor status				
Negative	67	34 (50.7)	33 (49.3)	0.866
Positive	73	39 (53.4)	34 (46.6)	
Her2 neu status				
Negative	70	37 (52.8)	33 (47.2)	1.0
Positive	70	36 (51.4)	34 (48.6)	
Tumor size				
<5	64	35 (54.6)	29 (45.4)	0.613
≥5	76	38 (50)	38 (50)	
Lymph node status				
Positive	119	59 (49.5)	60 (50.5)	0.164
Negative	21	14 (66.6)	7 (33.4)	
TNM staging				
Stage (I + II)	45	25 (55.5)	20 (44.5)	0.50
Stage (III + IV)	95	48 (50.5)	47 (49.5)	
Histological grade				
(I + II)	120	58 (48.3)	62 (51.7)	0.031*
(III)	20	15 (75)	5 (25)	
Molecular subtypes				
Luminal A	51	29 (56.8)	22 (43.2)	0.136
Luminal B	53	24 (45.2)	29 (54.8)	
Her2neu Enr	18	13 (72.2)	5 (27.8)	
TNBC	18	7 (38.8)	11 (61.2)	

(0.031). Though no significant associations were seen with other parameters included in the study (**Table 3**), we did note a significantly higher number of methylated cases in metastatic lymph node (59/73, 80.8%), estrogen receptor-positive (52/73, 71.2%), and menopausal female patients (51/73, 69.8%) (**Table 3**).

Convincing Association of FOXP3 Protein Expression With FOXP3 Promoter Methylation in the Cases of Breast Carcinoma

The correlation study of methylated *FOXP3* gene and its respective protein expression displayed the significant link, in which out of 95 cases with protein loss, 70 cases possessed methylation at the promoter region (70/95, 73.68%) (**Table 4**), whereas 25 cases were completely unmethylated (25/95, 26.3%) (**Table 4**). To add more, in 67 unmethylated samples (67/140, 47.8%), noticeably 62.6% (42/67) cases showed the presence of protein. Therefore, a potential statistical relation was observed between the *FOXP3* protein expression and its promoter methylation ($p = 0.0001$) (**Table A4**).

Association of Methylated FOXP3 Gene Exhibiting Loss of Protein With Numerous Clinicopathological Variables

The methylated promoter region of *FOXP3* cases demonstrating either the absence or presence of protein exhibited a statistically significant relation with Her 2 neu receptor ($p = 0.004$) and metastatic lymph node tumors ($p = 0.01$) (**Table 5**). Moreover, 95.7% of methylated cases (67/70) with lymph node metastasis have protein loss. (**Table 5**). Additionally, the cases having protein loss exhibiting either a methylated or unmethylated *FOXP3* promoter region shows a convincing association with positive Her-2 receptor ($p = 0.03$) and tumors of grades 1 and 2 ($p = 0.01$) (**Table 5**). Furthermore, it is seen that there is a strong statistical relation between *FOXP3* protein loss and the promoter methylation with the various clinicopathological parameters (**Table 6**), where most of the features were associated in a highly significant manner (p-value < 0.05).

DISCUSSION

FOXP3 expression is identified in tumors of the breast, prostate, lung, gastric, and thyroid (Liu et al., 2015; Yang et al., 2017; Ma et al., 2014; the Chu et al., 2015) suggesting its crucial role in the biology of cancer. The previous study demonstrated the inverse correlation between breast cancer angiogenesis and nuclear *FOXP3* expression. Adding to the same observation, the significant downregulation of *FOXP3* also resulted in the reduced survival in breast cancer (Li et al., 2018) *FOXP3* has been reported to modulate the expression of various genes involved in the process of carcinogenesis to exert its suppressing role in tumor development (Szylberg et al., 2016). At the same time, we cannot forget that significant studies have suggested the positive association between *FOXP3* expression and better survival in patients and the tumor-suppressive role of the *FOXP3* gene in breast cancer. Therefore, the present work investigated the *FOXP3* expression pattern and its correlation with various clinicopathological variables to strengthen its prognostic value and tumor-suppressive property. An earlier study demonstrated a quantitative method to assess the methylation status of *FOXP3* to understand the role of Treg cells in immunomodulation (Wieczorek et al., 2009). In our study transcription factor, *FOXP3* promoter methylation and expression were studied and analyzed in breast cancer patients of the northern region of India using methylation-specific PCR, real-time PCR, and immunohistochemistry to assess its role as a potential biomarker. The study correlated the findings with the clinicopathological variables (age, histological grade, ER status, HER2 status, *etc.*) of the procured cases.

In our study, at the mRNA level, nearly 63.5% (89/140) of cases were found to be downregulated (5.09 fold), and interestingly 86.5% (77/89) were linked with the histological grades I and II, suggesting the possible role of *FOXP3* in the early development of the disease. The study is supported by the previous studies on *FOXP3* expression at the transcript level (Zhang and Sun, 2010; Hinz et al., 2007). Furthermore, apart

TABLE 4 | Correlation study between FOXP3 protein expression and its promoter methylation in patients with breast cancer.

FOXP3 promoter	*FOXP3* protein expression		Total (%)	*p* value	OR value (95% CI)
—	Absent	Present	—	0.0001*	0.039 (0.011-0.13)
Methylated	70	3	73 (52.1)		
Unmethylated	25	42	67 (47.9)		
Total	95 (67.8)	45 (32.2)	140		

OR, odds ratio; CI, confidence interval; p-*value, Fischer's exact test*

TAble 5 | Significant association of promoter methylation and its protein expression in patients with methylated *FOXP3* promoter or *FOXP3* Protein expression loss with various clinicopathological features of breast carcinoma.

Clinical characteristics	Total (N)	Methylated *FOXP3*		*p*-value	Total (N)	*FOXP3* loss		*p*-value
		FOXP3 absent	*FOXP3* present			Methylated *FOXP3FOXP3*	Unmethylated	
Age								
Age < 50	30	29	1	1.0	36	29	7	0.3
Age ≥ 50	43	41	2	—	59	41	18	—
Menopausal status								
Premenopausal	22	22	0	0.54	27	22	5	0.3
Postmenopausal	51	48	3	—	68	48	20	—
ER status								
Negative	21	21	0	0.55	26	21	5	0.4
Positive	52	49	3	—	69	49	20	—
PR status								
Negative	34	33	1	1.0	46	33	10	0.6
Positive	39	37	2	—	49	37	15	—
HER2 status								
Negative	13	10	3	0.004*	19	10	9	0.03*
Positive	60	60	0	—	76	60	16	—
Tumor size								
<5	35	33	2	0.6	46	33	13	0.8
≥5	38	37	1	—	49	37	12	—
Lymph node status								
Positive	68	67	1	0.01*	90	67	23	0.6
Negative	5	3	2	—	5	3	2	—
Clinical stage								
Stage (1 + 2)	22	20	2	0.20	25	20	5	0.5
Stage (3 + 4)	51	50	1	—	70	50	20	—
Histological stage/grade								
Stage (1 + 2)	62	60	2	0.39	75	60	15	**0.01***
Stage (3)	11	10	1	—	20	10	10	—
Molecular subtypes								
Lum A	29	27	2	—	34	27	7	—
Lum B	24	23	1	0.6	36	23	13	0.4
HER2 Neu	13	13	0	—	16	13	3	—
TNBC	7	7	0	—	9	7	2	—

from breast cancer, one of the studies pointed out that the upregulation of *FOXP3* in gastric cancer cells put a brake on GC cell growth in both *in vivo* and *in vitro* studies (Hao et al., 2014), unraveling the crucial role of *FOXP3* expression in different carcinomas.

The protein expression profile exhibited low or no expression in nearly 69% (95/140) of breast cancer cases, followed by either moderate or high expression in 32% (45/140) of the cases. The expression was either cytoplasmic or nuclear, which was demonstrated in different types of cancer (Merlo et al., 2009; Karanikas et al., 2008; Zhang and Sun, 2010; Hinz et al., 2007; Ladoire et al., 2011; Tao et al., 2012). We did find a significant association between *FOXP3* protein expression and the TNM stage (*p*-value, 0.035). Interestingly almost 74% (95/140) of the cases of stage (III and IV) harbored protein loss. Because of the above mentioned condition, it has been observed earlier that, in the most aggressive cancer of epithelial tissues, *FOXP3* may help in the suppression of cancer as these aggressive cancer tissues harbored very low or no expression of *FOXP3* at the transcript and protein levels (Wang et al., 2009; Jiang and Ruan, 2014; Li et al., 2011).

The promoter methylation of *FOXP3* was observed in 52% of the cases (73/140) and significantly associated

TABLE 6 | Association study between methylated *FOXP3* and FOXP3 protein expression in stratification with clinicopathological features.

Clinical characteristics	Total (N)	*FOXP3* methylation status	FOXP3 expression		*p*-value
			Absent	Present	
Age					
Age < 50	30	M	29	1	0.0001*
		U	7	13	
Age ≥ 50	43	M	41	2	0.0001*
		U	18	29	
Menopausal status					
Premenopausal	22	M	22	0	0.0001*
		U	5	13	
Postmenopausal	51	M	48	3	0.0001*
		U	20	29	
ER status					
Negative	21	M	21	0	0.0001*
		U	5	11	
Positive	52	M	49	3	0.0001*
		U	20	31	
PR status					
Negative	34	M	33	1	0.0001*
		U	10	23	
Positive	39	M	37	2	0.001*
		U	15	19	
HER2 status					
Negative	37	M	35	2	0.0001*
		U	9	24	
Positive	36	M	35	1	0.0001*
		U	16	18	
Tumor size					
<5	35	M	33	2	0.0001*
		U	13	16	
≥5	38	M	37	1	0.0001*
		U	12	26	
Lymph node status					
Positive	59	M	57	2	0.0001*
		U	23	37	
Negative	14	M	13	1	0.005
		U	2	5	
Clinical stage					
Stage (1 + 2)	22	M	20	2	0.07
		U	15	8	
Stage (3 + 4)	51	M	50	1	0.0001*
		U	10	34	
Histological stage/grade					
Stage (1 + 2)	62	M	60	2	0.0001*
		U	20	38	
Stage (3)	11	M	10	1	0.12
		U	5	4	
Molecular subtypes					
Lum A	29	M	27	2	0.0001*
		U	7	15	
Lum B	24	M	23	1	0.0001*
		U	13	16	
Her2 neu	13	M	13	0	0.06
		U	3	2	
TNBC	7	M	7	0	0.002
		U	2	9	

with Nottingham histological grades 1 and 2 (*p*-value 0.031). The finding depicts a strong association between promoter hyper-methylation and mRNA expression in deactivating or down-regulating the possible role of *FOXP3* in the suppression of breast cancer (Esteller, 2007; Li et al., 2007).

While analyzing an association between protein loss and hypermethylated promoter cases, we found a compelling association as out of 95 protein loss cases, 70 cases possessed methylation at the promoter region (70/95, 73.68%), whereas in a total of 67 unmethylated cases, 62.6% (42/67) exhibited the presence of protein. More intriguing results came out while

analyzing methylation and protein loss with each other, as 95.7% methylated cases (67/70) with lymph node metastasis displayed protein loss. The findings are strongly supported by previous studies that have shown the association to be significantly lower in tumor grade and lymph node involvement in the breast tumor cells with positive *FOXP3* expression (Ladoire et al., 2012). Thus, the loss in *FOXP3* due to epigenetic change like methylation may serve as a potential biomarker in cases with lymph node metastasis.

In summary, our data provide some intriguing findings of *FOXP3* expression and its association with different clinicopathological parameters. However, the present study on a smaller sample size may weaken the statistical power. Therefore, further investigation on different sets of the population with a larger sample size is required to establish *FOXP3* as a potential cancer biomarker for diagnostic and prognostics purposes.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

ETHICS STATEMENT

This study was in agreement with the ethical standards of the host institute Jamia Millia Islamia University and AIIMS, New Delhi, India, and also as per the guidelines of the Helsinki Declaration in 1964 and its amendments. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Sa and SH took charge of the experimental design and execution. The experiments were executed by Sa. The reagents, materials, and analysis tools were contributed by Sa, MN, MK, and SD. Manuscript preparation was carried out by Sa and co-authors.

ACKNOWLEDGMENTS

UGC-MANF for Higher Studies (India).

REFERENCES

Brunkow, M. E., Jeffery, E. W., Hjerrild, K. A., Paeper, B., Clark, L. B., Yasayko, S.-A., et al. (2001). Disruption of a New Forkhead/winged-helix Protein, Scurfin, Results in the Fatal Lymphoproliferative Disorder of the Scurfy Mouse. *Nat. Genet.* 27 (1), 68–73. doi:10.1038/83784

Chen, Y., Chen, C., Zhang, Z., Liu, C.-C., Johnson, M. E., Espinoza, C. A., et al. (2015). DNA Binding by FOXP3 Domain-Swapped Dimer Suggests Mechanisms of Long-Range Chromosomal Interactions. *Nucleic Acids Res.* 43 (2), 1268–1282. doi:10.1093/nar/gku1373

Chu, R., Liu, S. Y. W., Vlantis, A. C., van Hasselt, C. A., Ng, E. K. W., Fan, M. D., et al. (2015). Inhibition of Foxp3 in Cancer Cells Induces Apoptosis of Thyroid Cancer Cells. *Mol. Cell. Endocrinol.* 399, 228–234. doi:10.1016/j.mce.2014.10.006

Esteller, M. (2007). Epigenetic Gene Silencing in Cancer: the DNA Hypermethylome. *Hum. Mol. Genet.* 16 (R1), R50–R59. doi:10.1093/hmg/ddm018

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., et al. (2015). Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012. *Int. J. Cancer* 136 (5), E359–E386. doi:10.1002/ijc.29210

Fiori Lopes, L., Losi Guembarovski, R., Guembarovski, A. L., Okuyama Kishima, M., Campos, C. Z., Oda, J. M. M., et al. (2014). FOXP3 Transcription Factor: a Candidate Marker for Susceptibility and Prognosis in Triple Negative Breast Cancer. *Biomed. Research International* 2014, 1–7. doi:10.1155/2014/341654

Hao, Q., Zhang, C., Gao, Y., Wang, S., Li, J., Li, M., et al. (2014). FOXP3 Inhibits NF-Kb Activity and Hence COX2 Expression in Gastric Cancer Cells. *Cell Signal.* 26 (3), 564–569. doi:10.1016/j.cellsig.2013.11.030

He, Y.-Q., Bo, Q., Yong, W., Qiu, Z.-X., Li, Y.-L., and Li, W.-M. (2013). FoxP3 Genetic Variants and Risk of Non-small Cell Lung Cancer in the Chinese Han Population. *Gene* 531 (2), 422–425. doi:10.1016/j.gene.2013.08.066

Hinz, S., Pagerols-Raluy, L., Oberg, H.-H., Ammerpohl, O., Grüssel, S., Sipos, B., et al. (2007). Foxp3 Expression in Pancreatic Carcinoma Cells as a Novel Mechanism of Immune Evasion in Cancer. *Cancer Res.* 67 (17), 8344–8350. doi:10.1158/0008-5472.CAN-06-3304

Hori, S., Nomura, T., and Sakaguchi, S. (2003). Control of Regulatory T Cell Development by the Transcription Factor Foxp3. *Science* 299 (5609), 1057–1061. doi:10.1126/science.1079490

Jiang, L.-L., and Ruan, L.-W. (2014). Association between FOXP3 Promoter Polymorphisms and Cancer Risk: A Meta-Analysis. *Oncol. Lett.* 8 (6), 2795–2799. doi:10.3892/ol.2014.2585

Karanikas, V., Speletas, M., Zamanakou, M., Kalala, F., Loules, G., Kerenidi, T., et al. (2008). Foxp3 Expression in Human Cancer Cells. *J. Transl Med.* 6 (1), 1–8. doi:10.1186/1479-5876-6-19

Ladoire, S., Arnould, L., Mignot, G., Coudert, B., Rébé, C., Chalmin, F., et al. (2011). Presence of Foxp3 Expression in Tumor Cells Predicts Better Survival in HER2-Overexpressing Breast Cancer Patients Treated with Neoadjuvant Chemotherapy. *Breast Cancer Res. Treat.* 125 (1), 65–72. doi:10.1007/s10549-010-0831-1

Ladoire, S., Mignot, G., Dalban, C., Chevriaux, A., Arnould, L., Rébé, C., et al. (2012). FOXP3 Expression in Cancer Cells and Anthracyclines Efficacy in Patients with Primary Breast Cancer Treated with Adjuvant Chemotherapy in the Phase III UNICANCER-PACS 01 Trial. *Ann. Oncol.* 23 (10), 2552–2561. doi:10.1093/annonc/mds028

Li, B., Samanta, A., Song, X., Iacono, K. T., Bembas, K., Tao, R., et al. (2007). FOXP3 Interactions with Histone Acetyltransferase and Class II Histone Deacetylases Are Required for Repression. *Proc. Natl. Acad. Sci.* 104 (11), 4571–4576. doi:10.1073/pnas.0700298104

Li, L.-C., and Dahiya, R. (2002). MethPrimer: Designing Primers for Methylation PCRs. *Bioinformatics* 18 (11), 1427–1431. doi:10.1093/bioinformatics/18.11.1427

Li, W., Wang, L., Katoh, H., Liu, R., Zheng, P., and Liu, Y. (2011). Identification of a Tumor Suppressor Relay between the FOXP3 and the Hippo Pathways in Breast and Prostate Cancers. *Cancer Res.* 71 (6), 2162–2171. doi:10.1158/0008-5472.CAN-10-3268

Li, X., Gao, Y., Li, J., Zhang, K., Han, J., Li, W., et al. (2018). FOXP3 Inhibits Angiogenesis by Downregulating VEGF in Breast Cancer. *Cell Death Dis* 9 (7), 1–12. doi:10.1038/s41419-018-0790-8

Liu, R., Liu, C., Chen, D., Yang, W.-H., Liu, X., Liu, C.-G., et al. (2015). FOXP3 Controls an miR-146/nf-Kb Negative Feedback Loop that Inhibits Apoptosis in Breast Cancer Cells. *Cancer Res.* 75 (8), 1703–1713. doi:10.1158/0008-5472.CAN-14-2108

Ma, G.-F., Miao, Q., Liu, Y.-M., Gao, H., Lian, J.-J., Wang, Y.-N., et al. (2014). High FoxP3 Expression in Tumour Cells Predicts Better Survival in Gastric Cancer and its Role in Tumour Microenvironment. *Br. J. Cancer* 110 (6), 1552–1560. doi:10.1038/bjc.2014.47

Martin, F., Ladoire, S., Mignot, G., Apetoh, L., and Ghiringhelli, F. (2010). Human FOXP3 and Cancer. *Oncogene* 29, 4121–4129. doi:10.1038/onc.2010.174

Merlo, A., Casalini, P., Carcangiu, M. L., Malventano, C., Triulzi, T., Mènard, S., et al. (2009). FOXP3 Expression and Overall Survival in Breast Cancer. *Jco* 27 (11), 1746–1752. doi:10.1200/JCO.2008.17.9036

Müller, S., Poehnert, D., Müller, J. A., Scheumann, G. W. F., Koch, M., and Lück, R. (2010). Regulatory T Cells in Peripheral Blood, Lymph Node, and Thyroid Tissue in Patients with Medullary Thyroid Carcinoma. *World J. Surg.* 34 (7), 1481–1487. doi:10.1007/s00268-010-0484-6

Presson, A. P., Yoon, N. K., Bagryanova, L., Mah, V., Alavi, M., Maresh, E. L., et al. (2011). Protein Expression Based Multimarker Analysis of Breast Cancer Samples. *BMC cancer* 11 (1), 1–14. doi:10.1186/1471-2407-11-230

Redpath, M., Xu, B., van Kempen, L. C., and Spatz, A. (2011). The Dual Role of the X-Linked FoxP3 Gene in Human Cancers. *Mol. Oncol.* 5 (2), 156–163. doi:10.1016/j.molonc.2011.03.001

Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F., and Ziegler, S. F. (2001). Scurfin (FOXP3) Acts as a Repressor of Transcription and Regulates T Cell Activation. *J. Biol. Chem.* 276 (40), 37672–37679. doi:10.1074/jbc.M104521200

Szylberg, Ł., Karbownik, D., and Marszałek, A. (2016). The Role of FOXP3 in Human Cancers. *Anticancer Res.* 36 (8), 3789–3794.

Tao, H., Mimura, Y., Aoe, K., Kobayashi, S., Yamamoto, H., Matsuda, E., et al. (2012). Prognostic Potential of FOXP3 Expression in Non-small Cell Lung Cancer Cells Combined with Tumor-Infiltrating Regulatory T Cells. *Lung cancer* 75 (1), 95–101. doi:10.1016/j.lungcan.2011.06.002

Wang, L., Liu, R., Li, W., Chen, C., Katoh, H., Chen, G.-Y., et al. (2009). Somatic Single Hits Inactivate the X-Linked Tumor Suppressor FOXP3 in the Prostate. *Cancer cell* 16 (4), 336–346. doi:10.1016/j.ccr.2009.08.016

Weigelt, B., and Reis-Filho, J. S. (2010). Molecular Profiling Currently Offers No More Than Tumour Morphology and Basic Immunohistochemistry. *Breast Cancer Res.* 12 (4), 1–4. doi:10.1186/bcr2734

Wieczorek, G., Asemissen, A., Model, F., Turbachova, I., Floess, S., Liebenberg, V., et al. (2009). Quantitative DNA Methylation Analysis of FOXP3 as a New Method for Counting Regulatory T Cells in Peripheral Blood and Solid Tissue. *Cancer Res.* 69 (2), 599–608. PMID: 19147574. doi:10.1158/0008-5472.CAN-08-2361

Wildin, R. S., Ramsdell, F., Peake, J., Faravelli, F., Casanova, J.-L., Buist, N., et al. (2001). X-linked Neonatal Diabetes Mellitus, Enteropathy and Endocrinopathy Syndrome Is the Human Equivalent of Mouse Scurfy. *Nat. Genet.* 27 (1), 18–20. doi:10.1038/83707

Winerdal, M. E., Marits, P., Winerdal, M., Hasan, M., Rosenblatt, R., Tolf, A., et al. (2011). FOXP3 and Survival in Urinary Bladder Cancer. *BJU Int.* 108 (10), 1672–1678. doi:10.1111/j.1464-410x.2010.10020.x

Yang, S., Liu, Y., Li, M.-Y., Ng, C. S. H., Yang, S.-l., Wang, S., et al. (2017). FOXP3 Promotes Tumor Growth and Metastasis by Activating Wnt/β-Catenin Signaling Pathway and EMT in Non-small Cell Lung Cancer. *Mol. Cancer* 16 (1), 1–12. doi:10.1186/s12943-017-0700-1

Zhang, H.-Y., and Sun, H. (2010). Up-regulation of Foxp3 Inhibits Cell Proliferation, Migration and Invasion in Epithelial Ovarian Cancer. *Cancer Lett.* 287 (1), 91–97. doi:10.1016/j.canlet.2009.06.001

Zuo, T., Wang, L., Morrison, C., Chang, X., Zhang, H., Li, W., et al. (2007). FOXP3 Is an X-Linked Breast Cancer Suppressor Gene and an Important Repressor of the HER-2/ErbB2 Oncogene. *Cell* 129 (7), 1275–1286. doi:10.1016/j.cell.2007.04.034

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fevo.2021.781498

DNA Methylation of ABC Transporters Differs in Native and Non-native Populations of *Conyza canadensis* L.

Ambreen Tahir Shah[1,2], *Zafar A. Reshi*[2] *and Mohammad Altaf*[1,3]*

[1] *Chromatin and Epigenetics Lab, Department of Biotechnology, University of Kashmir, Srinagar, India,* [2] *Department of Botany, University of Kashmir, Srinagar, India,* [3] *Centre for Interdisciplinary Research and Innovations, University of Kashmir, Srinagar, India*

While differences in the methylation patterns of ABC transporters under different environmental conditions and their role in plant growth, development, and response to biotic and abiotic stresses are well documented, less is known about the variation in the methylation patterns of ABC transporters in plant species in the native and non-native ranges. In this study, we present the results of differences in methylation of ABC transporters of *Conyza canadensis* L. in its native (North America) and non-native (Kashmir Himalaya) ranges. Our data show that ABC transporter genes have reduced DNA methylation in Kashmir Himalaya than in North America. Furthermore, in the non-native range of Kashmir Himalaya, we found that ABC transporter genes have enriched RNA Pol-II binding and reduced nucleosome occupancy, both hallmarks of transcriptional activity. Taken together, our study showed differential DNA methylation in the ABC transporter genes in the native range of North America and non-native range of Kashmir Himalaya in *Conyza canadensis* and that the reduced DNA methylation and increased RNA Pol-II binding is one of the possible mechanisms through which this species in the non-native range of Kashmir Himalaya may show greater gene expression of ABC transporter genes. This increased ABC transporter gene expression may help the plant to grow in different environmental conditions in the non-native range. Furthermore, this study could pave way for more studies to better explain the enigmatic plant invasions of *C. canadensis* in the non-native range of Kashmir Himalaya.

Keywords: ABC transporters, chromatin, epigenetics, DNA methylation, plant invasion, phenotypic plasticity, adaptation

OPEN ACCESS

Edited by:
Tanvir Ul Hassan Dar,
Baba Ghulam Shah Badshah University, India

Reviewed by:
Arif Tasleem Jan,
Baba Ghulam Shah Badshah University, India
Mohd Dar,
Indian Institute of Integrative Medicine (CSIR), India

***Correspondence:**
Mohammad Altaf
altafbhat@uok.edu.in

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Ecology and Evolution

Received: *22 September 2021*
Accepted: *16 December 2021*
Published: *03 February 2022*

Citation:
Shah AT, Reshi ZA and Altaf M (2022) DNA Methylation of ABC Transporters Differs in Native and Non-native Populations of Conyza canadensis L. Front. Ecol. Evol. 9:781498. doi: 10.3389/fevo.2021.781498

INTRODUCTION

DNA methylation, one of the most common epigenetic modifications, contributes to the overall growth, development, and tolerance to both biotic and abiotic factors (Alonso et al., 2018; Agarwal et al., 2020; Eriksson et al., 2020; Gallego-Bartolom, 2020; Liu and He, 2020; Kumar and Mohapatra, 2021). DNA methylation and other epigenetic modifications are now known to adjust phenotypes instantaneously and/or generate new phenotypes in a reversibly heritable manner (Flatscher et al., 2012; Miryeganeh and Saze, 2019). These heritable epigenetic variations could directly or indirectly

influence the course of evolution in plants, as they can affect the processes of adaptation (Rapp and Wendel, 2005; Jablonka and Raz, 2009).

Epigenetic processes, including DNA methylation, have emerged as important mechanisms contributing to adaptation and spread of alien plants in new habitats in the non-native ranges (Perez et al., 2015), despite several bottlenecks, such as low propagule pressure, low genetic variability, climate incompatibility, inadequate or inappropriate resources, or small population size (Banerjee et al., 2019). *Conyza canadensis,* commonly called as horseweed, has successfully adapted to several non-native habitats across the world and is considered a highly invasive problematic annual weed species of many agronomic crops in about 70 countries (Travlos and Chachalis, 2010; Moretti et al., 2017). This species has a very prolific growth and is reported to produce up to 200,000 seeds, which are easily dispersed by wind to infest new territories (Dauer et al., 2007). To control this weed species, commercial herbicides are in use for a long time but *Conyza* species have developed resistance against these herbicides, particularly glyphosate (Moretti et al., 2017). Vacuolar sequestration of herbicides has been identified as the predominant resistance mechanism that neutralizes its activity mediated through tonoplast localized ABC transporters (Peng et al., 2010; Douglas and Gaines, 2014; Ge et al., 2014). ABC transporter genes such as M10 and M11 are highly expressed in herbicide-resistant *Conyza* species (Peng et al., 2010; Nol et al., 2012; Moretti et al., 2017). Owing to environmental influence, there is a decreased expression of ABC transporters such as M10 and M11 under conditions of low temperature resulting in a lack of glyphosate resistance (Tani et al., 2016). The possible mechanisms of ABC transporter gene expression in invasive populations could be DNA methylation-induced alterations in chromatin structure, which, in turn, may be responsible for higher adaptability and invasiveness. Keeping in view the differential methylation patterns across various habitats and the potential role played by epigenetic diversity in influencing the invasive traits of a plant (Zhang et al., 2016; Banerjee et al., 2019), we hypothesized differential epigenetic signatures of *C. canadensis* in its native range of North America and non-native range of Kashmir Himalaya. The study, the first of its kind on *C. canadensis*, is likely to open new vistas of investigating the invasiveness of alien species in the non-native ranges, which is a prerequisite for their better management.

MATERIALS AND METHODS

Sampling of Leaf Material

Leaf samples at the flowering stage were collected from both the native range of North America and the non-native range of Kashmir Himalaya from 5 different sites. Each sample was collected along with three biological replicates and three technical replicates. The sites, namely, Kangan, Gulmarg, Budgam, KUBG1 (Kashmir University botanical garden), and KUBG2, were selected for the collection of non-native leaf samples of Kashmir Himalaya, while Turah, Clinton, Clarke, Butte, and Lolo were selected for the native leaf samples of North America.

S.No.	Site name	Geo-coordinates	Part collected	Sampling time
United States				
1.	Turah		Leaf	Flowering stage
2.	Clarke	46° 40′ 8″ N 111° 44′ 0″ W	Leaf	Flowering stage
3.	Butte	46°0′23″N 112°31′47″W	Leaf	Flowering stage
4.	Lolo	46°45′55″N 114°5′9″W	Leaf	Flowering stage
5.	Clinton	46°46′33″N 113°42′54″W	Leaf	Flowering stage
Kashmir				
1.	KUBG1	34°07′57″N 74°50′15″E	Leaf	Flowering stage
2.	KUBG2	34.1589887°N 74.8316992°E	Leaf	Flowering stage
3.	Gulmarg	34.05°N 74.38°E	Leaf	Flowering stage
4.	Kangan	34.263°N 74.903°E	Leaf	Flowering stage
5.	Budgam	34°1′12″N 74°46′48″E	Leaf	Flowering stage

Chromatin Extraction Followed by Methylated DNA Immunoprecipitation

Chromatin was extracted from the leaf samples of both the native range of North America and the non-native range of Kashmir Himalaya of *C. canadensis*. After quantification, approximately 6 μg of DNA was taken for sonication (20% amplitude, pulse: 30 s, and time: 3 min). Samples were normalized and incubated at 4°C using an antimethyl cytosine antibody overnight with end-to-end rotation. For pull-down, equilibrated Dynabeads were added to each sample. Then samples were treated with proteinase K and glycogen, incubated at 55°C, 1,200 rpm overnight (Fabio et al., 2009). DNA was extracted using PCI (phenol:chloroform:isoamyl) and resuspended in 1 × TE buffer, followed by quantitative PCR (qPCR).

Chromatin Immunoprecipitation Assay

Approximately 2 g leaf samples of *C. canadensis* from both the native range of North America and the non-native range of Kashmir Himalaya were taken and submerged in 37% formaldehyde and 2.5 M glycine for crosslinking of tissue using a vacuum infiltrator (Saleh et al., 2008). The tissue was then washed two times with double distilled water and then resuspended in extraction buffer 1 (0.4 M sucrose, 10 mM Tris-HCl, 5 mM βmerc, 0.1 mM PMSF, PIC) and filtered through Mira cloth (Merck Millipore, Burlington, MA, United States). After centrifugation for 20 min at 5,000 rpm, the pellet was resuspended in 1 ml of extraction buffer 2 (0.25 M sucrose, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1% Triton x-100, 5 mM βmerc, 0.1 mM PMSF, and PIC) followed by 300 μl of extraction buffer 3 (1.7 M sucrose, 10 mM Tris-HCl, 0.15% Triton x-100, 2 mM $MgCl_2$, 5 mM βmerc, 0.1 mM PMSF, and PIC). Isolated chromatin was resuspended in nuclei lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS, PMSF, and PIC) followed by sonication (30 s ON, 30 s OFF, 4 cycles, amplitude 20%) for chromatin fragments approximately to the size of 250–500 bp. After centrifugation for 30 min, the supernatant of samples was normalized and the chromatin solution was subjected to O/N incubation at 4°C on a rotor with antibody (anti-POL II/anti-H3K14ac). Equilibrated Dynabeads (protein A and G) were added to each sample and incubated at 4°C for 3 h. Samples were washed with high salt buffer, low salt buffer, and LiCl buffer. The

samples were eluted in TE/SDS at 65°C for 15 min. Samples were kept for reverse crosslinking at 65°C. DNA extraction was carried out using PCI and the pellet obtained was resuspended in TE and followed by qPCR.

MNase Titration Assay

Briefly, 2 g leaf samples of *C. canadensis* from both the native (North America) and non-native (Kashmir Himalaya) ranges were collected and ground using liquid nitrogen. Nuclei extraction buffer (0.25 M sucrose, 60 mM KCl, 0.8% Triton x-100, 15 mM $MgCl_2$, 1 mM $CaCl_2$, 15 mM HEPES, 0.1 mM PMSF) was added to leaf samples for nucleus isolation. The isolated nucleus was then resuspended into MNase digestion buffer and incubated with different units of MNase (1U, 2U, 3.5U, and 5U) at 30°C for 10 min (Pajoro et al., 2018). The reaction was then stopped using a stop buffer (250 mM EDTA and 10% SDS). All the samples were then subjected to 3 μl RNase treatment and 1/2 h incubation. This was followed by proteinase K treatment followed by 1 h incubation. DNA was then extracted, purified, and resuspended in TE buffer. The pattern of digested chromatin was confirmed on a 2% agarose gel, while its quantity and purity analyzed using NanoDrop™ (Thermo Fisher Scientific, Waltham, MA, United States). The digested chromatin was quantified using ImageJ software (NIH Java-based).

RESULTS

ABC Transporter Genes Show Reduced DNA Methylation in the Non-native Range of Kashmir Himalaya

Using methylated DNA immunoprecipitation (MeDIP) with a 5-methylcytidine antibody followed by quantitative real-time (qRT-PCR), we compared the DNA methylation level at the ABC transporter genes of *C. canadensis* in the native range of North America and non-native range of Kashmir Himalaya. MeDIP is an immunocapturing approach for unbiased detection of methylated DNA, wherein genomic DNA is sheared by sonication followed by immunoprecipitation using an antimethyl cytosine antibody (Fabio et al., 2009). The *C. canadensis* within the non-native range of Kashmir Himalaya showed significantly reduced DNA methylation at ABC transporter genes as compared to the native range of North America ($p = 0.0072$). The qRT-PCR results showed an IP/input value between 2.16 and 6.88 in samples from the native range of North America, while samples from the non-native range of Kashmir Himalaya showed significantly reduced IP/input values between 0.11 and 0.75 (**Figure 1**). Quite interestingly, the DNA methylation patterns of ABC transporter genes also varied across different populations within the native range of North America and the non-native range of Kashmir Himalaya. These results confirm that DNA methylation at ABC transporter genes could be one of the major factors regulating the expression of ABC transporter genes in the native range of North America and the non-native range of Kashmir Himalaya and that the reduced DNA methylation in the non-native range of Kashmir Himalaya may lead to increased ABC gene expression and hence more adaptability.

ABC Transporter Genes Show Enriched Pol II Occupancy in the Non-native Range of Kashmir Himalaya

Using chromatin immunoprecipitation with RNA Pol II antibody followed by qRT-PCR, we looked at RNA polymerase-II occupancy at ABC transporter genes in the native (North America) and non-native (Kashmir Himalaya) leaf samples. Pol II ChIP detects RNA polymerase II within the coding region of genes using polymerase antibodies analyzed using PCR (Sandoval et al., 2004). As expected, we observed increased RNA Pol II occupancy at ABC transporter genes in the non-native range of Kashmir Himalaya as compared to the native range of North America (**Figure 2**). The qRT-PCR results showed Pol II occupancy value between 0.128 and 0.296 in leaf samples of the native range of North America, while the leaf samples from the non-native range of Kashmir Himalaya showed Pol II occupancy value between 0.8 and 1.206. Furthermore, like DNA methylation, RNA Pol-II binding at ABC transporter genes also showed a variation across the different sites both within the native range of North America and the non-native range of Kashmir Himalaya. Reduced DNA methylation and subsequent increased RNA Pol-II binding may bring about the transcriptional activation of ABC transporter genes in the non-native range of Kashmir Himalaya and hence their increased expression.

ABC Transporter Genes Show Loss of Nucleosomes in the Non-native Range of Kashmir Himalaya

In this study, we observed reduced acetylation levels of H3K14ac in the non-native range of Kashmir Himalaya as compared with the native range of North America, which could be due to changes in nucleosome occupancy (**Figure 3**). Chromatin immunoprecipitation analysis of histone H3 showed reduced nucleosome occupancy under non-native conditions. To get further insights into the changes in chromatin structure, we carried out micrococcal nuclease (MNase titrations) of chromatin from the native range of North America and the non-native range of Kashmir Himalaya. The MNase assay involves the digestion of chromatin using an enzyme MNase and understands the nucleosomal pattern by quantifying the protection of DNA from enzymatic digestion (Pass et al., 2017). Using the different concentrations of enzymes in an MNase assay, the non-native samples of Kashmir Himalaya showed more mononucleosomes as compared with the native ones of North America, indicating more open chromatin in the non-native *C. canadensis* within Kashmir Himalaya (**Figure 4**). This result suggests that the ABC locus in *C. canadensis* undergoes chromatin changes to provide phenotypic plasticity to the plant in the non-native range of Kashmir Himalaya. Collectively, reduced DNA methylation associated with increased Pol-II binding at the ABC transporter genes and open chromatin in *C. canadensis* within the non-native range of Kashmir Himalaya increases

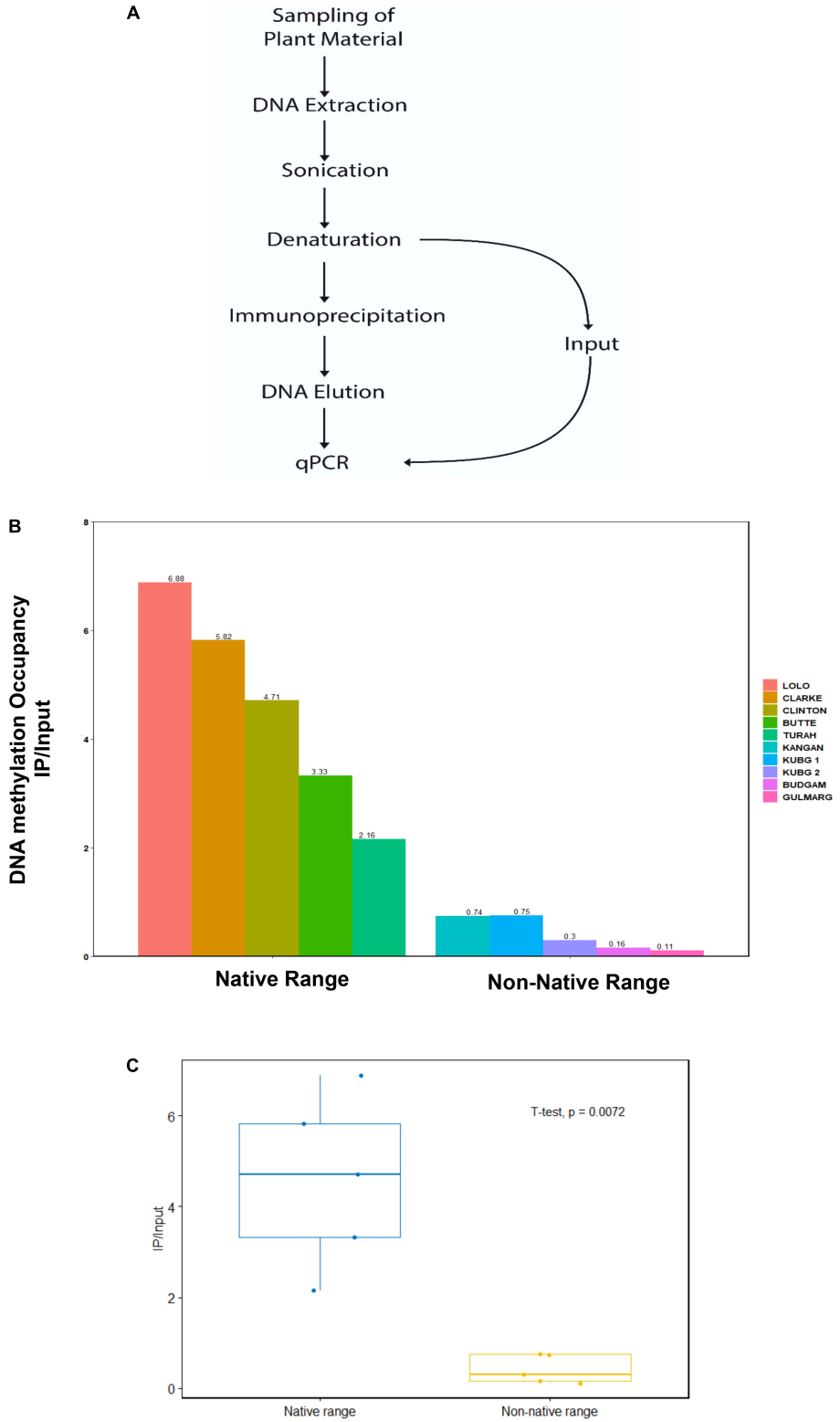

FIGURE 1 | DNA methylation levels at ABC transporter gene in the native (North America) and non-native (Kashmir Himalaya) range plants. **(A)** The plant material was collected at different geographical locations as mentioned in "Materials and Methods" section. The DNA was isolated, sonicated, and subjected to immunoprecipitation using an antibody against 5-methyl cytosine and subjected to PCR with specific primers. **(B)** The IP/input ratio for 5-methyl cytosine in DNA of native and non-native range plants across different sites at ABC transporter gene. **(C)** The IP/input values at ABC transporter gene in native and non-native range obtained using *R* software, $n = 5$.

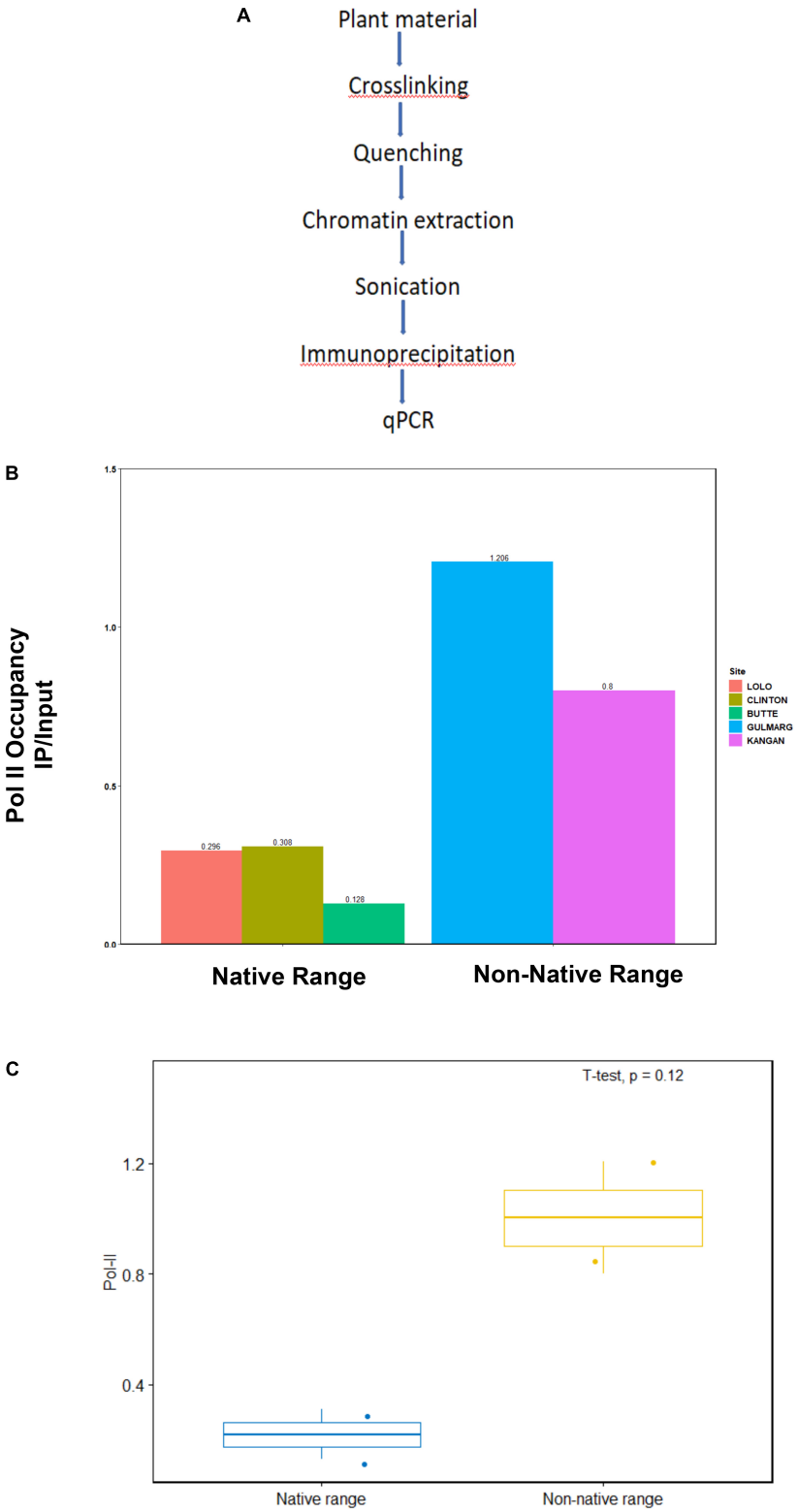

FIGURE 2 | Pol-II occupancy at ABC transporter genes in the native (North America) and non-native (Kashmir Himalaya) range plants. **(A)** The plant material was collected at different geographical locations as mentioned in "Materials and Methods" section. The DNA-protein interaction was cross-linked with a suitable cross-linking agent. Subsequently, the DNA was isolated, sonicated, and then subjected to immunoprecipitation using specific antibodies. **(B)** The bar graphs represent the RNA Pol-II occupancy at ABC transporter genes in native and non-native range plants. **(C)** RNA Pol-II occupancy at ABC transporter gene in the native and non-native range obtained using *R* software, $n = 3$.

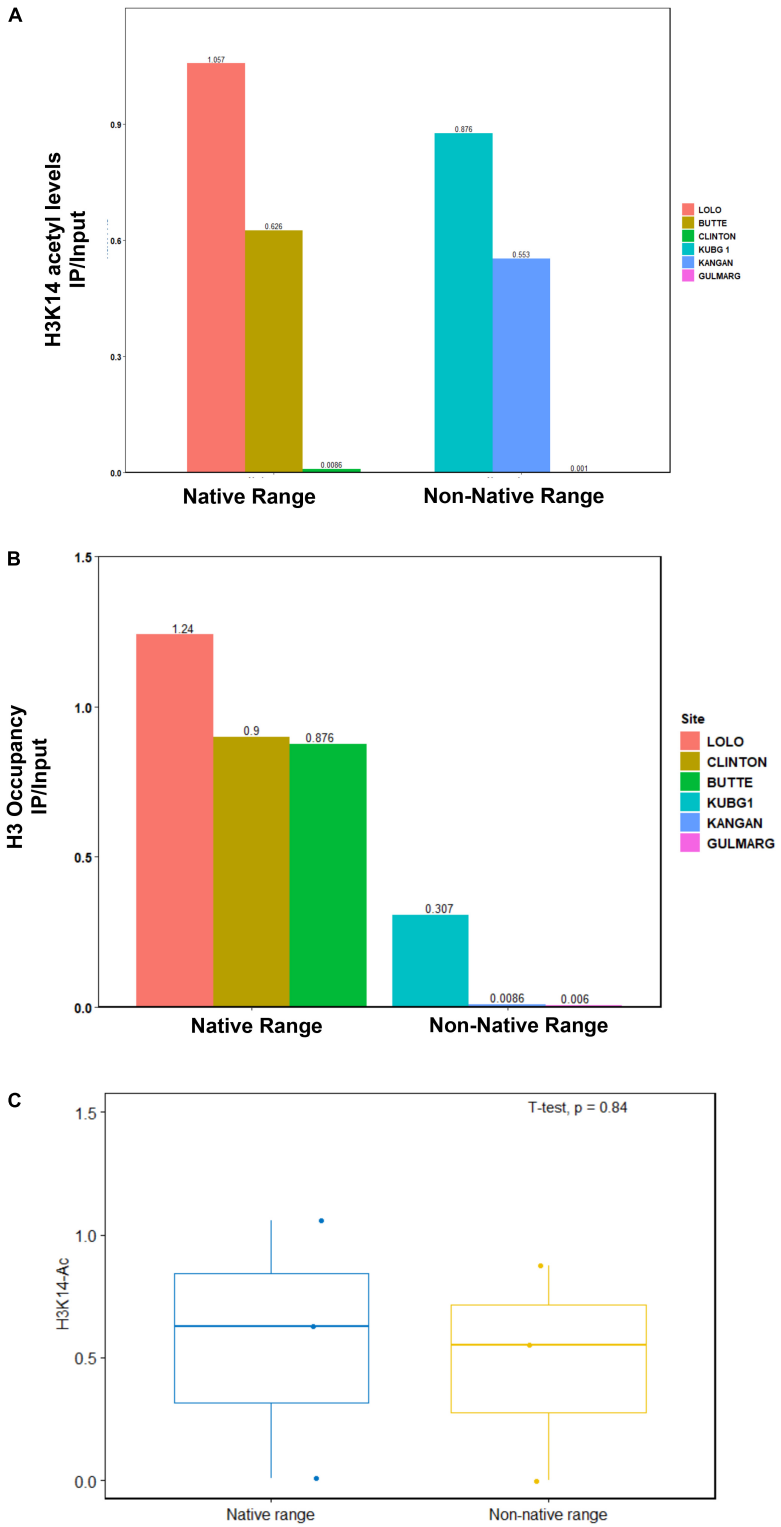

FIGURE 3 | H3 and H3K14 acetyl occupancy in the native (North America) and non-native (Kashmir Himalaya) range plants. The plant material was collected at different geographical locations as mentioned in "Materials and Methods" section. The DNA-protein interaction was cross-linked with a suitable cross-linking agent. Subsequently, the DNA was isolated, sonicated, and then subjected to immunoprecipitation using specific antibodies. **(A)** The bar graphs represent the IP/Input ratio of H3K14ac occupancy at ABC transporter genes in the native and non-native range plants. **(B)** The bar graphs represent the IP/Input ratio of H3 occupancy at ABC transporter genes in the native and non-native range plants. **(C)** H3K14 acetyl occupancy at ABC transporter gene in the native and non-native range obtained using *R* software, $n = 3$.

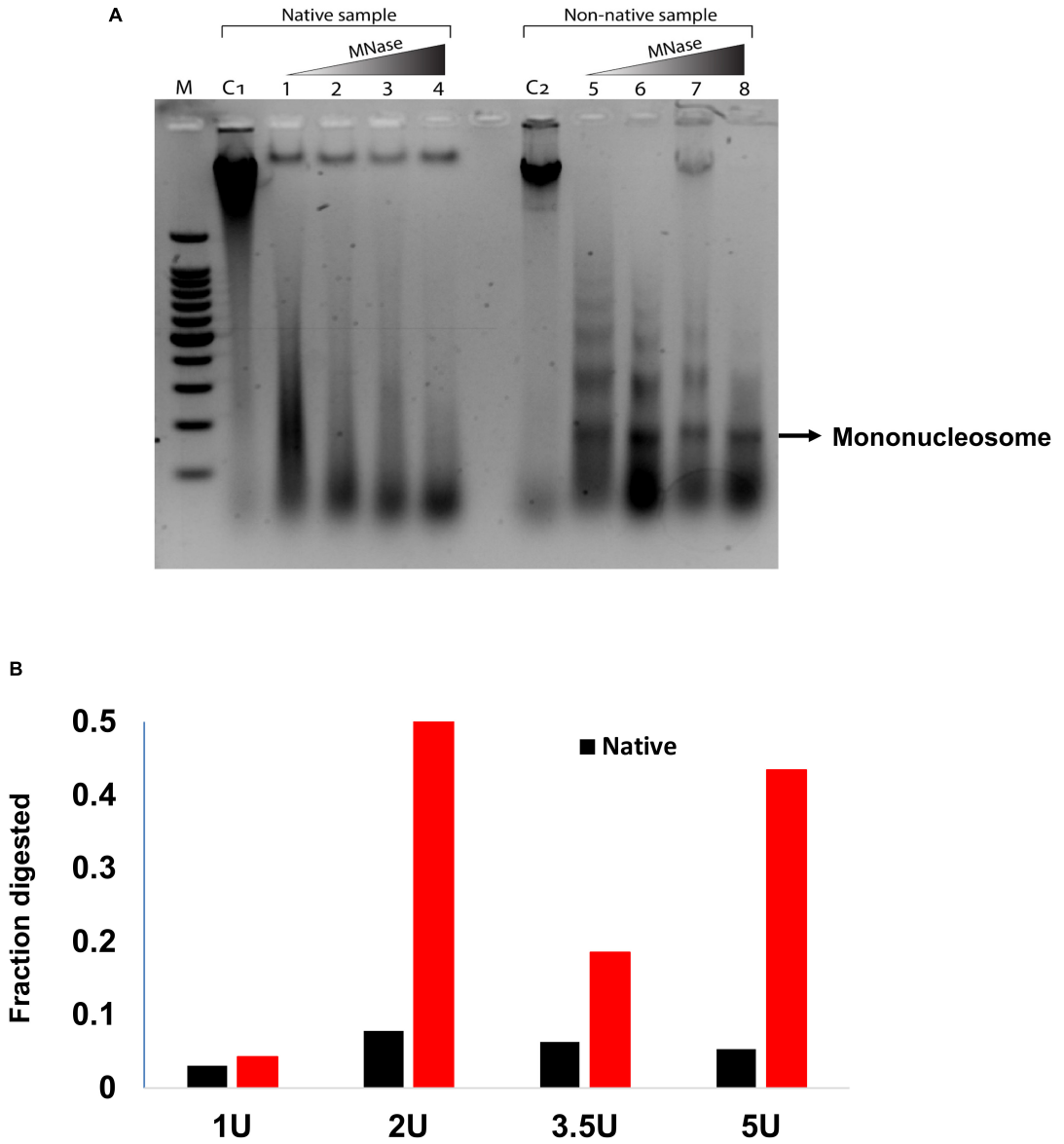

FIGURE 4 | MNase digestion assay in the native (North America) and non-native (Kashmir Himalaya) range plants. **(A)** The chromatin was isolated from plant samples of both native and non-native range *Conyza canadensis* and subjected to MNase digestion with different enzyme concentrations as mentioned in "Materials and Methods" section. **(B)** The bar graph represents the percentage of mononucleosomal fractions digested in the native and non-native range plants.

the transcriptional output that may increase the adaptability of these plants.

DISCUSSION

In this study, we observed a differential DNA methylation pattern at ABC transporter genes in the leaf samples of *C. canadensis* collected from its native range of North America and the non-native range of Kashmir Himalaya. The samples from the non-native range of Kashmir Himalaya showed a significantly lower DNA methylation level at ABC transporter genes as compared with the native samples of North America (**Figure 1**). Furthermore, DNA methylation patterns also showed variation across different populations within the native and non-native ranges. Studies have shown the involvement of differential methylation patterns in the widespread distribution of species in the non-native ranges (Ardura et al., 2018; Hawes et al., 2019). These studies revealed that epigenetic changes vary in heterogeneous habitats and these changes are heritable. Studies in invasive plant species have shown a positive correlation between differential methylation status and phenotypic plasticity (Zhang et al., 2016; Banerjee et al., 2019). This shows that epigenetic diversity may contribute to the phenotypic plasticity in invasive species independent of genetic diversity. As a consequence of DNA methylation, chromatin compresses into

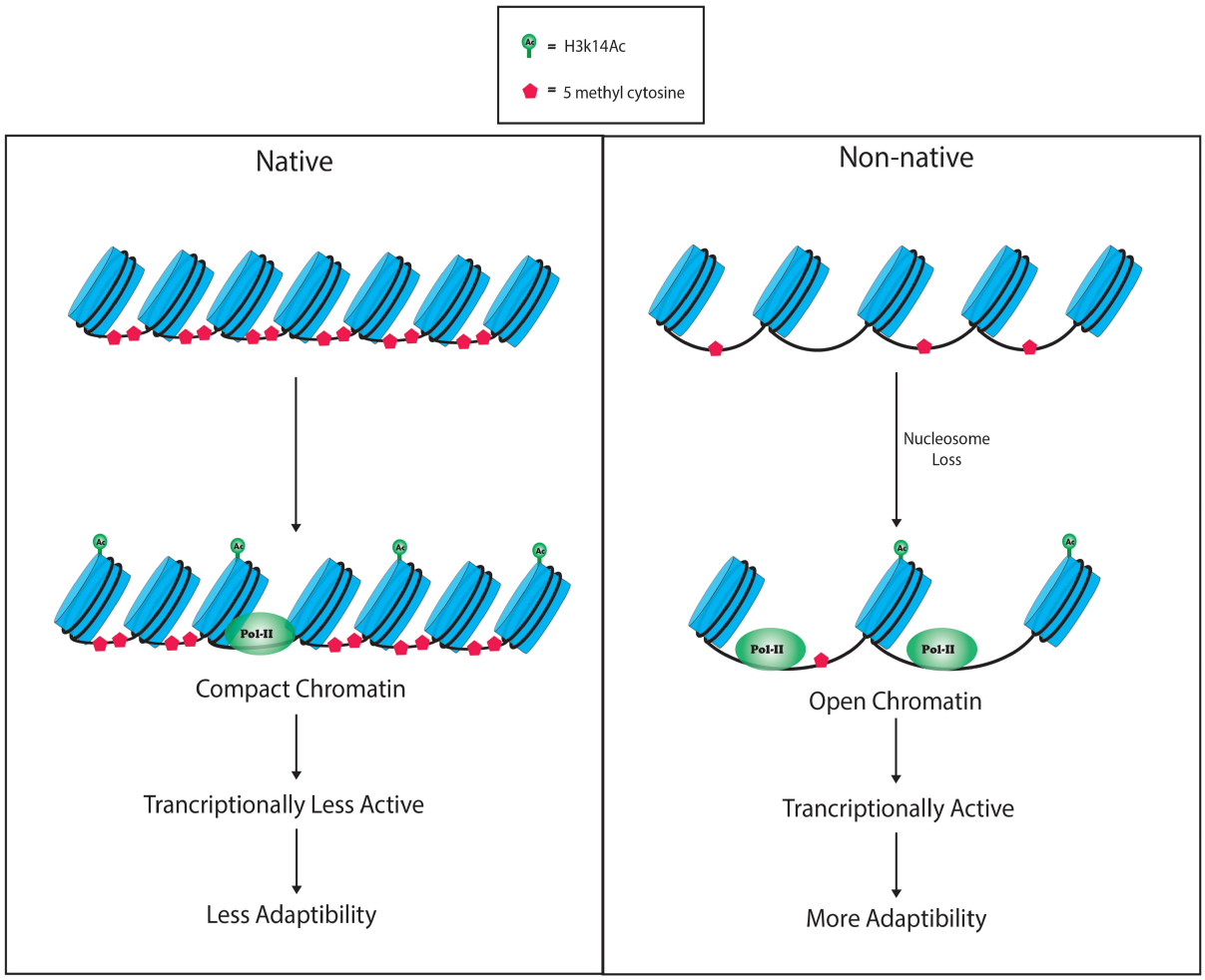

FIGURE 5 | Model showing changes in chromatin organization in the native (North America) and non-native (Kashmir Himalaya) samples of *C. canadensis*. The model represents the compact and open chromatin at ABC transporter genes in the native and non-native range plants of *C. canadensis*. The non-native range plants showed open chromatin possibly due to decreased DNA methylation and increased H3K14ac occupancy at the ABC transporter gene and hence making the ABC transporter gene transcriptionally more active in the non-native ones as compared with native range plants of *C. canadensis*.

heterochromatin at the transcription start site and hence reduces the gene activity and its transcription (Buitrago et al., 2021). The reduced DNA methylation observed in *C. canadensis* in the non-native range of Kashmir Himalaya might explain the increased expression of ABC transporter genes. Various studies have shown a direct link between the environmental condition and induction levels of key genes including ABC transporters and that these ABC transporters could be playing a role in the establishment and spread of *C. canadensis* in Kashmir Himalaya (Tani et al., 2016). A link has been deciphered between DNA demethylation and invasion during the stage of expansion (Ardura et al., 2017). This may happen because of gene activation and reactivation of transposable elements redefining gene expression. As a result, demethylation may confer varying phenotypic plasticity and fitness. Plants have at least three different DNA methyltransferases, also called cytosine methyltransferases, which differ in structure and function and methylates DNA at cytosine (m5C) and adenine (m6A) (Finnegan and Kovac, 2000; Vanyushin, 2006). These various classes of DNMTS within the plant system, unlike the mammalian system, show differential methyltransferase activity and context-dependent effect on cytosine base (Finnegan and Kovac, 2000). Given that we observed the differential DNA methylation patterns in native (North America) and non-native (Kashmir Himalaya) *C. canadensis*, it will be an interesting area of research to profile the *C. canadensis* in other native and non-native ranges for DNA methylation and differential DNA methylation patterns. Since, we hypothesize that DNA associated Pol-II elongation may increase the RNA Pol-II binding at ABC transporter genes (Jonkers and Lis, 2015), we observed increased RNA Pol-II binding in non-native (Kashmir Himalaya) range *C. canadensis* as compared with native (North America) samples (**Figure 2**). This result is in line with the fact that reduced DNA methylation is mostly associated with increased RNA Pol-II binding and hence results in increased gene expression (Bender, 2004; Takeshima et al., 2009). Furthermore, we observed the

loss of nucleosome at the ABC gene transporters in non-native samples of Kashmir Himalaya as indicated by comparative ChIP analysis of H3K14ac and H3 (**Figure 3**). This loss of nucleosomes is an indication of open chromatin in the non-native range of Kashmir Himalaya. Therefore, we also compared the chromatin dynamics in *C. canadensis* within its native (North America) and non-native range (Kashmir Himalaya) using MNase assay. In this study, nucleosomal occupancy and positioning have a critical impact on the expression and regulation of genes (Zhang et al., 2015). The MNase assay performed on the leaf samples of *C. canadensis* from native (North America) and non-native (Kashmir Himalaya) range revealed an interesting phenomenon of chromatin changes in non-native range plants of Kashmir Himalaya. The non-native plants of Kashmir Himalaya showed an increased mononucleosomal fraction as compared with native ones of North America (**Figure 4**). The increased mononucleosomal fraction is an indication of open and more accessible chromatin in non-native plants of Kashmir Himalaya that may modulate gene expression of these plants in the non-native range of Kashmir Himalaya (Kenchanmane Raju, 2020).

Taken collectively, our study has identified the differential DNA methylation patterns at ABC transporter genes in the native (North America) and non-native (Kashmir Himalaya) samples of *C. canadensis*. Non-native leaf samples of Kashmir Himalaya showed reduced DNA methylation at these genes, which was further associated with enhanced Pol-II binding. The combinatorial effect of reduced DNA methylation and increased RNA Pol-II binding along with open chromatin may enhance the transcriptional activity of ABC transporter genes and finally may increase the plant growth in different environmental conditions and may be one of the possible mechanisms through which plant becomes invasive in its non-native range of Kashmir Himalaya (**Figure 5**). Although our study deals with the chromatin dynamics at the ABC transporter genes, understanding the downstream events could be an exciting area of research.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AS carried out the experiments. AS, ZR, and MA designed the experiments, interpreted the results, and wrote the manuscript. All authors edited and approved it.

FUNDING

ZR acknowledged the financial support of the University Grants Commission under the CPEPA scheme. The work was supported by grants from the Department of Biotechnology, Government of India (BT/PR22348/BRB/10/1625/2017), DST-SERB (CRG/2020/003632), RUSA 2.0, and DST-PURSE (SR/PURSE/2020/31) to MA.

ACKNOWLEDGMENTS

We thank Ishfaq Ahamd Pandith for his help in the preparation of figures and tables. AS acknowledges CSIR India for Senior Research fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.781498/full#supplementary-material

REFERENCES

Agarwal, G., Kudapa, H., Ramalingam, A., Choudhary, D., Sinha, P., Garg, V., et al. (2020). Epigenetics and epigenomics: underlying mechanisms, relevance, and implications in crop improvement. *Funct. Integr. Genomics* 20, 739–761. doi: 10.1007/s10142-020-00756-7/Published

Alonso, C., Becker, C., and Ramos-Cruz, D. (2018). The role of plant epigenetics in biotic interactions. *New Phytol.* 221, 731–737. doi: 10.1111/nph.15408

Ardura, A., Clusa, L., Zaiko, A., Garcia-Vazquez, E., and Miralles, L. (2018). Stress related epigenetic changes may explain opportunistic success in biological invasions in Antipode mussels. *Sci. Rep.* 8:10793. doi: 10.1038/s41598-018-29181-4

Ardura, A., Zaiko, A., Morán, P., Planes, S., and Garcia-Vazquez, E. (2017). Epigenetic signatures of invasive status in populations of marine invertebrates. *Sci. Rep.* 7:42193. doi: 10.1038/srep42193

Banerjee, A. K., Guo, W., and Huang, Y. (2019). Genetic and epigenetic regulation of phenotypic variation in invasive plants–linking research trends towards a unified framework. *Neo Biota* 103, 77–103. doi: 10.3897/neobiota.49.33723

Bender, J. (2004). DNA methylation and epigenetics. *Annu. Rev. Plant Biol.* 55, 41–68. doi: 10.1146/ANNUREV.ARPLANT.55.031903.141641

Buitrago, D., Labrador, M., Arcon, J. P., Lema, R., Flores, O., Esteve-Codina, A., et al. (2021). Impact of DNA methylation on 3D genome structure. *Nat. Commun.* 12:3243. doi: 10.1038/s41467-021-23142-8

Dauer, J. T., Mortensen, D. A., and Vangessel, M. J. (2007). Temporal and spatial dynamics of long-distance *Conyza canadensis* seed dispersal. *J. Appl. Ecol.* 44, 105–114. doi: 10.1111/j.1365-2664.2006.01256.x

Douglas, R., and Gaines, T. A. (2014). Glyphosate resistance?: state of knowledge. *Pest Manag. Sci.* 70, 1367–1377. doi: 10.1002/ps.3743

Eriksson, M. C., Szukala, A., Tian, B., and Paun, O. (2020). Current research frontiers in plant epigenetics: an introduction to a virtual issue. *New Phytol.* 226, 285–288. doi: 10.1111/nph.16493

Fabio, M., Michael, W., Dirk, S., and Tim-Christoph, R. (2009). Methylated DNA immunoprecipitation (MeDIP). *Methods Mol. Biol.* 507, 55–64. doi: 10.1007/978-1-59745-522-0_5

Finnegan, E. J., and Kovac, K. A. (2000). Plant DNA methyltransferases. *Plant Mol. Biol.* 43, 189–201.

Flatscher, R., Frajman, B., Schönswetter, P., and Paun, O. (2012). Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation? *Genet. Res. Int.* 2012, 1–9. doi: 10.1155/2012/698421

Gallego-Bartolom, J. (2020). DNA methylation in plants: mechanisms and tools for targeted manipulation. *New Phytol.* 227, 38–44. doi: 10.1111/nph.16529

Ge, X., Avignon, D. A., Ackerman, J. J. H., and Sammons, R. D. (2014). In Vivo 31 P-nuclear magnetic resonance studies of glyphosate uptake, vacuolar sequestration, and tonoplast pump activity in glyphosate- resiatant Horseweed. *Plant Physiol.* 166, 1255–1268. doi: 10.1104/pp.114.247197

Hawes, N. A., Amadorou, A., Tremblay, L. A., Pochon, X., Dunphy, B., Fidler, A. E., et al. (2019). Epigenetic patterns associated with an ascidian invasion: a comparison of closely related clades in their native and introduced ranges. *Sci. Rep.* 9:14275. doi: 10.1038/s41598-019-49813-7

Jablonka, E., and Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* 84, 131–176. doi: 10.1086/598822

Jonkers, I., and Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16:167. doi: 10.1038/NRM3953

Kenchanmane Raju, S. K. (2020). Comparative profiling examines roles of DNA regulatory sequences and accessible chromatin during cold stress response in grasses. *Plant Cell* 32, 2451–2452. doi: 10.1105/TPC.20.00471

Kumar, S., and Mohapatra, T. (2021). Dynamics of DNA methylation and its functions in plant growth and development. *Front. Plant Sci.* 12:596236. doi: 10.3389/fpls.2021.596236

Liu, J., and He, Z. (2020). Small DNA methylation, big player in plant abiotic stress responses and memory. *Front. Plant Sci.* 11:595603. doi: 10.3389/fpls.2020.595603

Miryeganeh, M., and Saze, H. (2019). Epigenetic inheritance and plant evolution. *Popul. Ecol.* 62, 17–27. doi: 10.1002/1438-390X.12018

Moretti, M. L., Ala, R., Pearce, S., Morran, S., and Hanson, D. (2017). Transcription of putative tonoplast transporters in response to glyphosate and paraquat stress in *Conyza bonariensis* and *Conyza canadensis* and selection of reference genes for qRT-PCR. *PLoS One* 12:e0180794. doi: 10.1371/journal.pone.0180794

Nol, N., Tsikou, D., Eid, M., Livieratos, I. C., and Giannopolitis, C. N. (2012). Shikimate leaf disc assay for early detection of glyphosate resistance in *Conyza canadensis* and relative transcript levels of EPSPS and ABC transporter genes. *Weed Res.* 52, 233–241. doi: 10.1111/J.1365-3180.2012.00911.X

Pajoro, A., Muiño, J. M., Angenent, G. C., and Kaufmann, K. (2018). Profiling nucleosome occupancy by MNase-seq: experimental protocol and computational analysis. *Methods Mol. Biol.* 1675, 167–181. doi: 10.1007/978-1-4939-7318-7_11

Pass, D. A., Sornay, E., Marchbank, A., Crawford, M. R., Paszkiewicz, K., Kent, N. A., et al. (2017). Genome-wide chromatin mapping with size resolution reveals a dynamic sub-nucleosomal landscape in *Arabidopsis*. *PLoS Genet.* 13:e1006988. doi: 10.1371/JOURNAL.PGEN.1006988

Peng, Y., Abercrombie, L. G., Yuan, J. S., Riggins, C. W., Sammons, R. D., Tranel, J., et al. (2010). Characterization of the horseweed (*Conyza canadensis*) transcriptome using GS-FLX 454 pyrosequencing and its application for expression analysis of candidate non-target herbicide resistance genes. *Pest Manag. Sci.* 66, 1053–1062. doi: 10.1002/ps.2004

Perez, J. E., Alfonsi, C., Salazar, S., and Munoz, C. (2015). Ecological, genetic and epigenetic basis for bioinvasions. *Articulo De Revision Agrobiologia* 27, 167–177.

Rapp, R. A., and Wendel, J. F. (2005). Epigenetics and plant evolution. *New Phytol.* 168, 81–91. doi: 10.1111/j.1469-8137.2005.01491.x

Saleh, A., Alvarez-Venegas, R., and Avramova, Z. (2008). An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nat. Protoc.* 3, 1018–1025. doi: 10.1038/nprot.2008.66

Sandoval, J., Rodríguez, J. L., Tur, G., Serviddio, G., Pereda, J., Boukaba, A., et al. (2004). RNAPol-ChIP: a novel application of chromatin immunoprecipitation to the analysis of real-time gene transcription. *Nuclear Acids Res.* 32:e88. doi: 10.1093/nar/gnh091

Takeshima, H., Yamashita, S., Shimazu, T., Niwa, T., and Ushijima, T. (2009). The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. *Genome Res.* 19, 1974–1982. doi: 10.1101/GR.093310.109

Tani, E., Chachalis, D., Travlos, I. S., and Bilalis, D. (2016). Environmental conditions influence induction of key ABC-transporter genes affecting glyphosate resistance mechanism in *Conyza canadensis*. *Int. J. Mol. Sci.* 17:342. doi: 10.3390/ijms17040342

Travlos, I. S., and Chachalis, D. (2010). Glyphosate-resistant hairy fleabane (*Conyza bonariensis*) is reported in Greece. *Weed Technol.* 24, 569–573. doi: 10.1614/wt-d-09-00080.1

Vanyushin, B. F. (2006). DNA methylation in plants. *Curr. Top. Microbiol. Immunol.* 301, 67–122. doi: 10.1007/3-540-31390-7_4

Zhang, T., Zhang, W., and Jiang, J. (2015). Genome-wide nucleosome occupancy and positioning and their impact on gene expression and evolution in plants. *Plant Physiol.* 168, 1406–1416. doi: 10.1104/PP.15.00125

Zhang, Y.-Y., Parepa, M., Fischer, M., and Bossdorf, O. (2016). "Epigenetics of colonizing species? A study of Japanese knotweed in central Europe," in *Invasion Genetics: The Baker and Stebbins Legacy*, eds S. C. H. Barrett, R. Colautti, K. M. Dlugosch, and L. H. Rieseberg (Hoboken, NJ: Wiley).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fgene.2022.842943

FOXO1 Gene Downregulation and Promoter Methylation Exhibits Significant Correlation With Clinical Parameters in Indian Breast Cancer Patients

*Mohammad Aasif Khan[1], Sheersh Massey[1], Irfan Ahmad[2], Sadaf[1], Naseem Akhter[3], Maria Habib[1], Saad Mustafa[1], S. V. S. Deo[4] and Syed Akhtar Husain[1]**

[1]Human Genetics Laboratory, Department of Biosciences, Jamia Millia Islamia, New Delhi, India, [2]Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Khalid University, Abha, Saudi Arabia, [3]Department of Laboratory Medicine, Faculty of Applied Medical Sciences, Albaha University, Albaha, Saudi Arabia, [4]Department of Surgical Oncology BRA-IRCH, All India Institute of Medical Sciences (AIIMS), New Delhi, India

OPEN ACCESS

Edited by:
Kaiser Jamil,
Bhagwan Mahavir Medical Research Centre, India

Reviewed by:
Jogendra Pawar,
Purdue University, United States
Kanchan Bhatia,
Arizona State University, United States

***Correspondence:**
Syed Akhtar Husain
shusain@jmi.ac.in

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: 24 December 2021
Accepted: 10 February 2022
Published: 03 March 2022

Citation:
Khan MA, Massey S, Ahmad I, Sadaf, Akhter N, Habib M, Mustafa S, Deo SVS and Husain SA (2022) FOXO1 Gene Downregulation and Promoter Methylation Exhibits Significant Correlation With Clinical Parameters in Indian Breast Cancer Patients. Front. Genet. 13:842943. doi: 10.3389/fgene.2022.842943

Background: Forkhead box "O" one which is member of Forkhead box family of transcription factors is known to play key role in different physiological processes including cell cycle arrest, autophagy, and apoptosis. *FOXO1* is defined to play tumor suppressive role in various malignancies including breast cancer and its Dysregulation is frequently reported. However, the evaluation of *FOXO1* promoter methylation and its expression at mRNA and protein level in different stages of breast cancer and its association with different clinical parameters is still not studied. Therefore, for better understanding the role of *FOXO1* in breast cancer, in our study we examined the *FOXO1* mRNA and protein expression in Breast cancer samples of Indian breast cancer patients.

Results: Total 127 breast cancer samples along with adjacent normal tissue ($n = 127$) were analyzed through methylation specific PCR (MS-PCR), mRNA expression (Real-time PCR) and Immunohistochemistry (IHC). We detected 69.29% cases to be downregulated at the mRNA level, and 77.95% of cases exhibited no or low protein expression. In our data we report a significant association ($p = 0.0001$) between the downregulated protein expression and promoter hypermethylation of *FOXO1* gene. We also found a significant correlation of *FOXO1* mRNA level with Age ($p = 0.008$), age at first live birth ($p = 0,003$), tumor size ($p = 0.05$) and lymph node status ($p = 0.01$).

Conclusion: we in our study report the tumor suppressive role of *FOXO1* in case of Indian breast cancer patients and our data suggest it to exhibit prognostic importance. However, further research is needed to evaluate *FOXO1* significance in diagnostic and therapeutic targeting in breast cancer cases.

Keywords: methylation, immunohistochemistry, clinical, tumor, diagnosis

INTRODUCTION

According to Global Cancer statistics 2020, there were 684,996 deaths reported due to breast malignancy and it showed the highest incidence among all cancers (Globocan 2020). Breast cancer is a multifaceted disease exhibiting diverse morphological and histopathological features (Viale, 2012). To recognize the potential genes and their molecular mechanism associated with the pathogenesis of the disease still needs to be investigated (Tang et al., 2018). Lack of early diagnosis and inadequate personalized approach in treatment are chief factors in terms of poor survival of the patients (Shi et al., 2018). Hence, there is a necessity in the current time to search for more reliable molecular targets to develop a better diagnostic and therapeutic approach in the treatment of breast cancer (Chan et al., 2017; Jiang et al., 2018). Therefore, our study focuses on molecular profiling of the Forkhead box O 1 (*FOXO1*) gene which is a potent molecule and can exhibit promising results in the development of diagnostic and therapeutic strategies.

FOXO1 is one of the key members of the *FOXO* transcription factors subfamily, which is located on chromosome 13 (13q14.11) and is a chief target of insulin signalling. It is known to have a major role in the regulation of metabolic homeostasis, autophagy, apoptosis, cell cycle arrest genes, and immune regulators (Xing et al., 2018; Jiang et al., 2018; Kousteni, 2012, Gene cards). The activation of *FOXO1* via binding of insulin or several growth factors to their receptors consequently activates PI3K (phosphoinositide kinase) that further triggers the activity of other kinases including Akt and SGK (serum glucocorticoid inducible kinase). However, in the absence of insulin, the *FOXO1* is found to have nuclear localization and leads to cell cycle arrest. Thus, in presence of insulin or IGF-1, PI3K/Akt/SGK pathway is directly activated while *FOXO1* is inhibited resulting in cell survival (Cantley, 2002; Li et al., 2015). The low levels of *FOXO1* have been linked with tumor progression in several recent studies (Kaymaz et al., 2017; Procaccia et al., 2017). The decreased nuclear and cytoplasmic expressions of *FOXO1* in the case of breast cancer have been reported in the previous studies (Wu et al., 2012). However, the correlation of lower levels of *FOXO1* at mRNA and protein level with clinical parameters is not well known. The current study proposes to determine the correlation between methylation and expression of FOXO1 in breast cancer biopsy as compared to adjacent normal tissue.

METHODOLOGY AND MATERIALS

Collection of Biological Specimens

In our study, 127 participants were enrolled, and the cancerous tissue along with the adjacent non-cancerous tissue of the breast was obtained and stored at −20° for further experiments and analysis. The inclusion criteria for the specimens in the study included the histopathologically confirmed breast cancer patients of age group 20–79 years having at least 6 months of life expectancy.

Following clinical parameters were included for the study such as tumor size, age at diagnosis, histological grade, the status of lymph node (LN), reproductive history, and information (age at menarche and menopausal status), clinical staging or TNM stages, Estrogen receptor (ER) (+ or -), Progesterone receptor (PR) (+ or -) and Human Epidermal Growth Factor Receptor 2 (HER2) (+ or -).

The females ($n = 127$) included in the study were clinically confirmed with sporadic breast cancer and were genetically unrelated. Normal adjacent breast tissue was taken as control.

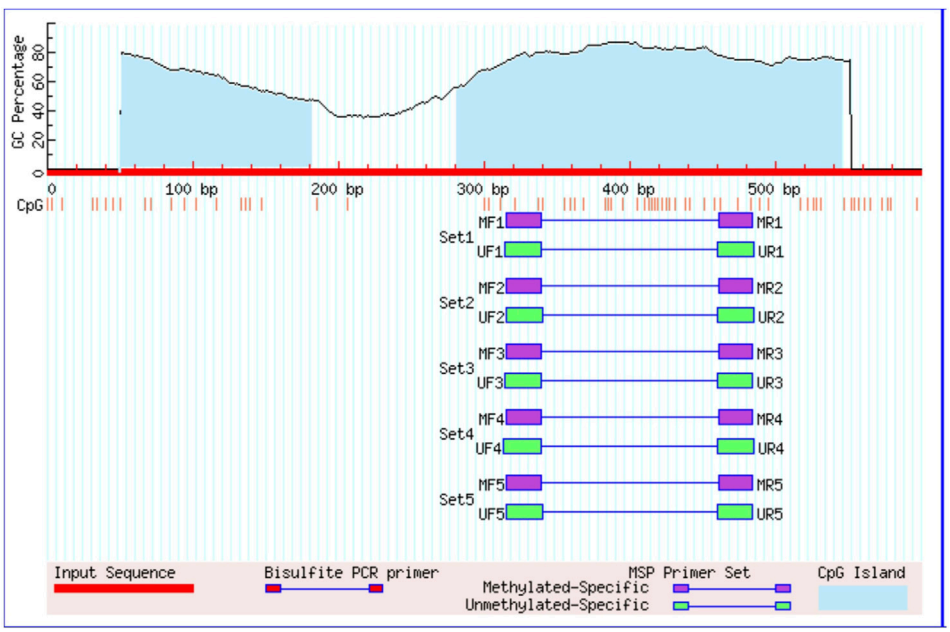

FIGURE 1 | Graphical represntation of CpG islands in the foxo1 promoter region taken from Meth Primer. Criteria used: Island size>100, GC Percent >50.0, Obs/Exp>0.60/.

TABLE 1 | Characteristics of study subjects (*n* =*127*).

S.no	Characteristic	Cases (%)
1	Age (years)	
	≤50	44 (34.65)
	>50	83 (65.35)
2	Geographic location	
	Rural	33 (25.98)
	Urban	94 (74.02)
3	Age at menarche	
	≤12	20 (15.75)
	>12	107 (84.25)
4	Age at first live birth	
	≤25	100 (78.75)
	>25	27 (21.25)
5	Breast feeding	
	Yes	122 (96.06)
	No	5 (3.94)
6	Use of exogenous hormone	
	Yes	6 (4.72)
	No	121 (95.28)
7	Family history of cancer	
	Yes	21 (16.54)
	No	106 (83.46)
8	Menopausal status	
	Premenopausal	36 (28.35)
	Postmenopausal	91 (71.65)
9	Age at menopause	
	≤45	39 (42.86)
	>45	52 (57.14)
10	ER status	
	Positive	92 (72.44)
	Negative	35 (27.56)
11	PR status	
	Positive	64 (50.39)
	Negative	63 (49.61)
12	Her2 status	
	Positive	61 (48.03)
	Negative	66 (51.97)
13	Molecular subtypes (Breast cancer)	
	Luminal A	45 (35.43)
	Luminal B	51 (40.16)
	Her2 enriched	17 (13.38)
	Triple negative breast cancer (TNBC)	14 (11.03)
14	Tumor size	
	≤5	59 (46.46)
	>5	68 (53.54)
15	Lymph node status	
	Positive	109 (85.83)
	Negative	18 (14.17)
16	TNM stage	
	I + II	36 (28.35)
	III + IV	91 (71.65)
17	Histological grade	
	I + II	102 (80.31)
	III	25 (19.69)

Inclusion Criteria

The study involved females with histopathologically confirmed primary breast cancer and having at least 6 months of life expectancy, lying between the age group 20–79 years. The participants provided consent to abide by the procedures of the study. All the females included in the study were registered in the medical record book of AIIMS, New Delhi, and their medical records were evaluated for studying various clinical and pathological parameters of the patients.

Real-Time Polymerase Chain Reaction

For the isolation of RNA, the breast cancer tissues, and the collected normal tissues were preserved in the RNAlater (Qiagen) kit, and afterward, RNA isolation was done by using TRIzo1Reagent (Invitrogen) by following the instruction provided by the manufacturer in the protocol. Thereafter, the complementary DNA (cDNA) from total RNA was synthesized using a cDNA kit (verso Thermo Fisher Scientific) and later storage was done at −20°C for further analysis. Subsequently, the above-prepared cDNA was used in the (qPCR) where amplification was carried out using Roche Light Cycler® 96 SYBR Green I Master mix. By applying the primers for *FOXO1*: sense 5′- CCACATTCAACAGGCAGCAG-3′ antisense 5′- GACGGAAACTGGGAGGAAGG-3′ which amplified a 152-bp. product. β actin gene was taken as an internal control and amplified in the same qPCR reaction. The primers used for qPCR reaction were sense 5′-AGATAGTGGATCAGCAAGCAG-3′ and antisense 5′-GCGAAGTTAGGTTTTGTCA-3′, which amplified a 160 bp. product. Standardized protocol of our laboratory (Real et al., 2018; Sadaf et al., 2018; Khan et al., 2020) was used to perform PCR. Measurements were taken in triplicates. The calculation was done for the relative amount of mRNA using Light Cycler 96 (Roche) equipped with Software 1.5. The calibrated normalized ratio was estimated as per the given standard formula: RQ = 2-ΔΔCq = [(Cq targeted gene–Cq βactin) calibration sample].

DNA Extraction

Phenol-chloroform Isoamyl (PCI) method was used for isolation of gDNA from Breast cancer and adjacent normal tissue (Russell and Sambrook, 2001). The quantity and quality evaluation of isolated genomic DNA was done using a Nanodrop spectrophotometer (ND1000), and agarose gel electrophoresis was further performed for validation.

Methylation Through MS-PCR

EZ DNA Methylation-Gold™ Kit was used to carry out Bisulfite conversion following the instruction given by the manufacturer. The converted product was amplified using dual sets of methylated and unmethylated *FOXO1* primers. Eukaryotic promoter database was used to retrieve the *FOXO1* gene promoter sequence, and MethPrimer software was used for primer designing (**Figure 1**). When searched by MethPrimer, the promoter region of the *FOXO1* gene was found to contain two CpG islands of 600 bp. The primer pairs that were used for the detecting methylation in the promoter region of *FOXO1* were as follows: sense 5′- GGAAAATCGGGT TTTATTTAGTTC-3′ and antisense 5′- GACTACTACGACTAC CAAACCGC-3′, for the unmethylated detection: sense 5′- TGG AAAATTGGGTTTTATTTAGTTT-3 and antisense 5′- CAACTA

TABLE 2 | Correlation study of *FOXO1* mRNA expression levels with clinical parameters of Breast Cancer case.

Characteristics	Total (*n= 127*)	*FOXO1* mRNA expression relative to beta actin (Mean ± S.E)	*p-Value*	Chi-squared
Age				
<50	44 (34.65)	0.58 ± 0.01		
≥50	83 (65.35)	0.94 ± 0.02	0.008*	6.93
Geographical location				
Rural	33 (25.98)	1.66 ± 0.02		
Urban	94 (74.02)	1.70 ± 0.21	0.95	0.003
Age of menarche				
≤12	20 (15.75)	1.21 ± 0.04		
>12	107 (84.25)	1.67 ± 0.01	0.13	2.27
Age at first live birth				
≤25	100 (78.74)	1.32 ± 0.05		
>25	27 (21.26)	1.20 ± 0.06	0.0003*	13.13
Breast feeding				
Yes	122 (96.06)	1.30 ± 0.02		
No	5 (3.94)	0.71 ± 0.01	0.64	0.21
Use of exogenous hormone				
Yes	6 (4.72)	1.37 ± 0.07		
No	121 (95.28)	1.19 ± 0.08	0.88	0.20
Family history of cancer				
Yes	21 (16.54)	0.91 ± 0.07		
No	106 (83.46)	1.70 ± 0.09	0.42	0.64
Menopausal Status				
Premenopausal	36 (28.34)	0.92 ± 0.00		
Postmenopausal	91 (71.66)	1.23 ± 0.01	0.40	0.68
Age at Menopausal				
≤45	39 (42.85)	1.78 ± 0.08		
>45	52 (57.15)	1.23 ± 0.02	0.94	0.004
Estrogen receptor status				
Negative	35 (27.56)	0.83 ± 0.02		
Positive	92 (72.44)	1.38 ± 0.21	0.58	0.29
Progesterone receptor status				
Negative	63 (49.61)	1.47 ± 0.33		
Positive	64 (50.39)	1.11 ± 0.04	0.36	0.81
Her2 neu Status				
Negative	66 (51.97)	1.18 ± 0.03		
Positive	61 (48.03)	1.74 ± 0.12	0.77	0.07
Tumor Size				
<5	68 (53.54)	1.17 ± 0.07		
≥5	59 (46.46)	1.21 ± 0.09	0.05*	3.54
Lymph Node Status				6.05
Positive	109 (85.83)	1.50 ± 0.01	0.01*	
Negative	18 (14.17)	0.88 ± 0.06		
TNM Staging				
Stage (I + II)	36 (28.35)	1.17 ± 0.001		
Stage (III + IV)	91 (71.65)	1.98 ± 0.02	0.20	1.58
Histological Grade				
(I + II)	102 (80.31)	2.01 ± 0.01	0.74	0.10
(III)	25 (19.69)	1.47 ± 0.07		
Molecular Subtypes				
Luminal A	45 (35.43)	1.52 ± 0.12		
Luminal B	51 (40.16)	1.33 ± 0.30	0.89	0.59
Her2neu Enriched	17 (13.38)	0.47 ± 0.01		
TNBC	14 (11.03)	1.21 ± 0.11		

TNBC, triple negative breast cancer; FOXO1, *Forkhead Box O1* [a] *Only Downregulated Cases were included.*

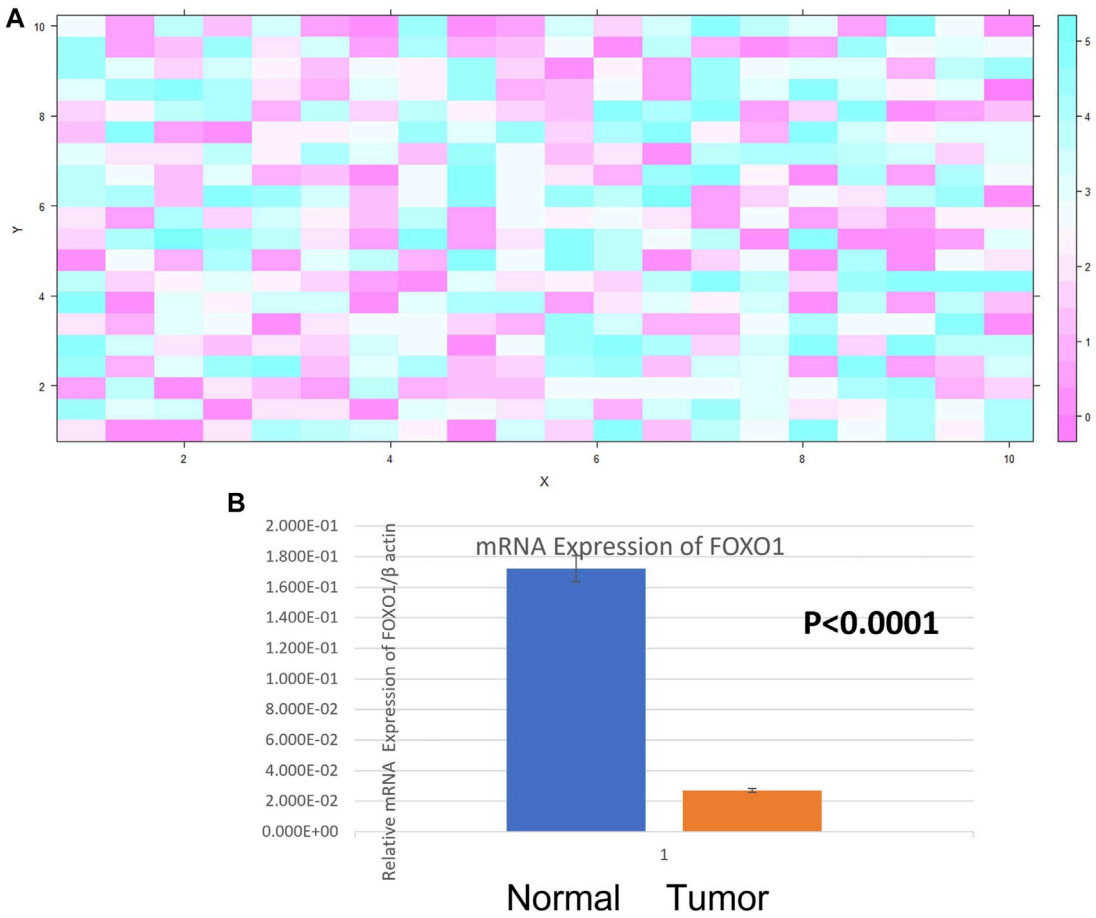

FIGURE 2 | (A) Heat Map plot (analyzed by R platform version 3.6.3 64-bit) of FOXO1 mRNA relative expression (fold change) in Breast cancer cases. *X*-axis depicts Δ Cq target against *Y*-axis Δ Cq control at default parameters. **(B)** Relative mRNA expression of FOXO1/*β* ACTIN in Breast tumor and adjacent normal tissue.

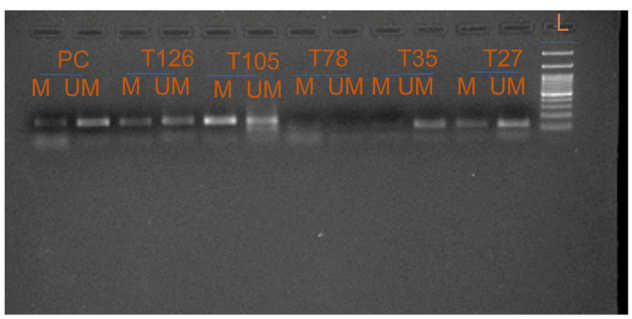

FIGURE 3 | Representative gel picture of Methylation—specific PCR analysis of FOXO1 gene in Breast cancer patients: DNA methylation was assessed using two specifically designed primers to amplify either methylated DNA (M) or unmethylated DNA (UM) (L: 100 bp DNA ladder; number indicates the case number; PC: Positive Control; T: Tumour tissue).

CTACAACTACCAAACCACC-3′. The product size for unmethylation was 172 bp, and for methylation, it was 170 bp. MS-PCR was done by the following condition: initially denaturation at 95°C for 5 min, 35 cycles of amplification at 95°C for the 30 s, annealing at 53.9°C (methylation) and 52.7°C (unmethylation) for 30 s, 72°C for 30 s and final extension at 72°C for 7 min. The pictures of the amplified product were obtained on a 2% agarose gel having EtBr, visualization was done under ultraviolet (UV) illumination with the Gel Doc (Bio-Rad Molecular Imaging System). The experiments were executed in triplicate without any disparity observed among the replicates.

Immunohistochemistry

Breast cancer tissue and adjacent normal tissue stored in formalin were used for Block preparation. Poly-L-lysine slides were used, and the sections of the block were taken on slides. Subsequently, different grades of xylene were used for deparaffinization, and rehydration was done using ethanol. The quenching of internal peroxide activity was done by 0.3% hydrogen peroxide, and citrate buffer boiling resulted in Ag withdrawal. To non-specific interaction of protein was ceased using serum solution as a blocking agent, and incubation at 4°C with primary antibody (CST#2880 *FOXO1*, 1:100) was done overnight. Furthermore, anti-rabbit biotinylated secondary antibody and streptavidin HRP incubation was done for 20–30 min, respectively. 3,3′-3,3′-Diaminobenzidine (DAB) method was used to visualize

TABLE 3 | Correlation study of *FOXO1* Promoter Methylation status with clinical parameters of Breast Cancer Patients.

Characteristics	Total cases (*n = 127*)	Methylated	Unmethylated	*p-Value*	Chi-squared
Age					
<50	44 (34.65)	29 (65.90)	15 (34.10)	0.44	0.57
≥50	83 (65.35)	49 (59.04)	34 (40.96)		
Geographical location					
Rural	33 (25.98)	19 (57.57)	14 (42.43)	0.59	0.27
Urban	94 (74.02)	59 (62.76)	35 (37.24)		
Age of menarche					
≤12	20 (15.75)	13 (65.00)	07 (35.00)	0.71	0.12
>12	107 (84.25)	65 (60.74)	42 (39.26)		
Age at first live birth					
≤25	100 (78.74)	57 (100)	43 (43)	0.04	3.87
>25	27 (21.26)	21 (77.77)	6 (22.22)		
Breast feeding					
Yes	122 (96.06)	75 (61.47)	47 (38.53)	0.94	0.004
No	5 (3.94)	03 (60)	2 (40)		
Use of exogenous hormone					
Yes	6 (4.72)	04 (66.66)	02 (33.33)	0.78	0.07
No	121 (95.28)	74 (61.15)	47 (38.85)		
Family history of cancer					
Yes	21 (16.54)	14 (66.66)	07 (33.33)	0.58	0.29
No	106 (83.46)	64 (60.37)	42 (39.63)		
Menopausal Status					
Premenopausal	36 (28.34)	27 (75)	09 (25)	0.04*	3.91
Postmenopausal	91 (71.66)	51 (56.04)	40 (43.96)		
Age at Menopausal					
≤45	39 (42.85)	23 (58.97)	16 (41.03)	0.66	0.19
>45	52 (57.15)	33 (63.46)	19 (36.54)		
Estrogen receptor status					
Negative	35 (27.56)	25 (71.42)	10 (28.58)	0.15	2.04
Positive	92 (72.44)	53 (57.60)	39 (42.40)		
Progesterone receptor status					
Negative	63 (49.61)	40 (63.49)	23 (36.51)	0.63	0.22
Positive	64 (50.39)	38 (59.37)	26 (40.63)		
Her2 neu Status					
Negative	66 (51.97)	43 (65.15)	23 (34.85)	0.36	0.80
Positive	61 (48.03)	35 (57.38)	26 (42.62)		
Tumor Size					
<5	68 (53.54)	46 (67.65)	22 (32.35)	0.12	2.39
≥5	59 (46.46)	32 (54.23)	27 (45.7)		
Lymph Node Status					
Positive	109 (85.83)	65 (59.63)	44 (40.37)	0.30	1.03
Negative	18 (14.17)	13 (72.23)	05 (27.77)		
TNM Staging					
Stage (I + II)	36 (28.35)	20 (55.55)	16 (44.46)	0.39	0.72
Stage (III + IV)	91 (71.65)	58 (63.74)	33 (36.26)		
Histological Grade					
(I + II)	102 (80.31)	60 (58.82)	42 (41.18)	0.22	1.47
(III)	25 (19.69)	18 (72)	07 (28)		
Molecular Subtypes					
Luminal A	45 (35.43)	26 (57.78)	19 (42.22)	0.66	1.59
Luminal B	51 (40.16)	30 (58.83)	21 (41.17)		
Her2neu Enriched	17 (13.38)	12 (70.59)	05 (29.41)		
TNBC	14 (11.03)	10 (71.43)	04 (28.57)		

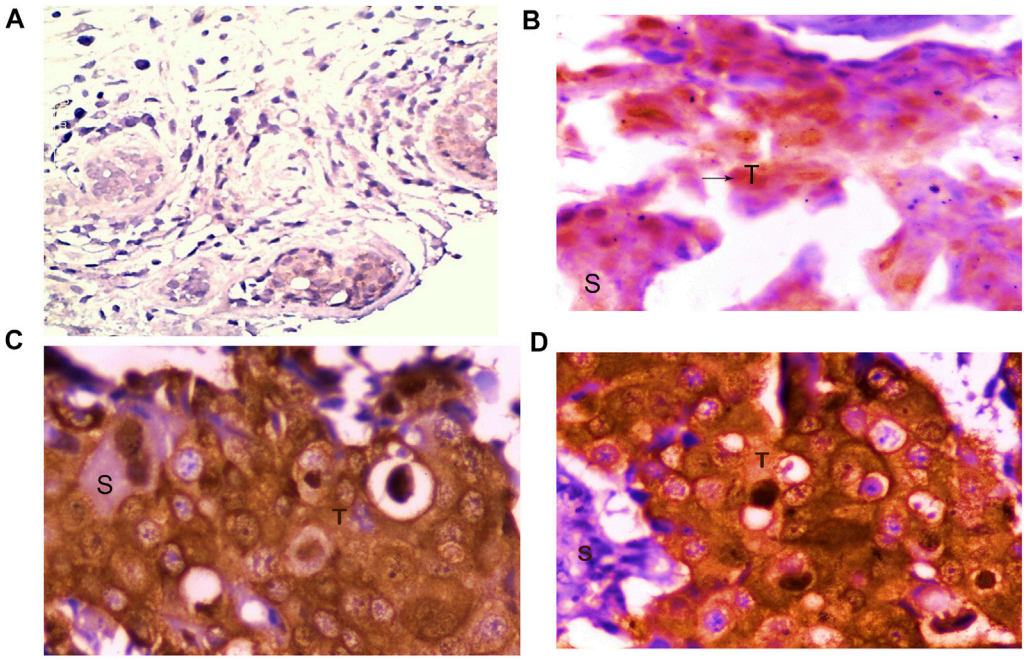

FIGURE 4 | Representative picture of Immunohistochemical staining of human breast cancer tissue samples by anti-FOXO1 antibody (magnification: ×400) showing **(A)** no expression, **(B)** low (+) expression, **(C)** moderate (++) expression, and **(D)** high (+++) expression of FOXO1. S: stromal tissue, T: tumor tissue.

antibody binding sites. Also, Hematoxylin counterstaining was carried out. For positive control, normal breast tissue was considered, while for the negative control, the primary antibody was skipped following the same protocol resulting in no staining. Interpretation of the staining was carried out under the guidance expert histopathologists using light microscope (magnification ×400), and the grading was done as follows: [1] 0% tumor staining with no expression, [2] 1%–10% tumor staining with mild expression (+), [3] 10%–50% tumor staining denoting moderate expression (++) [4] >50% tumor staining indicating high expression (+++ or ++++).

Statistical Analysis

SPSS-IBM (version 22.0) was utilized to find the relevant association with the clinicopathological parameters. The current study data are represented as mean ± standard error (SE). the *p*-value of less than 0.005 was considered significant. To evaluate the significance of differential *FOXO1* mRNA expression levels, a non-parametric test, i.e., Wilcoxon signed-rank test was used.

RESULTS

Downregulated *FOXO1* mRNA Expression in Breast Cancer Cases and its Correlation With Clinicopathological Parameters

The expression of *FOXO1* at the mRNA level was detected in case of breast cancer and adjacent normal tissue. Its expression was normalized against beta-actin expression. *FOXO1* mRNA expression was found to be downregulated in 69.29% cases (88/127) and out of which 73.80% cases (65/88) were categorized under histological grade I and II of breast cancer. The fold change of 88 downregulated cases was examined to be 5.16 the expression of *FOXO1* in breast cancer tissue was 1.16 ± 0.02 (Mean ± SE) and in the normal tissue was 1.95 ± 0.07 (Mean ± SE) ($p < 0.0001$). Correlating the *FOXO1* mRNA expression with different clinic pathological parameters of patients indicated significant association with Age, Age at first live birth, Tumor Size, Lymph Node Status (**Table 1**, **Table 2**, **Figure 2**).

FOXO1 Promoter Methylation and its Correlation With Clinical Parameters of Patients

Promoter methylation study of *FOXO1* promoter region was done through Methylation Specific PCR, the hypermethylated promoter region of *FOXO1* was found in 61.41% (78/127) cases. The correlation of promoter methylation with clinical parameters revealed a significant association with the menopausal status of breast cancer patients. In progressive stages III and IV of breast cancer, 58/91 cases were found to be methylated (**Figure 3**; **Table 3**.)

Low or No Expression of *FOXO1* Protein in Breast Cancer Tissue

FOXO1 expression analysis at the protein level was done by IHC and it was found to be absent in 77.95% (99/127) cases. However,

TABLE 4 | Correlation of *FOXO1* Protein Expression level with clinical parameters of Breast Cancer Patients.

Characteristics	Total cases (*n = 127*)	*FOXO1* absent	*FOXO1* present	*p Value*	Chi-squared
Age					
<50	44 (34.65)	36 (81.81)	08 (18.19)	0.44	0.58
≥50	83 (65.35)	63 (75.90)	20 (24.10)		
Geographical location					
Rural	33 (25.98)	29 (87.87)	04 (12.13)	0.10	2.56
Urban	94 (74.02)	70 (74.46)	24 (25.54)		
Age of menarche					
≤12	20 (15.75)	16 (80)	04 (20)	0.80	0.05
>12	107 (84.25)	83 (77.57)	24 (22.43)		
Age at first live birth					
≤25	100 (78.74)	78 (78)	22 (22)	0.98	0.001
>25	27 (21.26)	21 (77.77)	06 (22.23)		
Breast feeding					
Yes	122 (96.06)	96 (78.68)	26 (21.32)	0.32	0.97
No	5 (3.94)	03 (60)	02 (40)		
Use of exogenous hormone					
Yes	6 (4.72)	02 (33.33)	04 (66.66)	0.006*	7.29
No	121 (95.28)	97 (80.16)	24 (19.84)		
Family history of cancer					
Yes	21 (16.54)	18 (85.72)	03 (14.28)	0.34	0.88
No	106 (83.46)	81 (76.42)	25 (23.58)		
Menopausal Status					
Premenopausal	36 (28.34)	32 (88.88)	04 (11.12)	0.06	3.49
Postmenopausal	91 (71.66)	67 (73.62)	24 (26.38)		
Age at Menopausal					
≤45	39 (42.85)	25 (64.10)	14 (35.90)	0.80	0.06
>45	52 (57.15)	32 (61.53)	20 (38.47)		
Estrogen receptor status					
Negative	35 (27.56)	28 (80)	07 (20)	0.73	011
Positive	92 (72.44)	71 (77.17)	21 (22.83)		
Progesterone receptor status					
Negative	63 (49.61)	50 (79.36)	13 (20.64)	0.70	0.14
Positive	64 (50.39)	49 (76.56)	15 (23.44)		
Her2 neu Status					
Negative	66 (51.97)	58 (87.87)	08 (13.79)	0.005*	7.87
Positive	61 (48.03)	41 (67.21)	20 (32.79)		
Tumor Size					
<5	68 (53.54)	59 (86.76)	09 (13.24)	0.01*	6.61
≥5	59 (46.46)	40 (67.80)	19 (32.20)		
Lymph Node Status					
Positive	109 (85.83)	84 (77.06)	25 (22.94)	0.55	0.35
Negative	18 (14.17)	15 (83.33)	03 (16.67)		
TNM Staging					
Stage (I + II)	36 (28.35)	25 (69.44)	11 (30.56)	014	2.11
Stage (III + IV)	91 (71.65)	74 (81.31)	17 (18.69)		
Histological Grade					
(I + II)	102 (80.31)	76 (74.50)	26 (25.50)	0.05*	3.72
(III)	25 (19.69)	23 (92.00)	02 (8.00)		
Molecular Subtypes					
Luminal A	45 (35.43)	35 (77.77)	10 (22.23)	0.50	2.31
Luminal B	51 (40.16)	42 (82.35)	09 (17.65)		
Her2neu Enriched	17 (13.38)	11 (64.70)	06 (35.30)		
TNBC	14 (11.03)	11 (78.57)	03 (21.43)		

TABLE 5 | Correlation study of Promoter Methylation with Protein expression in Breast Cancer Patients from North India.

FOXO1 promoter	*FOXO1* protein expression		Total (%)	*p* Value	Chi-squared
	Absent	Present			
Methylated	74 (94.87)	04 (5.13)	78 (61.41)		
Unmethylated	25 (51.02)	24 (48.98)	49 (38.59)	0.0001	33.67
Total	99 (77.95)	28 (22.05)	127		

p *Value (Fischer's Exact Test).*

TABLE 6 | Correlation study of methylation and protein expression in samples having methylated *FOXO1* promoter or *FOXO1* expression loss with clinical parameters of Breast cancer patients from North Indian population.

Clinical characteristics		Total methylated (*n* = 78)	Methylated *FOXO1*		*p* Value	Chi-squared	Total (N)	*FOXO1* loss		*p* Value	Chi-squared
			FOXO1 absent	*FOXO1* present				Methylated *FOXO1*	Unmethylated *FOXO1*		
Age											
<50	44 (34.65)	29	26	03	0.10	2.58	36	26	10	0.66	0.91
≥50	83 (65.35)	49	48	01			63	48	15		
Geographical location											
Rural	33 (25.98)	19	17	02	0.22	1.50	29	17	12	0.01*	5.65
Urban	94 (74.02)	59	57	02			70	57	13		
Age of menarche											
≤12	20 (15.75)	13	10	03	0.001*	10.33	16	10	06	0.21	1.51
>12	107 (84.25)	65	64	01			83	64	19		
Age at first live birth											
≤25	100 (78.74)	57	56	01	0.02*	4.95	78	56	22	0.19	1.69
>25	27 (21.26)	21	18	03			21	18	03		
Breast feeding											
Yes	122 (96.06)	75	72	03	0.02*	5.10	96	72	24	0.74	0.10
No	5 (3.94)	03	02	01			03	02	01		
Use of exogenous hormone											
Yes	6 (4.72)	04	02	02	0.0001*	17.44	02	02	00	0.40	0.69
No	121 (95.28)	74	72	02			97	72	25		
Family history of cancer											
Yes	21 (16.54)	14	11	03	0.002	9.31	18	11	07	0.14	2.16
No	106 (83.46)	64	63	01			81	63	18		
Menopausal Status											
Premenopausal	36 (28.34)	27	26	01	0.67	0.17	32	26	06	0.30	1.05
Postmenopausal	91 (71.66)	51	48	03			67	48	19		
Age at Menopausal											
≤45	39 (42.85)	23	21	02	0.95	0.003	25	21	04	0.23	1.41
>45	52 (57.15)	33	30	03			32	30	02		
Estrogen receptor status											
Negative	35 (27.56)	25	24	01	0.75	0.09	28	24	04	0.11	2.48
Positive	92 (72.44)	53	50	03			71	50	21		
Progesterone receptor status											
Negative	63 (49.61)	40	39	01	0.28	1.16	50	39	11	0.45	0.56
Positive	64 (50.39)	38	35	03			49	35	14		

(Continued on following page)

TABLE 6 | (*Continued*) Correlation study of methylation and protein expression in samples having methylated *FOXO1* promoter or *FOXO1* expression loss with clinical parameters of Breast cancer patients from North Indian population.

Clinical characteristics		Total methylated (*n* = 78)	Methylated *FOXO1*		*p* Value	Chi-squared	Total (N)	*FOXO1* loss		*p* Value	Chi-squared
			FOXO1 absent	*FOXO1* present				Methylated *FOXO1*	Unmethylated *FOXO1*		
Her2 neu Status											
Negative	66 (51.97)	43	42	01	0.21	1.54	58	42	16	0.52	0.40
Positive	61 (48.03)	35	32	03			41	32	09		
Tumor Size											
<5	68 (53.54)	46	44	02	0.70	0.14	59	44	15	0.96	0.002
≥5	59 (46.46)	32	30	02			40	30	10		
Lymph Node Status											
Positive	109 (85.83)	65	62	03	0.64	0.21	84	62	22	0.61	0.25
Negative	18 (14.17)	13	12	01			15	12	03		
TNM Staging											
Stage (I + II)	36 (28.35)	20	19	01	0.97	0.001	25	19	06	0.86	0.02
Stage (III + IV)	91 (71.65)	58	55	03			74	55	19		
Histological Grade											
(I + II)	102 (80.31)	60	58	02	0.18	1.72	76	58	18	0.51	0.42
(III)	25 (19.69)	18	16	02			23	16	07		
Molecular Subtypes											
Luminal A	45 (35.43)	26	25	01	0.79	1.02	35	25	10	0.07	6.89
Luminal B	51 (40.16)	30	28	02			42	28	14		
Her2neu Enriched	17 (13.38)	12	11	01			11	11	00		
TNBC	14 (11.03)	10	10	00			11	10	01		

TABLE 7 | Correlation analysis between *FOXO1* methylation and *FOXO1* protein expression in stratification by various clinical characteristics of Breast cancer patients from North India.

Clinical characteristics	Total methylated (*n* = 78)	*FOXO1* methylation status	*FOXO1* expression		*p* Value	Chi-squared
			Absent	Present		
Age						
<50 44(34.65)	29	M	26	03		
		U	10	05	0.06	3.51
≥50 83(65.35)	49	M	48	01		
		U	15	19	0.0001*	31.84
Geographical location						
Rural 33(25.98)	19	M	17	02		
		U	12	02	0.74	0.10
Urban 94(74.02)	59	M	57	02		
		U	13	22	0.12	2.31
Age of menarche						
>12,107(84.25)	65	M	64	01		
		U	19	23	0.0001*	41.54
≤12 20(15.75)	13	M	10	03		
		U	06	01	0.63	0.22
Age at first live birth						
≤25,100(78.74)	57	M	56	01		
		U	22	21	0.0001*	31.66
>25 27(21.26)	21	M	18	03		
		U	03	01	0.59	0.28
Breast feeding						
Yes 122(96.06)	75	M	72	03		35.26
		U	24	23	0.0001*	
No 5(3.94)	03	M	02	01		0.13
		U	01	01	0.70	

(Continued on following page)

TABLE 7 | (*Continued*) Correlation analysis between *FOXO1* methylation and *FOXO1* protein expression in stratification by various clinical characteristics of Breast cancer patients from North India.

Clinical characteristics	Total methylated (*n* = 78)	*FOXO1* methylation status	*FOXO1* expression		*p* Value	Chi-squared
			Absent	Present		
Use of exogenous hormone						
Yes 6(4.72)	04	M	02	02		1.50
		U	00	02	0.22	
No 121(95.28)	74	M	72	02		
		U	25	22	0.0001*	35.16
Family history of cancer						
Yes 21(16.54)	14	M	11	03	0.18	1.75
		U	07	00		
No 106(83.46)	64	M	63	01	0.0001*	43.46
		U	18	24		
Menopausal Status						
Premenopausal 36(28.34)	27	M	19	01		
		U	06	10	0.0002*	13.85
Postmenopausal 91(71.66)	51	M	55	03		
		U	18	15	0.0001*	21.50
Age at Menopausal						
≤45 39(42.85)	23	M	21	02		18.02
		U	04	12	0.0001*	
>45 52(57.15)	33	M	30	03		
		U	02	17	0.0001*	32.91
Estrogen receptor status						
Negative 35(27.56)	25	M	24	01	0.0002*	14.00
		U	04	06		
Positive 92(72.44)	53	M	50	03		
		U	21	18	0.0001*	20.91
Progesterone receptor status						
Negative 63(49.61)	40	M	39	01	0.0001*	22.00
		U	11	12		
Positive 64 (50.39)	38	M	35	03		
		U	14	12	0.0004*	12.59
Her2 neu Status						
Negative 66(51.97)	43	M	42	01		
		U	16	07	0.0009*	11.11
Positive 61(48.03)	35	M	32	03		
		U	09	17	0.0001*	21.85
Tumor Size						
<5 68(53.54)	46	M	44	02		
		U	15	07	0.001*	9.78
≥5 59(46.46)	32	M	30	02		
		U	10	17	0.0001*	21.57
Lymph Node Status						
Positive 109(85.83)	65	M	63	03		
		U	22	21	0.0001*	29.74
Negative 18(14.17)	13	M	12	01		
		U	03	02	0.09	2.71
TNM Staging						
Stage (I + II) 36(28.35)	20	M	19	01		
		U	06	10	0.0002*	13.85
Stage (III + IV) 91 (71.65)	58	M	55	03		
		U	19	14	0.0001*	19.21
Histological Grade						
(I + II) 102(80.31)	60	M	58	02		
		U	18	24	0.0001*	37.66

(Continued on following page)

TABLE 7 | (*Continued*) Correlation analysis between *FOXO1* methylation and *FOXO1* protein expression in stratification by various clinical characteristics of Breast cancer patients from North India.

Clinical characteristics	Total methylated (*n = 78*)	*FOXO1* methylation status	*FOXO1* expression		*p* Value	Chi-squared
			Absent	Present		
(III) 25(19.69)	18	M	16	02		
		U	07	00	0.35	0.84
Molecular Subtypes						
		M	25	01	0.0005*	12.03
Luminal A 45(35.43)	26	U	10	09		
		M	28	02		
Luminal B 51(40.16)	30	U	14	07	0.01*	6.04
		M	11	01		
Her2neu Enriched 17(13.38)	12	U	00	05	0.0003*	12.98
		M	10	00		
TNBC 14(11.03)	10	U	01	03	0.002*	9.54

28 Cases exhibited the moderate or high expression of *FOXO1* protein. Also, the *FOXO1* protein expression pattern substantiated the mRNA expression. Furthermore, the percentage of *FOXO1* protein downregulation was significant with Her2 neu status, tumor size, and histological grade of breast cancer (**Figure 4**; **Table 4**.)

FOXO1 Promoter Methylation and its Association With Protein Expression

The results represented a strong correlation of *FOXO1* protein expression with the promoter methylation and 74 out of 78 hypermethylated cases showed low or no protein expression and 04 cases showed protein expression. In 51.02% (25/49) cases that showed no methylation had no protein expression. The cases which had downregulation of *FOXO1* showed 78.40% (69/88) hypermethylation while 23.07% (9/39) cases had moderate to high-level protein expression. highly significant *p*-value (*p= 0.0001*) was found between *FOXO1* methylation in the promoter region and protein expression, which represented a strong correlation **Tables 5**,**6**,**7**

DISCUSSION

FOXO subfamily of forkhead box transcription factor comprises four *FOXO* isoforms. *FOXO1* which is a member of this subfamily is the key target of insulin that inhibits its transcriptional events through nuclear exclusion (Kousteni, 2012). The prominent role of *FOXO1* is studied in the maintenance of tissue homeostasis at the time of various physiological as well as pathological conditions (Xing et al., 2018). In previous studies, the close association of lower *FOXO1* levels with human cancers such as hepatocellular carcinoma, colorectal cancer, pancreatic cancer, prostate cancer, and lung cancer have been demonstrated (Wu et al., 2012; Schwartz and Cote, 2016; Lou et al., 2017; Moeinifard et al., 2017).

In the current study, we examined the expression level of *FOXO1* in 127 breast cancer tissues taken along with adjacent normal tissues from the Indian female breast cancer patients. To analyze the *FOXO1* mRNA expression, we performed real-time PCR. Further, we studied *FOXO1* protein expression and its subcellular localization through immunohistochemistry and the epigenetic modulation in the promoter region of the *FOXO1* was analyzed using MS-PCR. While investigating our data for *FOXO1* mRNA expression we found 88 out of 127 cases (69.29%) exhibiting the downregulation at the *FOXO1* mRNA level. The downregulation of *FOXO1* mRNA in our data links positively to the previous studies which state the tumor suppressive role of the *FOXO1* gene in cancer progression (Myatt et al., 2010; Prasad et al., 2014). Further, on correlating our results we found a strong correlation of *FOXO1* mRNA expression with the age of the patient (*p = 0.008*), age at first live birth (*p = 0.0003*), tumor size (*p = 0.05*), and lymph node status (*p = 0.01*) of the breast cancer patients. *FOXO1* mRNA downregulation was earlier reported to show a significant association with the lymph node status and age of the patients in prostate cancer cases (Yang et al., 2021), our study also reveals this strong correlation of *FOXO1* mRNA expression with these clinical parameters in case of breast cancer patients.

The results of our study to detect the protein expression and localization of *FOXO1* protein reveals 77.95% (99/127) cases having low or no expression of *FOXO1* protein, whereas 22.05% (28/127) show moderate or high expression. The anti-proliferative role of *FOXO1* has been reported in case of cervical and prostate cancer, moreover, the enforced expression of *FOXO1* in endometrioid endometrial cancer cells and SiHa cells blocked the cell proliferation and decreased the tumorigenic activity (Goto et al., 2008; Zhang et al., 2015; Yang et al., 2021). Our results also suggest the tumor suppressive role of *FOXO1* in breast cancer and show a strong association of low protein expression with the histological grade (*p = 0.05*) and tumor size (*p = 0.01*) of breast cancer patients. The downregulated expression of *FOXO1* protein in the advanced stages (III and IV) of breast cancer suggests its repressive role in tumor progression and can be considered as the prognostic marker. Further, our investigation of *FOXO1* gene promoter methylation represented the hypermethylation (61.41%) of the *FOXO1* promoter region in most cases and a significant association (*p = 0.0001*) was found between the promoter methylation and protein expression of the *FOXO1*

gene. (**Table 5**, **6**). Also, the *FOXO1* promoter methylation exhibited significant association ($p = 0.04$) with the menopausal status of the female breast cancer patients, 65.38% (51/78) cases that showed promoter methylation were post-menopausal. Recent studies on Dysregulation of gene expression revealed the prominent role of epigenetics in gene silencing other than mutation, our study reveals *FOXO1* promoter methylation to be associated with low or no expression of *FOXO1* protein in breast cancer tissue in comparison to the adjacent normal tissue.

This study reports the tumor-suppressive role of FOXO1 in the case of Indian breast cancer patients and our data also suggested FOXO1 exhibited prognostic importance. Its downregulation is closely associated with the prognosis of the disease and different clinical parameters of the patients. We suggest that *FOXO1* can be taken as a biomarker in the case of breast cancer and further research can be carried out to find therapeutic strategies in targeting the *FOXO1* gene.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee Jamia Millia islamia and AIIMS New Delhi. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MK and SH took charge of the experimental design and execution. The experiments were executed by MK. The reagents, materials, and analysis tools were contributed by MK, Sa, SM, SD, MH, IA, NA, and SM. Manuscript preparation was carried out by MK and co-authors.

ACKNOWLEDGMENTS

Author MK acknowledge ICMR for providing fellowship (Fileno 3/2/2/23/2019/NCD-III.) and the author IA, is grateful to Scientific Research Deanship at King Khalid University, Abha, Saudi Arabia for their financial support through the Small Research Group Project under grant number (RGP.01-48-42).

REFERENCES

Cantley, L. C. (2002). The Phosphoinositide 3-kinase Pathway. *Science* 296 (5573), 1655–1657. doi:10.1126/science.296.5573.1655

Chan, C., Law, B., So, W., Chow, K., and Waye, M. (2017). Novel Strategies on Personalized Medicine for Breast Cancer Treatment: an Update. *Ijms* 18 (11), 2423. doi:10.3390/ijms18112423

Goto, T., Takano, M., Albergaria, A., Briese, J., Pomeranz, K. M., Cloke, B., et al. (2008). Mechanism and Functional Consequences of Loss of FOXO1 Expression in Endometrioid Endometrial Cancer Cells. *Oncogene* 27 (1), 9–19. doi:10.1038/sj.onc.1210626

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Kaymaz, Y., Oduor, C. I., Yu, H., Otieno, J. A., Ong'echa, J. M., Moormann, A. M., et al. (2017). Comprehensive Transcriptome and Mutational Profiling of Endemic Burkitt Lymphoma Reveals EBV Type-specific Differences. *Mol. Cancer Res.* 15 (5), 563–576. doi:10.1158/1541-7786.mcr-16-0305

Khan, M. A., Tiwari, D., Dongre, A., Sadaf, S., Mustafa, S., Das, C. R., et al. (2020). Exploring the P53 Connection of Cervical Cancer Pathogenesis Involving north-east Indian Patients. *PloS one* 15 (9), e0238500. doi:10.1371/journal.pone.0238500

Kousteni, S. (2012). FOXO1, the Transcriptional Chief of Staff of Energy Metabolism. *Bone* 50 (2), 437–443. doi:10.1016/j.bone.2011.06.034

Li, X., Kover, K. L., Heruth, D. P., Watkins, D. J., Moore, W. V., Jackson, K., et al. (2015). New Insight into Metformin Action: Regulation of ChREBP and FOXO1 Activities in Endothelial Cells. *Mol. Endocrinol.* 29 (8), 1184–1194. doi:10.1210/me.2015-1090

Lou, W., Liu, J., Gao, Y., Zhong, G., Chen, D., Shen, J., et al. (2017). MicroRNAs in Cancer Metastasis and Angiogenesis. *Oncotarget* 8 (70), 115787–115802. doi:10.18632/oncotarget.23115

Moeinifard, M., Hassan, Z. M., Fallahian, F., Hamzeloo-Moghadam, M., and Taghikhani, M. (2017). Britannin Induces Apoptosis through AKT-FOXO1 Pathway in Human Pancreatic Cancer Cells. *Biomed. Pharmacother.* 94, 1101–1110. doi:10.1016/j.biopha.2017.08.025

Myatt, S. S., Wang, J., Monteiro, L. J., Christian, M., Ho, K.-K., Fusi, L., et al. (2010). Definition of microRNAs that Repress Expression of the Tumor Suppressor Gene FOXO1 in Endometrial Cancer. *Cancer Res.* 70 (1), 367–377. doi:10.1158/0008-5472.can-09-1891

Prasad, S. B., Yadav, S. S., Das, M., Govardhan, H. B., Pandey, L. K., Singh, S., et al. (2014). Down Regulation of FOXO1 Promotes Cell Proliferation in Cervical Cancer. *J. Cancer* 5 (8), 655–662. doi:10.7150/jca.6554

Procaccia, S., Ordan, M., Cohen, I., Bendetz-Nezer, S., and Seger, R. (2017). Direct Binding of MEK1 and MEK2 to AKT Induces FOXO1 Phosphorylation, Cellular Migration and Metastasis. *Sci. Rep.* 7 (1), 1–16. doi:10.1038/srep43078

Real, S. A. S., Parveen, F., Rehman, A. U., Khan, M. A., Deo, S. V. S., Shukla, N. K., et al. (2018). Aberrant Promoter Methylation of YAP Gene and its Subsequent Downregulation in Indian Breast Cancer Patients. *BMC cancer* 18 (1), 711–715. doi:10.1186/s12885-018-4627-8

Russell, D. W., and Sambrook, J. (2001). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 112.

Sadaf, M., Habib, M., Khan, M. A., Najm, M. Z., Mallick, M. N., Sunita, K., et al. (2018). Hypermethylated LATS2 Gene with Decreased Expression in Female Breast Cancer: A Case Control Study from North India. *Gene* 676, 156–163. doi:10.1016/j.gene.2018.07.033

Schwartz, A. G., and Cote, M. L. (2016). Epidemiology of Lung Cancer. *Lung Cancer Personalized Med.* 79, 21–41. doi:10.1007/978-3-319-24223-1_2

Shi, M., Jiang, Y., Yang, L., Yan, S., Wang, Y. G., and Lu, X. J. (2018). Decreased Levels of Serum Exosomal miR-638 Predict Poor Prognosis in Hepatocellular Carcinoma. *J. Cell. Biochem.* 119 (6), 4711–4716. doi:10.1002/jcb.26650

Tang, C.-y., Zhu, L.-x., Yu, J.-d., Chen, Z., Gu, M.-c., Mu, C.-f., et al. (2018). Effect of β-elemene on the Kinetics of Intracellular Transport of D-Luciferin Potassium Salt (ABC Substrate) in Doxorubicin-Resistant Breast Cancer Cells and the Associated Molecular Mechanism. *Eur. J. Pharm. Sci.* 120, 20–29. doi:10.1016/j.ejps.2018.04.037

Viale, G. (2012). The Current State of Breast Cancer Classification. *Ann. Oncol.* 23, x207–x210. doi:10.1093/annonc/mds326

Wu, Z., Sun, H., Zeng, W., He, J., and Mao, X. (2012). Upregulation of MircoRNA-370 Induces Proliferation in Human Prostate Cancer Cells by Downregulating the Transcription Factor FOXO1. *PLoS One* 7 (9), e45825. doi:10.1371/journal.pone.0045825

Xing, Y.-q., Li, A., Yang, Y., Li, X.-x., Zhang, L.-n., and Guo, H.-c. (2018). The Regulation of FOXO1 and its Role in Disease Progression. *Life Sci.* 193, 124–131. doi:10.1016/j.lfs.2017.11.030

Yang, N., Wu, J., Zhang, T., Yang, F., Shao, J., He, C., et al. (2021). Clinical Evaluation of FOXO1 as a Tumor Suppressor in Prostate Cancer. *Comput. Math. Methods Med.* 2021 (1), 810. doi:10.1155/2021/8773423

Zhang, B., Gui, L. S., Zhao, X. L., Zhu, L. L., and Li, Q. W. (2015). FOXO1 Is a Tumor Suppressor in Cervical Cancer. *Genet. Mol. Res.* 14 (2), 6605–6616. doi:10.4238/2015.june.18.3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 09 May 2022
doi: 10.3389/fgene.2022.849834

Exploration of Potential Biomarker Genes and Pathways in Kawasaki Disease: An Integrated *in-Silico* Approach

Priyanka Srivastava[1]*, *Chitra Bamba*[1], *Rakesh Kumar Pilania*[2], *Anu Kumari*[1], *Rajni Kumrah*[2], *Archan Sil*[2] *and Surjit Singh*[2]

[1]*Genetic Metabolic Unit, Department of Pediatrics, Advanced Pediatric Centre, Post Graduate Institute of Medical Education & Research, Chandigarh, India,* [2]*Allergy Immunology Unit, Department of Pediatrics, Advanced Pediatric Centre, Post Graduate Institute of Medical Education & Research, Chandigarh, India*

Kawasaki disease (KD) is a common childhood systemic vasculitis with a special predilection for coronary arteries. Even after more than five decades of the initial description of the disease, the etiology of KD remains an enigma. This transcriptome data re-analysis study aimed to elucidate the underlying pathogenesis of KD using a bioinformatic approach to identify differentially expressed genes (DEGs) to delineate common pathways involved in KD. Array datasets from the Gene Expression Omnibus database were extracted and subjected to comparative meta-analysis for the identification of prominent DEGs. Fifteen hub genes with high connectivity were selected from these DEGs (*IL1B, ITGAM, TLR2, CXCL8, SPI1, S100A12, MMP9, PRF1, TLR8, TREM1, CD44, UBB, FCER1G, IL7R, and FCGR1A*). Of these 15 genes, five genes (*CXCL8, FCGR1A, IL1B, TLR2, and TLR8*) were found to be involved in neutrophil degranulation. To gain further insight into the molecular mechanism, a protein–protein network was established. Significantly enriched pathways based on the above-mentioned genes were mainly centered on biological regulation and signaling events. In addition, the pathway analysis also indicated that the majority of the DEGs in KD were enriched in systemic lupus erythematosus, suggesting a strong interplay between immunological and genetic factors in the pathogenesis of KD. These findings could significantly aid in identifying therapeutic targets and understanding KD biosignatures to design a biomarker panel for early diagnosis and severity of the disease.

Keywords: bioinformatics, biomarkers, hub genes, in-silico analysis, Kawasaki disease, microarray, transcriptomics analysis

OPEN ACCESS

Edited by:
Sajad Ahmad Dar,
Jazan University, Saudi Arabia

Reviewed by:
Arshad Jawed,
GE Healthcare Life Sciences, India
Mohd Wahid,
Jazan University, Saudi Arabia

****Correspondence:***
Priyanka Srivastava
srivastavapriy@gmail.com

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *06 January 2022*
Accepted: *11 April 2022*
Published: *09 May 2022*

Citation:
Srivastava P, Bamba C, Pilania RK, Kumari A, Kumrah R, Sil A and Singh S (2022) Exploration of Potential Biomarker Genes and Pathways in Kawasaki Disease: An Integrated in-Silico Approach. Front. Genet. 13:849834. doi: 10.3389/fgene.2022.849834

INTRODUCTION

Kawasaki Disease (KD) is an acute childhood febrile illness predominately affecting children below 5 years. It is a systemic vasculitis with a special predilection for coronary arteries (Newburger et al., 2004). It is now recognized as the most common cause of acquired heart disease in children in the developed world (Banday et al., 2021). The incidence of the KD ranges from 10 to 25 per 100,000 children below five years in North America and Europe. The highest incidence of KD is reported in children of Japanese ancestry; 350 per 100,000 children below 5 (Singh et al., 2015). Approximately

TABLE 1 | Search result GEO datasets using the ARGEOS web tool.

Accession	Organism	Type	Platform
GSE73461	*Homo sapiens*	Expression profiling by array	GPL10558
GSE73463	*Homo sapiens*	Expression profiling by array	GPL10558
GSE63881	*Homo sapiens*	Expression profiling by array	GPL10558
GSE68004	*Homo sapiens*	Expression profiling by array	GPL10558
GSE73462	*Homo sapiens*	Expression profiling by array	GPL6947
GSE73464	*Homo sapiens*	Expression profiling by array	GPL6947
GSE48498	*Homo sapiens*	Expression profiling by array	GPL570
GSE16797	*Homo sapiens*	Expression profiling by array	GPL570
GSE9864	*Homo sapiens*	Expression profiling by array	GPL6270
GSE9863	*Homo sapiens*	Expression profiling by array	GPL6271
GSE18606	*Homo sapiens*	Expression profiling by array	GPL6480
GSE109351	*Homo sapiens*	Expression profiling by array	GPL17586
GSE73577	*Homo sapiens*	Expression profiling by array	GPL4133

1% of all Japanese children would develop KD by the age of 10. Although many centers in India are now diagnosing children with KD on a regular basis, the vast majority of patients are perhaps still going undiagnosed and unreported (Singh et al., 2015). KD has the potential to cause severe complications and significant morbidity and mortality. We have reported that the mortality rate in our cohort has been 0.8%. This is significantly higher than mortality figures of ≤0.04% reported in developed countries (Singh et al., 2016).

Even after more than five decades of the initial description of the disease, the diagnosis of KD is still clinical and based on a constellation of clinical features. There is no pathognomonic laboratory feature. Differential diagnosis of KD includes several febrile illnesses in children, including viral infections (e.g., measles, Epstein Barr virus, and adenovirus), scarlet fever, toxic shock syndrome, and drug reactions like Stevens-Johnson syndrome or serum sickness. Clinical symptoms of KD include persistent fever for more than five days, rash, swelling of the dorsum of hands and feet, red strawberry tongue, conjunctival nonsuppurative injection, cervical lymphadenopathy, periungual peeling, and diffuse mucosal inflammation (McCrindle et al., 2017; CDC, 2020). Children with scarlet fever do not have lip changes and eye changes, which are present in KD. While rash in scarlet fever is characteristically sandpaper, KD can have a pleomorphic rash. Elevated antistreptolysin O (ASO) titer, positive throat culture for β hemolytic streptococcal group A, and brisk response to antimicrobials are other characteristic features of scarlet fever. Similarly, one can differentiate KD from other common differentials based on the constellation of clinical features (Jindal et al., 2019). Standard of care involves high-dose of intravenous immunoglobulin (IVIG) along with oral aspirin in the acute phase of the disease, which helps to resolve inflammation and minimize the risk of coronary arteries abnormalities (CAAs) (Oates-Whitehead et al., 2003). Approximately 10%–20% of children with KD doses did not respond to first-line therapy and persisted in having inflammation, known as IVIG resistant KD. Despite timely treatment, 3%–5% of patients with KD can develop CAAs. In patients having IVIG resistant KD, the chances of development of CAAs remain high (Ogata et al., 2013; Skochko et al., 2018; Brogan et al., 2020). KD remains the most common cause of acquired heart disease among children in Japan, North America, and Europe (Banday et al., 2021).

Multiple theories regarding the etiology of KD have been hypothesized such as infectious theory, infectious plus autoimmunity, RNA virus theory, and superantigen theory. However, none of them has been independently able to provide an exact mechanism for the initiation and progression of pathogenic mechanisms in KD (Ramphul and Mejias, 2018; Nakamura et al., 2019). The present consensus strongly indicates a complex interplay of infectious triggers in children with a genetic predisposition of KD, followed by an abnormal immune response (Dietz et al., 2017; Nakamura et al., 2019).

Over the last few years, high-throughput platforms such as microarray have emerged as potent tools to detect differential gene expression profiles and have been recognized as significant clinical approaches with efficient diagnosis at the molecular level, prognostic prediction, stratification of patients, and identification of therapeutic targets. To date, quite a number of genes have been examined in KD, which were usually selected based on information regarding their function in inflammation, immune response, and other biological mechanisms. Recently, gene expression profiling studies have been conducted to understand the pathophysiology of KD (Hoang et al., 2014; Dietz et al., 2017; Wu et al., 2019; Rahmati et al., 2020; Gao et al., 2021); however, the prognostic power of identified genes still remains unclear. Therefore, understanding the pathogenesis of KD is essential for identifying novel pathways that can be targeted for therapy. The current study was conducted to find out the most important hub genes among the known ones.

In our recent study, we explored the role of epigenetic factors in modulating the gene expression in KD (Sharma et al., 2021). Transcriptomic profiles in KD can reveal perturbations caused by inflammation or infection, which could help to delineate the pathogenesis of KD. In the present study, we extracted a large-scale expression dataset of a similar microarray-based platform to investigate the pathogenesis of KD. Using an integrated transcriptomic approach, we identified differentially expressed genes (DEGs) in KD. The underlying pathways were recognized using functional annotation and *via* constructing protein–protein interaction (PPI) network. This data from the present study could help to conceptualize the molecular events underpinning KD and identify diagnostic biosignature and design a biomarker panel.

MATERIALS AND METHODS

Microarray Data

Data of microarray samples with suitable gene expression was obtained from the NCBI Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) (Edgar et al., 2002). Using the keywords "Kawasaki Disease and *Homo sapiens*," a thorough search of the GEO database was performed. Furthermore, the ARGEOS web tool (https://argeos.org) was also used for the selection of datasets from various public databases. After a systematic review, we included GSE73461, GSE63881, GSE73463, and GSE68004, GEO datasets that belong to the same array platform

TABLE 2 | List of the datasets included in the study.

Accession	Samples	Type	Summary	References
GSE73461	459	Illumina HumanHT-12 V4.0 expression beadchip Platform: GPL10558	Genome-wide analysis of transcriptional profiles in children <17 years of age with inflammatory diseases, bacterial or viral infections, or with clinical features suggestive of an infection	Wright et al. 2018
GSE73463	233			
GSE63881	341		Transcriptional profiles in KD patients at acute and convalescent phases with different clinical outcomes were investigated	Hoang et al., 2014
GSE68004	162		1) To define the transcriptional signature of KD that can aid in the diagnosis of complete and incomplete KD in children; 2) to identify specific biomarkers that objectively discriminate between KD and other mimicking conditions, including HAdV and 3) to test the prognostic utility of GEP to determine response to IVIG therapy and development of coronary artery abnormalities (CAAs)	Jaggi et al., 2018

TABLE 3 | Top 25 upregulated genes of the microarray meta-analyses along with their fold change values.

Upregulated genes				FDR
Gene symbol	Gene name	Log ratio combined	Fold change combined	
PPM1A	Protein phosphatase, Mg^{2+}/Mn^{2+} dependent 1A	1.5612	36.408	0.001274
KIR2DL5A	Killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 5A	1.5063	32.085	0.004013
PPFIBP1	PTPRF interacting protein, binding protein 1 (liprin beta 1)	1.4856	30.594	0.000258
MBD3L5	Methyl-CpG binding domain protein 3-like 5	1.4677	29.354	0.0185
OSGIN1	Oxidative stress–induced growth inhibitor 1	1.3628	23.057	0.0319
FCAR	Fc fragment of IgA receptor	1.0824	12.089	0.000323
MCEMP1	Mast cell-expressed membrane protein 1	1.0476	11.158	0
HP	Haptoglobin	1.0464	11.127	0
FAM177A1	Family with sequence similarity 177 member A1	1.0097	10.227	0.000461
ANXA3	Annexin A3	1.0074	10.171	0
KCNJ15	Potassium channel, inwardly rectifying subfamily J, member 15	0.9882	9.731	5.19E-13
CEP170	Centrosomal protein 170 kDa	0.896	7.87	0.0192
FCGR1A	Fc fragment of IgG, high-affinity Ia, receptor (CD64)	0.8823	7.625	0
HMX3	H6 family homeobox 3	0.8658	7.342	0
SOD2	Superoxide dismutase 2, mitochondrial	0.8651	7.33	2.84E-06
S100A12	S100 calcium-binding protein A12	0.8571	7.196	0
H2AC20	Histone cluster 2, H2ac	0.8223	6.642	0
LIMK2	LIM domain kinase 2	0.8072	6.415	0.0275
C19orf38	Chromosome 19 open reading frame 38	0.7805	6.033	0.0396
RBMS1	RNA binding motif, single-stranded interacting protein 1	0.7598	5.751	0.00032
CYSTM1	Cysteine-rich transmembrane module containing 1	0.7549	5.687	0
ALPL	Alkaline phosphatase, liver/bone/kidney	0.7503	5.627	0
IL1R2	Interleukin 1 receptor, type II	0.7208	5.258	0
C1QB	Complement component 1, q subcomponent, B chain	0.7202	5.251	0
CCPG1	Cell cycle progression 1	0.6979	4.988	0.0239

GPL10558. The PRISMA checklist was followed for this study (**Supplementary Figure S1**).

Differentially Expressed Genes Screening

For microarray expression data analysis ExAtlas meta-analysis software was used (Sharov et al., 2015). The four GEO datasets included in the present study were extracted from the GEO database. Data normalization was carried out using the quantile method. From the sample file, unpaired samples were removed, followed by the generation of the gene expression matrix file. Using correlation of log10-transformed expression level with other data, the quality of individual samples and level of global standard deviation were evaluated with respect to a set of pre-selected housekeeping genes. Low-quality samples were removed from the datasets.

For statistical analysis, ExAtlas uses the algorithm applied in the NIA Array Analysis (Zhou et al., 2019). The significance of gene expression change was assessed by false discovery rate (FDR) instead of *p*-values.

Standard Meta-Analysis

In the pairwise comparison section of ExAtlas, one of the KD expression profiles was added as a sample for examination, and for baseline control, its adjacent comparative pair was added. Then, more gene expression profile pairs were added using the meta-analysis section. Furthermore, to perform the meta-analysis, a random-effect model was used to take into account the variance of heterogeneity between the studies. For each, gene symbol analysis was performed, and their effect

TABLE 4 | Top 25 downregulated genes of the microarray meta-analyses along with their fold change values.

Downregulated genes				FDR
Gene symbol	**Gene name**	**Log ratio combined**	**Fold change combined**	
RPL17P43	Ribosomal protein L17 pseudogene 43	–2.1741	149.314	7.33E-05
LRRC4B	Leucine-rich repeat-containing 4B	–1.9794	95.367	0.00244
OXNAD1	Oxidoreductase NAD-binding domain containing 1	–1.8402	69.215	0.006463
NR1I2	Nuclear receptor subfamily 1 group I member 2	–1.6318	42.835	0.008047
ASS1P13	Argininosuccinate synthetase 1 pseudogene 13	–1.2125	16.312	0.0476
GNLY	Granulysin	–0.9048	8.032	0.003075
SAMD3	Sterile alpha motif domain containing 3	–0.8114	6.477	0.006021
ITM2C	Integral membrane protein 2C	–0.7768	5.981	0.008414
SIRPG	Signal-regulatory protein gamma	–0.7626	5.789	0.00083
PYHIN1	Pyrin and HIN domain family member 1	–0.7397	5.492	0.0246
IL7R	Interleukin 7 receptor	–0.6883	4.879	0.000688
TLE5	Amino-terminal enhancer of split	–0.594	3.927	0.000246
TMEM204	Transmembrane protein 204	–0.5689	3.706	0.0214
TXNDC5	Thioredoxin domain containing 5 (endoplasmic reticulum)	–0.5602	3.633	0.0247
GNLY	Granulysin	–0.5543	3.583	1.47E-10
FAM83A-AS1	FAM83A antisense RNA 1	–0.5483	3.534	3.61E-06
RNF213	Ring finger protein 213	–0.5095	3.232	6.75E-05
SLAMF6	SLAM family member 6	–0.5078	3.219	0.0473
WASH3P	WAS protein family homolog 3 pseudogene	–0.488	3.076	0.0071
ADGRG1	Adhesion G protein-coupled receptor G1	–0.4751	2.986	1.35E-09
GZMK	Granzyme K	–0.4613	2.893	3.27E-07
POLR3GL	Polymerase (RNA) III (DNA directed) polypeptide G (32kD)-like	–0.4545	2.848	2.25E-05
SERPINA13P	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 13, pseudogene	–0.4539	2.844	0.000964
IL32	Interleukin 32	–0.4476	2.803	1.8E-06
RUNDC3A	RUN domain containing 3A	–0.4169	2.612	0.0124

was represented in terms of combined log-10 ratios and fold changes. FDR <0.05, p-value ≤0.05 and change of ≥2-folds in gene expression were considered significant.

Protein–Protein Interaction Network Construction and Hub Gene Identification

Protein–protein interaction (PPI) network analysis was performed using the Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/). To assess possible PPI correlations, previously identified DEGs were mapped to the STRING database, followed by the extraction of PPI pairs. Cytoscape software v3.9.0 (https://cytoscape.org/) with the CytoHubba plugin was then employed to visualize the PPI network. In our study, the top 15 genes were considered as hub genes.

Pathway Enrichment Analysis

The biological processes that are involved with the DEGs, along with the functional enrichment analysis, were studied using the BINGO app (Maere et al., 2005) of Cytoscape. A hypergeometric test was carried out using Benjamini and Hochberg FDR correction. The Gene Ontology (GO) biological process was selected as the ontology file for executing enrichment analyses. Furthermore, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (Kanehisa and Goto, 2000) enrichment analysis was performed using the online tool Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Version 6.8; https://david.ncifcrf.gov/home.jsp) (Huang et al., 2009).

RESULTS

A systematic search of the studies was carried out up to December 2021. The search with the keywords "Kawasaki Disease" in *Homo sapiens* and "expression profiling by array" resulted in 13 microarray gene expression datasets (**Table 1**). We have combined those studies which implement the same methodology/platform and where >2 datasets were available. From 13 data sets, we have chosen four datasets which belong to the same platform GPL10558 (**Table 2**).

Meta-Analysis

The meta-analysis identified overall 74 differentially over-expressed and 52 repressed genes in KD compared to the matched adjacent control samples. A list of the top 25 upregulated genes and 25 downregulated genes is given in **Tables 3** and **4**.

Protein–Protein Interaction Network

The PPI network analysis was performed to look for the physical and functional links of proteins encoded by the identified DEGs in KD. A confidence score cut-off >0.9 was considered to construct the PPI network (**Figure 1**).

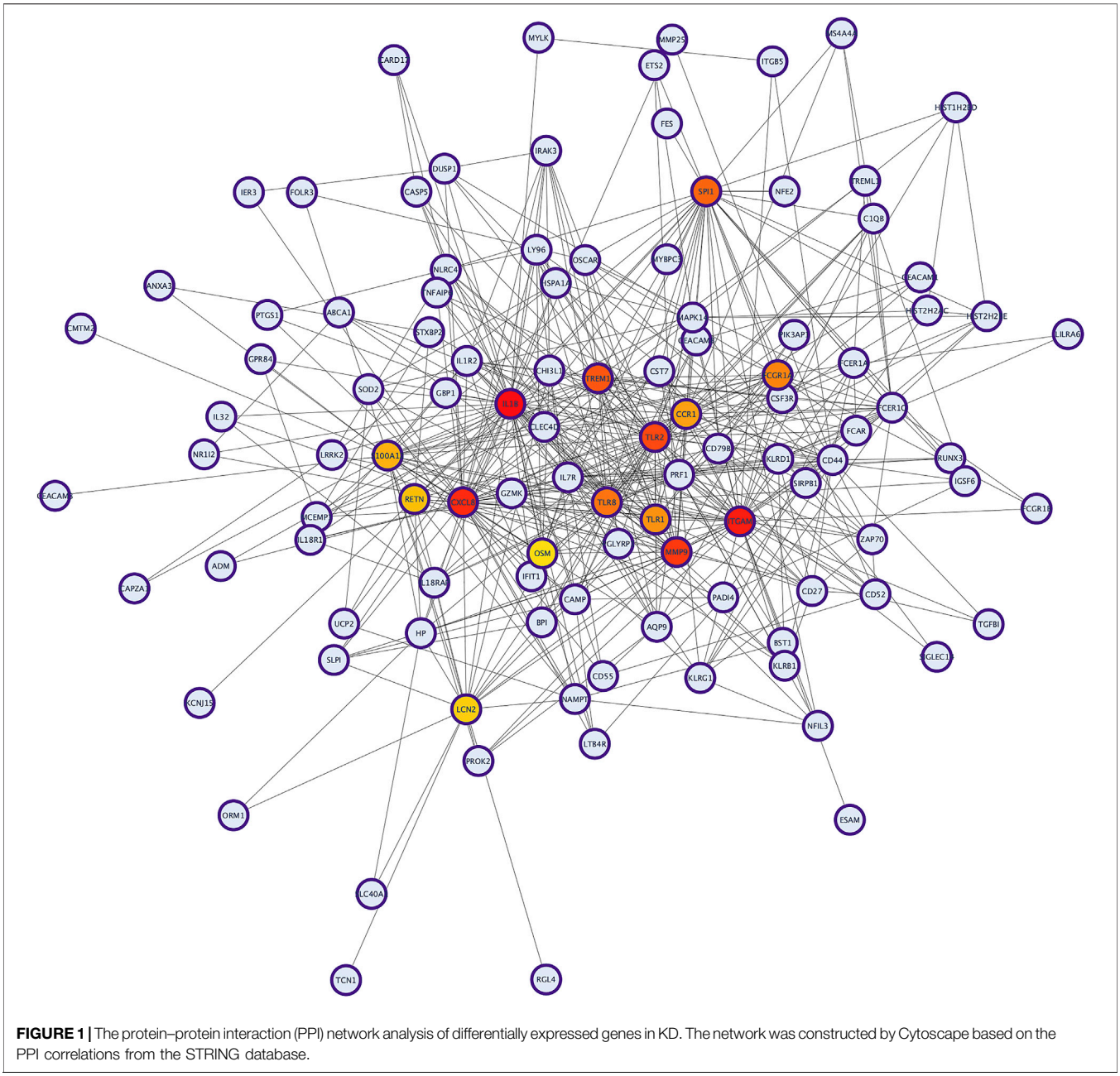

FIGURE 1 | The protein–protein interaction (PPI) network analysis of differentially expressed genes in KD. The network was constructed by Cytoscape based on the PPI correlations from the STRING database.

To identify hub proteins in the PPI network, the degree of each node in the network was calculated by using the Network Analyzer tool of Cytoscape software. The degree of a node is defined as the number of edges connected to the node. Nodes with higher degrees play a crucial role in the organization of the PPI network, and therefore they might be more crucial and relevant than non-hub genes (Vallabhajosyula et al., 2009). In this study, nodes with degrees >15 are considered to indicate "hub proteins" and are presented in **Table 5**. Interleukin 1-Beta (*IL1β*) had the highest degree of node (54), followed by Integrin αM (*ITGAM*) and Toll-like receptor-2 (*TLR2*), with their degrees of nodes being 44 and 42, respectively.

Pathway Enrichment Analyses

The BINGO plugin of Cytoscape was used to perform Gene Ontology (GO) functional enrichment analyses (**Figure 2**). Yellow nodes are significantly over-represented, while the white nodes are not significantly over-represented and are included only to show the yellow nodes in reference to the GO hierarchy. The size of a node is proportional to the number of query genes that are annotated to the corresponding GO category. The top 20 GO categories based on their respective node sizes, which are significantly over-represented in our study, are listed in **Table 6**. Among these significantly over-represented categories, the highest node size was reported for the biological regulation pathway

TABLE 5 | Top 15 hub genes with high degrees of connectivity in patients with KD.

Genes	Betweenness centrality	Closeness centrality	Degree
IL1B	0.19793249	0.43856333	54
ITGAM	0.08124394	0.40347826	44
TLR2	0.10349151	0.41281139	42
CXCL8	0.07645807	0.40773286	42
SPI1	0.08767911	0.38538206	34
S100A12	0.07193145	0.36535433	29
MMP9	0.0400542	0.38410596	27
PRF1	0.05125956	0.37001595	27
TLR8	0.01113809	0.36535433	26
TREM1	0.01784325	0.35474006	25
CD44	0.08380828	0.38157895	25
UBB	0.10610165	0.33237822	24
FCER1G	0.01477434	0.33819242	23
IL7R	0.01549726	0.3625	23
FCGR1A	0.01234507	0.35474006	21

followed by a response to stimulus and signaling. It is interesting to find that 25 genes (*ANXA3, BPI, C1QB, CCR1, CD55, CLEC4D, CXCL8, FCGR1A, IL18R1, IL18RAP, IL1B, IL32, LCN2, LILRA6, LTB4R, LY96, PGLYRP1, PRDX2, RAB27A, SH2D1A, TLR1, TLR2, TLR8, TREML1, and TRIM25*) were found to be common in top 5 pathways. This shows the importance of these genes in the flow of information in reference to the pathophysiology of the diseases. Of these 25 genes, five genes are common in between hub genes; *CXCL8, FCGR1A, IL1B, TLR2,* and *TLR8*.

Gene Set Overlap for Upregulated Genes

As shown in **Figure 3** and **Table 7**, the analysis identified significantly overrepresented categories (FDR<0.05) of GO molecular function for the upregulated genes. These include "neutrophil degranulation," "tertiary granule membrane," and "IgG binding." Among them, "neutrophil degranulation" and "tertiary granule membrane" were the most significantly enriched biological processes (**Figure 3**), while most of the proteins encoded by the upregulated genes (number = 50) were found to be located in the "neutrophil degranulation" (**Table 7**).

Kyoto Encyclopedia of Genes and Genomes Pathway Analysis

KEGG pathway analysis indicated that the majority of the DEGs in KD was enriched in systemic lupus erythematosus (SLE) (**Table 8**).

DISCUSSION

Despite recent advancements in clinical research, the exact etiology of KD is still poorly understood, and there is no laboratory test for confirmation of diagnosis, especially when the clinical presentation is incomplete or atypical. As a result, early diagnosis of KD may be different to establish. Growing evidence from the immunological, molecular, genetic, and epigenetic studies reveal that epidemiology might differ based on the infectious trigger for KD, especially in some children with a predilection based on genetic factors (Agarwal and Agrawal, 2017; McCrindle et al., 2017; Noval Rivas and Arditi, 2020; Sharma et al., 2021). In such individuals, a putative primary immune response may be activated in mucosal lymphoid tissues. This is then followed by activation of a cytokine cascade resulting in an uncontrolled systemic immune response. Inflammation of medium-sized vessels (especially the coronary arteries) probably results from the extravasation of immune cells in the sub-endothelium. Involvement of coronary arteries results in the development of coronary artery abnormalities (CAAs) (Senzaki, 2006). Therefore, early diagnosis of KD is essential for management decisions in children with KD.

In the last decade, several studies have been conducted to identify the genetic markers for KD; however, a consistent molecular marker having a significant prognostic value has not yet been determined. Furthermore, it is known that despite treatment, approximately 3%–5% of patients with KD would still develop CAAs. Therefore, understanding the pathogenesis of KD is essential for identifying novel pathways that can be targeted for therapy. The current study used an integrated transcriptomic analysis approach to identify DEGs with significant expression change in KD. These molecular biomarkers may be used as a novel multifunctional biomarker panel for diagnosis and therapeutic targets in patients with KD.

Recently, several studies have been conducted on gene expression and genome-wide association in patients with KD. However, there is no consensus on specific genes or biological pathways responsible for the pathogenesis of KD (Popper et al., 2007; Burgner et al., 2009; Onouchi et al., 2012; Xie et al., 2018; Chen et al., 2020; Johnson et al., 2021). The present study aims to pin down specific genes using large-scale datasets on microarray profiles from the GEO database and with the aid of bioinformatics provide key hub genes. Genome-wide expression data in whole-blood samples of KD patients and in controls were obtained using four different GSE datasets with the same array platform, and an integrated bioinformatic analysis was performed. As a result, 126 shared DEGs were screened out and these were mainly enriched in immune and inflammatory responses. The identified 15 hub genes were all significantly upregulated in KD. After careful inspection, we found that five of the 15 hub genes, including *ITGAM, MMP9, S100A12, TLR2,* and *FCER1G,* were involved in the top GO item neutrophil degranulation.

The top 10 hub genes from the analyzed data sets in our study were *IL-1β, ITGAM, TLR2, CXCL8, SPI1, S100A12, MMP9, PRF1, TLR8,* and *TREM1*. Of these, IL-1β had the highest degree of a node (Lau et al., 2008). IL-1β is a potent proinflammatory cytokine that has been responsible for chronic inflammatory conditions such as cardiovascular disease, coronary artery lesions, and vasculopathy, relevant to the pathogenesis of KD (Lee et al., 2012). Enhanced circulating levels of IL-1β have been reported in KD patients as compared to controls (Hoang et al., 2014; Porritt et al., 2020). Upregulated IL-1β expression is associated with IVIG resistance (Weng et al., 2010; Fu et al., 20192019). Studies have shown that administration of IL-1 receptor

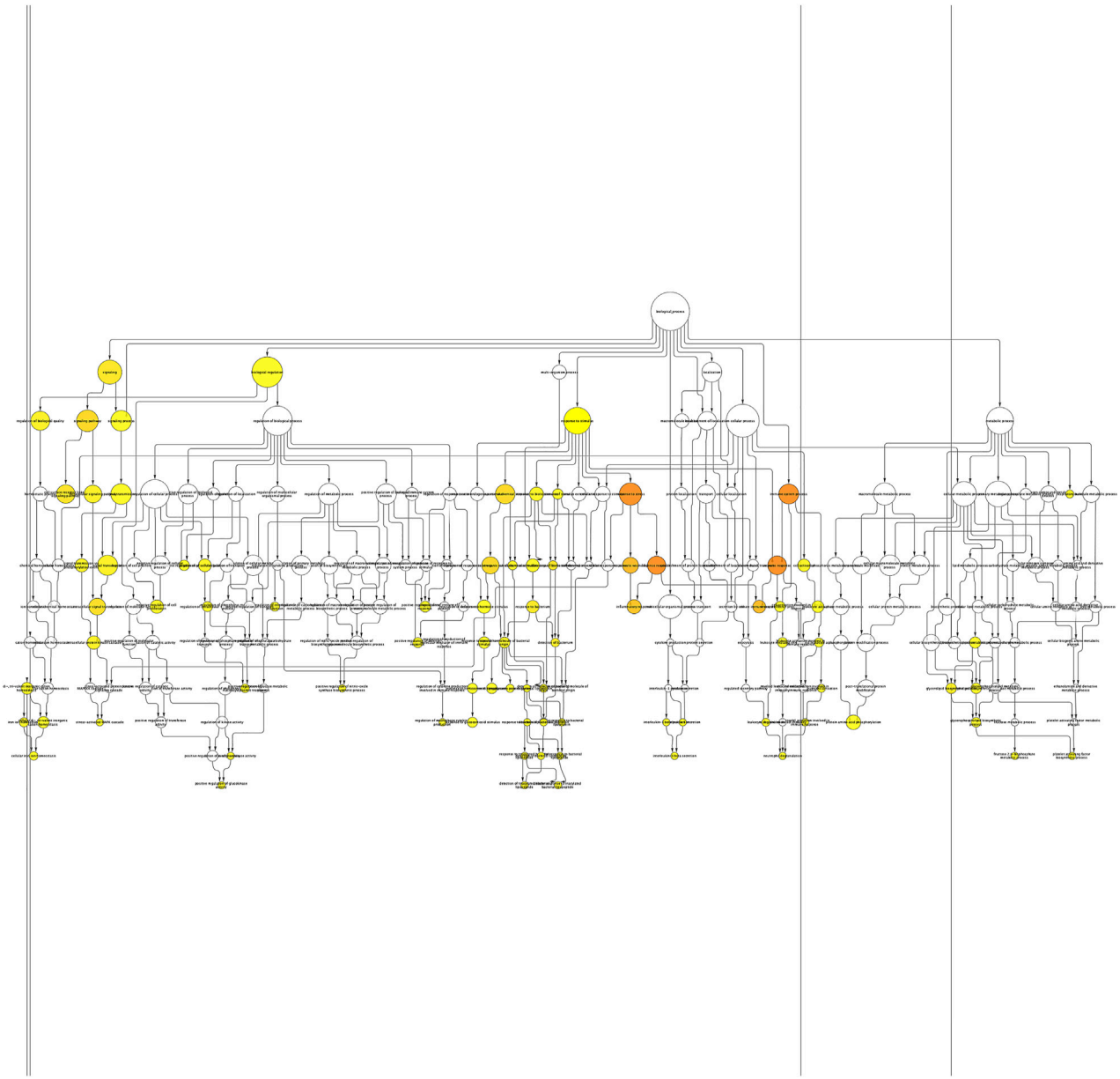

FIGURE 2 | Enrichment network of the shared DEGs based on biological processes. Biological process network of differentially expressed genes of KD patients using the BINGO app of Cytoscape. Large nodes indicate more genes involved, and the size of a node is proportional to the number of targets in the GO category. Yellow nodes indicate the genes playing a significant role in KD: p-value < 0.05.

antagonist (anakinra) in the KD mice model effectively prevents the development of coronary artery aneurysm, vasculitis, and myocarditis (Lee et al., 2012; Wakita et al., 2016). Based on the genetic and transcriptomic studies and evidence from the mice-model studies, anakinra is undergoing clinical trials (clinicaltrials.gov: NCT02179853) for the treatment of patients with KD.

The KEGG pathway analysis indicated that most DEGs were enriched in SLE and may thus be immune response-associated genes. While overlap of SLE and KD has been occasionally reported in the literature, this is extremely unusual. Both KD and SLE are immune-mediated disorders characterized by distinctive clinical features. However, some clinical findings are common to both conditions. These include fever, lymphadenopathy, arthritis or arthralgia, ocular and mucosal manifestations, rash, and multisystemic involvement. In addition, there appears to be a strong interplay between immunological and genetic factors in the pathogenesis of both KD and SLE (Guleria et al., 2020; Saez-de-Ocariz et al., 2020). *ITGAM* (Integrin αM) was one such gene with the second highest degree of nodes (44) in our analysis. *ITGAM* is associated with SLE (Han et al., 2009) and is found to be upregulated in KD

TABLE 6 | List of top 20 significantly overrepresented GO categories derived from the BINGO analysis output, based on our data. The list has been arranged in descending order of node size.

Name	Description GO	Average shortest path length	Betweenness centrality	Closeness centrality	Node size	Adjusted *p* value	No. of genes
65,007	Biological regulation	4.07075472	0.41100745	0.24565469	23.7486842	0.0394	43
50,896	Response to stimulus	4.27830189	0.24678813	0.2337376	20.4939015	0.0000	40
23,052	Signaling	4.65566038	0.02528647	0.2147923	17.8885438	0.0043	50
6,950	Response to stress	5.17924528	0.03215335	0.19307832	16.3707055	0.0000	67
23,033	Signaling pathway	5.52358491	0.01811721	0.18104184	15.7480157	0.0009	105
23,060	Signal transmission	5.16509434	0.02117959	0.19360731	14.832397	0.0314	28
23,046	Signaling process	4.58018868	0.02435299	0.21833162	14.832397	0.0314	20
7,165	Signal transduction	5.52358491	0.02548615	0.18104184	14.4222051	0.0164	34
2,376	Immune system process	4.55660377	0.07892457	0.2194617	14.1421356	0.0000	62
42,221	Response to chemical stimulus	5.14622642	0.01607496	0.19431714	13.8564065	0.0009	48
65,008	Regulation of biological quality	4.95283019	0.03262659	0.20190476	13.4164079	0.0162	32
6,955	Immune response	5.0990566	0.03425987	0.19611471	13.114877	0.0000	42
7,166	Cell surface receptor linked signaling pathway	6.51886792	0	0.15340087	12.9614814	0.0031	80
6,952	Defense response	6.0754717	0.00547133	0.16459627	12.6491106	0.0000	37
23,034	Intracellular signaling pathway	6.10377358	0.00418892	0.16383308	12.1655251	0.0132	45
10,033	Response to organic substance	5.94811321	0.00434888	0.16812054	11.6619038	0.0009	52
35,556	Intracellular signal transduction	6.33490566	0.01587904	0.15785555	11.3137085	0.0019	55
9,611	Response to wounding	6.16509434	0.00429051	0.16220352	10.5830052	0.0000	55
6,468	Protein amino acid phosphorylation	8.49528302	3.87E-04	0.11771238	9.38083152	0.0438	141
6,954	Inflammatory response	7.06132075	4.47E-05	0.14161657	8.94427191	0.0001	22

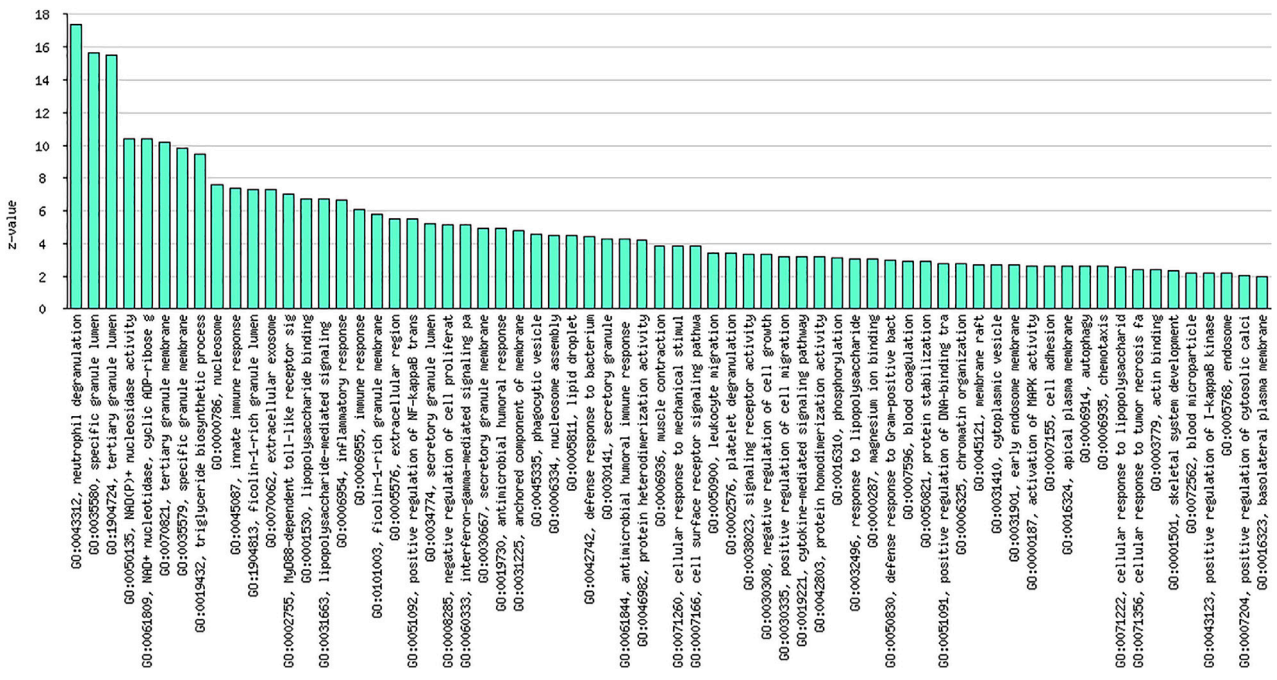

FIGURE 3 | Graphical representation of significantly overrepresented individual categories of GO molecular function for the upregulated genes.

patients with vasculopathy (Hom et al., 2008; Reindel et al., 2013). *ITGAM*, also known as *CD11b*, belongs to the integrin family and is involved in the regulation of neutrophils, monocyte activation, adhesion, and migration to damaged endothelium associated extracellular matrix. Moreover, studies have shown increased *ITGAM* levels in peripheral blood and upregulated protein expression in patients with KD (Reindel et al., 2013; Wu et al., 2019).

Immune system activation is one of the major pathogenic mechanisms of KD. These pathways play a crucial role in the

TABLE 7 | Significantly enriched GO Molecular functions of DEGs.

Title	z-value	FDR	Fold enrichment	N genes	Genes
GO: 0043312, neutrophil degranulation	17.3639	0	7.6743	50	STXBP2, FCAR, CRISPLD2, LILRA3, RETN, S100P, CDA, TCN1, ITGAM, BST1, CAMP, OSCAR, PGLYRP1, CD55, MMP9, CEACAM8, LCN2, SLPI, CLEC4D, CYSTM1, FOLR3, MCEMP1, MMP25, QSOX1, RAB27A, VNN1, PGM2, GPR84, FTH1, CKAP4, S100A11, MAPK14, S100A12, ORM1, SIGLEC14, CANT1, TLR2, HP, QPCT, SIRPB1, CEACAM3, TNFAIP6, BPI, HSPA1A, FCER1G, CEACAM1, GYG1, ANXA3, RAB31, and CD44
GO: 0035580, specific granule lumen	15.6351	0	17.9727	15	ORM1, CANT1, HP, QPCT, BPI, RETN, TCN1, CAMP, OSCAR, PGLYRP1, LCN2, SLPI, FOLR3, QSOX1, and RAB27A
GO: 1904724, tertiary granule lumen	15.536	0	18.9095	14	MMP9, PGLYRP1, OSCAR, QSOX1, FOLR3, CDA, CAMP, TCN1, TNFAIP6, ORM1, FTH1, QPCT, HP, and CANT1

TABLE 8 | Significantly enriched KEGG pathway of DEGs.

Title	z-value	FDR	Fold enrichment	N genes	Genes
Systemic lupus erythematosus	4.758	0	3.9336	10	H2AC21, HLA-DRB5, H2BC18, C1QB, H2AC20, FCGR1A, H2BC21, H2AC18, H2BC5, and H3C4

innate immune response by recognizing pathogen-associated molecular patterns (PAMPs) of infectious agents. Huang et al. (2009) found that mRNA levels of *TLR2* and *TLR8* were significantly elevated in patients with KD as compared to controls. Furthermore, after treatment with IVIG therapy, the level of TLR mRNA expression in KD patients was decreased. Studies have shown that augmented expression of *TLR2* on $CD14^+$ monocytes is observed in patients with KD and in coronary arteritis mice model, suggesting their role in immunopathogenesis (Rosenkranz et al., 2005; Lin et al., 2012; Mitsui et al., 2014). Therefore, these TLRs may be used as inflammatory biomarkers for the identification of patients with KD.

CXCL2, also known as interleukin-8 (*IL8*), is a member of the CXC chemokine family, known to possess tumorigenic properties. It is involved in proinflammatory activities, such as neutrophil degranulation, amelioration of adhesion molecule expression on neutrophils' surfaces and directional migration of neutrophils (Brat et al., 2005). In the acute phase of KD, *IL-8* is overexpressed in mononuclear cells and polymorphonuclear neutrophils of patients with KD as compared to controls (Lin et al., 1992; Asano and Ogawa, 2001). Furthermore, Asano et al. showed that IL-8 protein and the neutrophil chemoattractant activity within plasma were increased in the acute phase of KD and were significantly elevated following IVIG therapy (Asano and Ogawa, 2001).

Spi-1 proto-oncogene (*SPI1*) is a protein-coding gene mainly involved in the activation, differentiation, activation, and migration of macrophages or B cells. This gene was also identified as one of the hub genes in a study by Gao et al. (2021). The role of the *SPI1* gene is yet not explored in KD.

Activation of *TREM-1* (triggering receptor expressed on myeloid cells-1), trigger receptor is found to be associated with the pathogenesis of KD. Zhao et al. have shown that serum soluble *TREM-1* protein concentrations were significantly higher in the acute phase of KD as compared to controls, indicating its involvement in vasculitis and CAAs in patients with KD (Zhao and Wang, 2016). Our results confirm that the expression of the *TREM* gene was significantly upregulated in patients with KD.

S100 calcium-binding protein A12 (S100A12) belongs to the S100 protein family. Serum SAA00A12 levels were significantly higher in patients with KD, and the levels declined significantly after treatment with IVIG (Foell et al., 2003; Abe et al., 20051950; Armaroli et al., 2019; Lech et al., 2019; Gao et al., 2021). Furthermore, expression of S100A12 increased on the surface of circulating endothelial cells of patients with KD and persisted for a longer duration in KD patients with CAAs, indicating the role of S100A12 in the development of CAAs. In addition, S100A12 stimulates monocytes to produce IL-8, which in turn induces coronary artery endothelial cell dysfunction (Armaroli et al., 2019). Li et al. have shown that IVIG therapy and S100A12 antibody had similar effects on reducing neutrophil infiltration (Lin et al., 1992). Data from these studies highlight the potential of S100A12 as an imperative biomarker for monitoring patients with KD. Despite this, detailed mechanisms through which S100A genes regulate the pathogenesis of KD have not yet been well studied. We hope that our results will enhance our understanding of the role of the S100A gene family in the pathogenesis of KD.

Matrix metalloproteinases-9 (*MMP-9*) has been implicated in the pathogenesis of several disorders, including tumor metastasis, inflammatory disorders, and atherosclerosis (Sakata et al., 2010). The role of *MMP-9* in the pathogenesis of KD has also been highlighted. Overexpression of MMPs induces degradation of the

extracellular matrix. Takeshita et al. (2001) hypothesized that in patients with KD, activated neutrophils and monocytes produce *MMP-9* in large amounts, that upon migration leads to the breakdown of basement membrane that resulting in vascular damage and coronary artery inflammation. This hypothesis was validated in the mice model by Lau et al. (2008). Studies have shown that serum levels of *MMP-9* are elevated in KD patients and were found to play an important role in the development of CAAs (Takeshita et al., 2001; Chua et al., 2003; Gavin et al., 2003; Wu et al., 2019). Perforin-1 (*PRF1*) is one of the major cytolytic proteins of cytolytic granules. One of the main pathways of lymphocyte-mediated cytolysis entails the secretion onto target membranes of cytolytic granules contained in cytolytic effector lymphocytes of T-cell or NK-cell type (https://omim.org/entry/170280). This gene is related to hemophagocytic lymphohistiocytosis (HLH) but has no reports in KD. We assumed that together all these hub genes may play a key role in the pathogenesis of KD. However, future studies need to be conducted to verify this hypothesis.

CONCLUSION

The integrated transcriptomic approach, along with the bioinformatic analysis used in the present study, helped to reveal deregulated molecular mechanisms explaining the underlying etiology of KD. Furthermore, using data from more than one microarray dataset and their healthy controls helped eliminate the potential influences of clinical, demographic, and environmental factors on transcriptomic analysis. The 10 hub genes may provide clues to understanding the pathogenesis of KD pathogenesis and could be used to design a biomarker panel to monitor patients with KD.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

PS, CB, and AK: data analysis and preparation of the manuscript. RP, RK, AS, and SS: editing and critical review of the manuscript. All authors approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.849834/full#supplementary-material

REFERENCES

Abe, J., Jibiki, T., Noma, S., Nakajima, T., Saito, H., and Terai, M. (20051950)., 174. Baltimore, Md, 5837–5845. doi:10.4049/jimmunol.174.9.5837Gene Expression Profiling of the Effect of High-Dose Intravenous Ig in Patients with Kawasaki Disease*J. Immunol.*9

Agarwal, S., and Agrawal, D. K. (2017). Kawasaki Disease: Etiopathogenesis and Novel Treatment Strategies. *Expert Rev. Clin. Immunol.* 13 (3), 247–258. doi:10.1080/1744666X.2017.1232165

Armaroli, G., Verweyen, E., Pretzer, C., Kessel, K., Hirono, K., Ichida, F., et al. (2019). Monocyte-Derived Interleukin-1β as the Driver of S100A12-Induced Sterile Inflammatory Activation of Human Coronary Artery Endothelial Cells: Implications for the Pathogenesis of Kawasaki Disease. *Arthritis Rheumatol.* 71 (5), 792–804. doi:10.1002/art.40784

Asano, T., and Ogawa, S. (2001). Expression of IL-8 in Kawasaki Disease. *Clin. Exp. Immunol.* 122 (3), 514–519. doi:10.1046/j.1365-2249.2000.01395.x

Banday, A. Z., Mondal, S., Barman, P., Sil, A., Kumrah, R., Vignesh, P., et al. (2021). What Lies Ahead for Young Hearts in the 21st Century - Is it Double Trouble of Acute Rheumatic Fever and Kawasaki Disease in Developing Countries? *Front. Cardiovasc. Med.* 8, 694393. doi:10.3389/fcvm.2021.694393

Brat, D. J., Bellail, A. C., and Van Meir, E. G. (2005). The Role of Interleukin-8 and its Receptors in Gliomagenesis and Tumoral Angiogenesis. *Neuro-oncology* 7 (2), 122–133. doi:10.1215/S1152851704001061

Brogan, P., Burns, J. C., Cornish, J., Diwakar, V., Eleftheriou, D., Gordon, J. B., et al. (2020). Lifetime Cardiovascular Management of Patients with Previous Kawasaki Disease. *Heart* 106, 411–420. doi:10.1136/heartjnl-2019-315925

Burgner, D., Davila, S., Breunis, W. B., Ng, S. B., Li, Y., Bonnard, C., et al. (2009). A Genome-wide Association Study Identifies Novel and Functionally Related Susceptibility Loci for Kawasaki Disease. *Plos Genet.* 5 (1), e1000319. doi:10.1371/journal.pgen.1000319

Cdc (2020). *About Kawasaki Disease*. Available at: https://www.cdc.gov/kawasaki/about.html (Accessed January 1, 2022).

Chen, M.-R., Chang, T.-Y., Chiu, N.-C., Chi, H., Yang, K. D., Chang, L., et al. (2020). Validation of Genome-wide Associated Variants for Kawasaki Disease in a Taiwanese Case-Control Sample. *Sci. Rep.* 10 (1), 11756. doi:10.1038/s41598-020-68673-0

Chua, P. K., Melish, M. E., Yu, Q., Yanagihara, R., Yamamoto, K. S., and Nerurkar, V. R. (2003). Elevated Levels of Matrix Metalloproteinase 9 and Tissue Inhibitor of Metalloproteinase 1 during the Acute Phase of Kawasaki Disease. *Clin. Vaccin. Immunol* 10 (2), 308–314. doi:10.1128/cdli.10.2.308-314.2003

Dietz, S. M., van Stijn, D., Burgner, D., Levin, M., Kuipers, I. M., Hutten, B. A., et al. (2017). Dissecting Kawasaki Disease: a State-Of-The-Art Review. *Eur. J. Pediatr.* 176 (8), 995–1009. doi:10.1007/s00431-017-2937-5

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucleic Acids Res.* 30 (1), 207–210. doi:10.1093/nar/30.1.207

Foell, D., Ichida, F., Vogl, T., Yu, X., Chen, R., Miyawaki, T., et al. (2003). S100A12 (EN-RAGE) in Monitoring Kawasaki Disease. *The Lancet* 361 (9365), 1270–1272. doi:10.1016/S0140-6736(03)12986-8

Fu, L. Y., Qiu, X., Deng, Q. L., Huang, P., Pi, L., Xu, Y., et al. (20192019). The IL-1B Gene Polymorphisms Rs16944 and Rs1143627 Contribute to an Increased Risk of Coronary Artery Lesions in Southern Chinese Children with Kawasaki Disease. *J. Immunol. Res.* 2019, 1–7. doi:10.1155/2019/4730507

Gao, S., Ma, W., Lin, X., Huang, S., and Yu, M. (2021). Identification of Key Genes and Underlying Mechanisms in Acute Kawasaki Disease Based on Bioinformatics Analysis. *Med. Sci. Monit.* 27, e930547. doi:10.12659/MSM.930547

Gavin, P. J., Crawford, S. E., Shulman, S. T., Garcia, F. L., and Rowley, A. H. (2003). Systemic Arterial Expression of Matrix Metalloproteinases 2 and 9 in Acute Kawasaki Disease. *Atvb* 23 (4), 576–581. doi:10.1161/01.atv.0000065385.47152.fd

Guleria, S., Jindal, A., Pilania, R., Bhattarai, D., Kaur, A., Naganur, S., et al. (2020). An Uncommon Overlap of Two Common Rheumatological Disorders. *Lupus* 29 (9), 1121–1125. doi:10.1177/0961203320930101

Han, S., Kim-Howard, X., Deshmukh, H., Kamatani, Y., Viswanathan, P., Guthridge, J. M., et al. (2009). Evaluation of Imputation-Based Association in and Around the Integrin- -M (ITGAM) Gene and Replication of Robust Association between a Non-synonymous Functional Variant within ITGAM and Systemic Lupus Erythematosus (SLE). *Hum. Mol. Genet.* 18 (6), 1171–1180. doi:10.1093/hmg/ddp007

Hoang, L. T., Shimizu, C., Ling, L., Naim, A. N. M., Khor, C. C., Tremoulet, A. H., et al. (2014). Global Gene Expression Profiling Identifies New Therapeutic Targets in Acute Kawasaki Disease. *Genome Med.* 6 (11), 541. doi:10.1186/s13073-014-0102-6

Hom, G., Graham, R. R., Modrek, B., Taylor, K. E., Ortmann, W., Garnier, S., et al. (2008). Association of Systemic Lupus Erythematosus withC8orf13-BLKandITGAM-ITGAX. *N. Engl. J. Med.* 358 (9), 900–909. doi:10.1056/NEJMoa0707865

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211

Jindal, A. K., Pilania, R. K., Prithvi, A., Guleria, S., and Singh, S. (2019). Kawasaki Disease: Characteristics, Diagnosis, and Unusual Presentations. *Expert Rev. Clin. Immunol.* 15 (10), 1089–1104. doi:10.1080/1744666X.2019.1659726

Johnson, T. A., Mashimo, Y., Wu, J.-Y., Yoon, D., Hata, A., Kubo, M., et al. (2021). Association of an *IGHV3-66* Gene Variant with Kawasaki Disease. *J. Hum. Genet.* 66, 475–489. doi:10.1038/s10038-020-00864-z2

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27

Lau, A. C., Duong, T. T., Ito, S., and Yeung, R. S. M. (2008). Matrix Metalloproteinase 9 Activity Leads to Elastin Breakdown in an Animal Model of Kawasaki Disease. *Arthritis Rheum.* 58 (3), 854–863. doi:10.1002/art.23225

Lech, M., Guess, J., Duffner, J., Oyamada, J., Shimizu, C., Hoshino, S., et al. (2019). Circulating Markers of Inflammation Persist in Children and Adults with Giant Aneurysms after Kawasaki Disease. *Circ. Genomic Precision Med.* 12 (4), e002433. doi:10.1161/CIRCGEN.118.002433

Lee, Y., Schulte, D. J., Shimada, K., Chen, S., Crother, T. R., Chiba, N., et al. (2012). Interleukin-1β Is Crucial for the Induction of Coronary Artery Inflammation in a Mouse Model of Kawasaki Disease. *Circulation* 125 (12), 1542–1550. doi:10.1161/CIRCULATIONAHA.111.072769

Lin, C.-Y., Lin, C.-C., Hwang, B., and Chiang, B. (1992). Serial Changes of Serum Interleukin-6, Interleukin-8, and Tumor Necrosis Factor Alpha Among Patients with Kawasaki Disease. *J. Pediatr.* 121 (6), 924–926. doi:10.1016/s0022-3476(05)80343-9

Lin, I.-C., Kuo, H.-C., Lin, Y.-J., Wang, F.-S., Wang, L., Huang, S.-C., et al. (2012). Augmented TLR2 Expression on Monocytes in Both Human Kawasaki Disease and a Mouse Model of Coronary Arteritis. *PloS one* 7 (6), e38635. doi:10.1371/journal.pone.0038635

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks. *Bioinformatics* 21 (16), 3448–3449. doi:10.1093/bioinformatics/bti551

McCrindle, B. W., Rowley, A. H., Newburger, J. W., Burns, J. C., Bolger, A. F., Gewitz, M., et al. (2017). Diagnosis, Treatment, and Long-Term Management of Kawasaki Disease: a Scientific Statement for Health Professionals from the American Heart Association. *Circulation* 135, e927–e999. doi:10.1161/CIR.0000000000000484

Mitsui, K., Yusa, T., Miyazaki, S., Ohara, A., and Saji, T. (2014). Increased TLR2 and TLR4 Expression in Peripheral Neutrophils Isolated from Kawasaki Disease. *Pediatr. Allergy Immunol. Pulmonology* 27 (1), 24–29. doi:10.1089/ped.2013.0315

Nakamura, A., Ikeda, K., and Hamaoka, K. (2019). Aetiological Significance of Infectious Stimuli in Kawasaki Disease. *Front. Pediatr.* 7, 244. doi:10.3389/fped.2019.00244

Newburger, J. W., Takahashi, M., Gerber, M. A., Gewitz, M. H., Tani, L. Y., Burns, J. C., et al. (2004). Diagnosis, Treatment, and Long-Term Management of Kawasaki Disease. *Circulation* 110 (17), 2747–2771. doi:10.1161/01.CIR.0000145143.19711.78

Noval Rivas, M., and Arditi, M. (2020). Kawasaki Disease: Pathophysiology and Insights from Mouse Models. *Nat. Rev. Rheumatol.* 16 (7), 391–405. doi:10.1038/s41584-020-0426-0

Oates-Whitehead, R. M., Baumer, J. H., Haines, L., Love, S., Maconochie, I. K., Gupta, A., et al. (2003). Intravenous Immunoglobulin for the Treatment of Kawasaki Disease in Children. *Cochrane database Syst. Rev.* 2021 (4), CD004000. doi:10.1002/14651858.CD004000

Ogata, S., Shimizu, C., Franco, A., Touma, R., Kanegaye, J. T., Choudhury, B. P., et al. (2013). Treatment Response in Kawasaki Disease Is Associated with Sialylation Levels of Endogenous but Not Therapeutic Intravenous Immunoglobulin G. *PLoS ONE* 8, e81448. doi:10.1371/journal.pone.0081448

Onouchi, Y., Ozaki, K., Ozaki, K., Burns, J. C., Shimizu, C., Terai, M., et al. (2012). A Genome-wide Association Study Identifies Three New Risk Loci for Kawasaki Disease. *Nat. Genet.* 44 (5), 517–521. doi:10.1038/ng.2220

Popper, S. J., Shimizu, C., Shike, H., Kanegaye, J. T., Newburger, J. W., Sundel, R. P., et al. (2007). Gene-expression Patterns Reveal Underlying Biological Processes in Kawasaki Disease. *Genome Biol.* 8 (12), R261. doi:10.1186/gb-2007-8-12-r261

Porritt, R. A., Markman, J. L., Maruyama, D., Kocaturk, B., Chen, S., Lehman, T. J. A., et al. (2020). Interleukin-1 Beta-Mediated Sex Differences in Kawasaki Disease Vasculitis Development and Response to Treatment. *Atvb* 40 (3), 802–818. doi:10.1161/ATVBAHA.119.313863

Rahmati, Y., Mollanoori, H., Kakavandi, N., Nateghian, A., Sayyahfar, S., Babaei, V., et al. (2020). Gene Expression Analysis in Kawasaki Disease; Bioinformatics and Experimental Approach. *Inform. Med. Unlocked* 20, 100423. doi:10.1016/j.imu.2020.100423

Ramphul, K., and Mejias, S. G. (2018). Kawasaki Disease: a Comprehensive Review. *amsad* 3, 41–45. doi:10.5114/amsad.2018.74522

Reindel, R., Baker, S. C., Kim, K.-Y., Rowley, C. A., Shulman, S. T., Orenstein, J. M., et al. (2013). Integrins α4 and αM, collagen1A1, and Matrix Metalloproteinase 7 Are Upregulated in Acute Kawasaki Disease Vasculopathy. *Pediatr. Res.* 73 (3), 332–336. doi:10.1038/pr.2012.185

Rosenkranz, M. E., Schulte, D. J., Agle, L. M. A., Wong, M. H., Zhang, W., Ivashkiv, L., et al. (2005). TLR2 and MyD88 Contribute to Lactobacillus Casei Extract-Induced Focal Coronary Arteritis in a Mouse Model of Kawasaki Disease. *Circulation* 112 (19), 2966–2973. doi:10.1161/CIRCULATIONAHA.105.537530

Saez-de-Ocariz, M., Pecero-Hidalgo, M. J., Rivas-Larrauri, F., García-Domínguez, M., Venegas-Montoya, E., Garrido-García, M., et al. (2020). A Teenager with Rash and Fever: Juvenile Systemic Lupus Erythematosus or Kawasaki Disease? *Front. Pediatr.* 8, 149. doi:10.3389/fped.2020.00149

Sakata, K., Hamaoka, K., Ozawa, S., Niboshi, A., Yahata, T., Fujii, M., et al. (2010). Matrix Metalloproteinase-9 in Vascular Lesions and Endothelial Regulation in Kawasaki Disease. *Circ. J.* 74 (8), 1670–1675. doi:10.1253/circj.cj-09-0980

Senzaki, H. (2006). The Pathophysiology of Coronary Artery Aneurysms in Kawasaki Disease: Role of Matrix Metalloproteinases. *Arch. Dis. Child.* 91 (10), 847–851. doi:10.1136/adc.2005.087437

Sharma, K., Vignesh, P., Srivastava, P., Sharma, J., Chaudhary, H., Mondal, S., et al. (2021). Epigenetics in Kawasaki Disease. *Front. Pediatr.* 9, 673294. doi:10.3389/fped.2021.673294

Sharov, A. A., Schlessinger, D., and Ko, M. S. H. (2015). ExAtlas: An Interactive Online Tool for Meta-Analysis of Gene Expression Data. *J. Bioinform. Comput. Biol.* 13 (6), 1550019. doi:10.1142/S0219720015500195

Singh, S., Bhattad, S., Gupta, A., Suri, D., Rawat, A., and Rohit, M. (2016). Mortality in Children with Kawasaki Disease: 20 Years of Experience from a Tertiary Care centre in North India. *Clin. Exp. Rheumatol.* 34 (3 Suppl. 97), S129–S133.

Singh, S., Vignesh, P., and Burgner, D. (2015). The Epidemiology of Kawasaki Disease: A Global Update. *Arch. Dis. Child.* 100, 1084–1088. doi:10.1136/archdischild-2014-307536

Skochko, S. M., Jain, S., Sun, X., Sivilay, N., Kanegaye, J. T., Pancheri, J., et al. (2018). Kawasaki Disease Outcomes and Response to Therapy in a Multiethnic Community: A 10-Year Experience. *J. Pediatr.* 203, 408–415. doi:10.1016/j.jpeds.2018.07.090

Takeshita, S., Tokutomi, T., Kawase, H., Nakatani, K., Tsujimoto, H., Kawamura, Y., et al. (2001). Elevated Serum Levels of Matrix Metalloproteinase-9 (MMP-9) in Kawasaki Disease. *Clin. Exp. Immunol.* 125 (2), 340–344. doi:10.1046/j.1365-2249.2001.01608.x

Vallabhajosyula, R. R., Chakravarti, D., Lutfeali, S., Ray, A., and Raval, A. (2009). Identifying Hubs in Protein Interaction Networks. *PloS one* 4 (4), e5344. doi:10.1371/journal.pone.0005344

Wakita, D., Kurashima, Y., Crother, T. R., Noval Rivas, M., Lee, Y., Chen, S., et al. (2016). Role of Interleukin-1 Signaling in a Mouse Model of Kawasaki Disease-Associated Abdominal Aortic Aneurysm. *Atvb* 36 (5), 886–897. doi:10.1161/ATVBAHA.115.307072

Weng, K.-P., Hsieh, K.-S., Ho, T.-Y., Huang, S.-H., Lai, C.-R., Chiu, Y.-T., et al. (2010). IL-1B Polymorphism in Association with Initial Intravenous Immunoglobulin Treatment Failure in Taiwanese Children with Kawasaki Disease. *Circ. J.* 74 (3), 544–551. doi:10.1253/circj.cj-09-0664

Wu, W., You, K., Zhong, J., Ruan, Z., and Wang, B. (2019). Identification of Potential Core Genes in Kawasaki Disease Using Bioinformatics Analysis. *J. Int. Med. Res.* 47 (9), 4051–4058. doi:10.1177/0300060519862057

Xie, X., Shi, X., and Liu, M. (2018). The Roles of Genetic Factors in Kawasaki Disease: A Systematic Review and Meta-Analysis of Genetic Association Studies. *Pediatr. Cardiol.* 39 (2), 207–225. doi:10.1007/s00246-017-1760-0

Zhao, J. M., and Wang, Q. (2016). Zhongguo Dang Dai Er Ke Za Zhi = Chinese Journal of Contemporary Pediatrics 18 (6), 522–526. doi:10.7499/j.issn.1008-8830.2016.06.011

Zhou, G., Soufan, O., Ewald, J., Hancock, R. E. W., Basu, N., and Xia, J. (2019). NetworkAnalyst 3.0: a Visual Analytics Platform for Comprehensive Gene Expression Profiling and Meta-Analysis. *Nucleic Acids Res.* 47 (W1), W234–W241. doi:10.1093/nar/gkz240

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fevo.2022.868826

Genetic and Epigenetic Differentiation Across Intertidal Gradients in the Foundation Plant *Spartina alterniflora*

*Jeannie M. Mounger[1], Isolde van Riemsdijk[2], M. Teresa Boquete[1,3], Cornelis A. M. Wagemaker[4], Samar Fatma[5], Marta H. Robertson[1], Sandy A. Voors[1], Jenna Oberstaller[6], Fleur Gawehns[7], Torrance C. Hanley[8], Ivo Grosse[5,9], Koen J. F. Verhoeven[7], Erik E. Sotka[10], Catherine A. Gehring[11], A. Randall Hughes[8], David B. Lewis[1], Marc W. Schmid[12] and Christina L. Richards[1,2]**

[1] Department of Integrative Biology, University of South Florida, Tampa, FL, United States, [2] Plant Evolutionary Ecology Group, University of Tübingen, Tübingen, Germany, [3] Department of Evolutionary Ecology, Consejo Superior de Investigaciones Cientificas, Estación Biológica de Doñana, Seville, Spain, [4] Department of Experimental Plant Ecology, Radboud University, Nijmegen, Netherlands, [5] Institute of Computer Science, Martin Luther University of Halle-Wittenberg, Halle (Saale), Germany, [6] Center for Global Health and Infectious Diseases Research and USF Genomics Program, College of Public Health, University of South Florida, Tampa, FL, United States, [7] Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands, [8] Marine Science Center, Northeastern University, Nahant, MA, United States, [9] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, [10] Department of Biology and Grice Marine Laboratory, College of Charleston, Charleston, SC, United States, [11] Department of Biological Sciences, Center for Adaptable Western Landscapes, Northern Arizona University, Flagstaff, AZ, United States, [12] MWSchmid GmbH, Glarus, Switzerland

OPEN ACCESS

Edited by:
Tanvir Ul Hassan Dar, Baba Ghulam Shah Badshah University, India

Reviewed by:
Shailendra Goel, University of Delhi, India
Cristina Cruz, University of Lisbon, Portugal

***Correspondence:**
Christina L. Richards
clr@usf.edu

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Ecology and Evolution

Received: *03 February 2022*
Accepted: *15 March 2022*
Published: *12 May 2022*

Citation:
Mounger JM, van Riemsdijk I, Boquete MT, Wagemaker CAM, Fatma S, Robertson MH, Voors SA, Oberstaller J, Gawehns F, Hanley TC, Grosse I, Verhoeven KJF, Sotka EE, Gehring CA, Hughes AR, Lewis DB, Schmid MW and Richards CL (2022) Genetic and Epigenetic Differentiation Across Intertidal Gradients in the Foundation Plant Spartina alterniflora. Front. Ecol. Evol. 10:868826. doi: 10.3389/fevo.2022.868826

Ecological genomics approaches have informed us about the structure of genetic diversity in natural populations that might underlie patterns in trait variation. However, we still know surprisingly little about the mechanisms that permit organisms to adapt to variable environmental conditions. The salt marsh foundation plant *Spartina alterniflora* exhibits a dramatic range in phenotype that is associated with a pronounced intertidal environmental gradient across a narrow spatial scale. Both genetic and non-genetic molecular mechanisms might underlie this phenotypic variation. To investigate both, we used epigenotyping-by-sequencing (epiGBS) to evaluate the make-up of natural populations across the intertidal environmental gradient. Based on recent findings, we expected that both DNA sequence and DNA methylation diversity would be explained by source population and habitat within populations. However, we predicted that epigenetic variation might be more strongly associated with habitat since similar epigenetic modifications could be rapidly elicited across different genetic backgrounds by similar environmental conditions. Overall, with PERMANOVA we found that population of origin explained a significant amount of the genetic (8.6%) and epigenetic (3.2%) variance. In addition, we found that a small but significant amount of genetic and epigenetic variance (<1%) was explained by habitat within populations. The interaction of population and habitat explained an additional 2.9% of the genetic variance and 1.4% of the epigenetic variance. By examining genetic and epigenetic variation within the same fragments (variation in close-cis), we found that population explained epigenetic variation in 9.2% of 8,960 tested loci, even after accounting for differences in the DNA

sequence of the fragment. Habitat alone explained very little (<0.1%) of the variation in these close-cis comparisons, but the interaction of population and habitat explained 2.1% of the epigenetic variation in these loci. Using multiple matrix regression with randomization (MMRR) we found that phenotypic differences in natural populations were correlated with epigenetic and environmental differences even when accounting for genetic differences. Our results support the contention that sequence variation explains most of the variation in DNA methylation, but we have provided evidence that DNA methylation distinctly contributes to plant responses in natural populations.

Keywords: epigenomic analysis, foundation plant, *Spartina alterniflora*, non-genetic inheritance, salt marsh ecology, ecological genomics

INTRODUCTION

The importance of genetically based variation within species has been well accepted in the field of evolutionary biology, but further analyses are required to understand the mechanisms that enable organisms to adapt to conditions in novel or challenging environments. In addition to patterns of genetic variation that have been associated with different environmental challenges (Bock et al., 2015; Hodgins et al., 2015; Neinavaie et al., 2021), increasing evidence suggests that epigenetic variation (e.g., alterations to DNA methylation, small RNAs, and chromatin remodeling) plays a role in ecology, and that this variation can be both environmentally induced and contribute to phenotypic plasticity (Cortijo et al., 2014; Medrano et al., 2014; Robertson and Richards, 2015; Banta and Richards, 2018; Ashe et al., 2021; Mounger et al., 2021a; Stajic and Jansen, 2021). This additional source of variation may be particularly important for sessile organisms such as plants, given that individuals are unable to migrate away from stressors (Dodd and Douhovnikoff, 2016; Balao et al., 2018). Many plant species persist across broad ranges in environmental conditions, and phenotypes are known to vary in response (Schlichting, 1986; Des Marais et al., 2013; Gratani, 2014). The association of phenotypic variation with environmental variation is particularly evident in coastal ecosystems, where plant species are exposed to substantial fluctuations in inundation, salinity, and soil conditions (Pennings and Bertness, 2001; Richards et al., 2005; Lewis et al., 2021). As plants increasingly face habitat loss, competition from invasive species, and extreme weather conditions resulting from anthropogenic global change, rapid phenotypic modifications moderated by epigenetic mechanisms may be an important component of response (Nicotra et al., 2010). This source of variation could be especially important in ecosystems impacted by the most dramatic intensifications or shifts in climatic and habitat conditions (Nicotra et al., 2015; Burggren, 2016; Jueterbock et al., 2020; Mounger et al., 2021b).

Our understanding of the role of epigenetic regulatory mechanisms in plant trait response is primarily based on research with model organisms or from agriculture (Niederhuth and Schmitz, 2017; Balao et al., 2018; Miryeganeh and Saze, 2020). In these contexts, epigenetic mechanisms such as DNA methylation and histone modifications have been shown to play a role in stress response to salinity, drought, temperature, fluctuating nutrient levels and UV radiation (reviewed in Miryeganeh (2021)). However, the examination of these relationships in more natural settings is expanding (Richards et al., 2017; Jueterbock et al., 2020; Sarma et al., 2020, 2021; Mounger et al., 2021a,b; Boquete et al., 2022). For instance, several population epigenetics studies have demonstrated that DNA methylation varies with population structure and habitat (Lira-Medeiros et al., 2010; Paun et al., 2010; Schulz et al., 2014; Xie et al., 2015; Foust et al., 2016; Herrera et al., 2017; Gao et al., 2019; Gáspár et al., 2019; Mounger et al., 2021b). Studies of non-model plant species have also shown changes in DNA methylation that are associated with community composition (van Moorsel et al., 2019), abiotic stress response (Verhoeven et al., 2010; Nicotra et al., 2015; Li et al., 2020; Boquete et al., 2022), biotic interactions such as herbivory (Herrera and Bazaga, 2011; Alonso et al., 2019), and transgenerational plasticity (Herman and Sultan, 2016; Puy et al., 2021a,b). Despite this body of work, few studies have simultaneously analyzed genetic, epigenetic, and environmental factors to establish if and how much they each contribute to functional variation (e.g., Herrera et al., 2017; Wang et al., 2017).

Spartina alterniflora L. (Poaceae) populations provide a unique opportunity to explore the mechanisms that underlie phenotypic response to environmental challenges. Throughout the native range of *S. alterniflora*, a predictable change in phenotype is associated with repeated instances of the intertidal environmental gradient (**Figure 1**). On the leading edge of the shoreline, plants grow tall, sometimes exceeding 2 m in height. As elevation climbs away from the water, plants decrease in height and may not reach 0.25 m at the highest elevations (Pennings and Bertness, 2001; Richards et al., 2005; Voors, 2018). Plants along this gradient are exposed to tides at varying frequency, with plants in the low marsh (closest to tidal creeks and bays) experiencing regular tidal inundation and plants in the high marsh receiving less frequent exposure to flooding. Salt deposits accumulate to a greater degree in the high marsh soil as a combined result of evapotranspiration and diminished tidal flushing (Bertness and Ellison, 1987; Pennings et al., 2005). In addition, differences in competition, edaphic conditions, water content, elevation, and herbivory can drive phenotypic variation in this species (Richards et al., 2005; Wiêski and Pennings, 2014; Zerebecki et al., 2017; Voors, 2018). Previous studies on whether the phenotypic differences in so called short-form versus tall-form *S. alterniflora* are genetically based have been equivocal

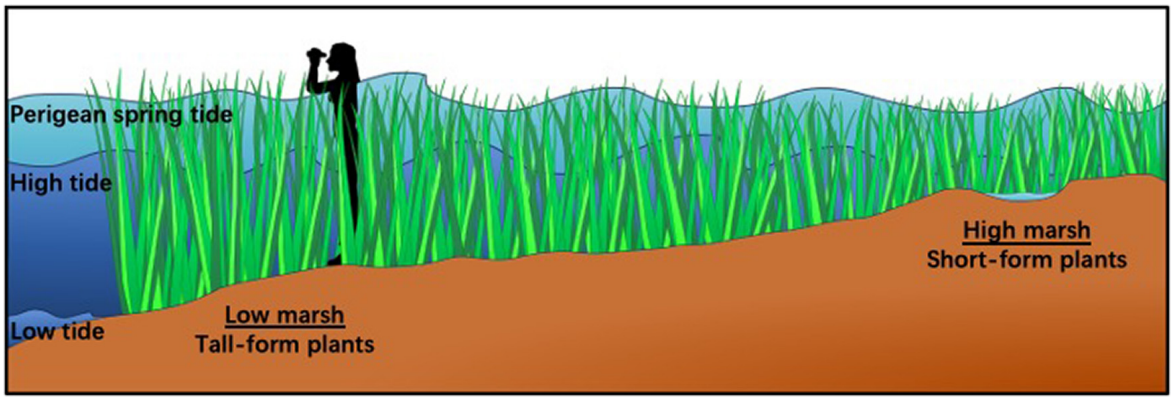

FIGURE 1 | Trait variation in *Spartina alterniflora* is correlated with tidal inundation across the marsh. Tall-form plants (>2 m) grow adjacent to tidal creeks in the low marsh where they are inundated daily. Plants decrease in height across increasing elevation such that short-form plants (<0.25 m) are found in the high marsh where exposure to tidal flooding is less often and for less duration. Perigean spring tides are extremely high tides often referred to as "King Tides" which reach the highest elevation in the marsh. They happen 6–8 times a year and inundate the entire salt marsh community (see: https://oceanservice.noaa.gov/facts/perigean-spring-tide.html).

(Stalter and Batson, 1969; Shea et al., 1975; Valiela et al., 1978; Gallagher et al., 1988). However, a recent reciprocal transplant study demonstrated that plants from the high and low marsh habitats were differentiated genetically and in several plant traits, which translated into fitness differences, supporting the hypothesis of local adaptation (Zerebecki et al., 2021).

In addition to well-documented differences in plant traits that are associated with environmental variation, native populations of *S. alterniflora* harbor high genetic diversity within natural populations (Richards et al., 2004; Travis et al., 2004; Hughes and Lotterhos, 2014; Foust et al., 2016; Zerebecki et al., 2021). This high genetic diversity could allow for fine-scale differentiation across the landscape considering the persistent differences in habitats. This species is known to be resilient to both natural (Pennings and Bertness, 2001; Edwards et al., 2005; Pennings et al., 2005) and anthropogenic stressors (Lin and Mendelssohn, 2012; Lin et al., 2016; Robertson et al., 2017; Alvarez et al., 2018). Response to these environmental conditions may be attributable to both genetic and non-genetic variation. For instance, recent work in *S. alterniflora* has shown that differences in DNA methylation were associated with different natural habitats (Foust et al., 2016), exposure to crude oil (Robertson et al., 2017) and with hybridization and allopolyploidization events that occurred during the process of invasion (Salmon et al., 2005). However, linking molecular mechanisms to trait variation and fine-scale environmental variation remains a daunting task.

Strategic application of multivariate statistics allows us to gain some insight into the relationships between genetic, epigenetic, environmental and phenotypic variation (Herrera et al., 2017; Wang et al., 2020). In this study, we used epigenotyping-by-sequencing (epiGBS) combined with measures of environmental factors and plant phenotypic variation to associate patterns of genomic and epigenomic diversity with environmental and phenotypic variation in natural *S. alterniflora* populations. We expected that DNA sequence variation would be explained by source population and habitat type (low and high marsh) within populations. We also predicted that DNA methylation patterns would be explained by both source population and habitat type, but more strongly associated with habitat type, since across genetic backgrounds similar epigenetic modifications could be elicited in response to similar environmental conditions. Finally, we used multiple matrix regression analysis with randomization (MMRR) to evaluate associations of a phenotypic distance matrix with matrices of genetic, epigenetic, and environmental distances. We predicted that the epigenetic differences and environmental differences would associate with differences in traits, even when accounting for genetic distances.

MATERIALS AND METHODS

Study Species and Site Description

Spartina alterniflora is a coastal salt marsh foundation species (*sensu* Ellison, 2019) that alters environmental conditions, builds land through the accumulation of sediments and organic matter, provides nursery grounds for numerous fish species, regulates trophic interactions, and serves as a crucial primary producer within subtropical and temperate estuarine environments (Pennings and Bertness, 2001; Lewis and Eby, 2002; Hughes and Lotterhos, 2014; Bertness, 2020; Hays et al., 2021). This species reproduces clonally by rhizomes (Pennings and Callaway, 2000), and sexually through completely outcrossed seeds (Somers and Grant, 1981). The relative contribution of each type of reproduction is unknown, but these populations have average levels of genetic diversity compared to non-clonal outcrossing species with otherwise similar life history characteristics (*sensu* Hamrick and Godt, 1996; Zerebecki et al., 2021).

We collected *S. alterniflora* individuals in two habitat types (low, or seaward, and high, or landward, parts of the salt marsh) at three sites near Charleston, South Carolina: Fort Johnson (N 32.74910; W 79.89822), Folly Beach (N 32.64436; W 79.96564), and Bowens Island (N 32.68027; W 79.95389; **Figure 2**). Mean

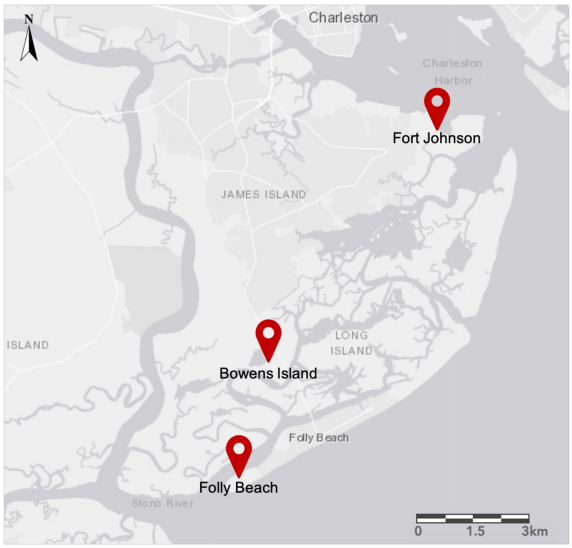

FIGURE 2 | Map of the James Island region near Charleston, South Carolina, United States. We sampled from two habitats (low and high marsh) in three populations: Fort Johnson (FJ), Bowens Island (BI), and Folly Beach (FB).

tidal range at these sites is between 1.59 and 1.65 m (U.S. NOAA National Ocean Service, Charleston, Cooper River Entrance, SC – Station ID: 8665530; Highway 171 Bridge, Folly Creek, SC – Station ID: 8665424; Folly River Bridge, Folly River, SC – Station ID: 8666652). The salt marshes within each site were typical of those in the southeastern United States, with *S. alterniflora* as the dominant plant species occurring across a wide range of salinity and soil conditions (Pennings and Bertness, 2001; Richards et al., 2005; Voors, 2018). We defined each site as a distinct population of potentially interbreeding individuals based on previous work in these locations (Zerebecki et al., 2021). The low marsh habitat occurred near tidal creek banks, was tidally inundated daily, and was characterized by the so-called tall-form of *S. alterniflora* (average height of 131 cm; Voors, 2018). The high marsh habitat was flooded less frequently, had higher soil salinity, and supported the short-form of *S. alterniflora* (average height of 53 cm; Voors, 2018).

During a 3-day period in August 2016, we established low- and high-marsh transects that ran parallel to the marsh-open water interface, and were separated by a minimum distance of 20 m. We ran a 50 m transect in each habitat at each site and laid down flags at randomly selected positions along each transect to establish ten 0.5 m × 0.5 m plots. The low marsh transects were within 2 m of tidal creeks. In a 0.25 m × 0.25 m quadrat in the center of each plot, we recorded densities of live and dead *S. alterniflora* stems. We then haphazardly flagged five stems in each plot and measured the stem height and the number of live leaves on each flagged stem. We collected a leaf sample from each flagged stem (i.e., 3 populations × 2 habitats × 10 plots × 5 stems = 300 individual plant samples) for (epi)genomic analyses and flash froze them in the field in liquid nitrogen before storing them long-term in −80°C. We collected an additional leaf tissue sample from each of the same 300 individuals for stable isotope analysis and kept them on ice in the field before freezing them in the lab. We measured redox, temperature, and salinity in a subset (min = 3) of plots along each transect. We also counted the number of snails present, measured porewater salinity and sediment respiration, and collected sediment samples from each plot. From the sediment samples obtained in the field, we recorded percent soil moisture, percent organic matter by loss-on-ignition, soil extractable ammonium (NH_4^+), soil extractable nitrate (NO_3^-), porewater NH_4^+, porewater NO_3^-, microbial biomass carbon (C), and microbial biomass nitrogen (N) in the lab as described in Voors (2018).

DNA Extractions and Library Prep

We extracted genomic DNA from the 300 samples using the Qiagen DNeasy plant mini kit according to the manufacturer's protocol with some modifications (*sensu* Foust et al., 2016). We prepared epiGBS libraries for 288 samples using three sets of 96 barcodes to be sequenced on three lanes of the Illumina HiSeq platform following methods outlined in van Gurp et al. (2016) and modified in Boquete et al. (2020). We digested 400 ng of genomic DNA from each sample with the enzymes *Nsi*I and *Ase*I. We ligated methylated, non-phosphorylated, barcoded adapters with variable barcodes of lengths from four to six base pairs to either end of the resulting fragments. The barcoded adaptors were designed so that we could identify forward ("Watson") and reverse ("Crick") strands for each fragment within each individual. Having the strand information allows for differentiating between C/T polymorphisms and methylation polymorphisms because we can recreate when unmethylated cytosines were present in either strand before bisulfite treatment without sequencing untreated samples (for details see van Gurp et al. (2016)).

For each group of 96 samples, we allocated the samples across eight pools and cleaned and size-selected genomic DNA fragments greater than 200 bp using 0.8 × SPRI beads (Agencourt AMPure XP; Beckman Coulter). We repaired nicks between the 3′ fragment overhangs and the 5′ non-phosphorylated adaptor nucleotides using a mix of dNTPs and DNA polymerase I. We used the Zymo EZ Lightning methylation kit to bisulfite-treat the DNA. We amplified libraries with the KAPA Uracil Hotstart Ready Mix with the following PCR conditions: an initial denaturation step at 95°C for 3 min followed by 18 cycles of 98°C for 10s, 65°C for 15s, and 72°C for 15s, with a final extension of 72°C for 5 min. The University of Florida Interdisciplinary Center for Biotechnology Research (UF ICBR) sequenced paired-end reads of the libraries on three lanes of the Illumina Platform HiSeq3000 (2 × 150 bp) in March 2017. Many of the libraries resulted in poor sequencing quality and depth, so we repeated library preparations. This time, we meta-pooled samples following ligation to reduce heterogeneity during the remaining library preparation process. Then, we sequenced 293 samples on four lanes of the Illumina Platform HiSeq X Ten System (2 × 150 bp) at the Novogene facility in Hong Kong in January 2019. We found that there were significantly fewer average reads per lane in the first set of libraries than the second set ($P < 10^{-16}$), but we could not detect significant differences in the overall read

count between populations, habitats or their interaction between the two library preparations (**Supplementary Figure 1**). We combined the sequencing from all seven lanes of data for analysis.

Data Pre-processing

We used the epiGBS2 pipeline (Gawehns et al., 2022) based on the original epiGBS pipeline (van Gurp et al., 2016) to process the raw sequence data. We demultiplexed, quality trimmed sequencing reads, and removed the barcode sequences with the original scripts as previously described (van Moorsel et al., 2019; Mounger et al., 2021b). Then, we used the processed reads for *de novo* reference construction using a minimal sequencing depth of 100 and an identity threshold of 95% during clustering. We used the default parameters of the epiGBS2 pipeline to call single nucleotide polymorphisms (SNPs) and single methylation polymorphisms (SMPs). We annotated *de novo* reference sequences with DIAMOND (protein coding genes; NCBI non-redundant proteins as reference; version 0.8.22; Buchfink et al., 2015) and RepeatMasker (transposons and repeats; Embryophyta as reference species collection; version 4.0.6; Smit et al., 2013–2015). We used this annotation to identify *de novo* reference contigs mapping to genes, transposons, repeats, or none of these classes (**Supplementary Table 1**).

We filtered SNPs and SMPs to keep only loci with a minimum coverage of five (i.e., five sequencing reads mapping to each locus in the *de novo* reference) and a maximum coverage of 1,000 within each individual. This filtering eliminated regions that did not have enough coverage for SNP or SMP identification and ruled out artifacts that could result from repetitive regions that were collapsed into a single contig (Schmid et al., 2018b). We then removed individuals that were missing coverage for more than 50% of the SNPs or SMPs found in the entire data set. This filtering resulted in a final number of 211 samples for analysis. With these 211 samples, we removed SNPs and SMPs that were not present in at least 20 individuals per subpopulation. We also applied an additional filter removing loci with a minor allele frequency of 1% or less. We considered that these loci were probably actually monomorphic given the inaccuracies of sequencing technology (Yoder and Tiley, 2021).

Genetic Variation and Genetic Structure

We characterized genetic variation with three approaches. First, we estimated the proportion of polymorphic loci as the number of fragments that contained SNPs and the average number of SNPs/kb of sequence. We then calculated allelic richness (defined as the rarefied mean number of alleles per locus within each population and habitat within populations as in Mounger et al. (2021b) using the function allelic.richness from hierfstat package (Goudet, 2005). Finally, we determined the mean and standard deviation of observed gene diversity, and heterozygosity per locus, with the function basic.stats within hierfstat, which is based on Nei (1987). We also estimated overall observed heterozygosity, total genetic diversity, total among-population genetic diversity, mean within-population genetic diversity, and overall population differentiation with the function basic.stats. For these analyses, we used 8,767 SNPs with no missing values.

We tested for genetic differentiation within and among *S. alterniflora* populations using three different approaches. First, we used a multivariate ANOVA (or PERMANOVA) with genetic distances between individuals (DIST) as a dependent variable. We modeled the contribution of populations, habitats, and the interaction between populations and habitats as explanatory variables for genetic distance with 9,999 permutations (package vegan, version 2.5-7, function adonis; Oksanen et al., 2020). To do so, we used the formula DIST $\sim$ population + habitat + population:habitat. Note, whereas the "habitat" term tests for differences between habitats (high vs. low marsh) that are common across populations, the interaction tests for additional differences between habitats that are not shared among populations.

Second, we used distance-based redundancy analysis (dbRDA) to quantify the relationship between genome-wide genetic variation and population, habitat and their interaction. This time we used the function capscale implemented within the vegan package (Oksanen et al., 2020). RDA is an ordination technique that summarizes the main patterns of variation in the response matrix, i.e., the scaled allele frequency matrix created from the SNP data [obtained using the function scaleGen from adegenet v. 2.1.3 (Jombart, 2008; Jombart and Ahmed, 2011) with NA.method set to "mean"], which can be explained by our explanatory variables. We fit the following models:

1) Genetic distance $\sim$ population
2) Genetic distance $\sim$ population + habitat
3) Genetic distance $\sim$ population + habitat + population $\times$ habitat

We tested the significance of the variation explained by our explanatory variables using a Monte Carlo permutation test with 999 permutations and obtained adjusted R^2 using the function RsquareAdj from the vegan package. We corrected *p*-values for multiple testing with false discovery rate (FDR; Benjamini and Hochberg, 1995) using the "fdr" method implemented with the p.adjust function in the base package of R.

Third, we obtained overall F_{ST} and pairwise F_{ST} values using the functions wc and genet.dist, respectively, from the package hierfstat. We calculated the confidence intervals of the pairwise F_{ST} values using the function boot.ppfst from the same package, with 999 permutations to determine whether F_{ST} values were significantly different from zero; that is, to find evidence of significant population differentiation. To identify candidate SNPs under selection, we used BayeScan v. 2.1 (Foll and Gaggiotti, 2008; Fischer et al., 2011). We corrected *p*-values for multiple testing with a significance threshold set to an FDR of 0.01.

Finally, we used a phylogenetic tree to assess the separation of populations and habitats based on genetic data. For this analysis, we used the same filtering process, but increased the MAF filter to include only MAF > 10% and excluded loci with any missing values. We concatenated the resulting 20,121 SNPs that survived this filtering and performed a maximum likelihood phylogenetic tree construction with IQ-tree (v 1.6.11) (Nguyen et al., 2015; Trifinopoulos et al., 2016; Hoang et al., 2018). We used the "general time reversible model with unequal rates and unequal base frequency," and optimized base frequencies by maximum likelihood (GTR + FO; Tavaré, 1986). We used the codon model for DNA, which is an ultrafast bootstrap of 1,000 on 3 CPU cores.

Other settings were left as default. We visualized the maximum likelihood tree with the R packages ape v. 5.5 (Paradis and Schliep, 2019) and scales v.1.1.1.9000 (Wickham and Seidel, 2020).

Epigenetic Variation and Epigenetic Structure

We calculated the DNA methylation level at each SMP within each individual sample (N = 211) as the number of reads mapping to one position that showed evidence of methylation (i.e., the bisulfite treatment did not convert the cytosine at that position) divided by the total number of reads mapping to that position. Then, we calculated the average level of DNA methylation for each sequence context (CG, CHG, and CHH) and across all contexts for plants grouped by habitat within each population (i.e., subpopulations) as well as the proportion of loci fixed for no methylation (i.e., methylation level $\leq$5% across $\geq$95% of the samples) and for full methylation (i.e., methylation level $\geq$95% across $\geq$95% of the samples).

We used a multivariate test for homogeneity of dispersions to estimate epigenetic diversity, i.e., variation in DNA methylation levels, following the approach of Anderson et al. (2006). This approach measures the average distance from each individual to their group centroid in a multivariate space using a dissimilarity measure. We argued previously that the distance from each individual sample to its population centroid in a multivariate space generated using an epigenetic distance matrix provides an estimate of the extent of the variation in DNA methylation, i.e., epigenetic variation (Mounger et al., 2021b). Then, the average distance of each population can be compared to test for significant differences in the amount of epigenetic variation among populations. Here, we applied the same logic to evaluate epigenetic variation within subpopulations. To do so, we generated pairwise epigenetic distance matrices by calculating the average difference in DNA methylation level across all cytosines between each pair of samples. Then, we used this matrix to calculate the distance between each individual sample and its subpopulation centroid using the function betadisper from the vegan package version 2.4-4 (Oksanen et al., 2020). We tested for differences in dispersion among subpopulations using a permutation-based test of multivariate homogeneity of group dispersions on the output of betadisper with 9,999 permutations. When this test was significant, we used the Tukey's Honest Significant Difference test to check which subpopulations differed in their average distance to the centroid, i.e., in their levels of epigenetic variation. Finally, to compare genetic and epigenetic diversity levels, we used this approach to calculate the distance from each sample to its population centroid using genetic distance matrices. Genetic distances were calculated as the average distance of all per−SNP differences between two individuals. For each SNP, the distance was set to 0 if both alleles were identical, 0.5 if one allele was different, and 1 if both alleles were different (*sensu* Mounger et al., 2021b).

We tested for epigenetic differentiation within and among *S. alterniflora* populations using similar approaches as for genetic differentiation: (1) PERMANOVA, (2) dbRDA, and (3) a close *cis* analysis for each cytosine. To assess the effect of population and habitat on genome-wide epigenetic variation across all sequence contexts, or each separate sequence context (i.e., CG, CHG, and CHH), we used the same PERMANOVA model and dbRDA. We also evaluated overall epigenetic variation with and without accounting for genetic structure, using partial constrained dbRDA. Partial constrained dbRDA allows for "conditioning" the analysis of epigenetic variation with genetic data which we summarized with principal component analysis (PCA).

For the dbRDA, we used only SMPs with complete data, i.e., no missing values across samples. First, we summarized the genetic data into principal components (PCs) based on 68,317 SNPs with alleles with a minor allele frequency (MAF) > 1%. We used the first four PCs which combined explained ∼14% of the genetic variation across all sequence contexts. Then, we ran the following models to predict DNA methylation in *S. alterniflora*:

1) Epigenetic distance ∼ population
2) Epigenetic distance ∼ habitat
3) Epigenetic distance ∼ population + habitat + population × habitat
4) Epigenetic distance ∼ population + habitat + population x habitat + Condition (PCs)
5) Epigenetic distance ∼ Condition (PCs)

Just as for the genetic data, we tested the significance of the variation explained by our explanatory variables using a Monte Carlo permutation test and obtained adjusted R^2, and adjusted *p*-values for multiple testing to reflect FDR.

We further evaluated how much of the variation in methylation is associated with populations and habitat before and after controlling for the underlying genetic variation *in cis* by comparing two models. We took advantage of the function "anova" in R (version 3.5.1) which uses type I (i.e., sequential) tests. We used the base functions lm and anova to calculate the model terms. We collected the terms for each reference sequence and adjusted P-values for each term for multiple testing to reflect FDR as above. First, we modeled the average DNA methylation level of individual reference sequences (i.e., ∼270 bp long contigs, n = 8,960) in response to the sequence context (CTXT), the population of origin (POP), the habitat (HAB), the interaction between the sequence context and the population (CTXT:POP) and habitat (CTXT:HAB), the interaction between population and habitat (POP:HAB) and the genotype of the reference sequence (GENO) and its interaction with sequence context (CTXT:GENO). We fit the terms first in this order, then compared this result to an alternative model in which POP and HAB were fitted after GENO.

1) percent methylation ∼ CTXT + POP + HAB + CTXT:POP + CTXT:HAB + POP:HAB + GENO + (CTXT:GENO).
2) percent methylation ∼ CTXT + GENO + CTXT:GENO + POP + HAB + CTXT:POP + CTXT:HAB + POP:HAB.

With this approach, the first model tested for epigenetic differentiation between populations of origin and habitats without first accounting for the sequence differences. The second model tested whether there was epigenetic differentiation

TABLE 1 | Number of alleles, allelic richness (i.e., average number of alleles per locus) and mean and standard deviation (SD) of observed gene diversity (H_s) and observed heterozygosity (H_o) per locus for each population.

Subpopulation(population × habitat)	No. alleles	Allelic richness	mean H_s	SD H_s	mean H_o	SD H_o
Bowens Island high marsh (27)	12,048	1.37	0.050	0.109	0.063	0.163
Bowens Island low marsh (40)	12,668	1.37	0.051	0.110	0.064	0.166
Folly Beach high marsh (28)	12,297	1.40	0.051	0.107	0.062	0.157
Folly Beach low marsh (37)	12,208	1.34	0.050	0.113	0.064	0.171
Fort Johnson high marsh (33)	12,670	1.40	0.054	0.113	0.069	0.174
Fort Johnson low marsh (46)	12,821	1.36	0.053	0.117	0.072	0.187
Overall (211)	17,553					

Overall number of alleles is also presented for the whole dataset. Calculated based on SNPs with no missing values (8,767).

between populations of origin or habitats that could not be explained by the underlying sequence differences since the effects of sequence differences were removed before evaluation of population and habitat. Results from this model cannot be directly compared with the results from our other models used to test for differential DNA methylation at individual cytosines, because this model uses an average level of DNA methylation across several cytosines within a reference contig. However, this model detects the dependency of epigenetic variation on genetic variation at the location of the DNA methylation (i.e., it only considers the variation within the ∼ 270 bp of a reference sequence). Such *cis* associations usually decay at relatively short distances (i.e., 200 bp in *Arabidopsis thaliana* or 1 kb in *Arabidopsis lyrata*; Hollister et al., 2011), and the model thus captures a good proportion of *cis* interactions. However, far-*cis* associations (for example an insertion variant which is close to, but not represented in the reference sequence) or trans dependencies (effects from loci anywhere else in the genome) will remain undetected (*sensu* van Moorsel et al., 2019; Mounger et al., 2021b). As a result, this model might overestimate the proportion of epigenetic variation which is unlinked to genetic variation.

We also analyzed variation in methylation at each individual cytosine with a linear model in R with the package DSS version 2.24.0 (Park and Wu, 2016) with a single factor summarizing population and habitat as separate levels (i.e., subpopulations) and using contrasts to compare levels of interest following the method for RNA-Seq in Schmid (2017) and the testing procedure in Schmid et al. (2018a). We compared groups using linear contrasts. We adjusted p-values for each contrast for multiple testing to reflect FDR. A cytosine was defined as differentially methylated ("DMC") if the FDR was below 0.01 for any of the contrasts [see also Schmid et al. (2018b)].

Correlation of Phenotype to Genetics, Epigenetics, and Environment

We compared phenotypic distance to genetic, epigenetic, and environmental distances using multiple MMRR adapted from Herrera et al. (2017). We created pairwise distance matrices for phenotypic [stem height (cm), number of leaves per stem, number of live and dead stems per quadrat] and environmental (soil salinity, temperature, oxygen, number of snails present, percent soil moisture, percent organic matter, respiration rate, soil extractable NH_4^+, soil extractable NO_3^-, porewater NH_4^+, porewater NO_3^-, microbial biomass C, and microbial biomass N) distances between subpopulations (population × habitat combinations) with function as.matrix in the R environment (R Core Team, 2020). We scaled and centered the phenotypic and environmental matrices using Euclidean distance (Herrera et al., 2017). We assessed the correlation between phenotypic and environmental distances using a Mantel test conducted in R with the package vegan as above (Oksanen et al., 2020). We regressed standardized genetic, epigenetic, and environmental distance matrices on the phenotypic distance matrix using the MMRR function in R as described in Wang and Bradburd (2014) and Herrera et al. (2017). We standardized each matrix to mean = 0 and SD = 1 to compare relative magnitude of effects of the predictor matrices. We ran this analysis on the full data set, subsetted by habitat, subsetted by population and subsetted by subpopulation to explore the patterns in more detail. For each subset, we restandardized the data.

RESULTS

Our *de novo* reference creation resulted in 183,319 fragments with a total length of 33,193,750 bp. About 10% of the fragments overlapped with functional proteins, while nearly 24% overlapped with repeats (1.94%) and transposons (21.72%; **Supplementary Table 1**). After filtering poorly covered loci and samples, we retained data for 211 out of 293 sequenced samples. We found 12,187 of the 183,319 fragments contained 414,449 SNPs before filtering for MAF <1%. After removing loci with MAF <1%, 9,429 fragments remained with 68,317 SNPs. For DNA methylation, we found 15,463 fragments with 512,559 cytosines with sufficient coverage.

Population Genetics

We estimated the proportion of polymorphic loci in this collection of plants to be 6.6% based on the number of fragments that contained SNPs before the MAF filter. This dropped to 5.1% after we removed the loci with MAF <1%. With this filter, the number of SNPs/kb of sequence dropped from 5.5 to 2.1. Mean allelic richness averaged 1.37 for subpopulations (population × habitat combinations; **Table 1**). We observed gene diversity with values per locus ranging between 0.050 and 0.054 and heterozygosity between 0.062 and 0.072 (**Table 1**). Overall

TABLE 2 | Results of the distance-based redundancy analysis (dbRDA) showing the percentage of genetic and epigenetic variance explained by population (P), habitat (H), and/or their interaction (P:H).

Model	df	F value	FDR	adj. R^2
GD ~ P	2	7.922	0.0001	6.4
GD ~ P + H	3	5.932	0.0001	6.8
GD ~ P + H + P:H	5	4.77	0.0001	8.5
ED ~ P	2	5.141	0.0001	3.8
ED ~ H	1	1.290	0.0145	0.14
ED ~ P + H	3	3.884	0.0001	3.96
ED ~ P + H + P:H	5	3.067	0.0001	4.7
ED ~ P + H + P:H + Condition(PCs)	5	1.148	0.0001	0.32
ED ~ Condition(PCs)	10	4.594	0.0001	14.6

In the case of the epigenetic data, the models include with and without adjusting for the variance explained by the genetic data (PCs). GD, genetic distance matrix; ED, epigenetic distance matrix; PC, first four principal components from genetic data (explaining 10.41% of the variation in scaled allele frequencies); df, degrees of freedom; F value, value of the test statistic; FDR, false discovery rate; adj. R^2, percent of variance explained, adjusted for multiple comparisons.

genetic diversity (H_t) across subpopulations was very similar, i.e., 0.0521 (**Supplementary Table 2**).

Each of our three methods to examine genetic structure provided evidence of significant genetic differentiation within and among populations of *S. alterniflora*. With PERMANOVA, we found that all effects in the model explained a significant (at FDR < 0.01) proportion of the genetic variation: population of origin explained 8.6%, habitat explained 0.9% and the interaction of population and habitat explained an additional 2.9%. Similarly, the RDA (**Table 2**) showed that population alone explained 6.4% of the genetic variation, while the addition of habitat and the interaction term added an additional 2.1% explanatory power (R^2 = 8.5; FDR = 0.0001).

Overall F_{ST} was low (0.026 across the 3 populations or 0.030 across the 6 subpopulations). Pairwise comparisons of the three populations ranged from (F_{ST} = 0.0187–0.0301; **Figure 3A**). Pairwise comparisons of the six subpopulations ranged from (F_{ST} = 0.0035–0.0565). Comparisons across subpopulations within the high marsh habitats revealed comparatively low F_{ST} ranging from 0.0035 between Bowens Island and Folly Beach and 0.0209 between Fort Johnson and Bowens Island (**Figure 3B**). On the other hand, comparisons among the low marsh habitats revealed on average more differentiation (F_{ST} = 0.0455–0.0565), particularly between Folly Beach and Fort Johnson (F_{ST} = 0.0565; **Figure 3B**).

We visualized the genetic data by means of PCA using the complete SNP dataset as well as the 5% most differentiated SNPs. This visualization suggested that there was a separation among populations, and partial separation of habitat within populations with the 5% most differentiated loci (**Figure 4**).

The maximum likelihood phylogeny generally showed three major genotype clusters that roughly correspond to each of the three sites (Folly Beach, Fort Johnson, and Bowens Island; **Figure 5**). These genotype clusters were not reciprocally monophyletic clades, indicating that sites have a mixture of the three clusters. It also appeared that genotypes were well mixed between habitats within sites.

Population Epigenetics

We found that DNA methylation across all contexts averaged 11.8% for all subpopulations, while DNA methylation levels in CG, CHG and CHH contexts averaged 53.3, 28.7, and 4.4%, respectively, for all subpopulations in the dataset (**Table 3**). We also found that about half of the cytosines were fixed for no methylation (50.1% across all contexts for all subpopulations, ranging between 36.8 and 60.6% within subpopulations; **Table 3**).

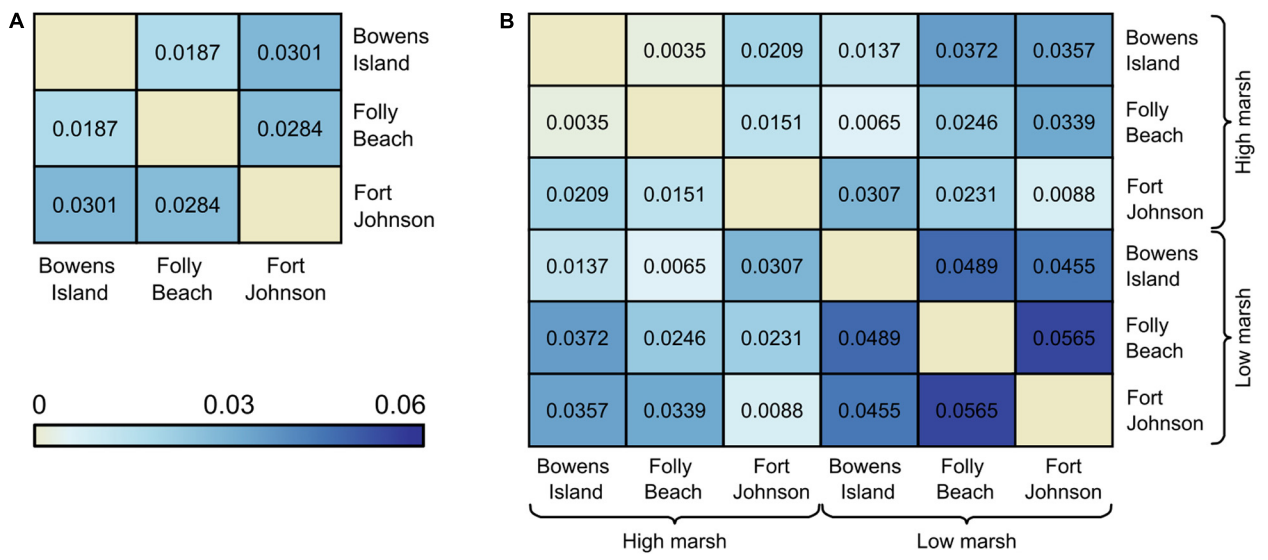

FIGURE 3 | Pairwise F_{ST} values **(A)** among field populations of *Spartina alterniflora*, and **(B)** between habitats (low marsh and high marsh) within populations. All values were significantly genetically differentiated pairs based on 95% CI.

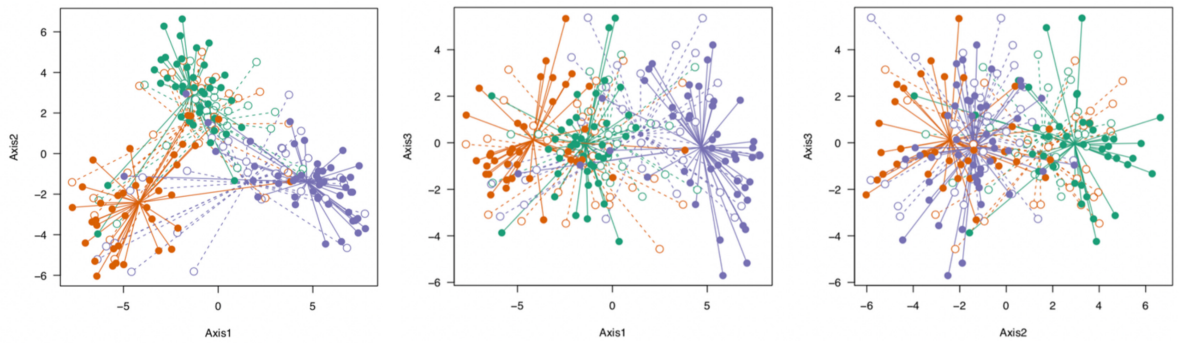

FIGURE 4 | Visualization of the genetic structure of *S. alterniflora* populations by PCA (PC1 explains 16.5% of the variation, PC2 explains 8%, and PC3 explains 4.6%). We used only the top 5% most differentiated SNPs based on Jost's D (Jost, 2008), and only SNPs without missing values ($N = 80$). Each dot represents one stem; populations are indicated by color (Green = Bowens Island, Red = Folly Beach, Blue = Fort Johnson) and habitats are filled (low marsh plants) or open (high marsh plants). See also **Supplementary Figure 2** for similar depiction of epigenetic structure of populations.

A substantially smaller proportion of cytosines were fixed for full methylation (1.4% across all contexts for all subpopulations, with a range of 0.5–2% within subpopulations; **Table 3**).

Using average distances from each sample to its subpopulation centroid, we estimated that epigenetic variation ranged between 0.08 and 0.10 (**Figure 6**). For comparison, the average distances to centroid estimated with the genetic data were between 0.015 and 0.020. The test for homogeneity of multivariate dispersions was highly significant ($F = 9.04$, $p < 0.001$), suggesting differences in epigenetic diversity among subpopulations.

With PERMANOVA, we found that population of origin explained a significant component of the overall epigenetic variance (3.2%; FDR < 0.0001). In addition, we found that a small but significant amount of epigenetic variance (0.6%; FDR = 0.006) was explained by habitat. The interaction of population and habitat explained an additional 1.4% (FDR = 0.0001) of the epigenetic variance. We found similar results in each of the cytosine contexts. Population explained 2.6–4.6% of the epigenetic variation within each of the CG, CHG and CHH contexts (FDR < 0.0001 for all). Habitat explained 0.6–0.7% for each context (FDR < 0.01). The interaction explained 1–2% for each context (FDR < 0.01). The RDA results were similar (**Table 2**): population alone explains 3.8% (FDR = 0.0001) of the epigenetic variation while the addition of habitat and the interaction term explains an additional 0.9% ($R^2 = 4.7$; FDR = 0.0001).

Correlation Between Population Genetics and Epigenetics

Using dbRDA we discovered that PCs of genetic data alone explained more of the variation in epigenetic distances than any other model (**Table 2**). In addition, we found that when conditioned on PCs of the genetic data, the partial constrained dbRDA only explained less than 1% of the variation (**Table 2**).

The genotype of the reference sequence explained much of the epigenetic diversity in close-cis even when fitted last (i.e., 28.5% for genotype and 23.5% for genotype x context; **Table 4**). However, among populations we found that epigenetic variation was different in around 9.2% of 8,960 tested fragments (of the 9,429 only 8,960 intersect with those that have methylation data), even after accounting for differences in DNA sequence of the underlying loci. Habitat only explained variation in < 1% of the cytosines when evaluated across populations, however, the interaction between populations and habitats explained variation in 2% of the cytosines when fitted last, but nearly 7% of the cytosines when fitted before genotype.

The pairwise comparison of differential methylation at individual cytosines yielded between 0.02 and 6.1% significantly differentially methylated cytosines (**Table 5**). We found that the

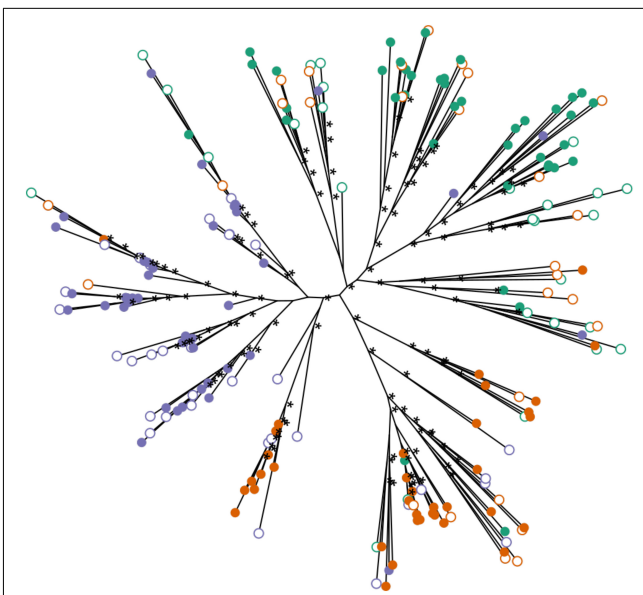

FIGURE 5 | Unrooted phylogenetic tree of *S. alterniflora* populations by maximum likelihood based on all SNPs filtered for MAF > 10% and no missing data. Each dot represents one stem; populations are indicated by color (Green = Bowens Island, Red = Folly Beach, Blue = Fort Johnson) and habitats are filled (low marsh plants) or open (high marsh plants). Asterisks indicate maximum likelihood support with bootstrap values greater than 95%.

TABLE 3 | Proportions of **(A)** methylated cytosines per context and for all contexts, **(B)** fixed cytosine positions for no methylation, and **(C)** fixed cytosine positions for methylation.

(A)	**All**			
Subpopulation (habitat x site)	**ALL**	**CG**	**CHG**	**CHH**
Bowens Island high marsh	11.718	53.543	28.534	4.261
Bowens Island low marsh	11.664	53.283	28.525	4.221
Folly Beach high marsh	11.713	53.274	28.62	4.268
Folly Beach low marsh	11.877	53.47	29.008	4.389
Fort Johnson high marsh	11.999	53.088	28.659	4.649
Fort Johnson low marsh	12.03	53.116	28.79	4.662
(B)	**Percent fixed unmethylated**			
	ALL	**CG**	**CHG**	**CHH**
Bowens Island high marsh	60.59	24.81	40.54	67.98
Bowens Island low marsh	52.65	22.35	34.83	59.06
Folly Beach high marsh	63.4	26.22	42.52	71.09
Folly Beach low marsh	47.89	21.16	32.62	53.46
Fort Johnson high marsh	39.26	17.6	26.74	43.81
Fort Johnson low marsh	36.77	16.66	24.62	41.08
(C)	**Percent fixed methylated**			
	ALL	**CG**	**CHG**	**CHH**
Bowens Island high marsh	2.01	11.83	0.83	1.18
Bowens Island low marsh	1.38	7.2	0.39	0.94
Folly Beach high marsh	2.36	13.91	1.35	1.32
Folly Beach low marsh	1.21	6.76	0.38	0.77
Fort Johnson high marsh	0.84	4.29	0.2	0.59
Fort Johnson low marsh	0.54	2.14	0.1	0.45

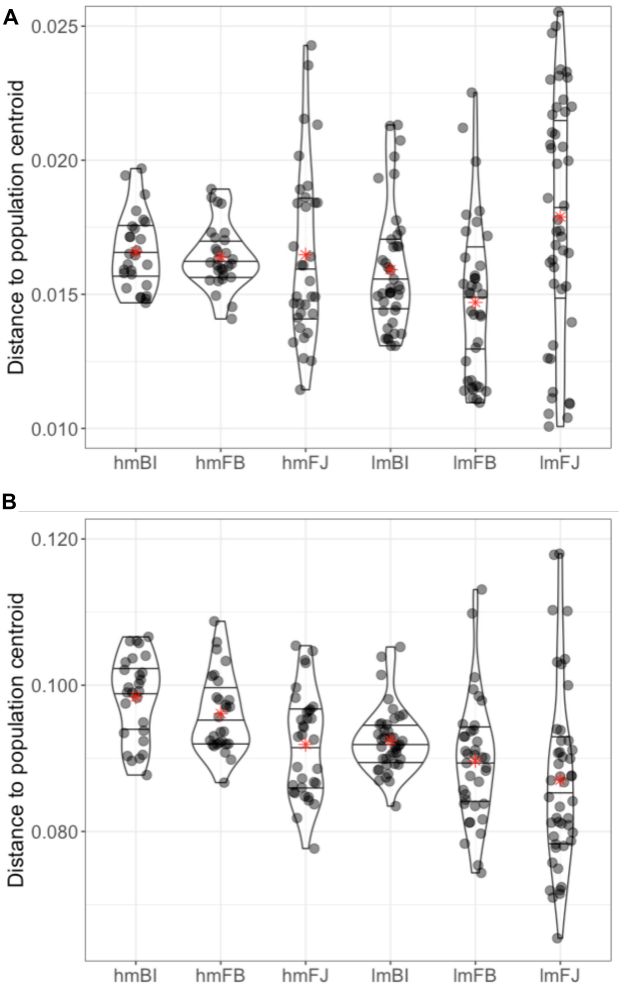

FIGURE 6 | Distance from each individual sample to its corresponding subpopulation centroid calculated using **(A)** genetic distance matrices and **(B)** epigenetic distance matrices for subpopulations (site × habitat). Lines within the violin plots represent the 25, 50, and 75% quartiles of the distribution and red asterisks denote the average distance to centroid for each subpopulation. Subpopulations are as follows: high and low marsh Bowens Island (hmBI, lmBI), Folly Beach (hmFB, lmFB), and Fort Johnson (hmFJ, lmFJ).

most pronounced differences were between Bowens Island vs. Fort Johnson (6.1% significant Cs) and in the comparison of low marsh plants in these two populations (3.99%) which contrasted with the comparison of the high marsh plants (0.95%). We found little support for differences between habitats at any level (≤0.5% significant Cs). Most of the SMPs were located in unannotated portions of the genome (63.0% on average among pairwise comparisons) or in TEs (24.1%). A few SMPs were located in genes and other repeats (on average, 16.6 and 2.9% respectively). Although most cytosines in our libraries were in the CHH context (78% in CHH compared to 8% in CG and 14% in CHG), the number of SMPs tended to be higher in the CG context (22, 17, and 60% in CG, CHG, and CHH, respectively).

Correlation of Phenotype to Genetics, Epigenetics, and Environment

In the MMRR analysis, the overall regression explained about 8.4% of the variation in phenotypic distance (p-value = 0.001; **Table 6A**). Genetic, epigenetic and environmental distances each explained a significant proportion of the variance in phenotypic distances (p-values ranged between 0.001 and 0.036). The standardized regression coefficients suggested that epigenetic and environmental distance matrices were stronger predictors of the variation in the phenotypic matrix than the genetic matrix. Subsetting the plants by habitat revealed similar, slightly stronger patterns in the high marsh plants, but none of the relationships held in the low marsh plants alone (**Table 6B**). By population, we found divergent patterns from no significant relationship for the overall model in Folly Beach plants, to the model that explained the most variation (R^2 = 0.312) in Fort Johnson plants (**Table 6C**. In this population, genetic distance explained more of the variation than environmental variation did, while epigenetic distance did not explain variation in phenotypic distance. Looking at each of the subpopulations separately, genetic, epigenetic and environmental distance were often not significantly associated with phenotypic distance (**Table 6D**).

TABLE 4 | Anova (type I) models of amount (%) of variation in methylation explained by population, habitat, and sequence context before (left) and after (right) accounting for genotype effects on methylation using MAF-filtered data only (8,960 contigs).

Population and habitat before GENO		GENO before population and habitat	
Context	96.33	Context	96.33
Population	19.56	Genotype	31.07
Habitat	0.49	Context × genotype	25.1
Context × population	17.89	Population	9.22
Context × habitat	0.42	Habitat	0.2
Population × habitat	6.9	Context × population	9.63
Genotype	28.45	Context × habitat	0.04
Context × genotype	23.54	Population × habitat	2.07

TABLE 5 | Summary of results of linear contrasts to evaluate variation in methylation at each individual cytosine with the package DSS (version 2.24.0; Park and Wu, 2016).

Comparison	Included levels	% DMC
Low marsh vs high marsh	across all sites	0.02
	within FB	0.31
	within BI	0.39
	within FJ	0.09
BI vs FB	with both habitats	0.42
	given low marsh plants	1.22
	given high marsh plants	0.14
FB vs FJ	with both habitats	2.28
	given low marsh plants	2.19
	given high marsh plants	0.34
BI vs FJ	with both habitats	6.07
	given low marsh plants	3.99
	given high marsh plants	0.95

DMC, differentially methylated cytosine [FDR < 0.01; see also Schmid et al. (2018b)]. FB, Folly Beach; BI, Bowens Island; FJ, Ford Johnson.

DISCUSSION

Several studies have shown that epigenetic changes can be induced by environmental variation and can be partially independent of underlying genetic variation, but very few have linked these patterns with phenotypic variation (Herrera et al., 2017; Wang et al., 2020). We addressed these questions with natural populations of *S. alterniflora*, a species known to have high genetic diversity and to vary dramatically in traits along a steep environmental gradient. We predicted that DNA sequence variation and epigenetic variation would be explained by source population and habitat (low and high marsh). Moreover, we expected that DNA methylation patterns would be more strongly associated with habitat since the same epigenetic modifications might be induced across genetic backgrounds in response to environmental conditions. Accordingly, we predicted that epigenetic and environmental differences would associate with variation in phenotypic traits, even after accounting for genetic differences. We found that populations, habitats and the interaction of populations and habitats explained variation in both genetic and epigenetic variation. However, these models, and the habitat effect, were better able to explain variation in DNA sequence than they were variation in DNA methylation. Several lines of evidence support the contention that sequence variation explains most of the variation in DNA methylation genome wide and in close-*cis*. However, a small percentage of epigenetic variation was explained by habitat or the interaction of population and habitat even after correcting for genetic variation in close-*cis*. Lastly, we show that epigenetic variation in this system contributes to phenotypic variation even when genetic and environmental distances are accounted for.

Population Genetics

Previous work on native populations of *S. alterniflora* has reported high levels of genetic diversity (Richards et al., 2004; Hughes and Lotterhos, 2014; Foust et al., 2016; Zerebecki et al., 2021). Richards et al. (2004) found high genetic diversity in stands on Sapelo Island, Georgia using allozymes, and these levels of diversity were later confirmed among the same populations using AFLP (Foust et al., 2016). Similarly, Hughes and Lotterhos (2014) found that high genetic diversity persisted even at small spatial scales across marshes in the Florida Panhandle (Hughes and Lotterhos, 2014). Travis et al. (2004) found that genetic diversity among populations along coastal Louisiana varied according to age, with younger populations (<16 years) supporting higher genetic diversity than older, established populations (Travis et al., 2004).

More recently, Zerebecki et al. (2021) found significant genetic differentiation among habitats within our populations of *S. alterniflora* near Charleston, South Carolina as well as across a much broader geographic range. They showed that the phenotypic divergence in stem height and biomass allocation observed across intertidal gradients was at least partly due to local genetic differentiation (Zerebecki et al., 2021). Further, their reciprocal transplant experiment combined with repeated patterns of molecular differentiation among habitats within individual marshes supported parallel microgeographic adaptation to intertidal gradients in natural populations of *S. alterniflora* (Zerebecki et al., 2021).

Estimates of heterozygosity and genetic diversity more generally are not easily compared across studies with different sample sizes, different types of molecular markers, and with different filtering approaches (Sunde et al., 2020; Schmidt et al., 2021). In this study on *S. alterniflora,* we found that H_o was approximately six times and H_t was five times that of *Rhizophora mangle* from natural populations using similar approaches (Mounger et al., 2021b), confirming previous reports of high genetic diversity in these Charleston populations of *S. alterniflora* (Zerebecki et al., 2021). Although habitat alone, and in interaction with population explained variation in genetic differences, population was the best predictor of genetic variation. In both RDA and PERMANOVA, habitat added very little explanatory power, except through the interaction with population. This result provides some insight into our previous work that examined the association between phenotypic variation and edaphic factors (Voors, 2018). In that study, habitat

TABLE 6 | Results of multiple matrix regression with randomization (MMRR) analysis relating the combined phenotypic distance matrix across all *S. alterniflora* samples with the matrices of genetic (SNP), epigenetic (SMPs), and environmental distances for **(A)** across all samples, **(B)** within each habitat type across populations, **(C)** within each population across habitats and **(D)** for each subpopulation. NS indicates p value > 0.10.

(A) Across all samples (*N* = 211 individuals)

			Predictor matrices					
Overall regression			Genetic distance		Epigenetic distance		Environmental distance	
F	*p*-Value	R^2	Coefficient	*p*-Value	Coefficient	*p*-Value	Coefficient	*p*-Value
681.22	0.001	0.084	−0.114	0.014	0.229	0.001	0.249	0.001

(B) Within high marsh (*N* = 88) and low marsh (*N* = 123) habitat across populations

				Predictor matrices					
	Overall regression			Genetic distance		Epigenetic distance		Environmental distance	
	F	*p*-Value	R^2	Coefficient	*p*-Value	Coefficient	*p*-Value	Coefficient	*p*-Value
High marsh	167.98	0.001	0.116	−0.127	0.07	0.314	0.006	0.311	0.001
Low marsh	12.70	NS	0.005	−0.056	NS	0.025	NS	0.062	NS

(C) Within each population across habitats (*N* = 65-79)

				Predictor matrices					
	Overall regression			Genetic distance		Epigenetic distance		Environmental distance	
	F	*p*-Value	R^2	Coefficient	*p*-Value	Coefficient	*p*-Value	Coefficient	*p*-Value
Bowens Island	116.69	0.001	0.137	0.032	NS	0.314	0.006	0.311	0.001
Folly Beach	19.84	NS	0.028	−0.230	NS	0.345	0.09	0.303	0.001
Fort Johnson	464.66	0.001	0.312	−0.102	0.008	0.056	NS	0.562	0.001

(D) Within each sub-population (*N* = 27-46 individuals)

				Predictor matrices					
	Overall regression			Genetic distance		Epigenetic distance		Environmental distance	
	F	*p*-Value	R^2	Coefficient	*p*-Value	Coefficient	*p*-Value	Coefficient	*p*-Value
BI high	4.38	NS	0.036	0.044	NS	0.088	NS	0.186	0.003
BI low	15.78	NS	0.057	0.413	0.058	−0.478	NS	0.017	NS
FB high	4.38	NS	0.034	−0.078	NS	0.737	NS	0.091	NS
FB low	3.96	NS	0.018	−0.237	NS	0.423	NS	0.042	NS
FJ high	23.17	0.003	0.117	−0.155	NS	−0.201	NS	0.326	0.001
FJ low	18.74	0.03	0.052	−0.209	0.044	−0.261	0.047	0.088	NS

explained a large portion of the variation in plant traits, but not in soil characteristics. The patterns in genetic marker diversity could reflect that the environmental differences between high-marsh and low-marsh habitats are not consistent among populations (i.e., among sites). Thus, different environmental conditions could be driving divergent genetic patterns between the two habitats, but not necessarily in the same way at the various populations, leading to population-by-habitat interactions.

We were able to visualize a clear differentiation between populations of origin when the 5% most differentiated loci were compared in a PCA. The most obvious difference was the separation of Fort Johnson. This differentiation was also supported by the pairwise F_{ST} analyses where the highest values were comparisons between Fort Johnson and the other populations, particularly in the low marsh. The Fort Johnson population of *S. alterniflora* is situated within Charleston Harbor, while Bowens Island and Folly Beach populations are geographically closer together, connected by the Folly River and situated along less intensely developed tidal creeks to the south of the main shipping harbor. Genetic differentiation at Fort Johnson relative to the other two populations may in part be driven by geographical distance (Herrera et al., 2017; Wang et al., 2020; Mounger et al., 2021b), population age (Travis et al., 2004), cryptic environmental variation (Zerebecki et al., 2021), or by anthropogenic impacts such as long-term or historic exposure to urban effluents and industrial

pollutants (Robertson et al., 2017). Additional research is required to investigate which of these factors may be driving the genetic differentiation we detected at Fort Johnson. However, we also found support for differences among populations using all SNPs in the phylogenetic tree. More specifically, the tree supported the contention made by Zerebecki et al. (2021) that the differences we found among habitats could have evolved within populations. An alternative hypothesis would be that the high marsh and low marsh habitat at each site was colonized by high marsh and low marsh genotypes that had diverged previously. Additional research is required to investigate which processes may explain the lack of genetic differentiation at Fort Johnson. Our study also supports the possibility of convergent processes particularly in the high marsh which could be contributing to less differentiation among subpopulations in the high marsh than we found among subpopulations in the low marsh.

Population Epigenetics

Although epigenetic variation can be dictated by underlying genetic variation (Becker et al., 2011; Dubin et al., 2015; Sasaki et al., 2019), we found a small but statistically significant amount of epigenetic variation was associated with habitat even after accounting for genetic variation. This finding is important because it supports the theoretical understanding that epigenetic mechanisms may provide an additional source of variation that is distinct from genetic variation in natural populations. Our results suggest that the habitat differences in natural populations of *S. alterniflora* may correspond to both genetic and epigenetic differentiation, but more fine scale analyses are needed to dissect these relationships. This work supported our previous studies of differences in these molecular mechanisms within this species. For instance, Foust et al. (2016) found that a small percentage of epigenetic variation across sites was associated with similar habitats after accounting for genetic variation. Similarly, a study of response to crude oil exposure found both genetic and epigenetic differentiation between oiled and non-oiled populations in Southern Louisiana following the Deepwater Horizon oil spill (Robertson et al., 2017).

Genetic, Epigenetic, and Environmental Variation Correlates With Phenotypic Variation

Herrera et al. (2017) compared the genetic, epigenetic, environmental (six habitat features) and phenotypic (20 traits) variation of ten subpopulations of the perennial herb *Helleborous foetidus* using MMRR. They determined that geographical distance explained genetic differentiation among subpopulations, while environmental distance explained epigenetic differentiation (Herrera et al., 2017; Wang et al., 2020). These results for *H. foetidus* supported earlier work that showed that epigenetic isolation by environment (IBE) can arise as a consequence of native epigenotypes having higher fitness in a given environment than epigenotypes from a different subpopulation (Wang and Bradburd, 2014; Herrera et al., 2017; Wang et al., 2020).

Following these methods, we used standardized genetic, epigenetic and environmental distance matrices to determine if these variables might explain the phenotypic variation we have observed within habitats or within populations. We found significant relationships between phenotypic distance across our *S. alterniflora* samples and all three predictor matrices of genetic, epigenetic, and environmental distances. Using standardized distance matrices allowed us to infer that environmental and epigenetic distances each explained more of the phenotypic distance than did genetic distance in the overall collection of plants. However, additional analyses revealed that these relationships varied within habitats and within populations. We found no relationship between these predictors and phenotypic variation in the low marsh plants, but the relationships between epigenetic and phenotypic distance and between environmental and phenotypic distance appeared to be even stronger when evaluating only the high marsh plants. Further work will be required to assess the contribution of these different components to phenotypic variation in the context of complex natural populations and habitats. These results indicate that genetic, epigenetic, and environmental variation contribute to phenotypic response in these plants even when evaluated simultaneously. Considering that we expected most of the phenotypic and environmental differences to occur largely between habitats or among populations, the lack of relationships within separate subpopulations may be the result of limited variance.

CONCLUSION AND FUTURE PERSPECTIVES

The importance of epigenetic mechanisms in ecological contexts is still an open question that is difficult to address (Boquete et al., 2021). This line of inquiry has a variety of constraints. In particular, plant populations with high levels of genetic diversity challenge our ability to clearly distinguish patterns of epigenetic variation (Foust et al., 2016). In addition, the lack of well annotated reference genomes for most plant species greatly limits our ability to make inferences about the functionality of genes or methylation patterns (Boutte et al., 2016; Richards et al., 2017; Paun et al., 2019; Mounger et al., 2021a). Reduced Representation Bisulfite Sequencing (RRBS) methods like epiGBS provide substantially more genome coverage than previous methods such as methylation-sensitive amplified fragment length polymorphism (MS-AFLP or MSAP; Schrey et al., 2013; Richards and Pigliucci, 2020). However, for large polyploid genomes like that of *S. alterniflora* this approach still only captures a small percentage of the genome (approximately 3%). In addition, polyploidization, or genome doubling, can provide additional sources of molecular variation which thereby contribute to plant responses to environmental conditions, but this phenomenon complicates our ability to analyze and draw conclusions about genomic mechanisms (Dufresne et al., 2014; Robertson et al., 2020; Mounger et al., 2021a). RRBS methods such as epiGBS do not yet provide adequate coverage to confidently estimate polyploid genotype likelihoods (Dufresne et al., 2014; Alvarez et al., 2020;

Robertson et al., 2020). Despite these limitations, this work adds to a growing body of evidence that shows that epigenetic differences, particularly variation in DNA methylation, are associated with responses to environmental variation in natural populations. This association is not explained simply by sequence based differences and requires continued investigation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the following online repositories: https://zenodo.org/record/5833452, doi: 10.5281/zenodo.5833452; https://www.ncbi.nlm.nih.gov/, PRJNA798549.

AUTHOR CONTRIBUTIONS

CR, AH, and CG conceived the study. CR, AH, CG, and DL designed the experiments. MS, MB, and CR designed the genomics analyses. JM, TH, CR, AH, ES, SV, and MR collected plants from the field. JM, SV, MR, and CW did the epiGBS laboratory work. MS, JM, MB, JO, FG, IG, and KV adapted the epiGBS pipeline for use. MS, JM, IR, MB, and ES analyzed the data. JM, IR, MS, and CR wrote the first draft of the manuscript. All authors provided input and revisions to the manuscript.

FUNDING

This work was supported by funding from the National Science Foundation (United States) DEB-1419960 and IOS-1556820 (CR), IOS-1556738 (AH), IOS-1556087 (CG); the Federal Ministry of Education and Research (BMBF; MOPGA Project ID 306055 to CR); the German Research Foundation (DFG) grant number: 431595342 (CR); the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 704141-BryOmics (MB).

ACKNOWLEDGMENTS

We thank Stephen Savage for help with the epiGBS laboratory work. We thank the Grice Marine Laboratory at the College of Charleston for providing materials and laboratory space to aid in field collections. We thank Carlos Herrera and Ian Wang for their insights and advice about the MMRR methods. We acknowledge the contribution of the University of South Florida Genomics Program and USF Omics Hub for completion of this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.868826/full#supplementary-material

REFERENCES

Alonso, C., Ramos-Cruz, D., and Becker, C. (2019). The role of plant epigenetics in biotic interactions. *New Phytol.* 221, 731–737. doi: 10.1111/nph.15408

Alvarez, M., Ferreira, de Carvalho, J., Salmon, A., Ainouche, M. L., Cavé-Radet, A., et al. (2018). Transcriptome response of the foundation plant Spartina alterniflora to the Deepwater Horizon oil spill. *Mole. Ecol.* 27, 2986–3000. doi: 10.1111/mec.14736

Alvarez, M., Robertson, M., van Gurp, T., Wagemaker, N., Giraud, D., Ainouche, M. L., et al. (2020). Reduced representation characterization of genetic and epigenetic differentiation to oil pollution in the foundation plant Spartina alterniflora. *bioRxiv* doi: 10.1101/426569

Anderson, M. J., Ellingsen, K. E., and McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecol. Lett.* 9, 683–693. doi: 10.1111/j.1461-0248.2006.00926.x

Ashe, A., Colot, V., and Oldroyd, B. P. (2021). How does epigenetics influence the course of evolution? *Philosoph. Trans. R. Soc. Lond. Series B, Biol. Sci.* 376:20200111. doi: 10.1098/rstb.2020.0111

Balao, F., Paun, O., and Alonso, C. (2018). Uncovering the contribution of epigenetics to plant phenotypic variation in Mediterranean ecosystems. *Plant Biol.* 20, 38–49. doi: 10.1111/plb.12594

Banta, J. A., and Richards, C. L. (2018). Quantitative epigenetics and evolution. *Heredity* 121, 210–224. doi: 10.1038/s41437-018-0114-x

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., et al. (2011). Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480, 245–249. doi: 10.1038/nature10555

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bertness, M. (2020). *A Brief Natural History of Civilization: Why a Balance Between Cooperation and Competition Is Vital to Humanity*. New Haven: Yale University Press, doi: 10.2307/j.ctvzpv6np

Bertness, M. D., and Ellison, A. M. (1987). Determinants of pattern in a New England salt marsh plant community. *Ecol. Monogr.* 57, 129–147. doi: 10.2307/1942621

Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., et al. (2015). What we still don't know about invasion genetics. *Mole. Ecol.* 24, 2277–2297. doi: 10.1111/mec.13032

Boquete, M. T., Muyle, A., and Alonso, C. (2021). Plant epigenetics: phenotypic and functional diversity beyond the DNA sequence. *Am. J. Bot.* 108, 553–558. doi: 10.1002/ajb2.1645

Boquete, M. T., Schmid, M. W., Wagemaker, N. C. A. M., Carey, S. B., McDaniel, S. F., Richards, C., et al. (2022). Molecular basis of intraspecific differentiation for heavy metal tolerance in the copper moss Scopelophila cataractae. *Authorea* 2022:91599576. doi: 10.22541/au.164476549.91599576/v1

Boquete, M. T., Wagemaker, N. C. A. M., Vergeer, P., Mounger, J., and Richards, C. L. (2020). "Epigenetic Approaches in Non-Model Plants," in *Plant Epigenetics and Epigenomics: Methods and Protocols*, Vol. 2093, ed. P. McKeown (New York, NY: Springer Science+Business Media, LLC), 203–215. doi: 10.1007/978-1-0716-0179-2_14

Boutte, J., Ferreira de Carvalho, J., Rousseau-Gueutin, M., Poulain, J., Da Silva, C., Wincker, P., et al. (2016). Reference Transcriptomes and Detection of Duplicated Copies in Hexaploid and Allododecaploid *Spartina Species* (Poaceae). *Gen. Biol. Evol.* 8, 3030–3044. doi: 10.1093/gbe/evw209

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Burggren, W. (2016). Epigenetic inheritance and its role in evolutionary biology: Re-evaluation and new perspectives. *Biology* 5:24. doi: 10.3390/biology5020024

Cortijo, S., Wardenaar, R., Colomé-Tatché, M., Gilly, A., Etcheverry, M., Labadie, K., et al. (2014). Mapping the epigenetic basis of complex traits. *Science* 343, 1145–1148. doi: 10.1126/science.1248127

Des Marais, D. L., Hernandez, K. M., and Juenger, T. E. (2013). Genotype-by-Environment Interaction and Plasticity: exploring Genomic Responses of Plants

to the Abiotic Environment. *Annu. Rev. Ecol. Evol. Syst.* 44, 5–29. doi: 10.1146/annurev-ecolsys-110512-135806

Dodd, R. S., and Douhovnikoff, V. (2016). Adjusting to Global Change through Clonal Growth and Epigenetic Variation. *Front. Ecol. Evol.* 4:86. doi: 10.3389/fevo.2016.00086

Dubin, M. J., Zhang, P., Meng, D., Remigereau, M.-S., Osborne, E. J., Paolo Casale, F., et al. (2015). DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *eLife* 4:e05255. doi: 10.7554/eLife.05255

Dufresne, F., Stift, M., Vergilino, R., and Mable, B. K. (2014). Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Mole. Ecol.* 23, 40–69. doi: 10.1111/mec.12581

Edwards, K. R., Travis, S. E., and Proffitt, C. E. (2005). Genetic effects of a large-scaleSpartina alterniflora (smooth cordgrass) dieback and recovery in the northern Gulf of Mexico. *Estuaries* 28, 204–214. doi: 10.1007/bf02732855

Ellison, A. M. (2019). Foundation species, non-trophic interactions, and the value of being common. *iScience* 13, 254–268. doi: 10.1016/j.isci.2019.02.020

Fischer, M. C., Foll, M., Excoffier, L., and Heckel, G. (2011). Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (Microtus arvalis). *Mole. Ecol.* 20, 1450–1462. doi: 10.1111/j.1365-294x.2011.05015.x

Foll, M., and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180, 977–993. doi: 10.1534/genetics.108.092221

Foust, C. M., Preite, V., Schrey, A. W., Alvarez, M., Robertson, M. H., Verhoeven, K. J. F., et al. (2016). Genetic and epigenetic differences associated with environmental gradients in replicate populations of two salt marsh perennials. *Mole. Ecol.* 25, 1639–1652. doi: 10.1111/mec.13522

Gallagher, J. L., Somers, G. F., Grant, D. M., and Seliskar, D. M. (1988). Persistent differences in two forms of Spartina alterniflora: a common garden experiment. *Ecology* 69, 1005–1008. doi: 10.2307/1941255

Gao, Z., Moorjani, P., Sasani, T. A., Pedersen, B. S., Quinlan, A. R., Jorde, L. B., et al. (2019). Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc. Natl. Acad. Sci. U S A* 116, 9491–9500. doi: 10.1073/pnas.1901259116

Gáspár, B., Bossdorf, O., and Durka, W. (2019). Structure, stability and ecological significance of natural epigenetic variation: a large-scale survey in Plantago lanceolata. *New Phytol.* 221, 1585–1596. doi: 10.1111/nph.15487

Gawehns, F., Postuma, M., van Antro, M., Nunn, A., Sepers, B., Fatma, S., et al. (2022). epiGBS2: improvements and evaluation of highly multiplexed, epiGBS-based reduced representation bisulfite sequencing. *Mole. Ecol. Resour.* 2022:13597. doi: 10.1111/1755-0998.13597

Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Mole. Ecol. Notes* 5, 184–186. doi: 10.1111/j.1471-8286.2004.00828.x

Gratani, L. (2014). Plant phenotypic plasticity in response to environmental factors. *Adv. Bot.* 2014:208747.

Hamrick, J. L., and Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosoph. Trans. R. Soc. Lond. Ser B, Biol. Sci.* 351, 1291–1298. doi: 10.1098/rstb.1996.0112

Hays, C. G., Hanley, T. C., Hughes, A. R., Truskey, S. B., Zerebecki, R. A., and Sotka, E. E. (2021). Local Adaptation in Marine Foundation Species at Microgeographic Scales. *Biol. Bull.* 241, 16–29. doi: 10.1086/714821

Herman, J. J., and Sultan, S. E. (2016). DNA methylation mediates genetic variation for adaptive transgenerational plasticity. *Proc. R. Soc. B* 283:1838. doi: 10.1098/rspb.2016.0988

Herrera, C. M., and Bazaga, P. (2011). Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. *Mole. Ecol.* 20, 1675–1688. doi: 10.1111/j.1365-294X.2011.05026.x

Herrera, C. M., Medrano, M., and Bazaga, P. (2017). Comparative epigenetic and genetic spatial structure of the perennial herb Helleborus foetidus: isolation by environment, isolation by distance, and functional trait divergence. *Am. J. Bot.* 104, 1195–1204. doi: 10.3732/ajb.1700162

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mole. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hodgins, K. A., Bock, D. G., Hahn, M. A., Heredia, S. M., Turner, K. G., and Rieseberg, L. H. (2015). Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa. *Mole. Ecol.* 24, 2226–2240. doi: 10.1111/mec.13026

Hollister, J. D., Smith, L. M., Guo, Y.-L., Ott, F., Weigel, D., and Gaut, B. S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2322–2327. doi: 10.1073/pnas.1018222108

Hughes, A. R., and Lotterhos, K. E. (2014). Genotypic diversity at multiple spatial scales in the foundation marsh species, Spartina alterniflora. *Mar. Ecol. Prog. Ser.* 497, 105–117. doi: 10.3354/meps10565

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinform.* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129

Jombart, T., and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070–3071. doi: 10.1093/bioinformatics/btr521

Jost, L. (2008). GST and its relatives do not measure differentiation. *Mol. Ecol.* 17, 4015–4026. doi: 10.1111/j.1365-294X.2008.03887.x

Jueterbock, A., Boström, C., Coyer, J. A., Olsen, J. L., Kopp, M., Dhanasiri, A. K. S., et al. (2020). The seagrass methylome is associated with variation in photosynthetic performance among clonal shoots. *Front. Plant Sci.* 11:571646. doi: 10.3389/fpls.2020.571646

Lewis, D. B., and Eby, L. A. (2002). Spatially heterogeneous refugia and predation risk in intertidal salt marshes. *Oikos* 96, 119–129. doi: 10.1034/j.1600-0706.2002.960113.x

Lewis, D. B., Jimenez, K. L., Abd-Elrahman, A., Andreu, M. G., Landry, S. M., Northrop, R. J., et al. (2021). Carbon and nitrogen pools and mobile fractions in surface soils across a mangrove saltmarsh ecotone. *Sci. Total Env.* 798:149328. doi: 10.1016/j.scitotenv.2021.149328

Lin, Q., and Mendelssohn, I. A. (2012). Impacts and Recovery of the Deepwater Horizon Oil Spill on Vegetation Structure and Function of Coastal Salt Marshes in the Northern Gulf of Mexico. *Environmental Science & Technology* 46, 3737–3743. doi: 10.1021/es203552p

Lin, Q., Mendelssohn, I. A., Graham, S. A., Hou, A., Fleeger, J. W., and Deis, D. R. (2016). Response of salt marshes to oiling from the Deepwater Horizon spill: Implications for plant growth, soil surface-erosion, and shoreline stability. *Sci. Total Env.* 557-558, 369–377. doi: 10.1016/j.scitotenv.2016.03.049

Lira-Medeiros, C. F., Parisod, C., Fernandes, R. A., Mata, C. S., Cardoso, M. A., and Ferreira, P. C. G. (2010). Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS One* 5:e10326. doi: 10.1371/journal.pone.0010326

Li, R., Hu, F., Li, B., Zhang, Y., Chen, M., Fan, T., et al. (2020). Whole genome bisulfite sequencing methylome analysis of mulberry (Morus alba) reveals epigenome modifications in response to drought stress. *Sci. Rep.* 10:8013. doi: 10.1038/s41598-020-64975-5

Medrano, M., Herrera, C. M., and Bazaga, P. (2014). Epigenetic variation predicts regional and local intraspecific functional diversity in a perennial herb. *Mole. Ecol.* 23, 4926–4938. doi: 10.1111/mec.12911

Miryeganeh, M. (2021). Plants' Epigenetic Mechanisms and Abiotic Stress. *Genes* 12:8. doi: 10.3390/genes12081106

Miryeganeh, M., and Saze, H. (2020). Epigenetic inheritance and plant evolution. *Populat. Ecol.* 62, 17–27. doi: 10.1002/1438-390x.12018

Mounger, J., Ainouche, M. L., Bossdorf, O., Cavé-Radet, A., Li, B., Parepa, M., et al. (2021a). Epigenetics and the success of invasive plants. *Philosoph. Trans. R. Soc. Lond. Ser.B Biol. Sci.* 376:20200117. doi: 10.1098/rstb.2020.0117

Mounger, J., Boquete, M. T., Schmid, M. W., Granado, R., Robertson, M. H., Voors, S. A., et al. (2021b). Inheritance of DNA methylation differences in the mangrove *Rhizophora mangle*. *Evol. Dev.* 23, 351–374. doi: 10.1101/2020.10.24.353482

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York, NY: Columbia University Press, doi: 10.7312/nei-92038

Neinavaie, F., Ibrahim Hashim, A., Kramer, A. M., Brown, J. S., and Richards, C. L. (2021). The genomic processes of biological invasions: from invasive species to cancer metastases and back again. *Front. Ecol. Evol.* 2021:681100.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mole. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nicotra, A. B., Atkin, O. K., Bonser, S. P., Davidson, A. M., Finnegan, E. J., Mathesius, U., et al. (2010). Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* 15, 684–692. doi: 10.1016/j.tplants.2010.09.008

Nicotra, A. B., Segal, D. L., Hoyle, G. L., Schrey, A. W., Verhoeven, K. J. F., and Richards, C. L. (2015). Adaptive plasticity and epigenetic variation in response to warming in an Alpine plant. *Ecol. Evol.* 5, 634–647. doi: 10.1002/ece3.1329

Niederhuth, C. E., and Schmitz, R. J. (2017). Putting DNA methylation in context: from genomes to gene expression in plants. *Biochim. Biophys. Acta* 1860, 149–156. doi: 10.1016/j.bbagrm.2016.08.009

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). *vegan: Community Ecology R package version 2.5-7*. Available online at: https://cran.r-project.org/web/packages/vegan/index.html (accessed date 2020-November-28).

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633

Park, Y., and Wu, H. (2016). Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* 32, 1446–1453. doi: 10.1093/bioinformatics/btw026

Paun, O., Bateman, R. M., Fay, M. F., Hedrén, M., Civeyrel, L., and Chase, M. W. (2010). Stable epigenetic effects impact adaptation in allopolyploid orchids (Dactylorhiza: Orchidaceae). *Mole. Biol. Evol.* 27, 2465–2473. doi: 10.1093/molbev/msq150

Paun, O., Verhoeven, K. J. F., and Richards, C. L. (2019). Opportunities and limitations of reduced representation bisulfite sequencing in plant ecological epigenomics. *New Phytol.* 221, 738–742. doi: 10.1111/nph.15388

Pennings, S. C., and Bertness. (2001). "Salt Marsh Communities," in *Marine Community Ecology*, eds S. D. Bertness and M. Hay (Sunderland,: Sinauer Associates), 289–316.

Pennings, S. C., and Callaway, R. M. (2000). The advantages of clonal integration under different ecological conditions: a community-wide test. *Ecology* 81, 709–716. doi: 10.1890/0012-96582000081[0709:taociu]2.0.co;2

Pennings, S. C., Grant, M.-B., and Bertness, M. D. (2005). Plant zonation in low-latitude salt marshes: disentangling the roles of flooding, salinity and competition. *J. Ecol.* 93, 159–167. doi: 10.1111/j.1365-2745.2004.00959.x

Puy, J., Carmona, C. P., Dvoøáková, H., Latzel, V., and de Bello, F. (2021a). Diversity of parental environments increases phenotypic variation in Arabidopsis populations more than genetic diversity but similarly affects productivity. *Ann. Bot.* 127, 425–436. doi: 10.1093/aob/mcaa100

Puy, J., de Bello, F., Dvoøáková, H., Medina, N. G., Latzel, V., and Carmona, C. P. (2021b). Competition-induced transgenerational plasticity influences competitive interactions and leaf decomposition of offspring. *New Phytol.* 229, 3497–3507. doi: 10.1111/nph.17037

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Richards, C. L., Alonso, C., Becker, C., Bossdorf, O., Bucher, E., Colomé-Tatché, M., et al. (2017). Ecological plant epigenetics: evidence from model and non-model species, and the way forward. *Ecol. Lett.* 20, 1576–1590. doi: 10.1111/ele.12858

Richards, C. L., Hamrick, J. L., and Donovan, L. A. (2004). Unexpectedly high clonal diversity of two salt marsh perennials across a severe environmental gradient. *Ecology* 2004:674. doi: 10.1111/j.1461-0248.2004.00674.x

Richards, C. L., Pennings, S. C., and Donovan, L. A. (2005). Habitat range and phenotypic variation in salt marsh plants. *Plant Ecol.* 2005:3. doi: 10.1007/s11258-004-0841-3

Richards, C. L., and Pigliucci, M. (2020). Epigenetic Inheritance. A Decade into the Extended Evolutionary Synthesis. *Paradigmi* 38, 463–494. doi: 10.30460/99624

Robertson, M., Alvarez, M., van Gurp, T., Wagemaker, C. A. M., Yu, F., Amador, D. M., et al. (2020). Combining epiGBS markers with long read transcriptome sequencing to assess differentiation associated with habitat in Reynoutria (aka Fallopia). *bioRxiv* 2020:317966. doi: 10.1101/2020.09.30.317966

Robertson, M., and Richards, C. (2015). Opportunities and challenges of next-generation sequencing applications in ecological epigenetics. *Mole. Ecol.* 2015:13277. doi: 10.1111/mec.13277

Robertson, M., Schrey, A., Shayter, A., Moss, C. J., and Richards, C. (2017). Genetic and epigenetic variation in Spartina alterniflora following the Deepwater Horizon oil spill. *Evol. Appl.* 10, 792–801. doi: 10.1111/eva.12482

Salmon, A., Ainouche, M. L., and Wendel, J. F. (2005). Genetic and epigenetic consequences of recent hybridization and polyploidy in Spartina (Poaceae). *Mole. Ecol.* 14, 1163–1175. doi: 10.1111/j.1365-294X.2005.02488.x

Sarma, R. R., Crossland, M. R., Eyck, H. J. F., DeVore, J. L., Edwards, R. J., Cocomazzo, M., et al. (2021). Intergenerational effects of manipulating DNA methylation in the early life of an iconic invader. *Philosoph. Transac. R. Soc. Lond. Ser. B Biol. Sci.* 2021:1098. yet assigned doi: 10.1098/not

Sarma, R. R., Edwards, R. J., Crino, O. L., Eyck, H. J. F., Waters, P. D., Crossland, M. R., et al. (2020). Do Epigenetic Changes Drive Corticosterone Responses to Alarm Cues in Larvae of an Invasive Amphibian? *Integr. Comp. Biol.* 60, 1481–1494. doi: 10.1093/icb/icaa082

Sasaki, E., Kawakatsu, T., Ecker, J. R., and Nordborg, M. (2019). Common alleles of *CMT2* and *NRPE1* are major determinants of CHH methylation variation in *Arabidopsis thaliana*. *PLoS Genet.* 15:e1008492. doi: 10.1371/journal.pgen.1008492

Schlichting, C. D. (1986). The evolution of phenotypic plasticity in plants. *Annu. Rev. Ecol. Syst.* 17, 667–693. doi: 10.1146/annurev.es.17.110186.003315

Schmid, M. W. (2017). RNA-Seq Data Analysis Protocol: combining In-House and Publicly Available Data. *Methods Mole.Biol.* 1669, 309–335. doi: 10.1007/978-1-4939-7286-9_24

Schmid, M. W., Heichinger, C., Coman Schmid, D., Guthörl, D., Gagliardini, V., Bruggmann, R., et al. (2018b). Contribution of epigenetic variation to adaptation in Arabidopsis. *Nat. Comm.* 9:4446. doi: 10.1038/s41467-018-06932-5

Schmid, M. W., Giraldo-Fonseca, A., Rövekamp, M., Smetanin, D., Bowman, J. L., and Grossniklaus, U. (2018a). Extensive epigenetic reprogramming during the life cycle of Marchantia polymorpha. *Gen. Biol.* 19:9. doi: 10.1186/s13059-017-1383-z

Schmidt, T. L., Jasper, M.-E., Weeks, A. R., and Hoffmann, A. A. (2021). Unbiased population heterozygosity estimates from genome-wide sequence data. *Methods Ecol. Evol.* 12, 1888–1898. doi: 10.1111/2041-210x.13659

Schrey, A. W., Alvarez, M., Foust, C. M., Kilvitis, H. J., Lee, J. D., Liebl, A. L., et al. (2013). Ecological Epigenetics: Beyond MS-AFLP. *Integr. Comp. Biol.* 53, 340–350. doi: 10.1093/icb/ict012

Schulz, B., Eckstein, R. L., and Durka, W. (2014). Epigenetic variation reflects dynamic habitat conditions in a rare floodplain herb. *Mole. Ecol.* 23, 3523–3537. doi: 10.1111/mec.12835

Shea, M. L., Warren, R. S., and Niering, W. A. (1975). Biochemical and transplantation studies of the growth form of Spartina alterniflora on Connecticut salt marshes. *Ecology* 56, 461–466. doi: 10.2307/1934977

Smit, A. F. A., Hubley, R., and Green, P. (2013–2015). *RepeatMasker Open-4.0 2013–2015*. Hyderabad: Indian School of Business.

Somers, G. F., and Grant, D. (1981). Influence of seed source upon phenology of flowering of Spartina alterniflora Loisel. And the likelihood of cross pollination. *Am. J. Bot.* 68, 6–9. doi: 10.1002/j.1537-2197.1981.tb06349.x

Stajic, D., and Jansen, L. (2021). Empirical evidence for epigenetic inheritance driving evolutionary adaptation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 376:20200121. doi: 10.1098/rstb.2020.0121

Stalter, R., and Batson, W. T. (1969). Transplantation of salt marsh vegetation, Georgetown, South Carolina. *Ecology* 50, 1087–1089. doi: 10.2307/1936901

Sunde, J., Yıldırım, Y., Tibblin, P., and Forsman, A. (2020). Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: analysis of Data for Pike (Esox lucius) and a Synthesis of Previous Studies. *Front. Genet.* 11:218. doi: 10.3389/fgene.2020.00218

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* 17, 57–86.

Travis, S. E., Proffitt, C. E., and Ritland, K. (2004). Population structure and inbreeding vary with successional stage in created Spartina alterniflora marshes. *Ecol. Applic.* 14, 1189–1202. doi: 10.1890/03-5135

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232–W235. doi: 10.1093/nar/gkw256

Valiela, I., Teal, J. M., and Deuser, W. G. (1978). The Nature of Growth Forms in the Salt Marsh Grass Spartina alterniflora. *Am. Natural.* 112, 461–470. doi: 10.1086/283290

van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., and Verhoeven, K. J. F. (2016). epiGBS: reference-free reduced representation bisulfite sequencing. *Nat. Methods* 13, 322–324. doi: 10.1038/nmeth.3763

van Moorsel, S. J., Schmid, M. W., Wagemaker, N. C. A. M., van Gurp, T., Schmid, B., and Vergeer, P. (2019). Evidence for rapid evolution in a grassland biodiversity experiment. *Mole. Ecol.* 28, 4097–4117. doi: 10.1111/mec.15191

Verhoeven, K. J. F., Jansen, J. J., Van Dijk, P. J., and Biere, A. (2010). Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol.* 185, 1108–1118. doi: 10.1111/j.1469-8137.2009.03121.x

Voors, S. E. (2018). *Ecosystem Function and Phenotypic Variation in Spartina alterniflora Salt Marshes*. Tampa, FL: University of South Florida.

Wang, I. J., and Bradburd, G. S. (2014). Isolation by environment. *Mole. Ecol.* 23, 5649–5662. doi: 10.1111/mec.12938

Wang, M.-Z., Li, H.-L., Li, J.-M., and Yu, F.-H. (2020). Correlations between genetic, epigenetic and phenotypic variation of an introduced clonal herb. *Heredity* 124, 146–155. doi: 10.1038/s41437-019-0261-8

Wang, Y.-J., Müller-Schärer, H., van Kleunen, M., Cai, A.-M., Zhang, P., Yan, R., et al. (2017). Invasive alien plants benefit more from clonal integration in heterogeneous environments than natives. *New Phytol.* 216, 1072–1078. doi: 10.1111/nph.14820

Wickham, H., and Seidel, D. (2020). *Scale Functions for Visualization [R package scales version 1.1.1]*. Comprehensive R Archive Network (CRAN). Available online at: https://cran.r-project.org/web/packages/scales/index.html (accessed date 2020-05-11).

Wiêski, K., and Pennings, S. C. (2014). Climate Drivers of Spartina alterniflora Saltmarsh Production in Georgia, USA. *Ecosystems* 17, 473–484. doi: 10.1007/s10021-013-9732-6

Xie, H. J., Li, H., Liu, D., Dai, W. M., He, J. Y., Lin, S., et al. (2015). ICE1 demethylation drives the range expansion of a plant invader through cold tolerance divergence. *Mole. Ecol.* 24, 835–850. doi: 10.1111/mec.13067

Yoder, A. D., and Tiley, G. P. (2021). The challenge and promise of estimating the de novo mutation rate from whole-genome comparisons among closely related individuals. *Mole. Ecol.* 2021:16007. doi: 10.1111/mec.16007

Zerebecki, R. A., Crutsinger, G. M., and Hughes, A. R. (2017). Spartina alterniflora genotypic identity affects plant and consumer responses in an experimental marsh community. *J. Ecol.* 105, 661–673. doi: 10.1111/1365-2745.12703

Zerebecki, R. A., Sotka, E. E., Hanley, T. C., Bell, K. L., Gehring, C., Nice, C. C., et al. (2021). Repeated Genetic and Adaptive Phenotypic Divergence across Tidal Elevation in a Foundation Plant Species. *Am. Natural.* 2021:716512. doi: 10.1086/716512

Conflict of Interest: MS was employed by MWSchmid GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 17 May 2022
doi: 10.3389/fgene.2022.867906

Rare Alleles and Signatures of Selection on the Immunodominant Domains of Pfs230 and Pfs48/45 in Malaria Parasites From Western Kenya

Kevin O. Ochwedo[1,2], Fredrick O. Ariri[3,2], Wilfred O. Otambo[3,2], Edwin O. Magomere[2,4], Isaiah Debrah[2,5], Shirley A. Onyango[2,6], Pauline W. Orondo[2], Harrysone E. Atieli[2], Sidney O. Ogolla[7], Antony C. A. Otieno[1], Wolfgang R. Mukabana[1,2], Andrew K. Githeko[2,7], Ming-Chieh Lee[8], Guiyun Yan[8], Daibin Zhong[8] and James W. Kazura[9]*

[1]Department of Biology, Faculty of Science and Technology, University of Nairobi, Nairobi, Kenya, [2]Sub-Saharan Africa International Centre for Excellence in Malaria Research, Homa Bay, Kenya, [3]Department of Zoology, School of Physical and Biological Sciences, Maseno University, Kisumu, Kenya, [4]Department of Biochemistry, Egerton University, Nakuru, Kenya, [5]West Africa Centre for Cell Biology of Infectious Pathogen, Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Accra, Ghana, [6]School of Zoological Sciences, Kenyatta University, Nairobi, Kenya, [7]Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya, [8]Program in Public Health, College of Health Sciences, University of California, Irvine, Irvine, CA, United States, [9]Centre for Global Health and Diseases, Case Western Reserve University, Cleveland, OH, United States

OPEN ACCESS

Edited by:
Sajad Ahmad Dar, Jazan University, Saudi Arabia

Reviewed by:
Muzamil Mahdi Abdel Hamid, University of Khartoum, Sudan
Mustafa Mohammad, Guru Teg Bahadur Hospital, University of Delhi, India

***Correspondence:**
Daibin Zhong
dzhong@hs.uci.edu

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *01 February 2022*
Accepted: *25 April 2022*
Published: *17 May 2022*

Citation:
Ochwedo KO, Ariri FO, Otambo WO, Magomere EO, Debrah I, Onyango SA, Orondo PW, Atieli HE, Ogolla SO, Otieno ACA, Mukabana WR, Githeko AK, Lee M-C, Yan G, Zhong D and Kazura JW (2022) Rare Alleles and Signatures of Selection on the Immunodominant Domains of Pfs230 and Pfs48/45 in Malaria Parasites From Western Kenya. Front. Genet. 13:867906. doi: 10.3389/fgene.2022.867906

Background: Malaria elimination and eradication efforts can be advanced by including transmission-blocking or reducing vaccines (TBVs) alongside existing interventions. Key transmission-blocking vaccine candidates, such as *Pfs230* domain one and *Pfs48/45* domain 3, should be genetically stable to avoid developing ineffective vaccines due to antigenic polymorphisms. We evaluated genetic polymorphism and temporal stability of *Pfs230* domain one and *Pfs48/45* domain three in *Plasmodium falciparum* parasites from western Kenya.

Methods: Dry blood spots on filter paper were collected from febrile malaria patients reporting to community health facilities in endemic areas of Homa Bay and Kisumu Counties and an epidemic-prone area of Kisii County in 2018 and 2019. *Plasmodium* speciation was performed using eluted DNA and real-time PCR. Amplification of the target domains of the two *Pfs* genes was performed on *P. falciparum* positive samples. We sequenced *Pfs230* domain one on 156 clinical isolates and *Pfs48/45* domain three on 118 clinical isolates to infer the levels of genetic variability, signatures of selection, genetic diversity indices and perform other evolutionary analyses.

Results: *Pfs230* domain one had low nucleotide diversity ($\pi = 0.15 \times 10^{-2}$) with slight variation per study site. Six polymorphic sites with nonsynonymous mutations and eight haplotypes were discovered. I539T was a novel variant, whereas G605S was nearing fixation. *Pfs48/45* domain three had a low π (0.063×10^{-2}), high conservation index, and three segregating sites, resulting in nonsynonymous mutation and four haplotypes. Some loci of *Pfs230* D1 were in positive or negative linkage disequilibrium, had negative or positive selection signatures, and

others (1813, 1955) and (1813, 1983) had a history of recombination. Mutated loci pairs in *Pfs48/45* domain three had negative linkage disequilibrium, and some had negative and positive Tajima's *D* values with no history of recombination events.

Conclusion: The two transmission blocking vaccine candidates have low nucleotide diversity, a small number of zone-specific variants, high nucleotide conservation index, and high frequency of rare alleles. With the near fixation a polymorphic site and the proximity of mutated codons to antibody binding epitopes, it will be necessary to continue monitoring sequence modifications of these domains when designing TBVs that include Pfs230 and Pfs48/45 antigens.

Keywords: Pfs230, Pfs48/45, transmission blocking vaccines, genetic diversity, evolutionary forces

INTRODUCTION

Genetic polymorphism of *Plasmodium falciparum* antigens has hampered efforts to develop an effective vaccine that is protective against pre-erythrocytic and asexual blood-stage parasites (Genton et al., 2002; Takala et al., 2007; Ogutu et al., 2009; Bergmann-Leitner et al., 2012; Neafsey et al., 2015; Ouattara et al., 2015). Recent efforts, however, have been made to develop vaccines that reduce and block *Plasmodium falciparum* transmission at the community level. Two of the existing transmission-blocking vaccine (TBV) candidates, *P. falciparum* surface protein 230 (Pfs230) (Sabeti et al., 2007; Lee et al., 2019, 2020; Singh et al., 2019, 2020; Tachibana et al., 2019; Huang et al., 2020; Healy et al., 2021) and *P. falciparum* surface protein 48/45 (Pfs48/45) (Singh et al., 2019, 2021; Lee et al., 2020) have been shown to elicit antibody responses in mice and people that block *P. falciparum* gametocyte fertilization in the mid-gut of the *Anopheles* vector.

Pfs230 is a cysteine-rich 230 kDa protein expressed by both male and female gametocytes (Rener et al., 1983; MacDonald et al., 2016). The antigen is thought to play a role in gamete fusion in the mosquito blood meal after forming a complex with another cysteine-rich protein, Pfs48/45 (Eksi et al., 2006). In comparison to antibodies elicited by immunization with other Pfs230 domains, Domain 1 (D1) has been shown to elicit transmission-blocking monoclonal antibodies with strong inhibitory activity against oocyst development in standard membrane feeding assays (Lee et al., 2019; Singh et al., 2019, 2020; Tachibana et al., 2019; Huang et al., 2020; Healy et al., 2021). Like Pfs230 D1, fusion with its counterpart Pfs48/45 D3 has good potential as a component of a TBV. The latter fused doublet antigen consists of three domains linked by disulphide bonds and contains 16 cysteine residues (Kocken et al., 1993; Lennartz et al., 2018). Unlike Pfs230, Pfs48/45 is anchored on the gamete surface membrane by glycophophatidylinositol (Kocken et al., 1993; Dijk et al., 2001; Gilson et al., 2006; Lennartz et al., 2018) and is essential for male gamete fertility. Domain 3 has been shown to elicit antibodies in the host (Graves et al., 1988; Roeffen et al., 1994; Dijk et al., 2001; Bousema et al., 2010; Jones et al., 2015; Acquah et al., 2017; Singh et al., 2019; Baptista et al., 2022). Pfs48/45 D3 is located at the C-terminus of the protein and contains binding sites for non-inhibitory and inhibitory human and mouse mAbs that reduce *P. falciparum* infection in mosquitoes (Vermeulen et al., 1986; Graves et al., 1988; Outchkourov et al., 2007; Chowdhury et al., 2009; Singh et al., 2017, 2019; Kundu et al., 2018; Lennartz et al., 2018; Lee et al., 2020).

Antigenic polymorphism of Pfs230 D1 and Pf48/45 D3 should be assessed in malaria endemic areas on a regular basis to support the successful development of these TBV candidates due to the fact that, if the targeted regions are genetically unstable, polymorphisms may cause critical codon changes within immunogenic epitopes, thereby reducing TBV efficacy. Several dimorphic sites on Pfs230 D1 and Pfs48/45 D3 have previously been identified (Kocken et al., 1995; Drakeley et al., 1996; Escalante et al., 1998; Conway et al., 2001; Jones et al., 2015; MacDonald et al., 2016; Kundu et al., 2018; Singh et al., 2020; Coelho et al., 2021); however, there is limited knowledge of the extent of genetic diversity, signatures of selection, and other evolutionary forces that may be shaping alleles in *P. falciparum* from different malaria transmission zones. We therefore performed an in-depth genetic analysis of *Pfs230* D1 and *Pfs48/45* D3 in parasites isolated from patients with uncomplicated falciparum malaria from three different areas in western Kenya.

MATERIALS AND METHODS

Study Site and Sampling

Dry blood spots (DBS) were collected on filter paper from febrile malaria patients at health clinics in Homa Bay County (Kochia), Kisumu County (Chulaimbo), and Kisii County (Eramba) in 2018 and 2019 (**Figure 1**). The study site in Homa Bay is characterized by perennial transmission. Vector control consists of universal distribution of long-lasting insecticidal bed nets with annual indoor residual spraying of insecticides. The study site in Kisumu County also has perennial transmission. Vector control consists of long-lasting insecticidal bed nets (LLINs) alone. The site in Kisii County is malaria epidemic-prone with low transmission and residents use LLINs (Kapesa et al., 2018). In brief, four

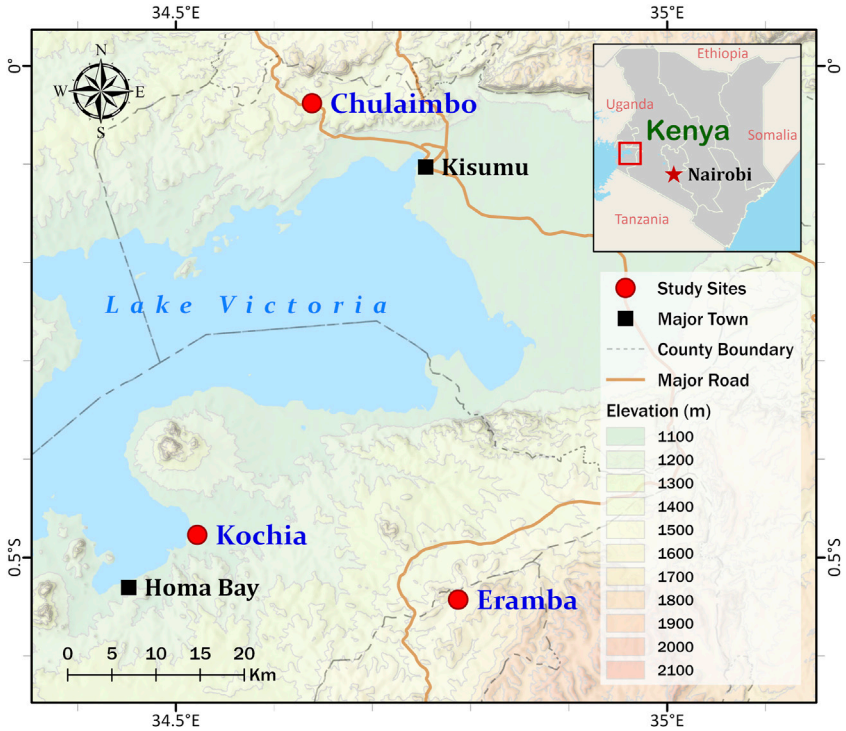

FIGURE 1 | Map showing Homa Bay (Kochia), Kisumu (Chulaimbo) and Kisii (Eramba) study sites in western Kenya. The sampling points are represented by the red circles on the map. This figure was prepared with ESRI ArcGIS Pro 2.8 with field survey results and publicly available datasets. The map contains information from OpenStreetMap and OpenStreetMap 115 Foundation, which is made available under the Open Database License.

drops of approximately 25 µL of blood from each patient were spotted on Whatman™ Blood Stain Cards (GE Healthcare WB100014) as previously described (Coombs and Fiscus, 2009). Each card was stored individually in silica gel-containing plastic bags before being transported to the joint International Centre of Excellence for Malaria Research (ICEMR) and Tom Mboya University College Laboratory in Homa Bay town for storage at -20°C. 150 of the 372 DBS collected came from Homa Bay; 120 and 102 came from Kisumu and Kisii, respectively.

Amplification and Sequencing of *Pfs230* Domain one and *Pfs48/45* Domain three

Genomic DNA was extracted from filter paper using the modified Chelex resin (Chelex -100) method and stored at -20°C. As a positive control, DNA from the cultured laboratory strain NF54 was extracted and stored. *Plasmodium* species-specific real-time PCR targeting 18S ribosomal RNA gene was used to confirm *P. falciparum* positive DNA samples before amplification of specific target fragments of each gene (Ochwedo et al., 2021; Onyango et al., 2021). Primer sets were designed using Primer3 version 0.4.0 for *Pfs230* D1 and *Pfs48/45* D3 and in silico validation of each set was performed using the Sequence Manipulation Suite (Stothard, 2000). Among the 372 samples, 332 (89.3%) tested positive for *P. falciparum* DNA (n = 150, 120 and 62 from Homa Bay, Kisumu and Kisii, respectively) and were used to amplify *Pfs230* D1 and *Pfs48/45* D3 in a T100™ Thermal Cycler (Bio-Rad, Hercules, CA, United States). Briefly, 3 µL of sample DNA was added to a mixture of 11.5 µL of DreamTaq Green PCR Master Mix (2X), 0.5 µL of *Pfs230* D1 forward (5′-TGGTGAAGCTGTCGAAGATG-3′) and reverse primers (5′-GTGTACCACAGGGGGAAGAG-3′) targeting 514 base pairs and 7.5 µL of double-distilled water. The thermal profile was set as follows 95 °C for 3 min, 34 cycles (94 °C for 30 s, 58 °C for 30 s and 72 °C for 45 s) and final extension at 72 °C for 6 min. For *Pfs48/45* D3, similar reaction volume was prepared using forward (5′-TTTTCAAGAAGGAAAAGAAAAAGC-3′) and reverse primers (5′-GCCAAAAATCCATAATATGCTGA-3′) targeting 600bp. The PCR conditions were set as follows 95°C for 3 min, 34 cycles (94°C for 30 s, 55 °C for 30 s and extension at 72°C for 45 s) final extension at 72 °C for 6 min. All the amplicons were assessed by gel electrophoresis in 1.5% w/v agarose gel before sequencing. For *Pfs230* D1, 82, 39, and 35 samples from Homa Bay, Kisumu, and Kisii, respectively, were amplified. For *Pfs48/45* D3, 36, 44, and 38 samples from Homa Bay, Kisumu, and Kisii, respectively, were amplified. All the PCR amplicons, together with positive controls, were purified using Exonuclease I and Shrimp Alkaline Phosphatase (ExoSAP-IT) and bi-directionally sequenced using 3730 BigDye® Terminator v3.1 Sequencing Standard kit on ABI PRISM® 3700 DNA Analyzer (Applied Biosystems, Foster City, CA, United States).

TABLE 1 | Polymorphic sites on *Pfs230* domain one and *Pfs48/45* domain two and three from Homa Bay, Kisumu and Kisii region in western Kenya n: number of sequences harbouring mutations; *: reference (3D7and NF54) allele only; D: Domain; Nsyn: Non-synonymous mutation; Syn: Synonymous mutation; A: Adenine; C: Cytosine; T: Thymine; G: Guanine.

***Pfs230* (N = 156)**								
Segregating Sites	**Domains**	**Allelic Frequency**			**Substituted Bases**	**Type of substitution**	**Codon Change**	**Type of Mutation**
		Homa Bay n (%)	**Kisumu n (%)**	**Kisii n (%)**				
1616	D1	-	2 (5.1)	-	*T/C	Transition	I539T	Nsyn
1813	D1	81 (98.8)	37 (94.9)	35 (100)	*G/A	Transition	G605S	Nsyn
1955	D1	2 (2.4)	1 (2.6)	2 (5.7)	*C/G	Transversion	T652R	Nsyn
1964	D1	-	-	1 (2.9)	*A/T	Transversion	E655V	Nsyn
1967	D1	-	1 (2.6)	-	*C/A	Transversion	T656N	Nsyn
1983	D1	35 (42.7)	19 (48.7)	16 (45.7)	*A/C	Transversion	K661N	Nsyn
Pfs48/45 (N = 118)								
753	D2	-	-	1 (2.6)	*T/C	Transition	Y251Y	Syn
757	D2	-	4 (9.1)	3 (7.9)	*A/G	Transition	K253E	Nsyn
762	D2	1 (2.8)	3 (6.8)	3 (7.9)	*C/G	Transversion	N254K	Nsyn
911	D3	1 (2.8)	-	-	*T/A	Transition	V304D	Nsyn
940	D3	3 (8.3)	3 (6.8)	9 (23.7)	*T/A	Transition	L314I	Nsyn
979	D3	-	-	1 (2.6)	*T/G	Transversion	C327G	Nsyn

Data Analysis

All sequences were assembled using Geneious version 11.1.5 software, and multiple sequence alignment was performed using ClustalW. Polymorphic locus and codons were inferred after comparing each sequence to the respective sequence of positive control (NF54) as well as 3D7 (PF3D7_0209000 for *Pfs230* and PF3D7_1346700 for *Pfs48/45*). DnaSP Version 6.12.03 (Rozas et al., 2017) and Arlequin version 3.5.2 (Excoffier and Lischer, 2010) were used to compute genetic diversity indices such as nucleotide diversity (π), haplotype diversity (Hd), number of haplotypes (h), number of segregating sites (S) and mean number of pairwise difference (k). Population Analysis with Reticulate Trees (Popart) version 1.7 software (Clement et al., 2000) was used to infer haplotype networks. Neutrality tests; Tajima's *D*, Fu and Li's D, Fu and Li's F and Fu's Fs statistics and test for the presence of Recombination events (Rm) and linkage disequilibrium (LD) were computed in DnaSP Version 6.12.03 and Arlequin version 3.5.2. Generated Tajima's *D* values were plotted using GraphPad version 8.3.0. Both antigen structural delineation was done using Protein Homology/analogY Recognition Engine (PHYRE2) version 2.0 and generated models visualized and edited in UCSF Chimera version 1.15 (Pettersen et al., 2004).

RESULTS

Analysis of Mutations Detected in *Pfs230* D1 and *Pfs48/45* D3

Six loci (1,616, 1813, 1955, 1964, 1967, and 1983) in *Pfs230* D1 were found to be polymorphic, resulting in nonsynonymous mutations (**Table 1**). The mutations were skewed toward transversion, with a transversion to transition ratio (Tv: Ts) greater than 0.5. Two polymorphic sites were singletons (1964 and 1967), whereas four dimorphic sites (1,616, 1813, 1955, and 1983) were parsimony informative. These polymorphisms resulted in I539T, G605S, T652R, E655V, T656N, and K661N codon changes. Nonsynonymous alterations T652R and K661N were on separate beta (β) pleated sheets connected by a loop containing mutated codons E655V and T656N (**Supplementary Figure S1**). G605S was also on the loop connecting two different β pleated sheets. In general, western Kenya parasites had a high allelic frequency of G605S (98.08%), followed by progressively lower frequencies of K661N, T652R, and I539T. E655R and T656N were each observed at a frequency of <1%. The prevalence of various alleles was almost similar across the various study sites. For example, as shown in **Table 1**, G605S was the most common codon change in the three study sites. Only two *P. falciparum* isolates from Kisumu and Homa Bay County lacked this mutation.

In contrast to the six nonsynonomous mutated sites observed in *Pfs230* D1, *Pfs48/45* D3 had three segregating sites (**Table 1**). Singleton sites were found at loci 911 and 979 in parasites isolated from patients residing in Homa Bay County and Kisii County, respectively. A low frequency polymorphism at locus 940 was observed across parasite populations in all three counties, and was parsimony-informative. These transition bias mutations at loci 911, 940, and 979 resulted in nonsynonymous mutations V304D, L314I, and C327G, respectively. The variants were in the Pfs48/45 D3 antigen loop connecting different β pleated sheets (**Supplementary Figure S1**). Codon change C327G in D3 was found in only one sequence in parasites isolated from a patient in Kisii County (**Table 1**). The ability of the designed primer set to cover *Pfs48/45* D3 also allowed for the discovery of a singleton site 753 (Y251Y) and parsimony-informative sites 757 and 762 (K253E and N254K) (**Table 1**; **Supplementary Figure S1**). Except for G605S, which was near dimorphic codons on

TABLE 2 | Summary of genetic diversity indices for *Pfs230* domain one and *Pfs48/45* domain three from parasites in western Kenya N: Sample size; C: Conservation index; S: Segregating sites; π: nucleotide diversity; Vars: Variants; Hd: Haplotype diversity.

Pfs230 **D1 Region**	**N**	**C (%)**	**S**	**π (×10^{-2})**	**h**	**Hd**	**Tajima's *D***	**Fu's F_S**	**FLD***	**FLF***
Homa Bay	82	99.40	3	0.12	4	0.52	−0.11	−0.21	−0.54	−0.45
Kisumu	39	98.90	5	0.18	6	0.63	−0.80	−1.94	−0.73	−0.82
Kisii	35	99.40	3	0.15	4	0.60	−0.15	−0.39	−0.31	−0.29
W. Kenya	156	98.70	6	0.14	8	0.56	−0.82	−3.33	−0.94	−1.03
Pfs48/45 D3										
	36	99.50	2	0.05	3	0.21	−1.09	−1.42	−0.80	−0.95
Homa Bay										
Kisumu	44	99.80	1	0.03	2	0.13	−0.60	−0.30	0.55	0.244
Kisii	38	99.50	2	0.10	3	0.41	−0.21	−0.12	−0.81	−0.74
W. Kenya	118	99.30	3	0.06	4	0.25	−0.94	−1.87	−2.06	−2.00

Pfs48/45 domain 2, the superimposed structure revealed dimorphic codons of Pfs48/45 D3 antigen close to those of Pfs230 D1 (**Supplementary Figure S2**).

Genetic Diversity of *Pfs230* and *Pfs48/45* Genes in Western Kenya

Pfs230 D1 from the three sites had π of 0.15×10^{-2}, k; 0.68 and haplotype diversity (Hd) of 0.57 (**Table 2**). The domain had a nucleotide conservation index of 98.7% with a total of eight haplotypes circulating in western Kenya (**Figure 2**). Kisumu had the highest π (0.18×10^{-2}) followed by Kisii (0.15×10^{-2}) and Homa Bay (0.12×10^{-2}). The site also had the most haplotypes 6) and the highest Hd (0.63) when compared to Kisii and Homa Bay, which had four haplotypes each and Hd of 0.60 and 0.52, respectively (**Figure 2**). Haplotype 2 (Hap_2) with the mutated codon G605S was the most common in western Kenya and at each study site. This was followed by Hap_4 (mutated codons G605S and K661N), Hap_7 (G605S, T652R, and K661N), and Hap_8 (T652R and K661N) (**Supplementary Table S1**). The remaining haplotypes (Hap_1, Hap_3, Hap_6, and Hap_5) were observed at a lower frequency. Only one sequence (from Chulaimbo) in western Kenya lacked a mutated site and had 100% sequence identity to the laboratory strain PF3D7_0209000 or NF54 sequence used as a positive control (**Figure 2**).

The *Pfs48/45* D3 from western Kenya had low π (0.06×10^{-2}) and Hd (0.25) (**Table 2**). The domain had a conservation index of 99.3%, with four haplotypes circulating in the study area (**Figure 2**). Kisii parasites had the highest π (0.10×10^{-2}) followed by Homa Bay (0.05×10^{-2}) and Kisumu (0.03×10^{-2}) (**Table 2**). Kisii and Homa Bay study sites each had three haplotypes in circulation, while Kisumu had four (**Figure 2**). The majority of haplotypes lacked a mutation (Hp_1) or had 100% sequence identity to the laboratory strain PF3D7_1346700 or the NF54 sequence used as a positive control (85.6%). This was followed by Hp_3 (mutated codon L314I), which had an overall frequency of 12.7%,

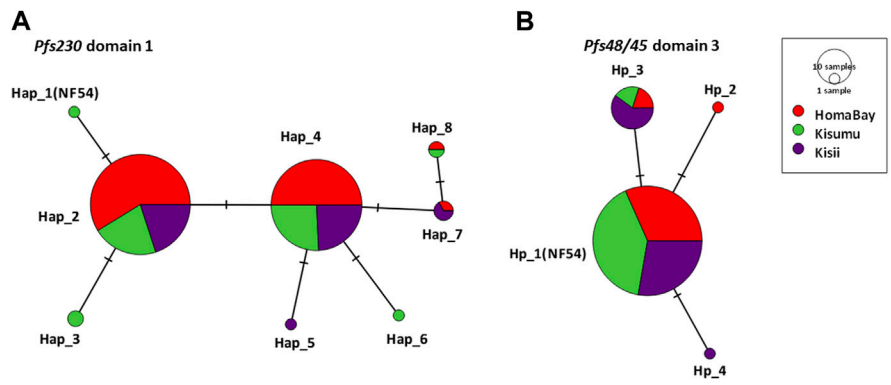

FIGURE 2 | TCS-network analysis of the relationship of haplotypes based on the of *Pfs230* domain one and *Pfs48/45* domain three from parasites in malaria-endemic and epidemic-prone regions of western Kenya. **(A)**. represents haplotypes based on observed sequence variation in *Pfs230* D1 whereas **(B)** represents haplotypes in *Pfs48/45* D3. The network shows the distribution of haplotype within malaria-endemic Homa Bay (red) and Kisumu (Chulaimbo) (green) as well as epidemic Kisii highland (purple). The hatch marks represent the number of mutations resulting in a specific haplotype whereas the size of the circle equates to the frequency of the observed haplotypes. Haplotypes labelled Hap_1(NF54) for *Pfs230* D1 and Hp-1 (NF54) for *Pfs48/45* D3 lacked mutated locus and have 100% sequence identity to laboratory strain NF54 or 3D7. Nucleotide sequences of all haplotypes were submitted to GenBank (accession: MW624857- MW625101).

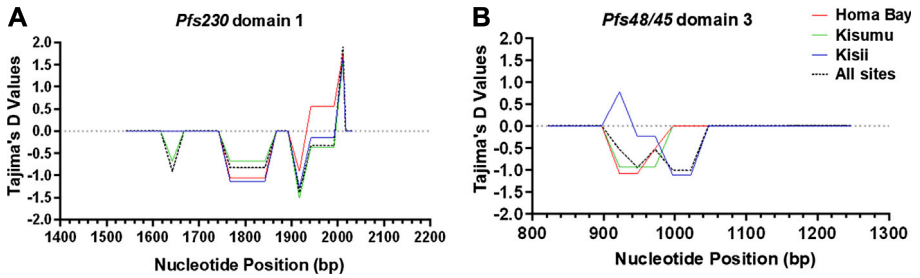

FIGURE 3 | Sliding window plot of Tajima's *D* values for *Pfs230* domain one and *Pfs48/45* domain three in western Kenya. The *X*-axis displays the nucleotide position (Window midpoint) whereas the Tajima's *D* values are represented on the *Y*-axis. **(A)** is a representation of *Pfs230* domain one whereas **(B)** represents of *Pfs48/45* domain 3. The blue curve represents computed Tajima's *D* plot for sequences of *P. falciparum* circulating in Kisii, the red colour is for Homa Bay, the green colour is for Kisumu (Chulaimbo) whereas the black dotted colour represents the population from the three study sites. The middle horizontal dotted line (intersecting the *Y*-axis at 0.0) represents a standard neutral model where the Tajima's *D* value is equal to zero. Positive deviation from the grey dotted line signifies balancing selection whereas negative deviation represents purifying selection.

while the rest (Hp_2 and Hp_4) had a frequency <1% (**Supplementary Table S2**).

Signatures of Selection, Linkage Disequilibrium and Recombination Events

Pfs230 D1 from *P. falciparum* isolates deviated from a standard neutral model. Tajima's *D* (-0.8), FLD* (-0.9) and FLF* (-1.0) tests were all negative and non-significant ($p > 0.05$) (**Table 2**). However, Fu's F_S result (-3.3), was significant ($p = 0.023$). Tajima's *D* test results were also non-significant ($p > 0.05$) in each site (**Table 2**). Despite the overall negative Tajima's D results, there was a slight variation among individual mutated loci on *Pfs230* D1. Locus 1983 (codon change K661N) had a significant ($p < 0.05$) positive (1.9) Tajima's *D* result, whereas the rest had negative results (**Figure 3**). Apart from deviating from a standard neutral model, some loci pairs (1813, 1955) had positive LD (D′) results with highly significant ($p < 0.001$) χ^2 values of 19.7 and 12.7 among Homa Bay and Kisumu sequences, respectively (**Supplementary Table S1**). Among Kisii sequences, loci pairs (1813 and 1983) had positive D′ results, despite the fact that the χ^2 values was non-significant and the r^2 value was low (0.03). Other loci with positive D′ results, low r^2 values and non-significant ($p > 0.05$) χ^2 values in Homa Bay and Kisumu parasites include loci pairs (1813,1983) (1955,1983) and (1967,1983). Some loci pairs on Kisumu and Kisii sequences had negative D′ results (**Supplementary Table S1**). On *Pfs230* D1 from parasites in Homa Bay and Kisumu, recombination events were detected across loci pairs (1813, 1955) and (1813, 1983) (**Supplementary Table S1**).

All of the *Pfs48/45* D3 sequences had non-significant ($p > 0.05$) negative Tajima's *D* -1.9, FLD*: 2.1 and FLF*: 2.0 results. The Fu's F_S (-1.9) result was, however, significant ($p = 0.096$). Locus 940 (L314I) among Kisii sequences had a significant ($p > 0.05$) positive Tajima's *D* (0.8) results (**Figure 3**). There was no evidence of Rm or positive D′ results at any of the dimorphic loci within *Pfs48/45* D3. However, some loci pairs from Homa Bay and Kisii had negative D′ results (**Supplementary Table S2**).

DISCUSSION

The Pfs230 D1 and Pfs48/45 D3 antigens are important candidate antigens in the development of an effective TBV. Despite having a high frequency of rare alleles in western Kenya, both targets had low nucleotide diversity. Two variants, each on Pfs230 D1 and Pfs48/45 D3, were novel and private to western Kenya. The study validated five previously described polymorphic sites on *Pfs230* D1 (Singh et al., 2020). In this study, G605S, one of the five mutated codons, was fixed in some study areas but not in others. *Pfs230* D1 had the most mutations, while the Pfs48/45 D3 was the most conserved. Mutated loci from both domains were either under purifying or balancing selections. Other genetic forces revealed to have shaped alleles on the two genes included inbreeding and genetic drift with recombination being discovered only *Pfs230* D1.

Pfs230 D1 from western Kenya had low nucleotide diversity, with significant Fu's F_S results indicating a high frequency of rare alleles. In addition to the previously reported 15 polymorphisms on *Pfs230* D1 from parasites in Asian (Bangladesh, Cambodia, Laos, Myanmar, Thailand, and Vietnam) and African (Democratic Republic of the Congo, Ghana, Guinea, Malawi, Mali, Nigeria, Senegal, and Gambia) countries (MacDonald et al., 2016; Singh et al., 2020), this study discovered one additional mutation (I539T). This novel variant was identified only at the Kisumu study site along with five other polymorphisms (G605S, T652R, E655V, T656N, and K661N). These findings validate five previously described polymorphisms reported by (Singh et al., 2020). We speculate that the relatively higher nucleotide diversity index and number of haplotypes in Kisumu compared to other sites in western Kenya is related to the region's slightly higher malaria transmission and absence of IRS activities (Oduma et al., 2021; Ochwedo et al., 2022).

Missense mutation G605S was found in parasites at a slightly higher allelic frequency (AF = 0.98) than in other geographical regions, as described by (Singh et al., 2020) (AF = 0.94) (MacDonald et al., 2016), (AF = 0.11), and (Coelho et al., 2021) (AF = 0.91). With only two clinical isolates in western Kenya lacking this mutation, G605S is almost completely fixed.

This indicates the presence of selection pressure from either host antibodies, vector immune response or genetic drift (decreased variation and increasing homozygosity), may be stronger on *P. falciparum* populations from Kisii (low parasite population size) or Homa Bay (endemic site with declining parasite population size) compared to Kisumu (Chulaimbo) (Hancock and Di Rienzo, 2008; Honnay, 2013; Oduma et al., 2021). In contrast, the second most common polymorphism, K661N, was found in Kisumu at a higher frequency than in Kisii and Homa Bay. This reversal in the observed G605S and K661N frequencies could be attributed to factors such as recombination events (Rm), which are known to interfere with linked loci and could be effective on linked dimorphic loci pair 1813 and 1983 (responsible for G605S and K661N mutations respectively), thus increasing diversity in Kisumu (Chulaimbo) (Mejia, 2012) as opposed to Kisii and Homa Bay parasites, which also lack Rm between the two sites. Since immunogenic epitope binding light chain of transmission-blocking 4F12 monoclonal antibodies (TB 4F12 mAb) is close to the dimorphic codon G605, selection pressure from host antibodies on the epitope may be affecting the surrounding codons (MacDonald et al., 2016; Singh et al., 2020). This codon is located within a disulphide loop (from 593 to 611) that is thought to be stabilizing the epitope binding of TB 4F12 mAb (Singh et al., 2020). With near-complete fixation, the mutation may be beneficial to parasites but have a negative effect on the epitope binding affinity of TB 4F12 mAb. This needs to be looked into further by immunoassays of haplotypes with this polymorphism. Polymorphism I539T was found near codons 542–592 that contain 3G2 and 5G3 mAb binding epitopes which were previously shown to have no detectable oocyst reduction activity (Singh et al., 2020). Other polymorphisms, T652R, K661N on different *ß* pleated sheets, and E655V, T656N on disulphide loops linking the two loops, were distally located from the epitope that binds TB 4F12 mAb (Singh et al., 2020). When the two fusion proteins were superimposed, these four polymorphic codons were closer to mutated codon V304D, L314I, and C327G on Pfs48/45 D3, supporting the hypothesis that antibodies could be sterically interfering with protein-protein interaction (Singh et al., 2020). *Plasmodium falciparum* may induce these mutations in response to antibody-induced pressure in order to circumvent the blockade of fusion between Pfs230 D1 and Pfs48/45 D3, resulting in an uninterrupted gametocyte fertilization process.

The novel missense polymorphism C327G on Pfs48/45 D3 has the potential to be very important because it can interfere with one of the six cysteine residue pairings (pairing between codon C298 and C327) on the 85RF45.1 mAb epitope (Kundu et al., 2018). Other polymorphisms (Y251Y, K253E, N254K in Pfs48/45 D2 and V304D, L314I in Pfs48/45 D3) have been observed in *P. falciparum* populations in other malaria endemic regions (Conway et al., 2001; Jones et al., 2015; Kundu et al., 2018). However, none of these polymorphisms had been previously reported by a study conducted in the Asembo Bay area of western Kenya (Escalante et al., 1998). Though not the focus of this study, polymorphisms on codon 254 is thought to influence the type of host antibody that binds at the epitope bearing this mutation on Pfs48/45 antigen (Kocken et al., 1995). The three polymorphic codons Y251Y, K253E, and N254K on Pfs48/45 D2, are close to the disulphide loop that stabilizes the epitope binding TB 4F12 mAb on Pfs230 D1, thus suggesting steric interference from the antibodies. *Pfs48/45* domain three is highly conserved, with low nucleotide and haplotype diversity when compared to *Pfs230* D1. The key polymorphism based on this domain was L314I, which has a higher allelic frequency in Kisii highlands than in Homa Bay and Kisumu. Despite the presence of a high frequency of rare alleles, the majority of parasites lacked polymorphic loci on *Pfs48/45* D3.

Inbreeding, recombination, and natural selection were identified as major drivers of the observed mutations in *Pfs230* D1 and *Pfs48/45* D3. The presence of linkage disequilibrium confirmed the history of selection pressure and inbreeding across various loci in *Pfs230* D1 and *Pfs48/45* D3 (Larrañaga et al., 2013). Some polymorphisms were considered intermediary because they had negative linkage disequilibrium (D′) values (Silvela et al., 1999). The negative D′ values also confirmed a history of random drift, which is decreasing the number of variants while increasing homozygosity that may play a role in the parasite's loss of favourable mutations if it persists (Barton, 2010).

The presence of natural selection was confirmed by the Tajima's D values. Overall negative Tajima's *D* results revealed that purifying selection was affecting the majority of loci within *Pfs230* D1 and *Pfs48/45* D3, reducing genetic diversity (Cvijović et al., 2018). The aforementioned selection was, however, weak because the computed negative Tajima's *D* values in both antigens were not significant. Individual Tajima's *D* results for each codon revealed all other dimorphic codons to be under purifying selection, with the exception of K661N on Pfs230 D1 from all study sites and V304D on Pfs48/45 D3 from Kisii, which are under strong and weak balancing selection, respectively. The two mutated loci under balancing selection may play an important role within the Pfs230 D1 and Pfs48/45 D3 fusion proteins, which may explain why they are maintained in the *P. falciparum* population from western Kenya (Escalante et al., 1998). These findings support previous postulation (Jones et al., 2015) that selection pressure is acting on immunogenic domains of Pfs48/45.

The presence of weak purifying selection acting on dimorphic sites may impact not only host mAb binding and functional activity but also be affected by selective pressure in the mosquito vector (Lombardo and Christophides, 2016). This pressure could be exerted on individual antigens before or after complex formation. Findings in the present study support future investigations that examine functional antibody responses such as the ability of PfS230 and Pfs48/45 antibodies that activate human plasma complement and reduce mosquito infectivity in membrane feeding assays.

CONCLUSION

The *Pfs230* D1 and *Pfs48/45* D3 in *P. falciparum* from western Kenya have low nucleotide diversity and a high conservation index with high frequency of rare alleles. Among the observed polymorphisms in *Pfs230* D1, G605S is nearly fixed in the population. Natural selection, inbreeding, and, to some extent, recombination are important driving forces in shaping these alleles in the two antigens. With the discovery of novel polymorphic sites, the two domains of the *Pfs230* and *Pfs48/*

45 from different malaria-prone regions, including areas where clinical trials have been conducted, should be monitored indefinitely. This will help track the genetic stability of the two TBV candidates.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Maseno University Ethics Review Committee (MUERC protocol No. 00456) and the University of California, Irvine Institutional Review Board (HS#2017–3512), as well as the Ministry of Health. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

KO conceptualised and designed the study, collected data, carried out the experiments, curated and analyzed the data, and drafted, edited, and revised the final manuscript. FA carried out the experiments and revised the final manuscript. WO collected data and revised the final manuscript. EM carried out the experiment and revised the final manuscript. ID revised the final manuscript. SO collected data. PO collected data, HA administration, SO administration and revised the final manuscript, AO revised the final manuscript. WM revised the final manuscript. AG revised the final manuscript. M-CL drew the map. GY developed the study's concept and revised the final manuscript. DZ developed the concept, carried out the experiment, curated the data, and revised the final manuscript. JK developed the study's concept and revised the final manuscript.

FUNDING

This research is supported by grants from the National Institutes of Health (U19 AI129326 and D43 TW001505).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.867906/full#supplementary-material

REFERENCES

Acquah, F. K., Obboh, E. K., Asare, K., Boampong, J. N., Nuvor, S. V., Singh, S. K., et al. (20172017). Antibody Responses to Two New *Lactococcus lactis*-Produced Recombinant Pfs48/45 and Pfs230 Proteins Increase with Age in Malaria Patients Living in the Central Region of Ghana. *Malar. J.* 16 (1 16), 1–11. doi:10.1186/S12936-017-1955-0

Baptista, B. O., de Souza, A. B. L., Riccio, E. K. P., Bianco-Junior, C., Totino, P. R. R., Martins da Silva, J. H., et al. (2022). Naturally Acquired Antibody Response to a *Plasmodium falciparum* Chimeric Vaccine Candidate GMZ2.6c and its Components (MSP-3, GLURP, and Pfs48/45) in Individuals Living in Brazilian Malaria-Endemic Areas. *Malar. J.* 21, 6. doi:10.1186/s12936-021-04020-6

Barton, N. H. (2010). Mutation and the Evolution of Recombination. *Phil. Trans. R. Soc. B* 365, 1281–1294. doi:10.1098/RSTB.2009.0320

Bergmann-Leitner, E. S., Duncan, E. H., Mease, R. M., and Angov, E. (2012). Impact of Pre-existing MSP142-Allele Specific Immunity on Potency of an Erythrocytic *Plasmodium falciparum* Vaccine. *Malar. J.* 11, 315. doi:10.1186/1475-2875-11-315

Bousema, T., Roeffen, W., Meijerink, H., Mwerinde, H., Mwakalinga, S., van Gemert, G.-J., et al. (2010). The Dynamics of Naturally Acquired Immune Responses to *Plasmodium falciparum* Sexual Stage Antigens Pfs230 & Pfs48/45 in a Low Endemic Area in Tanzania. *PLoS ONE* 5, e14114. doi:10.1371/journal.pone.0014114

Chowdhury, D. R., Angov, E., Kariuki, T., and Kumar, N. (2009). A Potent Malaria Transmission Blocking Vaccine Based on Codon Harmonized Full Length Pfs48/45 Expressed in *Escherichia coli*. *PLOS ONE* 4, e6352. doi:10.1371/JOURNAL.PONE.0006352

Clement, M., Posada, D., and Crandall, K. A. (2000). TCS: a Computer Program to Estimate Gene Genealogies. *Mol. Ecol.* 9, 1657–1659. doi:10.1046/J.1365-294X.2000.01020.X

Coelho, C. H., Tang, W. K., Burkhardt, M., Galson, J. D., Muratova, O., Salinas, N. D., et al. (20212021). A Human Monoclonal Antibody Blocks Malaria Transmission and Defines a Highly Conserved Neutralizing Epitope on Gametes. *Nat. Commun.* 12 (1 12), 1–12. doi:10.1038/s41467-021-21955-1

Conway, D., Machado, R. L. D., Singh, B., Dessert, P., Mikes, Z. S., Povoa, M. M., et al. (2001). Extreme Geographical Fixation of Variation in the *Plasmodium falciparum* Gamete Surface Protein Gene Pfs48/45 Compared with Microsatellite Loci. 115, 145–156.doi:10.1016/s0166-6851(01)00278-x

Coombs, R. W., and Fiscus, S. (2009). Processing of Dried Blood Spots Standard Operating ProcedureProcedure Title: Processing of Dried Blood Spots Standard Operating Procedure Approved by (Network): Network Name, Title Signature Date ACTG Reviewed by (Laboratory): Name, Title Signature Date Revision History Version Effective Date (Dd/mmm/yy) Comments.

Cvijović, I., Good, B. H., and Desai, M. M. (2018). The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics* 209, 1235–1278. doi:10.1534/GENETICS.118.301058

Drakeley, C. J., Duraisingh, M. T., Póvoa, M., Conway, D. J., Targett, G. A. T., and Baker, D. A. (1996). Geographical Distribution of a Variant Epitope of Pfs4845, a *Plasmodium falciparum* Transmission-Blocking Vaccine Candidate. *Mol. Biochem. Parasitol.* 81, 253–257. doi:10.1016/0166-6851(96)02718-1

Eksi, S., Czesny, B., van Gemert, G.-J., Sauerwein, R. W., Eling, W., and Williamson, K. C. (2006). Malaria Transmission-Blocking Antigen, Pfs230, Mediates Human Red Blood Cell Binding to Exflagellating Male Parasites and Oocyst Production. *Mol. Microbiol.* 61, 991–998. doi:10.1111/J.1365-2958.2006.05284.X

Escalante, A. A., Lal, A. A., and Ayala, F. J. (1998). Genetic Polymorphism and Natural Selection in the Malaria Parasite *Plasmodium falciparum*. *Genetics* 149, 189–202. doi:10.1093/genetics/149.1.189

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Genton, B., Betuela, I., Felger, I., Al-Yaman, F., Anders, R. F., Saul, A., et al. (2002). A Recombinant Blood-Stage Malaria Vaccine Reduces *Plasmodium falciparum* Density and Exerts Selective Pressure on Parasite Populations in a Phase 1-2b Trial in Papua New Guinea. *J. Infect. Dis.* 185, 820–827. doi:10.1086/339342

Gilson, P. R., Nebl, T., Vukcevic, D., Moritz, R. L., Sargeant, T., Speed, T. P., et al. (2006). Identification and Stoichiometry of Glycosylphosphatidylinositol-Anchored Membrane Proteins of the Human Malaria Parasite *Plasmodium falciparum*. *Mol. Cell. Proteomics* 5, 1286–1299. doi:10.1074/mcp.M600035-MCP200

Graves, P. M., Wirtz, R. A., Carter, R., Burkot, T. R., Looker, M., and Targett, G. A. (1988). Naturally Occurring Antibodies to an Epitope on *Plasmodium falciparum* Gametes Detected by Monoclonal Antibody-Based Competitive Enzyme-Linked Immunosorbent Assay. *Infect. Immun.* 56, 2818–2821. doi:10.1128/iai.56.11.2818-2821.1988

Hancock, A. M., and Di Rienzo, A. (2008). Detecting the Genetic Signature of Natural Selection in Human Populations: Models. *Annu. Rev. Anthropol.* 37, 197–217. doi:10.1146/annurev.anthro.37.081407.085141

Healy, S. A., Anderson, C., Swihart, B. J., Mwakingwe, A., Gabriel, E. E., Decederfelt, H., et al. (2021). Pfs230 Yields Higher Malaria Transmission-Blocking Vaccine Activity Than Pfs25 in Humans but Not Mice. *J. Clin. Investigation* 131. doi:10.1172/JCI146221

Honnay, O. (2013). *Genetic Drift. Brenner's Encyclopedia of Genetics*. Second Edition, 251–253. doi:10.1016/B978-0-12-374984-0.00616-1

Huang, W. C., Deng, B., Mabrouk, M. T., Seffouh, A., Ortega, J., Long, C., et al. (2020). Particle-based, Pfs230 and Pfs25 Immunization Is Effective, but Not Improved by Duplexing at Fixed Total Antigen Dose. *Malar. J.* 19, 309–312. doi:10.1186/S12936-020-03368-5/FIGURES/6

Jones, S., Grignard, L., Nebie, I., Chilongola, J., Dodoo, D., Sauerwein, R., et al. (2015). Naturally Acquired Antibody Responses to Recombinant Pfs230 and Pfs48/45 Transmission Blocking Vaccine Candidates. *J. Infect.* 71, 117–127. doi:10.1016/j.jinf.2015.03.007

Kapesa, A., Kweka, E. J., Atieli, H., Afrane, Y. A., Kamugisha, E., Lee, M.-C., et al. (2018). The Current Malaria Morbidity and Mortality in Different Transmission Settings in Western Kenya. *PLOS ONE* 13, e0202031. doi:10.1371/journal.pone.0202031

Kocken, C. H. M., Jansen, J., Kaan, A. M., Beckers, P. J. A., Ponnudurai, T., Kaslow, D. C., et al. (1993). Cloning and Expression of the Gene Coding for the Transmission Blocking Target Antigen Pfs48/45 of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 61, 59–68. doi:10.1016/0166-6851(93)90158-T

Kocken, C. H. M., Milek, R. L. B., Lensen, T. H. W., Kaslow, D. C., Schoenmakers, J. G. G., and Konings, R. N. H. (1995). Minimal Variation in the Transmission-Blocking Vaccine Candidate Pfs4845 of the Human Malaria Parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 69, 115–118. doi:10.1016/0166-6851(94)00193-Q

Kundu, P., Semesi, A., Jore, M. M., Morin, M. J., Price, V. L., Liang, A., et al. (2018). Structural Delineation of Potent Transmission-Blocking Epitope I on Malaria Antigen Pfs48/45. *Nat. Commun.* 9. doi:10.1038/s41467-018-06742-9

Larrañaga, N., Mejía, R. E., Hormaza, J. I., Montoya, A., Soto, A., and Fontecha, G. A. (20132013). Genetic Structure of *Plasmodium falciparum* Populations across the Honduras-Nicaragua Border. *Malar. J.* 12 (1 12), 1–10. doi:10.1186/1475-2875-12-354

Lee, S.-M., Hickey, J. M., Miura, K., Joshi, S. B., Volkin, D. B., King, C. R., et al. (2020). A C-Terminal Pfs48/45 Malaria Transmission-Blocking Vaccine Candidate Produced in the Baculovirus Expression System. *Sci. Rep.* 10. doi:10.1038/s41598-019-57384-w

Lee, S.-M., Wu, Y., Hickey, J. M., Miura, K., Whitaker, N., Joshi, S. B., et al. (2019). The Pfs230 N-Terminal Fragment, Pfs230D1+: Expression and Characterization of a Potential Malaria Transmission-Blocking Vaccine Candidate. *Malar. J.* 18, 356. doi:10.1186/s12936-019-2989-2

Lennartz, F., Brod, F., Dabbs, R., Miura, K., Mekhaiel, D., Marini, A., et al. (2018). Structural Basis for Recognition of the Malaria Vaccine Candidate Pfs48/45 by a Transmission Blocking Antibody. *Nat. Commun.* 9. doi:10.1038/s41467-018-06340-9

Lombardo, F., and Christophides, G. K. (2016). Novel Factors of *Anopheles gambiae* Haemocyte Immune Response to *Plasmodium berghei* Infection. *Parasites Vectors* 9. doi:10.1186/S13071-016-1359-Y

MacDonald, N. J., Nguyen, V., Shimp, R., Reiter, K., Herrera, R., Burkhardt, M., et al. (2016). Structural and Immunological Characterization of Recombinant 6-cysteine Domains of the *Plasmodium falciparum* Sexual Stage Protein Pfs230. *J. Biol. Chem.* 291, 19913–19922. doi:10.1074/jbc.M116.732305

Mejia, R. (2012). Natural Selection, Not Mutation: Recombination in Drosophila Increases Diversity. *PLoS Biol.* 10, e1001423. doi:10.1371/JOURNAL.PBIO.1001423

Neafsey, D. E., Juraska, M., Bedford, T., Benkeser, D., Valim, C., Griggs, A., et al. (2015). Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N. Engl. J. Med.* 373, 2025–2037. doi:10.1056/NEJMOA1505819

Ochwedo, K. O., Omondi, C. J., Magomere, E. O., Olumeh, J. O., Debrah, I., Onyango, S. A., et al. (20212021). Hyper-prevalence of Submicroscopic *Plasmodium falciparum* Infections in a Rural Area of Western Kenya with Declining Malaria Cases. *Malar. J.* 20 (1 20), 1–8. doi:10.1186/S12936-021-04012-6

Ochwedo, K. O., Onyango, S. A., Omondi, C. J., Orondo, P. W., Ondeto, B. M., Lee, M.-C., et al. (2022). Signatures of Selection and Drivers for Novel Mutation on Transmission-Blocking Vaccine Candidate Pfs25 Gene in Western Kenya. *PLoS One* 17, e0266394. doi:10.1371/journal.pone.0266394

Oduma, C. O., Ogolla, S., Atieli, H., Ondigo, B. N., Lee, M.-C., Githeko, A. K., et al. (20212021). Increased Investment in Gametocytes in Asymptomatic *Plasmodium falciparum* Infections in the Wet Season. *BMC Infect. Dis.* 21 (1 21), 1–10. doi:10.1186/S12879-020-05761-6

Ogutu, B. R., Apollo, O. J., McKinney, D., Okoth, W., Siangla, J., Dubovsky, F., et al. (2009). Blood Stage Malaria Vaccine Eliciting High Antigen-specific Antibody Concentrations Confers No Protection to Young Children in Western Kenya. *PLoS ONE* 4, e4708. doi:10.1371/journal.pone.0004708

Onyango, S. A., Ochwedo, K. O., Machani, M. G., Omondi, C. J., Debrah, I., Ogolla, S. O., et al. (2021). Genetic Diversity and Population Structure of the Human Malaria Parasite *Plasmodium falciparum* Surface Protein Pfs47 in Isolates from the Lowlands in Western Kenya. *PloS one* 16, e0260434. doi:10.1371/JOURNAL.PONE.0260434

Ouattara, A., Barry, A. E., Dutta, S., Remarque, E. J., Beeson, J. G., and Plowe, C. V. (2015). Designing Malaria Vaccines to Circumvent Antigen Variability. *Vaccine* 33, 7506–7512. doi:10.1016/j.vaccine.2015.09.110

Outchkourov, N., Vermunt, A., Jansen, J., Kaan, A., Roeffen, W., Teelen, K., et al. (2007). Epitope Analysis of the Malaria Surface Antigen Pfs48/45 Identifies a Subdomain that Elicits Transmission Blocking Antibodies. *J. Biol. Chem.* 282, 17148–17156. doi:10.1074/jbc.M700948200

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera?A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 25, 1605–1612. doi:10.1002/JCC.20084

Rener, J., Graves, P. M., Carter, R., Williams, J. L., and Burkot, T. R. (1983). Target Antigens of Transmission-Blocking Immunity on Gametes of *Plasmodium falciparum*. *J. Exp. Med.* 158, 976–981. doi:10.1084/JEM.158.3.976

Roeffen, W., Lensen, T., Mulder, B., Teelen, K., Sauerwein, R., Eling, W., et al. (1994). Transmission Blocking Immunity as Observed in a Feeder System and Serological Reactivity to Pfs 48/45 and Pfs230 in Field Sera. *Mem. Inst. Oswaldo Cruz* 89 (Suppl. 2), 13–15. doi:10.1590/S0074-02761994000600004

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* 34, 3299–3302. doi:10.1093/molbev/msx248

Sabeti, P. C., Varilly, P., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., et al. (2007). Genome-wide Detection and Characterization of Positive Selection in Human Populations. *Nature* 449, 913–918. doi:10.1038/nature06250

Silvela, L., de la Peña, I., and Gomez-Ruano, R. (19991999). Selection under Negative Linkage Disequilibrium. Random Mating versus Inbreeding. *Heredity* 826 82, 598–604. doi:10.1046/j.1365-2540.1999.00520.x

Singh, K., Burkhardt, M., Nakuchima, S., Herrera, R., Muratova, O., Gittis, A. G., et al. (2020). Structure and Function of a Malaria Transmission Blocking Vaccine Targeting Pfs230 and Pfs230-Pfs48/45 Proteins. *Commun. Biol.* 3, 1–12. doi:10.1038/s42003-020-01123-9

Singh, S. K., Plieskatt, J., Chourasia, B. K., Fabra-García, A., Garcia-Senosiain, A., Singh, V., et al. (2021). A Reproducible and Scalable Process for Manufacturing a Pfs48/45 Based *Plasmodium falciparum* Transmission-Blocking Vaccine. *Front. Immunol.* 11, 3369. doi:10.3389/FIMMU.2020.606266

Singh, S. K., Thrane, S., Chourasia, B. K., Teelen, K., Graumans, W., Stoter, R., et al. (2019). Pfs230 and Pfs48/45 Fusion Proteins Elicit Strong Transmission-

Blocking Antibody Responses against *Plasmodium falciparum*. *Front. Immunol.* 10, 1256. doi:10.3389/fimmu.2019.01256

Singh, S. K., Thrane, S., Janitzek, C. M., Nielsen, M. A., Theander, T. G., Theisen, M., et al. (2017). Improving the Malaria Transmission-Blocking Activity of a *Plasmodium falciparum* 48/45 Based Vaccine Antigen by SpyTag/SpyCatcher Mediated Virus-like Display. *Vaccine* 35, 3726–3732. doi:10.1016/J.VACCINE.2017.05.054

Stothard, P. (2000). The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *BioTechniques* 28, 1102–1104. doi:10.2144/00286IR01

Tachibana, M., Miura, K., Takashima, E., Morita, M., Nagaoka, H., Zhou, L., et al. (2019). Identification of Domains within Pfs230 that Elicit Transmission Blocking Antibody Responses. *Vaccine* 37, 1799–1806. doi:10.1016/j.vaccine.2019.02.021

Takala, S. L., Coulibaly, D., Thera, M. A., Dicko, A., Smith, D. L., Guindo, A. B., et al. (2007). Dynamics of Polymorphism in a Malaria Vaccine Antigen at a Vaccine-Testing Site in Mali. *PLoS Med.* 4, e93. doi:10.1371/journal.pmed.0040093

van Dijk, M. R., Janse, C. J., Thompson, J., Waters, A. P., Braks, J. A. M., Dodemont, H. J., et al. (2001). A Central Role for P48/45 in Malaria Parasite Male Gamete Fertility. *Cell.* 104, 153–164. doi:10.1016/S0092-8674(01)00199-4

Vermeulen, A. N., van Deursen, J., Brakenhoff, R. H., Lensen, T. H. W., Ponnudurai, T., and Meuwissen, J. H. E. T. (1986). Characterization of *Plasmodium falciparum* Sexual Stage Antigens and Their Biosynthesis in Synchronised Gametocyte Cultures. *Mol. Biochem. Parasitol.* 20, 155–163. doi:10.1016/0166-6851(86)90027-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

REVIEW
published: 18 May 2022
doi: 10.3389/fgene.2022.819941

Unravelling the Role of Epigenetic Modifications in Development and Reproduction of Angiosperms: A Critical Appraisal

Priyanka Kumari[1], Sajid Khan[1], Ishfaq Ahmad Wani[1], Renu Gupta[2], Susheel Verma[3], Pravej Alam[4] and Abdullah Alaklabi[5]*

[1]Conservation and Molecular Biology Lab., Department of Botany, Baba Ghulam Shah Badshah University, Rajouri, India, [2]Division of Soil Sciences & Agricultural Chemistry, Faculty of Agriculture Sher e Kashmir University of Agricultural Sciences and Technology, Chatha, India, [3]Department of Botany, University of Jammu, Jammu, India, [4]Department of Biology, College of Science and Humanities, Prince Sattam bin Abdulaziz University (PSAU), Alkharj, Saudi Arabia, [5]Department of Biology, College of Science, University of Bisha, Bisha, Saudi Arabia

Epigenetics are the heritable changes in gene expression patterns which occur without altering DNA sequence. These changes are reversible and do not change the sequence of the DNA but can alter the way in which the DNA sequences are read. Epigenetic modifications are induced by DNA methylation, histone modification, and RNA-mediated mechanisms which alter the gene expression, primarily at the transcriptional level. Such alterations do control genome activity through transcriptional silencing of transposable elements thereby contributing toward genome stability. Plants being sessile in nature are highly susceptible to the extremes of changing environmental conditions. This increases the likelihood of epigenetic modifications within the composite network of genes that affect the developmental changes of a plant species. Genetic and epigenetic reprogramming enhances the growth and development, imparts phenotypic plasticity, and also ensures flowering under stress conditions without changing the genotype for several generations. Epigenetic modifications hold an immense significance during the development of male and female gametophytes, fertilization, embryogenesis, fruit formation, and seed germination. In this review, we focus on the mechanism of epigenetic modifications and their dynamic role in maintaining the genomic integrity during plant development and reproduction.

Keywords: epigenetic modification, DNA methylation, histone proteins, stress response, plant hormones, transposon silencing, miRNA

OPEN ACCESS

Edited by:
Sajad Ahmad Dar,
Jazan University, Saudi Arabia

Reviewed by:
Veenu Kaul,
University of Jammu, India
Maryam Sarwat,
Amity University, India

****Correspondence:***
Susheel Verma
eremurus@rediffmail.com

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: *22 November 2021*
Accepted: *14 February 2022*
Published: *18 May 2022*

Citation:
Kumari P, Khan S, Wani IA, Gupta R, Verma S, Alam P and Alaklabi A (2022) Unravelling the Role of Epigenetic Modifications in Development and Reproduction of Angiosperms: A Critical Appraisal. Front. Genet. 13:819941. doi: 10.3389/fgene.2022.819941

INTRODUCTION

Epigenetics is the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in the DNA sequence (Waddington 1957; Waddington, 2012; Iwasaki and Paszkowski, 2014). Epigenetic changes are induced through behavioral or environmental factors that may affect the way the genes work (Niederhuth and Schmitz, 2017; Parker et al., 2021). These modifications are reversible and do not change the sequences of DNA but alters the way in which DNA sequences are read. Epigenetic modifications include DNA methylation (Bouyer et al., 2017), histone modifications (Zhou, 2009; Liu et al., 2010), ubiquitination of histone

N-tails, and posttranscriptional silencing through small noncoding RNAs and RNA-mediated mechanisms (Slotkin and Martienssen, 2007; Matzke and Mosher, 2014). Gene and transposon activity get affected by epigenetic changes in the DNA present within a chromatin (Lippman et al., 2004; Rodrigues and Zilberman, 2015). Epigenetics regulate flowering time in plants through transposon silencing, paramutation, and genomic imprinting (Yaish et al., 2011; Ay et al., 2014; Blüumel et al., 2015; Parker et al., 2021). Epigenetic changes are inherited through alleles or epialleles. Epialleles are the sites that get transmitted to the next generations after being retained stably in the chromatin state. Epialleles act as supplementary sources of variation to regulate phenotypic diversity. In plant species, epialleles affect floral morphology, time of flowering, resistance against diseases, and leaf senescence (Brukhin and Albertini, 2021). The present review is a summary of the information available on different epigenetic modifications that appear to be important in growth, development, and reproduction of plants.

Epigenetic Modifications in Plants

Plants being sessile in nature are invariably affected by changing environmental conditions. However, they have the ability to adapt their biological processes according to the changing environments. They interact with their environment through consistent adjustments at the molecular level by modifying the patterns of gene expression (Yaish et al., 2011). Epigenetic regulations assist plants in increased tolerance against different environmental stresses by reprogramming their developmental stages, such as flowering time (Barozai and Aziz, 2018). In transgenic plants, epigenetics helps to understand the problems related to suitable expression of newly introduced transgenic segments (Madhusudhan, 2015). Epigenetic changes are conserved in plants and influence the structure of the chromatin which in turn regulates the gene expression. Epigenetic mechanisms are important to regulate various biological processes and disruption of any one of the epigenetic mechanisms leads to developmental abnormalities in plants. Therefore, epigenetic changes play dynamic roles in the growth and development of plants (Zhang et al., 2018).

Different biological pathways such as phytohormone signaling, photoperiodism, and vernalization in combination with the environmental signals (epigenetic changes) regulate flowering time by integrating internal state of development of plants (An et al., 2004; He, 2009; Amasino, 2010; Andrés and Coupland, 2012; Sun et al., 2014; Burgarella et al., 2016). This network of flowering regulation involves FLOWERING LOCUS D (FLD) and FLOWERING LOCUS C (FLC) gene transcription which are controlled through epigenetic mechanisms such as ubiquitination, acetylation/deacetylation, and methylation/demethylation concealed by hormone signaling. HISTONE DEACETYLASE 6 (HDA6) protein increases the rate of expression of FLC gene but ethylene sets off HDA6 expression. FLC is suppressed by FLOWERING LOCUS D through demethylation of H3K4me2 which facilitates H4 deacetylation in the same locus. H3K27me3 is added by PRC2 marking into FLD chromatin. This also engages PICKLE to link with DELLA which in turn facilitates the repression of FLC (Bastow et al., 2004; De Lucia et al., 2008; Jiang et al., 2008; Choi et al., 2009).

Regulation of Epigenetic Modifications

Phenotypic plasticity within the plant species plays an important role in adaptation to different environmental conditions allowing different cultivars to adjust and grow. Plasticity maintains the homeostasis within changing environmental conditions that allows better gene expression to adapt to different biotic or abiotic constrains (disease, herbivory, plant–plant competition, altitude, soil type, seasonal, day length, rain, and ambient temperature) (Gratani, 2014). Genetic plasticity within the plants is inferred by different epigenetic modifications that are regulated by DNA methylation, histone modification, transposon modification, noncoding RNAs, and chromatin modulation (Zhang et al., 2018; Zhang et al., 2008; Zhang et al., 2013).

DNA methylation

DNA methylation is a chromatin modification in plants and is conceivably inherited mitotically or meiotically over generations. DNA methylation is catalyzed by cytosine methyltransferases. It involves the addition of a CH_3 group (methyl group) at the fifth carbon position on cytosine residue generating 5-methyl cytosine in a sequence-specific manner. The methyl group acts as a platform for various protein complexes to attach and modify the chromatin scaffolds causing altered gene expression (Niederhuth and Schmitz, 2017). On the basis of the target sequence, methylation is of two types: asymmetrical and symmetrical methylation. Symmetrical methylation is CG and CHG methylation and asymmetrical methylation is CHH methylation (where H denotes any nucleotide other than guanine). Both symmetric, i.e., CG and CHG, and asymmetric methylation, i.e., CHH, exist in plants (Jacobsen and Meyerowitz, 1997). Only some genes are methylated in plants within a gene body, and methylation is restricted only to CG sites (Niederhuth and Schmitz, 2017). DNA methylation is found to be higher at repetitive sequences than genic regions in case of plant species. To maintain genome stability, silencing of TEs is important which can be mediated through RNA-directed DNA methylation (RdDM) (Slotkin and Martienssen, 2007; Matzke and Mosher, 2014). Despite having a pivotal role in different biological processes, DNA methylation applications in crop improvement are not fully investigated.

DNA methylation is induced biochemically as an epigenetic heritable change initiated through enzymes. It entails a shift of a methyl group to the fifth position on the cytosine residue and is catalyzed by DNA methyltransferases utilizing S-adenosylmethionine (Thapa and Shrestha, 2020). DNA methylation occurs at cytosine regions, *viz*., CG, CHG, and CHH (H stands in for A, T, or C) (**Figure 1**). METHYLTRANSFERASE 1 (MET1) enzyme catalyzes the methylation of CG. After the completion of DNA replication process, MET1 identifies hemi-methylated CG dinucleotides and methylates the unchanged cytosine in the daughter strand (Kankel et al., 2003; He et al.,

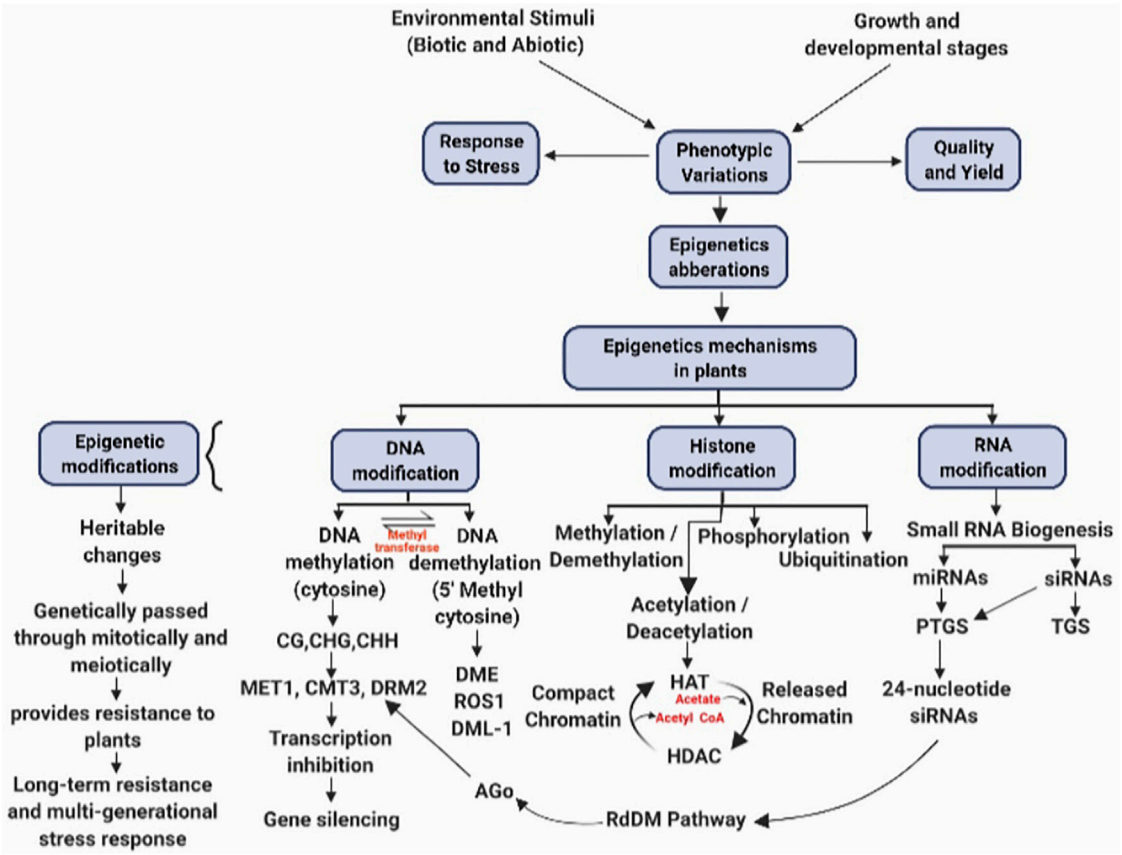

FIGURE 1 | Epigenetic changes in response to stress management during growth and development of plant.

2011). DNA methyltransferases CHROMOMETHYLASE 3 (CMT3) and CHROMOMETHYLASE 2 (CMT2) are said to catalyze CHG methylation (Lindroth et al., 2001; Stroud et al., 2014). Depending on the chromosomal region, CHH methylation is catalyzed through DOMAINS REARRANGED methyltransferase 2 (DRM2) or CMT2 methyltransferases. Methylation through CHROMOMETHYLASE 2 is catalyzed at histone H1–containing heterochromatin sites, while DRM2 catalyzes methylation at RdDM target areas (Zemach et al., 2013; Zhang et al., 2018) (**Figure 1**). There are two types of DNA methylation mechanisms: active and passive DNA methylation. A particular protein participates in the active process and demethylates the DNA sequence. The base excision repair pathway is involved. During DNA replication, methylation of cytosine is replaced with unmodified cytosine in a passive process. The reduction of activity of DNA methylases such as METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3) causes the addition of non-modified cytosine during DNA replication (Ibarra et al., 2012). In the case of plants, cytosine methylation has been extensively researched (Ruffini Castiglione et al., 2002). Cytosine alterations are not constant and vary greatly depending on the stages of development in plants and environmental factors (Burn et al., 1993).

DNA methylation is recruited *de novo via* DRM2 at all sequence contexts, and its role in CHH methylation is more prominent because CHH methylation depends only on *de novo* methylation. The RdDM pathway regulates the DRM2 activity. The RdDM pathway contains two plant-specific DNA-dependent RNA polymerases. The large subunit of these polymerases consists of proteins, i.e., NRPD1 and NRPE1, and functions specifically in transcriptional gene silencing and *de novo* DNA methylation. Single-stranded RNAs are produced through DNA-dependent RNA polymerase IV (Pol IV) in DNA repeat sites and in transposon regions. Pol IV transcription is facilitated by chromatin remodeling protein CLASSYs (CLSYs; CLSY1-4) (Yang et al., 2018; Zhou et al., 2018). Single-stranded RNA is converted to double-stranded RNAs *via* RNA-dependent RNA polymerase (RDR2). Double-stranded RNA is then cleaved by Dicer-like protein (DCL3) into 24-nt siRNA. This 24-nt siRNA binds to ARGONAUTE proteins (AGO4, AGO6, and AGO9). Loading of siRNA to ARGONAUTE proteins require KOW CONTAINING TRANSCRIPTION FACTOR 1 (KTF1)/SPT5-like protein. KTF1 (RNA-binding protein) binds to noncoding RNA transcripts produced from Pol V forming RNA-directed DNA methylation effector complex. Pol V generates single-stranded RNA transcripts in intergenic noncoding (IGN) regions. RNA transcripts generation *via* Pol V requires DRD1, DMS3, RDM1 and RDM4. RDM1 binds to single-stranded methylated DNA and leads to the recruitment of Pol V to

chromatin regions. To catalyze new DNA methylation, *de novo* DNA methyltransferases DRM2 are directed to specific chromatin regions *via* effector complex (He et al., 2011; Xie et al., 2012). It has been reported that six homologous proteins act in a redundant way in the RdDM pathway. These proteins are found in *Arabidopsis* and named as FACTOR of DNA METHYLATION 1–5 (FDM1-5) and INVOLVED IN DE NOVO (IDN2/RDM12). They belong to the SGS3-like plant-specific protein family, and their rice homolog is X1. They have an important role in transcriptional gene silencing like the SGS3 protein family (Xie et al., 2012).

Histone Protein Modifications

Histone modifications comprise an interesting part in epigenetics (Pfluger and Wagner, 2007). Histone proteins act as winder around which the segment of DNA gets wrapped and leads to the formation of a structural unit called as nucleosome. Nucleosomes comprise histone octamers consisting of two copies of each of the H2A, H2B, H3, and H4 histone proteins. The N-terminal tail of these histone proteins undergoes different modifications such as acetylation, methylation, sumoylation, ubiquitination, and phosphorylation. These histone modifications are correlated with either gene activation or repression. Several histone variants and enzymes are present in plants that modify histones posttranslationally and regulate gene expression (Zhou, 2009; Liu et al., 2010). Gene expression is brought about by the process of acetylation and phosphorylation, whereas gene expression is reduced through sumoylation and biotinylation. In the case of plants, deacetylation and methylation of H3K27 and H3K9 repress genes, while H3K4 and H3K36 acetylation and methylation activate gene expression. Epigenetic modifications participate in several biological processes like transcription, replication, and DNA repair by recruiting specific proteins needed in such processes (Jiang et al., 2009; Iwasaki and Paszkowski, 2014). Epigenetic modifications not only consist histone marks/modifications but also consist replacement with histone variants having different properties to influence gene expression (Liu et al., 2010). Chromatin immunoprecipitation applications following deep sequencing provide an insight for the genome-wide association studies regarding variants of histones and their posttranscriptional modifications (Butterbrodt et al., 2006). Responding to various biotic and abiotic stresses, histone modifications regulate the DNA transcription by interfering with the packaging structure either by activating the DNA to transcribe or making condensed structures thereby deactivating transcription machinery.

N-terminal tails of histones are the sites where most of the histone modifications take place. These modifications specify the function of chromatin and transcriptional activities (Jenuwein and Allis, 2001; Zhao et al., 2019). Histone modifications include methylation, acetylation, ubiquitination, and phosphorylation and occur at lysine and arginine residues. Histone methyltransferases (writers) are a group of SET domain which catalyzes histone lysine methylation, and downstream events are mediated by proteins (readers) which recognize specific lysine methylation. Two histone demethylases, i.e., Jumonji C (Jmj C) and lysine-specific demethylase 1 (LSD 1) catalyze the removal of histone lysine methylation (Zhao et al., 2019). In eukaryotes, the involvement of histone acetylation and methylation in gene expression regulation was first identified by Allfrey et al. (1964). It has been demonstrated that increase in histone acetylation leads to poor separation of sister chromatids in human fibroblasts and causes chromosomal defects during cell cycle in tobacco. Trichostatin A (TSA) is identified to have negative pleiotropic effects and has been found to decrease global histone deacetylation, resulting in an increase in acetylated histones (Cimini et al., 2003; Li et al., 2005).

The overexpression of antisense of the histone deacetylase gene (AtHD1) induces histone acetylation activity in *Arabidopsis thaliana*. The AtHD1 gene gives rise to pleiotropic phenotypes having a variety of developmental defects such as the suppression of apical dominance, ectopic expression of silenced genes, floral structure abnormalities, male and female sterility, heterochronic shift toward juvenility (Tian and Chen, 2001). In plant genome, the repressive state of heterochromatic regions is marked by H3K9me1 and H3K9me2. Heterochromatic regions are enriched with transposable elements and repetitive sequences. Heterochromatic regions in *Arabidopsis* are enriched with H3K27me1, and the association of H3K27me1 is catalyzed by plant-specific histone methyltransferases ARABIDOPSIS TRITHORAX-RELATED PROTEINS, i.e., ATXR 5 and ATXR 6. Mutation in H3K27me1 results in de-condensation of heterochromatin and the release of transposable silencing (Zhao et al., 2019). H3K27me3 deposition on chromatin is catalyzed by polycomb repressive complex 2 (PRC2) *via* histone methyltransferases. The subunits of PRC2 were first identified in *Drosophila melanogaster* (Shen et al., 2021). In plants, PRC2 deposits the H3K27me3 methylation and plays an important role in growth and developmental phases of plants. This modification is found in protein-coding genes and is regulated dynamically during the growth stages of plants. PRC2 consists of four components, *viz*., histone methyltransferases enhancer of zeste [E(z)], extra sex combs (Esc), suppressor of zeste 12 [Su(z)12], and the histone-binding nucleosome-remodeling factor 55 kDa (Nurf55, also called p55). One component of PRC2, i.e., E(z), specifically belongs to the SET [Su(var)3-9; E(z); trithorax] domain family and is responsible for histone H3 tri-methylation at Lys27 (Czermin et al., 2002; Butenko and Ohad, 2011; Simon and Kingston, 2013; Shen et al., 2021). In case of *Arabidopsis*, PRC2 components have multiple duplications, and there exists three homologs of E(z), *viz*., CURLY LEAF (CLF), SWINGER (SWN), and MEDEA (MEA); three homologs of Su(z), *viz*., EMBRYONIC FLOWER 2 (EMF2), VERNALIZATION 2 (VRN2), and FERTILIZATION-INDEPENDENT SEED 2 (FIS2); and one homolog of Esc, *viz*., COPY SUPPRESSOR OF IRA 1–5 (MSI1-5). FIS2 regulates mega-gametogenesis and endosperm development in plants during postfertilization events and EMF, and VRN polycomb repressive complexes regulate the development of sporophyte and phase transition, i.e., vegetative to reproductive in plants (Shen et al., 2021). Histone demethylases found in *Arabidopsis*, i.e., JUMONJI 13 (JMJ13), JUMONJI 30 (JMJ30), JUMONJI 32 (JMJ32), EARLY FLOWERING 6 (ELF6), and RELATIVE OF EARLY FLOWERING 6 (REF6), demethylate H3K27 and

depress genes temporally or spatially for processes like flowering, signaling of hormones, and circadian clock control (Sanchez et al., 2020).

Histone methylation is influenced by environmental factors (Boyko and Kovalchuk, 2008; Kim et al., 2010). Global gene expression analysis and chromatin immunoprecipitation (ChIP) tests have revealed that histone H3 Lys4 methylation (H3K4) patterns in *Arabidopsis* respond dynamically to dehydration stress (van Dijk et al., 2010). The floral initiator SHK1 kinase BINDING PROTEIN 1 (SKB1) mutant line *skb1* provides an example of the interaction between environmental stress and blooming. SKB1 attaches to chromatin and raises the quantity of histone 4 Arg3 (H4R3) symmetric dimethylation (H4R3sme2) and causes FLC expression and a number of stress-responsive genes to be downregulated. As a result, its mutant characteristics include salt hypersensitivity, late flowering, and stunted development (Zhang et al., 2011; Chen et al., 2013; Cheng et al., 2019). The standard ABC model determines flower architecture at the molecular level (Bowman et al., 1991; Bowman et al., 2012). The geographical bounds of each floral whorl are determined by precise union of gene expression and protein interactions in this model (sepals, petals, stamens, and carpel). The A class gene APETALA2 (AP2) regulates target gene expression as part of a complex that it forms with TOPLESS (TPL) and HISTONE DEACETYLASE 19 in *Arabidopsis* (HDA19). AGAMOUS (AG) and SEPALATA3 (SEP3), the C class and E class genes, respectively, are negatively regulated by the transcription repressor complex. Deacetylation of H4K16 in regulatory areas of AG and SEP3 mediates gene suppression (Krogan et al., 2012). Expression studies revealed and identified additional HDACs expressed in reproductive tissues in *Arabidopsis*, i.e., HDA5, HDA6, HDA7, HDA9, HDA15, and HDA18. Their function in fruit or flower development is unknown. Only the function of HDA6 has been reported, and it plays a role in the regulation of blooming time. Histone H3K4 demethylase, i.e., FLOWERING LOCUS D (FLD), interacts directly with HDA6. The complex represses the expression of three flowering repressors: FLD, MADS AFFECTING FLOWERING 4 (MAF4), and MAF5 by removing the acetyl and methyl groups from histone 3 at their loci (Yu et al., 2011). HDA6 is one of the HDACs engaged in RdDM. RdDM is a plant-specific epigenetic process and small interfering RNA (siRNA)–mediated epigenetic mechanism which regulates the chromatin silencing of developmental genes, transposable elements, and repetitive elements. The RdDM mechanism involves a large number of participants whose actions may be broken down into a few simple phases (Matzke and Mosher, 2014). The RdDM machinery involves two kinds of transcripts, *viz*., Pol IV and Pol V transcripts. Pol IV transcribes long noncoding RNAs (lncRNAs), and the lncRNAs gets transformed to double-stranded RNAs (dsRNAs) through RDR2 (Haag et al., 2012). dsRNAs then gets converted into siRNAs by DICER-like 3 (DCL3). The siRNAs are loaded into AGO4 and reimported into the nucleus after being exported to the cytoplasm. siRNA direct AGO4 to nascent scaffold transcripts of Pol V through precise base pairing. siRNA, AGO4, and lncRNA scaffold derived from Pol V recruit histone deacetylases (HDACs) and DNA methyltransferases, which in turn silence the genomic loci transcribed by Pol V by the process of histone deacetylation and DNA methylation. Histone deacetylation characterizes RdDM-silenced promoters, which is mediated by RPD3-type histone deacetylase AtHDA6 in *Arabidopsis*, which is homologous to SIHDA3 of tomato. Deacetylation is essential for subsequent methylation by histone methyltransferases (HMTs), and to control siRNA-dependent heterochromatin, there is a requirement of functional AtHDA6 (Li et al., 2005; Aufsatz et al., 2007). The mutants of AtHDA6 display the revival of RdDM-silenced promoters in spite of the presence of an RNA-silencing signal. Reduced cytosine methylation indicates that AtHDA6 plays an important role in methylation maintenance. The physical connection of AtHDA6 with DMTs, MET1, and CMT3 may facilitate this function. Acetylases and deacetylases of histones have an important role in flowering and fruit development of tomato plant (Aufsatz et al., 2002). Cigliano et al. (2013) identified potential histone modifiers of AU4 tomato genome using RNA sequencing data of tomato genome (Tomato Genome C 2012) generated by worldwide collaboration sequencing. Also, they analyzed the expression profiles of each histone modification in the sample tissues used by using RNA sequencing data from the same source. Two histone acetylases S1HAG18 and S1HAG6 showed peak expression in the floral samples used, which indicates their function in reproductive development of tomato plant. Recently, 15 histone deacetylases have been discovered in tomatoes. SlHDA3 was found to be the tomato homolog of AtHDA6 which is expressed in all tissues having the highest blossom stage expression (Zhao et al., 2014). Another tomato homolog of AtHDA19, i.e., S1HDA1, was found to be significantly expressed in the flowering stage, and its expression was repressed at the fruiting stage. In yeast two-hybrid tests, it has been found that histone deacetylases, including S1HDA1, S1HDA3, and S1HDA4, interact with MADS-box transcription factors, i.e., TOMATO AGAMOUS1 (TAG1) and TOMATO MADS-BOX (TM29) (Zhao et al., 2014). Transcription factor TAG1 is required for the expression of both ethylene-dependent and ethylene-independent ripening genes (Klee and Giovannoni, 2011). TM29 is a homolog of SEPALLATA, which when silenced, leads to the formation of aberrant flowers and parthenocarpic fruits (Ampomah-Dwamena et al., 2002).

Transposon Modifications

Most species have transposon elements (TEs) in their chromosomes, and multicellular eukaryotes have TEs as a key component of their genome. The majority of transposable elements are silenced epigenetically, although certain transposable elements have active transcription in epigenetic regulation mutants. Furthermore, environmental stress can trigger TE transcription, a mechanism that occurs across the evolutionary spectrum from bacteria to mammals (Capy et al., 2000). McClintock (1984) was the first to report that stress might cause TEs to shift, a result that has been widely corroborated in subsequent research (Grandbastien, 1998). Tnt1 and Tto1 are LTR-type Class I retroelements in tobacco, whose transposition is triggered by injury or through pathogen attack (Takeda et al.,

2001; Perez-Hormaeche et al., 2008). Also, Bs1 LTR-type Class I retroelement in maize was found to transpose after viral infection (Johns et al., 1985). In *Arabidopsis*, heat stress induces transcription of ONSEN (LTR-type Class I retroelement), and it transposes into siRNA-defective mutants (Ito et al., 2011). LTR of ONSEN has a heat-responsive region that is activated by transcriptional heat stress responses (Cavrak et al., 2014). As a result, genes near or containing newly inserted ONSEN copies become heat-responsive (Ito et al., 2011). All the above examples of transposons are of Class I DNA transposons, and these transpose *via* the "copy and paste" mechanism in response to stress. Class II DNA transposons transpose *via* a "cut and paste" process in response to stress. In *Antirrhinum majus*, low temperature increases the excision frequency of the Ac/Ds type transposon Tam3 (Harrison and Fincham, 1964; Carpenter et al., 1987). Transposable elements are a response of the genome toward environmental challenges and play a critical role in gene regulation and evolution of the genome (McClintock, 1984; Slotkin and Martienssen, 2007; Fedoroff, 2012). It has been proposed that TEs activation in response to environmental stress could provide epigenetic variability which could contribute toward the greater adaptive capacity of plants under stress conditions (Mirouze and Paszkowski, 2011; Bucher et al., 2012; Ashapkin et al., 2020). The active DNA transposon mPing has been found to preferentially insert into 50 flanking regions of genes rather than exons in rice. Cold and salt stress encourages transcription of a subset of genes by inserting mPing in the promoter region (Naito et al., 2009; Wang et al., 2016).

It has been reported that epigenetic reprogramming has an important role in transposon silencing and reprogramming in germ cells of plants (Feng et al., 2010). In *Arabidopsis thaliana*, one egg cell and one central cell containing two nuclei are produced during the process of female gametogenesis, and several accessory cells are also produced. The egg cell fuses with one sperm cell during double fertilization process forming an embryo, and the second sperm cell fuses with the central cell leading to the formation of an endosperm (triploid). A helix hairpin DNA glycosylase, DEMETER (*DME*) causes hypomethylation in the endosperm by removing methylated cytosine residues. Demethylation *via DME* activates expression of transposons through the RNAi pathway that introduces transcripts of transposons and produces additional siRNAs guiding DNA methylation. The siRNAs expression in the endosperm development indicates genome imprinting, and siRNAs production guides DNA methylation in egg cell reinforcing transposons silencing in the germ cells. Transposons silencing occurs in the germ cells, but mild activity of transposons in endosperm have no major effects as the endosperm is not inherited to next generation (Hsieh et al., 2009b; Gehring et al., 2009; Ibarra et al., 2012; Ito 2013). Transposons activity in the embryo has been suppressed *via* the RNAi pathway having a major role in reprogramming of paternal genome in *Arabidopsis* (Han et al., 2019). It has also been reported that sRNAs lead to hypomethylation of vegetative cells when they are transported to the sperm cell *via* the cytoplasm of the pollen grains. This accumulates siRNAs and activates *de novo* remethylation through the RdDM pathway leading to transposons silencing in the gametes. Genes responsible for biogenesis of siRNA and transposons silencing are expressed at a very low concentration in pollen. But the *DDM1* gene is an exception, which is specifically expressed in the sperm cells of mature pollen. In the vegetative nucleus of wild-type, DNA demethylation and activation of transposons occur by the downregulation of the *DDM1* gene. This activation of transposons in vegetative cells also has no effect on the fitness of the species, as vegetative cells are not inherited to the next generation and have no contribution of genetic material, i.e. DNA, to the fertilized embryo. This demonstrates that epigenetic inheritance and transposon silencing are contributed through genome reprogramming guided *via* RNAs (Ito, 2013; Dziegielewski and Ziolkowski, 2021).

miRNA Modifications

miRNAs comprise 20–24 noncoding nucleotides that regulate gene expression after transcription and are also involved in the age pathway by regulating the time of flowering in plants by using RNA-directed DNA methylation (RdDM) (Matzke et al., 2001; Matzke et al., 2007; Pikaard, 2006; Teotia and Tang, 2015; Dziegielewski and Ziolkowski, 2021) (**Figure 1**). The miRNA molecules with the help of the RdDM pathway can bring about DNA methylation on a specific location (Teotia and Tang, 2015). Global gene expression analysis in *Arabidopsis thaliana* having genetic disorder in photoperiodic signaling pathway and system integrate genes suggesting the role of miRNAs in mediating the effects of floral induction (Schmid et al., 2003; Kinoshita and Richter, 2020). miR156 and miR172 are the two key miRNAs acting as the main elements in controlling the age pathways in plants by downregulating target genes and also effecting flowering time in many plant species (**Figure 2**). The level of miR156 increases during the vegetative stage of plants and decreases as plants proceed toward the reproductive stage, and at this stage, miR172 increases (Tanaka et al., 2011; Luo et al., 2013; Teotia and Tang, 2015). The expression of miRNAs is determined by environmental factors. miRNAs affect the expression of certain genes of plants when exposed to abiotic and biotic stress by frequently reprogramming genes involved in the developmental pathways (Covarrubias and Reyes, 2010; Hirayama and Shinozaki, 2010; Urano et al., 2010; Tiwari and Rajam, 2022) (**Figure 1**). Stress-inducible miRNAs and their expected targets have been discovered to be preserved in *Arabidopsis* (Sunkar and Zhu, 2004). It has been reported through global gene expression in rice plants that when exposed to stress conditions such as cold, drought, excessive salt, and ABA treatment, miRNAs modulate gene expression in the rice plants (Shen et al., 2010). Correlation between miRNA biogenesis mechanism proteins, response to stress, and flowering has been found in many mutant lines of *Arabidopsis*. For example, ABH1 and CBP20 encode cap-binding factors which are required for maturation of RNA (Papp et al., 2004). The *abh1* mutant exhibits ABA hypersensitivity, and the *cbp20* line exhibits both drought tolerance and ABA hypersensitivity (Hugouvieux et al., 2001; Kwak et al., 2005). In addition to the role of miRNA in stress responses, they are also important in controlling the flowering in *Arabidopsis* (Aukerman and

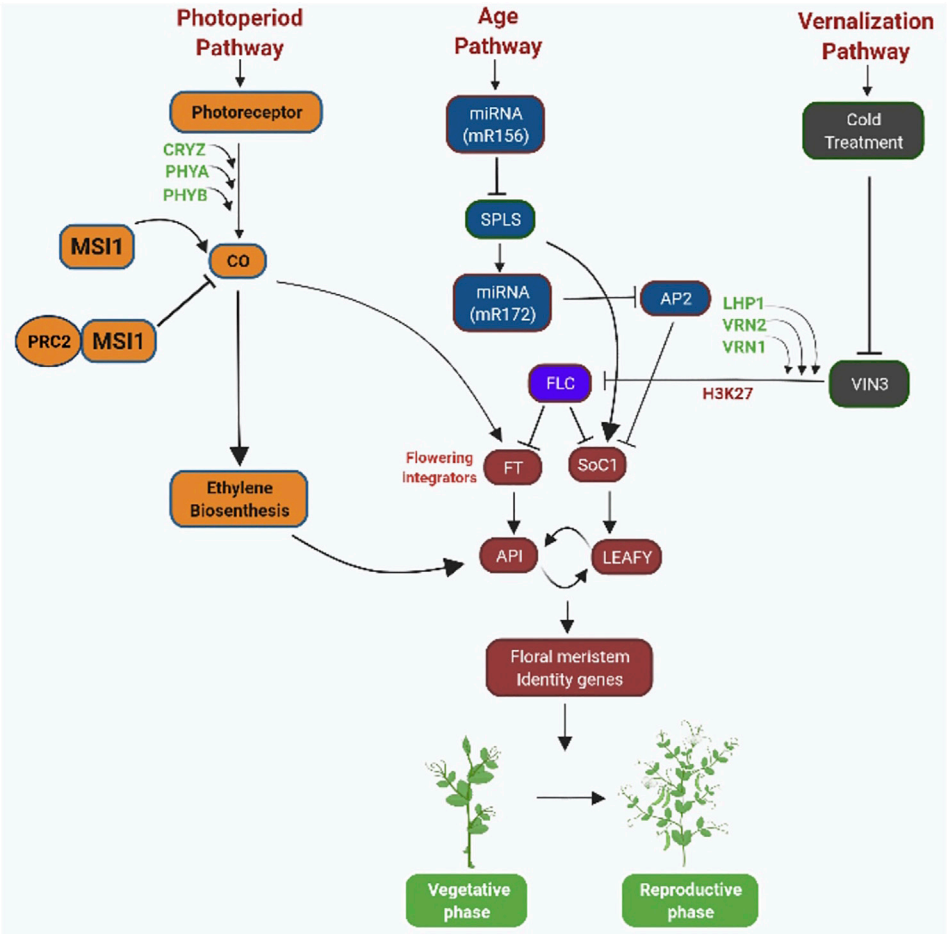

FIGURE 2 | Role of epigenetic modifications in controlling flowering time in co-relation with biological pathways: flowering is induced by a number of molecular pathways that respond to external and internal signals. Flower integrator genes such as FT and SOC1 are regulated by flowering activators and repressors. In the photoperiodic pathway, chromatin modifications involve the well-conserved histone binding protein MULTICOPY SUPRESSORS OF IRA1 (MSI1)–like protein family. MSI1-like protein forms a complex with ubiquitous protein, i.e., POLYCOMB REPRESSIVE COMPLEX 2 in *Arabidopsis* and controls the switch to flowering. MSI1 acts in the normal expression of CO in long day (LD) plants. Reduced expression of CO in *msi1* mutants leads to FT and SOC1 repression. In age pathway, miR156 and miR172 acts as positive regulators of SOC1 gene. Vernalization leads to the expression of VIN3. VIN3 represses FLC transcription by binding with PcG protein (VRN1, VRN2, LHP1). PcG proteins epigenetically modify chromatin of FLC by trimethylation of H3K27.

Sakai, 2003; Chen and Li, 2004; Sunkar and Zhu, 2004). Long intronic noncoding RNA (COLDAIR) mediates interaction of H3K27me3 at FLC. This interaction of COLDAIR with FLOWERING LOCUS C (FLC) gene targets PRC2 to interact with FLC, resulting in FLC suppression during cold treatment, i.e., vernalization (Baulcombe and Dean, 2014). miRNA partially regulates the FL, and mutations in the miRNA biogenesis genes DCL1 and DCL3 cause delayed flowering in these mutant backgrounds due to overly high FLC expression (Schmitz et al., 2007) (**Figure 1**). Another mutant line HYPONASTIC LEAVES 1 (*HYL1*) was also found to show late flowering characteristics (Lu and Fedoroff, 2000). HYPONASTIC LEAVES 1 (HYL1) gene produces a protein that binds to double-stranded RNA (dsRNA) and mediates gene control *via* miRNA (Han et al., 2004). Scientists have reported that in addition to *hyl1* mutants, many *Arabidopsis* mutants in miRNA biogenesis machinery genes have phenotype related to ABA and salt hypersensitivity, i.e., SERRATE (SE) gene, DCL1 gene, HUA-ENHANCER 1 (HEN1) gene, and HASTY gene (Lu and Fedoroff, 2000; Han et al., 2004; Rasia et al., 2010; Zhang et al., 2018).

Epigenetic Regulation in Plant Development and Morphogenesis

Stem cells are present in plant meristems and lead to the formation of all tissues and organs. The RdDM factor transcript levels are higher in tissues of meristem in *A. thaliana* than in cell expansion tissues, e. g., tissues of hypocotyl and differentiated leaves (Zhang et al., 2018). The DNA methylation level was found to be more in columella cells of root meristem because these cells are least condensed with pericentromeric chromatin. This allows more accessibility to RdDM factors. There have been no obvious reports of

TABLE1 | Role of epigenetically induced modifications in trait control, development, and morphogenesis of different plant species.

Species	Epigenetic modification	Plant developmental responses	References
Arabidopsis thaliana	DNA demethylation	Stomatal development	Yamamuro et al. (2014)
	DNA methylation	Diseases resistance	Zhang et al. (2018)
	Histone modification	Growth and development	Czermin et al. (2002)
	Histone modification	Development of sporophyte and phase transition	Shen et al. (2021)
	DNA methylation	Response to pathogen pathways and floral induction	Gao et al. (2003)
	mi RNA	Enhanced plant phenotypic vigor	Mondal et al. (2020)
	miRNA	Regulation of trichome distribution	Xue et al. (2014)
	Histone methylation	Response to drought stress	Van Dijk et al. (2010)
	Histone methylation	Response to necrotrophic fungi	Berr et al. (2010)
	Histone methylation	Response to salt stress	Kim et al. (2010)
	Histone methylation	Response to cold stress	Pavangadkar et al. (2010)
Aegilops tauschii	DNA hypomethylation	Plant defense response against pathogens	Geng et al. (2019)
Alternanthera philoxeroides	DNA methylation	Maintains leaf and stem morphology	Gao et al. (2010)
Antirrhinum majus	DNA methylation	Stem elongation	Gourcilleau et al. (2019)
Brassica rapa	DNA methylation/histone modifications	Increased crop yield	Hauben et al. (2009)
	DNA hypomethylation	Plant defense response against pathogens	Kellenberger et al. (2016)
Citrullus lanatus	DNA hypomethylation	Plant defense response against pathogens	Sun et al. (2019)
Elaeis guineensis	DNA methylation	Somaclonal variations	Jaligot et al. (2000)
Glycine max	DNA methylation and histone modifications	Response to salt stress	Song et al. (2012)
Ilex aquifolium	DNA methylation	Leaf development	Herrera and Bazaga (2012)
Lycopersicon esculentum	Small RNA	Enhanced plant vigor	Kundariya et al. (2020)
Nicotiana tabacum	DNA hypomethylation	Plant defense response	Wang et al. (2018)
Nicotiana tabacum	DNA methylation	Aluminum and salt stress	Choi et al. (2009)
	DNA methylation	More efficient genetic transformation of plants.	Yamagishi and Kikuta (2021)
	DNA methylation	Regulation of nutritive value	Quadrana et al. (2014)
Oryza sativa	DNA methylation	Leaf development	Zhang et al. (2018)
	miRNA	Floral abnormalities	Zhu et al. (2009)
	miRNA	Leaf development	Xie et al. (2012)
	Histone methylation	Gene expression under drought stress	Zong et al. (2012)
	Histone methylation	Stem elongation	Chen et al. (2013)
	DNA hypomethylation	Plant defense response	Atighi et al. (2020)
	Histone methylation	Control of transposon activity	Cui et al. (2013)
Panicum virgatum	miRNA	Morphological alterations	Fu et al. (2012)
Torenia fournieri	miRNA	Plant growth	Shikata et al. (2012)
Trifolium pratense	DNA methylation	Defense against stress	Yang et al. (2020)
Vitis vinifera	DNA methylation	Response to medium-high temperatures in regenerated plants	Baranek et al. (2015)
	DNA methylation	Authentication of plant origin	Xie et al. (2017)
Zea mays	DNA methylation	Improved yield	Tani et al. (2005)
	DNA methylation	Adaptive evolution	Xu et al. (2020)
	DNA methylation	Prediction of key phenotypes	Regulski et al. (2013)
	DNA methylation	Phenotypic predictor, independent of genetic polymorphism data	Xu et al. (2019)
	DNA methylation	Controls cell division in maize leaves	Candaele et al. (2014)
	Histone modifications	Enhance plant resilience to stress	Forestan et al. (2019)
	Small RNA	Used as complementary biomarkers in crops	Seifert et al. (2018)
Sugar beet	DNA methylation	Tolerance to bolting	Trap-Gentil et al. (2011)
Rubber trees	DNA methylation	Tolerance against cold stress	Tang et al. (2018)

apparent meristem abnormalities in RdDM mutants in *A. thaliana*, but rice and maize mutants show severe developmental defects, indicating that these components play critical roles in meristem function (Zhu et al., 2012; Kawakatsu et al., 2017; Zhang et al., 2018). After the emergence of leaves from shoot apical meristem, many developmental genes get suppressed

by the deposition of SET DOMAIN GROUP PROTEIN 711 (SDG711)–dependent H3K27me3 in rice. SDG711-dependent H3K27me3 synchronizes with DRM2-catalyzed non-CG DNA methylation in the gene body of *Oryza sativa*. SDG711 interacts physically with DRM2, and mutation in DRM2 decreases chromatin binding of SDG711 and deposition of H3K27me3 at repressed gene sites (Zhou et al., 2016; Zhang et al., 2018). During the growth of leaves in maize, DNA methyltransferases were found to be regulated variably resulting in distinct patterns of CG and CHG methylation in the division zone, transition, elongation, and mature zone reflecting the leaf's spatial gradient of cells (Zhou et al., 2016). DNA methylation is required for pattern development in some leaf epidermal cells of *Arabidopsis thaliana*. DNA demethylation in *Arabidopsis thaliana* is dependent on genes of the ROS1 subfamily encoding 5-methylcytosine DNA glycosylases/lyases. DNA demethylation initiated *via* ROS1 leads to the expression of EPF2 gene. EPIDERMAL PATTERNING FACTOR 2 (EPF2) is a peptide ligand that inhibits stomatal development, resulting in an excess of stomatal lineage cells. Malfunction of ROS1 results in promoter hypermethylation and suppression of the gene encoding EPF2, resulting in stomatal cell development (Yamamuro et al., 2014; Zhou et al., 2016; Zhang et al., 2018). Similarly, the loss of function of H3K9 demethylase IBM1 (increase in BONSAI methylation 1) causes elevated H3K9me2 and CHG DNA methylation, as well as the suppression of three LRR RECEPTOR-LIKE SERINE/THREONINE-PROTEIN KINASE ERECTA family genes that encode EPF2 receptors, resulting in stomatal pattern formation problems as seen in ROS1 mutant plants. In *ros1* plants, the mutation in RdDM factors, H3K9 methyltransferase SUVH4, and CMT3 in *ibm1* plants can rescue the stomatal pattern formation resulting through abnormal epigenetic regulation because DNA methylation of EPF2 promoter and silencing of EPF2 depends on RNA-directed DNA methylation. Also, ROS1 acts against RdDM action leading to the expression of EPF2. This indicates that two DNA methylation-mediated mechanisms are responsible for regulating leaf epidermal cell patterning in *A. thaliana* (**Table 1**) (Candaele et al., 2014; Yamamuro et al., 2014; Wang et al., 2016; Zhou et al., 2016; Zhang et al., 2018). FLC, a MADS box transcriptional repressor that keeps *Arabidopsis* apices in the vegetative stage, is downregulated by vernalization. As a result, epigenetic alterations at the FLC locus hasten flowering (Kim et al., 2009).

Epigenetic Regulation Under Environmental Stress

DNA methylation suppresses gene expression under stressful conditions, allowing the plant to conserve energy and strength for survival (Thapa and Shrestha, 2020). During water scarcity, CAM plants show the transition from C3 photosynthetic cycle to CAM pathways, which increases their resilience. This is accompanied by an increase in genomic methylation and hypermethylation of satellite DNA. Hypermethylation response is used to synthesize chromatin structure, which controls the expression of several genes and helps the plants to withstand stressful conditions. Hypermethylation was also discovered when the root tip of pea plants was exposed to water scarcity conditions (Thapa and Shrestha, 2020). The vernalization process, which involves prolonged exposure to cold conditions, initiates flowering in some plant species and is a well-studied example of how cold causes epigenetic changes that affect flowering. Epigenetic regulator, NRPD1, a DNA-binding bromodomain-containing protein, AtGCN5-related GNAT family 5 (acetyltransferase 5) and histone deacetylase were upregulated in *Arabidopsis* (Lee et al., 2005). Low temperature has been linked to DNA demethylation in *Arabidopsis* and other plant species like *Zea mays* (Steward et al., 2002), *Antirrhinum majus*, and *Triticum aestivum* (Sherman and Talbert, 2002; Hashida et al., 2003; Hashida et al., 2003) (**Table 1**). The *Arabidopsis* VERNALIZATION INSENSITIVE 3 gene (VIN3), a chromatin-remodeling plant homeodomain (PHD) finger protein that increases acetylation levels, is induced by cold exposure. This protein is essential for FLC repression and flowering enhancement. Because FLC expression is not lowered by cold treatment, the mutant lines for VIN3 do not respond to vernalization and so remain in a vegetative state for longer durations (Sung and Amasino, 2004; Soppe et al., 2021) (**Table 1**). During vernalization, this complex attaches to VIN3 locus chromatin (Schonrock et al., 2006). In *Arabidopsis*, on the other hand, a decrease in H3K27me3 modifications within the histones of the cold-responsive gene COR15A and the GALACTINOL SYNTHASE gene ATGOLS3 results in enhanced gene expression (Taji et al., 2002; Kwon et al., 2009). Similarly, during dehydration stress, the plant trithorax factor (ATX1) (Alvarez-Venegas et al., 2003) tri-methylates Lys4 residues of histone H3 (H3K4me3), regulating floral organ development and altering expression of transcription factor WRKY70 (Alvarez-Venegas et al., 2007). ATX1 mutations result in severe flaws in floral architecture (Alvarez-Venegas et al., 2003).

Various studies have reported that in *A. thaliana* and other plant species, including apples, *Pharbitis nil*, plant hormones such as auxin, cytokinin, Gibberellic acid, and abscisic acid interact to control flowering (Domagalska et al., 2010; Matsoukas, 2014). Salicylic acid (SA) is implicated in the control of CONSTANS, FLOWERING LOCUS C, FLOWERING LOCUS T, and MADS-box protein SOC1 transcription (Martinez et al., 2004). Interestingly, late-blooming phenotype of SA-deficient plants coincides with a 2- to 3-fold expression of FLC, lowering the FT levels in LD or SD circumstances as compared to wild-type plants. Furthermore, chromatin alterations are involved in the dynamic shift in the gene expression (Sun et al., 2014). For example, FLC and FT expression in *A. thaliana* is controlled epigenetically (Swiezewski et al., 2009; Ietswaart et al., 2012) (**Figures 2**, **3**). It has been reported that under cold stress, Polycom Repressive Complex 2 (PRC2) is involved in silencing the FLC locus (floral repressor) through H3K27me3 (Yuan et al., 2016) (**Figure 2**). Also, some studies reported that the silencing of FLC successfully brought about through reducing H3K4me2 levels in FLC gene (Liu et al., 2010). PRC2 and Flowering Locus D (FLD) work in coordination to silence FLC (Shafiq et al., 2016; Campos-Rivero et al., 2017). Sumoylation/

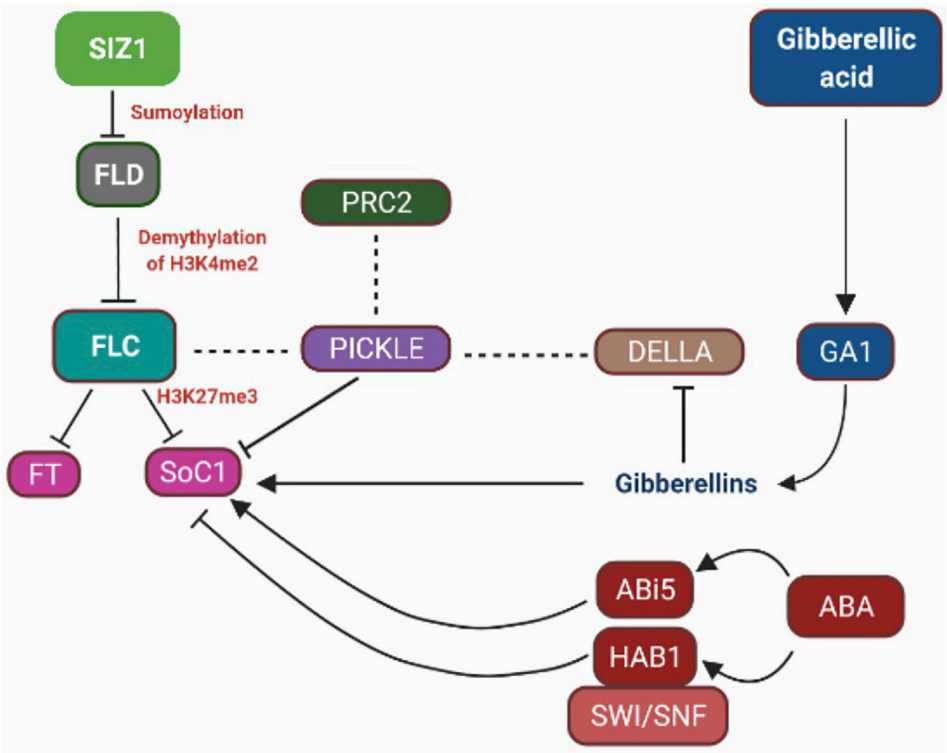

FIGURE 3 | Epigenetic control on phytohormones: DELLA proteins are negative regulators of SOC1. DELLA proteins in combination with FLC interact with PICKLE (chromatin-remodeling protein) and PRC2 and repress the expression of SOC1. PRC2 regulates GA signaling by increasing H3K27me3-repressing histone protein. ABA INSENSITIVE MUTANT 5 overexpression delays flowering initiation by upregulating the expression of FLC. ABA HYPERSENSITIVE 1 (HAB1) is a negative regulator of flowering when combine with the chromatin-remodeling complex SWI/SNF. FLC is negatively regulated by lysine-specific demethylase 1–type histone demethylase (FLD). FLD causes demethylation of H3 histone (H3K4me3) in FLC. FLC downregulates SOC1 expression by trimethylation of lysine 27 on histone H3 protein *via* the formation of heterochromatic regions. SUMO E3 ligase (SIZ1) promotes expression of FLC by facilitating repression of FLD through sumoylation.

TABLE 2 | Summary of the role of epigenetic regulation along with phytohormone action in plant species.

Species	Trait	Hormone action	References
Arabidopsis thaliana	Production of flowers	Auxin (IAA)	Wu et al. (2015)
	Function in the floral meristem	Cytokinin dehydrogenase enzymes CKX3 and CKX5	Bartrina et al. (2011); Meijon et al. (2011)
	Flowering	ABA INSENSITIVE 3 and ABA INSENSITIVE 5 genes, ABA HYPERSENSITIVE 1 (HAB1) by interconnecting with SWI/SNF chromatin-remodeling complex	Saez et al. (2008); Bond et al. (2009)
	Delay in bloom	Rise in ethylene concentration	Achard et al. (2007)
	Flowering *via* FLC expression	HDA6 and HDA19 expression by ethylene	Zhou et al. (2005); Wu et al. (2008)
	Flowering	Jasmonic acid signaling *via* HDA6, a HISTONE DEACETYLASE	Liu et al. (2012); Yu et al. (2011)
Brassica napus	Response to pathogen pathways and floral induction	HAD19 expression *via* ethylene	Gao et al. (2003)
Elaeis guineensis	Development of flower	Auxin/cytokinin ratio	Eeuwens et al. (2002); Jaligot et al. (2011)
Populus species	Bud dormancy	Abscisic acid (ABA) and ABA-responsive factors, i.e., PtAB13	Rohde et al. (2002); Ruttink et al. (2007); Rios et al. (2014)
Rosa hybrid	Control of floral development and blossoming time	changes in miRNA expression in response to ethylene	Ma et al. (2008); Pei et al. (2013)

desumoylation action of the FLD gene can regulate acetylation/deacetylation of histones through the unspecified procedure. Histone demethylase is encoded by the FLD gene, thereby mediating H3K4me2 demethylation and facilitating H4 histone deacetylation in FLC chromatin (Jin et al., 2008). Exposure to various amounts of synthetic auxins results in epigenetic alterations that affect flower growth (Jaligot et al., 2000). The mantled phenotype of *Elaeis guineensis* Jacq (African oil palm) is characterized by anomalies in the development of flowers, leading to alteration in the auxin/cytokinin ratio (**Table 2**) (Eeuwens et al., 2002, Jaligot et al., 2011; Campos-Rivero et al., 2017). Mantled blooms were enhanced by applying a high amount of cytokinin (kinetin) and a low amount of auxin [1-naphthaleneacetic acid (NAA)], and there were less number of mantled flowers when a high amount of NAA and low amount of kinetin were applied (Eeuwens et al., 2002). Mantled phenotype resulted from DNA hypomethylation caused by kinetin, and the opposite phenotype of the plant resulted from DNA hypermethylation caused by NAA (Jaligot et al., 2000; Eeuwens et al., 2002, Jaligot et al., 2011). Pin-shaped inflorescence in *Arabidopsis thaliana* resulted from failed floral primordial initiation caused by Pin-formed mutant *pin-1*. This mutant diminishes the polar auxin transport thereby producing inflorescence devoid of flowers in *A. thaliana* (Okada et al., 1991). When IAA is given exogenously, it causes the production of flowers, which can be reversed (Reinhardt et al., 2000). Histone alterations have also been shown to have a role in transcriptional control of auxin target genes (Wu et al., 2015). mRNA accumulation increases in LFY and FILAMENTOUS FLOWER (FIL) when auxin is applied, resulting in initiation of the floral primordium. Auxin treatment leads to elevation of H3K9ac levels in LFY and FIL gene loci boosting flower primordial (Wu et al., 2015). In the absence of auxin, TOPLESS and HDA19 were reported to repress LFY and FIL loci through binding at their MP-sites, resulting in transcription inhibition of genes (Wu et al., 2015). Cytokinins play a major role in the division and differentiation of cells in the floral meristem (Schaller et al., 2015). Accumulation of cytokinins in *A. thaliana* meristem regulates the size of shoot apical meristem and the activity of cells in the shoot meristem. Cytokinin degradation is catalyzed by two cytokinin dehydrogenase enzymes CKX3 and CKX5 performing a regulatory function in the floral meristem of *A. thaliana* (**Table 2**) (Bartrina et al., 2011). In the central WUSCHEL (WUS) domain, CKX3 is expressed, and CKX5 expresses in the broad region of the apical meristem. Double mutants, i.e., *ckx3* and *ckx5*, lead to the formation of large inflorescence and flower meristem. Phenotype developed by these double mutants indicates that cytokinin signaling precisely identifies the niche of stem cells and retards the development of cells (Bartrina et al., 2011). Meijon et al. (2011) reported that cytokinins work in coordination with epigenetic modifications and regulate flowering processes in plants. Cytokinin dihydrozeatin riboside and isopentenyladenine end the dormancy period and influence flowering through the DNA methylation process. The levels of DNA methylation decrease before the initiation of flowers, but after the formation of floral organs, the levels of DNA methylation increase (**Table 2**) (Meijon et al., 2011). Scientists state that cytokinin is an epigenetic component whose function is to regulate gene expression during the transition from the vegetative to reproductive form by instigating demethylation. Abscisic acid (ABA) acts as a floral repressor in *A. thaliana* where externally applied ABA affects the blooming time (Wang et al., 2013). ABSCISIC ACID-INSENSITIVE MUTANT 5 (*AB15*) overexpression retards floral initiation through upregulation of expression of FLC (Wang et al., 2013). ABA INSENSITIVE 3 and ABA INSENSITIVE 5 genes also regulate flowering by encoding two transcription factors, i.e., basic leucine zipper (bZIP)–type and B3-type (Hauser et al., 2011). ABA HYPERSENSITIVE 1 (HAB1) protein is also involved in controlling flowering in *Arabidopsis* by interconnecting with SWI/SNF chromatin-remodeling complex during transcription induced by ABA (**Figure 3**). In response to cold stress, chromatin-remodeling affects the histone core proteins by increasing the concentration of ABA, which leads to an increased level of histone H3 acetylation (Saez et al., 2004). Abscisic acid (ABA) and ABA-responsive factors play an important role in the maintenance of bud dormancy in perennial plants such as the *Populus* species. The PtAB13 gene is a homolog of AB13 of *Arabidopsis*, and its overexpression and downregulation regulate seed dormancy *via* ABA signaling. It causes alterations in bud developmental processes and misregulates the expression of genes during the process of bud dormancy (Rohde et al., 2002; Ruttink et al., 2007; Rios et al., 2014) (**Table 2**). There is some evidence that epigenetic regulation and ET synthesis are linked (Zhou et al., 2005). It is well known that during cold stress, ET levels rise, which may be associated with vernalization processes (**Figure 2**) (Chu and Lee, 1989; Zhao et al., 2014), which are often regulated by DNA methylation or demethylation (Burn et al., 1993; Sherman and Talbert, 2002). Cold temperatures cause a rise in ET in *Arabidopsis*, which delay bloom; nevertheless, if the temperature is again appropriate, blooming is stimulated. Ethylene, on the other hand, has been shown to promote the expression of HDA6 and HDA19 (*Arabidopsis* HDACs) (Zhou et al., 2005). HDA6 upregulates FLC expression in *A. thaliana* (Wu et al., 2008) (**Table 2**). HAD19 links the hormone response to pathogen pathways and floral induction in *Brassica napus* through a similar epigenetic mechanism in response to variations in ethylene sensitivity. HAD19 has been shown to interact with bnKCP1 (a putative factor with a kinase-inducible domain), a cold-inducible factor that is highly expressed in flowers (Gao et al., 2003). The function of ethylene in the control of floral development and blossoming time has been demonstrated in roses (*Rosa hybrida* Samantha). ET is involved in the regulation of petal cell expansion during the opening of a rose flower (**Table 2**) (Ma et al., 2008). In roses, changes in miRNA expression in response to ethylene have been documented, with five miRNAs (miR156, miR164, miR166, miR139, and rhy-miRC1) demonstrating a strong link between ethylene and petal growth control (**Table 2**) (Pei et al., 2013). Plants use Jasmonic Acid (JA) and jasmonate molecules as signaling molecules in response to

stress, such as mechanical or biotic injuries produced by ozone exposure, dehydration, or pathogen infection (Overmyer et al., 2000; Berger, 2002; Farmer et al., 2003; Loyola-Vargas et al., 2012). These, on the other hand, are involved in a variety of developmental processes, including nitrogen storage, fruit ripening, senescence, and blooming (Creelman and Mullet, 1995; Creelman and Mullet, 1997). The study of the epigenetic involvement in the control of JA during blooming has been less thorough than the study of the genetic role. However, significant breakthroughs have been made. In *Arabidopsis*, for example, HDA6, a HISTONE DEACETYLASE, is essential for JA response and flowering (Wu et al., 2008). This deacetylase, in collaboration with MET1, governs locus-directed heterochromatin silencing, potentially by recruiting MET1 to certain loci and therefore sets the groundwork for later non-CG methylation (To et al., 2011). HDA6 also physically binds with FLD (Yu et al., 2011; Liu et al., 2012), which contributes in the deacetylation of FLC chromatin and hence represses gene expression (He et al., 2003; Wu et al., 2008). This shows that HDA6 is involved in *Arabidopsis*' JA response and blooming (**Figure 3**).

Epigenetic Regulation in Response to Biological Virulence

In response to pathogen infection and symbiotic microbe colonization, plants show genome-wide DNA methylation alterations (Xiao et al., 2021). The first evidence of epigenetic regulations in response to biotic factors was reported to control virulence *via* posttranscriptional gene silencing (PTGS). Transcriptional gene silencing (TGS) is another permanent defense mechanism against DNA viruses *via* RdDM (**Figure 1**). It has been reported that virus infection in *Arabidopsis* is controlled *via* m^6A-specific methylation of the RNA genome in the alfalfa mosaic virus (AMV). ALKBH9B, an *Arabidopsis* protein having demethylase activity, removes m^6A from ssRNA molecules and accumulates in the cytoplasm of the siRNA bodies. This process suggests the role of m^6A demethylase in mRNA silencing (Ramirez-Prado et al., 2018; Martinez-Perez et al., 2017; Ashapkin et al., 2020) (**Table 1**). The demethylase DME is required for nodulation in *Medicago truncatula*. Several hundred genomic sites, including a small proportion of nodule-specific symbiosis genes, are variably methylated during nodule growth. In cyst nematode–infected soybean and *A. thaliana* roots, widespread DNA hypomethylation was detected. Differential DNA methylation in *A. thaliana* has been produced by *Pseudomonas syringae* pv. tomato str. DC3000 (Pst DC3000) (Dowen et al., 2012; Sahu et al., 2013). Differentially methylated cytosines have been found in CG and CHH contexts in gene-rich areas, specifically at the 5′ and 3′ ends of protein-coding genes. Also, Pst DC3000–responsive DNA methylation correlates negatively with neighboring gene expression levels across the genome, indicating that at the gene borders, DNA methylation is regulated dynamically and may contribute to differential gene expression in response to pathogens. In cucumber leaves and pollen grains, DNA methylation is caused by plant pathogenic ncRNAs promoter areas and transcriptional activation of some ribosomal RNA (rRNA) genes, leading in an abundance of short RNAs produced from rRNA. In *A. thaliana*, external application with pathogenic resistance phytohormone, i.e., salicylic acid, resulted in megabase-scale DNA hypomethylation in pericentromeric areas. DNA methylation has been proved to be a protective mechanism against unwanted transposition and a defense system against endonuclease digestion (Yaish, 2013). After infection with the virus, genomic methylation increases, and gene methylation related to resistance decreases. An increase in methylation promotes stability of the genome when plants are attacked by a virus, but genetic recombination is caused by a decrease in gene methylation levels and ultimately new genes produced help in resistance against pathogens (Engler et al., 1993).

Plant susceptibility to certain infectious pathogens can be altered by mutations in DNA methylation or demethylation regulators (Yu et al., 2013; Zhang et al., 2018). Plant tolerance to the biotrophic fungus *Hyaloperonospora arabidopsidis* is similarly increased in DNA hypomethylation mutants like *nrpe1* and diminished in DNA hypermethylation mutants like *ros1* (Sanchez et al., 2016; Zhang et al., 2018). POL V mutations lower susceptibility to the fungal necrotrophic pathogens *Botrytis cinerea* and *Plectosphaerella cucumerina*, in addition to enhancing resistance to biotrophic pathogens. An *nrpd1* (POL IV) mutant does not have altered resistance to Pst DC3000 or fungal infections, unlike POL V mutants, indicating that POL V can regulate plant immune responses independently of canonical RdDM (Lopez et al., 2011; Zhang et al., 2018). Plants with the AGO4 mutant alleles *ago4-1* and *ago4-2*, on the other hand, showed higher vulnerability to Pst DC3000, suggesting that AGO4 plays a distinct role in plant disease resistance compared to the other RdDM components. The DNA demethylase triple mutant *ros1–dml2–dml3* and wild-type *A. thaliana* were compared, and it was discovered that DNA hypermethylation occurred more frequently in the mutant at regions flanking the gene body, such as upstream promoter regions and three untranslated regions. In the *ros1–dml2–dml3* plants, over 200 genes are repressed, a large number of which have known or potential involvement in biotic stress response and are enriched with tiny transposons in their promoters. The *ros1–dml2–dml3* mutant is more susceptible to the fungal disease *Fusarium oxysporum*, which supports this theory (Zhang et al., 2018) (**Table 1**).

Epigenetic Regulation in Plant Reproduction and Meiosis

Epigenetic modifications display a significant contribution in the position and rate of crossovers; however, the mechanism of the molecular crossovers remains to be fully studied (Wibowo et al., 2016). The highly coordinated gene expression within the germ cells requires epigenetic reprogramming. Epigenetically induced molecular pathways play a vital role in essential chemical and physiological processes during plant meiosis. The transmission of epi-alleles produced in response to environmental pressures raises issues about how agronomic treatments and

environmental circumstances may influence the expression of critical crop characteristics that are evaluated in particular genetic improvement techniques. Indeed, from the standpoint of agricultural genetic improvement, understanding the epigenetic regulation of plant reproduction and meiosis is of great importance.

Meiotic siRNAs play an important role in epigenetic control of meiotic chromosome condensation, with essential implications for crop genetic improvement. In maize, OUTER CELL LAYER 4 (OCL4) encodes a transcription factor HD-ZIP IV which is required for the biogenesis of small RNAs and the other 21-nt phasiRNA (Zhai et al., 2015). This transcription factor also induces the synthesis of other proteins belonging to pentatricopeptide repeat (PPR) proteins, NB-LRR, and MYB families in other species (Howell et al., 2007; Zhu et al., 2012). It has been established that the presence of the histone variation H2A.Z is favorably linked to the occurrence of crossovers (Shilo et al., 2015). Plant crossover hotspots are suppressed by DNA methylation and H3K9me2 (Yelina et al., 2015). In *Arabidopsis*, the loss of DNA methylation has also been demonstrated to change crossover distribution in a chromatin type–dependent way (Mirouze et al., 2012).

SPO11-1-oligonucleotides and SPO11 topoisomerase-like transesterases produce DNA double-strand breaks to generate a high-resolution method to profile meiotic double-strand break patterns genome-wide. SPO11-1-oligonucleotides have been mapped in the *Arabidopsis* genome, and their role in regulating chromatin, DNA, and crossover frequency has been studied (Choi et al., 2018). The identification and mapping of these short DNA sequences in crop genomes will be important to discover epigenetic markers associated with key epigenetic modulators. Another important gene involved in the epigenetic modulation of plant reproduction is DECREASE IN DNA METHYLATION 1 (DDM1). DDM1 has ATPase activity that controls DNA methylation linked to crossover occurrence (Castiglione et al., 2002; Higo et al., 2012).

Male and Female Gametophyte Development

In plants, the primary germ cells do not directly enter spermatogenesis and oogenesis. Instead, the pollen mother cells (PMCs) in the anthers and the megaspore mother cells (MMCs) in the ovaries are generated in the floral meristem as a consequence of two meiotic divisions followed by a series of mitotic divisions, resulting in haploid male and female gametophytes, respectively (Manning et al., 2006). The male and female gametophytes are good models for investigating cell polarity, morphogenesis, and epigenetic regulation of cell development and specialization, and signaling pathways in angiosperms, despite their modest size and limited number of cells. Microsporogenesis and microgametogenesis are two phases of pollen formation that take place in the anthers. The sporogenous layer of the anther produces diploid microsporocytes, or PMCs. PMCs produce a tetrad of four haploid cells after two meiotic divisions. The tetrads then split up into individual microspores. Following that, two mitotic divisions occur: the first produces a big vegetative cell and a smaller generative cell, followed by the generative cell division, which produces two sperm cells, while the vegetative cell does not divide (Brownfield et al., 2009).

Epigenetic alterations are critical in the development of both male and female gametophytes, as well as in fertilization (Ingouff et al., 2017). It was discovered in *Arabidopsis* that the methylation level of PMCs in a normal environment (CG and CHG) was greater than in an adverse situation (CHH) (Kumar and Mohapatra, 2021). Symmetric methylation is generally seen in transposable elements, whereas asymmetric hypermethylation is mostly found in protein-coding genes. Increased methylation in a symmetric situation is believed to facilitate TE activity suppression, which guarantees genome stability before and throughout meiosis. Inactivation of methylation in an asymmetric setting, on the other hand, increases the activation of genes required for sperm cell development and conception. A substantial remodeling of chromatin occurs during PMC maturation, in addition to DNA methylation, promoting the start of meiosis.

The transition from the mitotic to meiotic phase is accompanied by a decrease in restrictive chromatin (H3K27me1 and H3K27me3) and an increase in permissive chromatin (H3K4me3) (Borg et al., 2020). The vegetative cell becomes roundish during meiosis and asymmetric during mitotic division of the haploid microspore. It has more methylation in the CHH areas, but it loses centromere-specific histone H3 (CENH3) due to decondensation of pericentromeric heterochromatin, local hypomethylation due to DME/ROS1 demethylases, and transposable element activation (Calarco et al., 2012).

During mitotic and meiotic cell division, the centromeric histone H3 (CENH3) variation is critical for the assembly and function of kinetochores. The inclusion of CENH3 into centromeric nucleosomes is the first step in kinetochore formation. The amount of CENH3 deposited on the centromeres changes depending on the stage of the cell cycle. CENH3 is also required for vegetative cell division and the removal of additional DNA (Lermontova et al., 2015). Hypomethylation of TEs leads in the production of 21–22 nt siRNAs, which are transferred to sperm cells and used by RdDM methylation to repress their TEs (Slotkin et al., 2009). During the whole time of pollen generation and development, whole-genome cell-specific methylation profiling indicated a high degree of CG and CHG methylation in the DNA of microspores, sperm, and vegetative cells.

During the development of a megasporocyte, DNA methylation level remains unchanged in context to CG, while it decreases temporarily in the context of CHH (Ingouff et al., 2017). The MMC, as well as the functional megaspore, are specified and differentiated through intercellular interactions mediated by mobile trans activating siRNAs (tasiRNAs), which are produced in the nucellus' surrounding cells and transported to the MMC, where they implement transcriptional and translational silencing (Baulcombe and Dean, 2014). The AGO9, RDR6, and SDS3 (a suppressor of genetic silencing 3) enzymes have been demonstrated to control the generation of

TABLE 3 | Role of epigenetically induced modifications in reproduction of different plant species.

Plant species	Plant characteristic/s	Epigenetic mechanism	References
Arabidopsis thaliana	Flowering phenology	DNA methylation	Borrdorf et al. (2010); Zhang et al. (2013)
	Control of flowering time	Chromatin modifications	He (2009)
	Control of flowering time	Histone methylation	Deal et al. (2007); Tamada et al. (2009)
	Control of flowering time	Epigenetic repression through vernalization	Sheldon et al., 2008
	Control of flowering and senescence	Histone methylation *via* Jasmonate signaling	Wu et al. (2008)
	Temperature response and flowering time	Histone modification	He et al. (2011)
	Increased stamen and carpel numbers	Cytosine methylation	Jacobsen and Meyerowitz (1997)
	Double-fertilization	DNA demethylation	Gehring et al. (2009); Hsieh et al. (2009b); Ibarra et al. (2012)
	Seed dormancy	DNA demethylation	Bouyer et al. (2017); Zhang et al. (2018)
	Fertilization of egg cells	Transposon silencing *via* RdDM pathway	Bouyer et al. (2017); Zhang et al. (2018)
	Transcriptional gene silencing	DNA methylation *via* RdDM pathway	Xie et al. (2012)
	Phenotype with late flowering	DNA methylation	Ghoshal and Gardiner, (2021)
	Transcriptional gene silencing (TGS)	DNA methylation	Ramirez-Prado et al. (2018), Ashapkin et al. (2020)
	Mega-gametogenesis and endosperm development	Histone modification	Shen et al. (2021)
	Flowering, signaling of hormones and circadian clock control	Histone modification	Zheng and Gehring (2019); Sanchez et al. (2020)
Brachypodium distachyon	Control of flowering time	miRNA	Wu et al. (2013)
Gossypium barbadense	Flowering time and seed dormancy	DNA methylation	Song et al. (2017)
Linaria vulgaris	Peloric flowers with abnormal actinomorphic flowers	Cytosine methylation	Cubas et al. (1999)
Lycopersicon esculentum	aberrant flowers and parthenocarpic fruits	Histone deacetylation	Ampomah-Dwamena et al. (2002)
	Fruit ripening	DNA methylation	Zhong et al. (2013)
Marchantia polymorpha	Spermatogenesis	DNA methylation	Kumar and Mohapatra, (2021)
Oryza sativa	Flowering time	Histone modification	Sun et al. (2014)
	Flowering and reproduction	Histone modification, DNA methylation	Shi et al. (2015)
Populus Spp.	Bud dormancy	DNA methylation	Rohde et al. (2002)
Prunus mume	Bud dormancy	DNA methylation	Zhong et al. (2013)
Pyrus pyrifolia	Bud dormancy	DNA methylation	Liu et al. (2012)
Rubus idaeus	Bud dormancy	DNA methylation	Mazzitelli et al. (2007)
Sinningia speciosa	Flowering time control	miRNA	Li et al. (2013)
Solanum lycopersicum	Fruit ripening	DNA methylation	Manning et al. (2006)
Solanum lycopersicum	Fruit ripening	Histone deacetylation	Zhao et al. (2014)
Solanum ruiz-lealii	Flower abnormalities	DNA methylation	Marfil et al. (2009)
Spinacia oleracea	Artificial induction of flowering	DNA methylation	Cheng et al. (2019)

such siRNAs in *Arabidopsis* (Olmedo-Movolif et al., 2010). The onset of mega gametogenesis in *Arabidopsis* nucellus is hampered when AGO5 expression is disrupted. Throughout mega gametogenesis, methylation in the CG and CHH contexts stays constant. CG methylation inside genes and transposons of the central cell of the embryo sac was lower than that of sperm cells, as shown in *Arabidopsis* (Park et al., 2016). This suggests that even before fertilization, the potential transcription of male genes is repressed. Mobile noncoding tasiRNAs regulate epigenetic control during gametogenesis in the embryo sac, thus siRNAs from the central cell penetrate the egg cell and decrease transposable element activity.

Fertilization and Embryogenesis

In both the embryo and endosperm, fertilization eliminates CHH hypomethylation of the paternal genome (Ibarra et al., 2012). Remethylation of the paternal DNA is most likely mediated by maternal siRNAs. One rationale for the epigenetic suppression of the male genome during early embryogenesis might include maternal regulation of embryo and endosperm size, as well as detection of self-pollen, which is important in interspecific crosses (Creasey et al., 2014). For appropriate embryo development, proper and consistent methylation of the dividing egg cell DNA is critical. In comparison to mature embryos, young embryos and endosperm tissues are

hypomethylated, which represents the high transcriptional activity of genes in the developing embryo and the preparation for the embryo's dormancy (Bouyer et al., 2017; Kawakatsu et al., 2017).

The levels of DNA methylation are strictly controlled in distinct tissues and cell types throughout the life cycle of a plant. In comparison to the embryos, endosperms of *Oryza sativa* and *A. thaliana* show worldwide DNA hypomethylation. It is caused by DME-dependent active demethylation in the central cell before fertilization in *A. thaliana* (Gehring et al., 2009; Hseih et al., 2009a; Ibarra et al., 2012). MET1 transcriptional repression also occurs during female gametogenesis, but it is unable to play a role in demethylation. Because genome-wide CG hypomethylation was not found in wild-type endosperm, DNA methylation is recovered in the dme mutant endosperm (Ibarra et al., 2012; Park et al., 2016). The vegetative cells experience DME-dependent DNA demethylation, which is accompanied by significant DDM1 downregulation (**Table 3**) (Slotkin et al., 2009; Ibarra et al., 2012). As a result, demethylated and de-silenced transposons create siRNAs. The siRNAs reach the vegetative cell after passing through the sperm cells, fortifying the RdDM pathway. POL V and DRM2 but not POL IV of *A. thaliana* was found in egg cells. These are the requirements for generation of siRNA through the conventional RdDM pathway. Therefore, transposon siRNAs accumulation in the sperm cells may enhance transposon silencing following fertilization of egg cells. During seed development, there is a rise in global levels of CHH methylation, and during seed germination, the levels fall due to passive demethylation.

During seed germination, however, the metabolic and, as a result, transcriptional-genetic activity of the embryonic tissues increases again, accompanied by a reduction in methylation in the CHH context, which is linked to the activation of protein-coding gene expression (Papareddy et al., 2020). The epigenetic "memory" associated with the histone marks is removed and is not transferred to the following generations of cells because the histones inherited from the egg and sperm are not reproduced in the cells of the embryo but are synthesized afresh (Papareddy et al., 2020). Thus, embryogenesis comes after the meiosis checkpoint or clearing box, which eliminates the maternal sporophyte's epigenetic markers from the DNA. Endosperm methylation is substantially lower than that of the embryo, reflecting its strong transcriptional and metabolic activity.

At the same time, paternal genomes (i.e., genomes transported by sperm cells into the egg cell and the central cell of the embryo sac) are more methylated than maternal genomes (Heish et al., 2009a). Endosperm demethylation appears to be necessary to diminish transposable element activity *via* the production of siRNAs, which are transported into neighboring embryo cells and methylate the terminal sections of transposons by RNA-directed DNA methylation, thus inactivating them (Ingouff et al., 2017).

Fruit Ripening

About 1% of the DNA methylation at cytosine–phosphate–guanine sites in the fruit pericarp of tomato gets changed during fruit development (Lang et al., 2017; Zhang et al., 2018). Many fruit-ripening genes have active DNA demethylation because their promoter regions contain binding sites for RIPENING-INHIBITOR (RIN), a prominent ripening transcription factor. Most known ripening genes whose expressions were adversely linked with promoter DNA methylation levels had confirmed RIN binding to target promoters. Premature ripening of tomato fruits was induced by treatment with a chemical inhibitor of DNA methylation, which caused promoter hypomethylation and expression of the gene encoding COLORLESS NON-RIPENING (CNR), which is a critical RIN-targeted gene for fruit ripening (Gao et al., 2010; Liu et al., 2015; Lang et al., 2017). The gradual DNA demethylation that occur during fruit ripening in *Solanum lycopersicum* is mediated by DML2. The expression of DNA demethylase DME-LIKE 2 (DML2) increases rapidly in ripening fruits. In *S. lycopersicum*, DML2 targets both ripening-induced and ripening-repressed genes, implying that active DNA demethylation is necessary for both ripening-induced gene activation and ripening-repressed gene suppression (Telias et al., 2011; El-Sharkaw et al., 2015; Daccord et al., 2017; Zhang et al., 2018) (**Table 3**).

Seed Dormancy

In perennial plants, epigenetic modifications regulate seasonal dormancy cycles. Meristem and bud growth are controlled by photoperiod, temperature, etc. Transcriptomic studies have revealed that bud dormancy events in plants such as *Populus* spp., *Rubus idaeus*, *Euphorbia esula*, *Prunus mume*, *Vitis* spp., *Prunus persica*, and many other perennial plant species have been triggered through changes in their gene expression. These changes affect regulation of cell cycle, perception of light, signaling of hormones, and response to stress (Rios et al., 2014). Characterization and identification of nondormant mutants of perennial plants, such as *evergrowing* (*evg*) mutant of *Prunus persica*, contributed toward increasing molecular work at the level of genes, thereby renewing the field of dormancy. Deletion in tandemly repeated sequences of MADS-box genes, i.e., DORMANCY-ASSOCIATED MADS-box gene (DAM1-6) leads to nondormant phenotype of *evg* mutant. DAM genes are expressed in buds and get affected by photoperiod and chilling temperatures, thereby affecting the developmental stages of plants (Lloret et al., 2021). The concentration of DAM genes, i.e., DAM5 and DAM6, was found to be high in dormant buds, and after chilling treatment, the level fell and released the dormancy of buds (Yamane et al., 2011). In many other perennial species, such as *Rubus idaeus*, *Prunus mume*, *Pyrus pyrifolia*, and *Actinidia deliciosa*, DAM-like genes have been found to have dormancy-dependent expressions. In transgenic plants, the heterologous expression of these genes has regulatory roles in flowering and dormancy. In *Arabidopsis*, the expression of DAM1 and of SVP1 and SVP3 in *Actinidia deliciosa* led to delay in the time of flowering (Horvath et al., 2010; Wu et al., 2012) (**Table 3**).

CONCLUSION AND FUTURE PROSPECTS

Plants being sessile in nature are invariably affected by changing environmental conditions. However, they have the ability to adapt their biological processes according to the changing

environments. They interact with their environment through consistent adjustments at the molecular levels. Epigenetic mechanisms contribute to these adjustments. These changes within the plant species modify the gene expression and help different plant species to withstand the extremes of different biotic or abiotic constrains. Such changes are induced due to DNA methylation, histone modification, e.g. acetylation/deacetylation, methylation/demethylation, ubiquitination, phosphorylation, and sRNA/modifications, which work in tandem with respect to environmental signals, transposon silencing, and hormone signaling to control the expression of genes in plants. These alterations help to maintain the survival of plants and maximize the chances of sexual reproduction under stress conditions. The role of DNA methylation in plant improvement is the preferred mechanism to investigate gene function and manipulate plants for creating novel varieties having capability to survive under stress conditions. Therefore, improved knowledge of epigenetic mechanisms *via* thorough study at the molecular level will be helpful. FLC controls flowering time in *Arabidopsis* that involves many genes in FLC expression *via* chromatin modifications. Further studies have to be undertaken to know how these diverse epigenetic modifications interact with one another to regulate the expression of FLC. CRISPR-based systems can be useful in altering the expression of genes for research applications as well as crop improvement efforts. The generation of genetically viable agricultural varieties that can endure a warmer world relies heavily on information generated from the plants' epigenomic profiles subjected to environmental challenges. Understanding the dynamic management of histone methylation and how histone methylation controls plant growth would be expanded through biochemical and genetic studies. Identification of intrinsic histone demethylases in plants, particularly lysine and/or arginine demethylases, and elucidation of their functions in regulating plant development and genome stability would require biochemical and genetic evidence. The roles of regulation of H3K27me3 in plants are limited and more in-depth knowledge will enable new researchers to enhance productivity of crops under limiting environmental conditions.

AUTHOR CONTRIBUTIONS

PK conceptualized the review and wrote it in consultation with SV and SK. PK, IAW, and SK contributed to writing and editing various sections on epigenetic modifications in angiosperms. PK, SK, and IW contributed to the figures and table. AA and PA contributed to revision of the manuscript. SV contributed to supervision and review of the manuscript. All authors have read and approved the submitted version of the manuscript.

REFERENCES

Achard, P., Baghour, M., Chapple, A., Hedden, P., Van Der Straeten, D., Genschik, P., et al. (2007). The Plant Stress Hormone Ethylene Controls floral Transition via DELLA-dependent Regulation of floral Meristem-Identity Genes. *Proc. Natl. Acad. Sci.* 104 (15), 6484–6489. doi:10.1073/pnas.06107171010.1073/pnas.0610717104

Aiese Cigliano, R., Sanseverino, W., Cremona, G., Ercolano, M. R., Conicella, C., and Consiglio, F. M. (2013). Genome-wide Analysis of Histone Modifiers in Tomato: Gaining an Insight into Their Developmental Roles. *BMC Genomics* 14, 57. doi:10.1186/1471-2164-14-57

Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. *Proc. Natl. Acad. Sci.* 51 (5), 786–794. doi:10.1073/pnas.51.5.786

Alvarez-Venegas, R., Abdallat, A. A., Guo, M., Alfano, J. R., and Avramova, Z. (2007). Epigenetic Control of a Transcription Factor at the Cross Section of Two Antagonistic Pathways. *Epigenetics* 2 (2), 106–113. doi:10.4161/epi.2.2.4404

Alvarez-Venegas, R., Pien, S., Sadder, M., Witmer, X., Grossniklaus, U., and Avramova, Z. (2003). ATX-1, an *Arabidopsis* Homolog of Trithorax, Activates Flower Homeotic Genes. *Curr. Biol.* 13 (8), 627–637. doi:10.1016/s0960-9822(03)00243-4

Amasino, R. (2010). Seasonal and Developmental Timing of Flowering. *Plant J. Cell Mol. Biol.* 61 (6), 1001–1013. doi:10.1111/j.1365-313x.2010.04148.x

Ampomah-Dwamena, C., Morris, B. A., Sutherland, P., Veit, B., and Yao, J.-L. (2002). Down-Regulation of TM29, a TomatoSEPALLATA Homolog, Causes Parthenocarpic Fruit Development and Floral Reversion. *Plant Physiol.* 130, 605–617. doi:10.1104/pp.005223

An, H., Roussot, C., Suárez-López, P., Corbesier, L., Vincent, C., Piñeiro, M., et al. (2004). CONSTANS Acts in the Phloem to Regulate a Systemic Signal that Induces Photoperiodic Flowering ofArabidopsis. *Development* 131 (15), 3615–3626. doi:10.1242/dev.01231

Andrés, F., and Coupland, G. (2012). The Genetic Basis of Flowering Responses to Seasonal Cues. *Nat. Rev. Genet.* 13 (9), 627–639. doi:10.1038/nrg3291

Ashapkin, V. V., Kutueva, L. I., Aleksandrushkina, N. I., and Vanyushin, B. F. (2020). Epigenetic Mechanisms of Plant Adaptation to Biotic and Abiotic Stresses. *Ijms* 21 (20), 7457. doi:10.3390/ijms21207457

Atighi, M. R., Verstraeten, B., De Meyer, T., and Kyndt, T. (2020). Genome-wide DNA Hypomethylation Shapes Nematode Pattern-triggered Immunity in Plants. *New Phytol.* 227, 545–558. doi:10.1111/nph.16532

Aufsatz, W., Mette, M. F., van der Winden, J., et al. (2002). HDA6, a Putative Histone Deacetylase Needed to Enhance DNA Methylation Induced by Double-Stranded RNA. *EMBO J.* 21, 6832–6841. doi:10.1371/journal.pbio.000006710.1093/emboj/cdf663

Aufsatz, W., Stoiber, T., Rakic, B., et al. (2007). Arabidopsis Histone Deacetylase 6: a green Link to RNA 839 Silencing. *Oncogene* 26, 5477–5488. doi:10.1038/sj.onc.1210615

Aukerman, M. J., and Sakai, H. (2003). Regulation of Flowering Time and floral Organ Identity by a MicroRNA and its APETALA2-like Target Genes. *Plant Cell* 15 (11), 2730–2741. doi:10.1105/tpc.016238

Ay, N., Janack, B., and Humbeck, K. (2014). Epigenetic Control of Plant Senescence and Linked Processes. *J. Exp. Bot.* 65 (14), 3875–3887. doi:10.1093/jxb/eru132

Baránek, M., Čechová, J., Raddová, J., Holleinová, V., Ondrušíková, E., and Pidra, M. (2015). Dynamics and Reversibility of the DNA Methylation Landscape of Grapevine Plants (Vitis vinifera) Stressed by *In Vitro* Cultivation and Thermotherapy. *PLoS ONE* 10, e0126638. doi:10.1371/journal.pone.0126638

Barozai, M. Y. K., and Aziz, A. N. (2018). Recent Plant Growth and Stress Management Related Significant Advancements in Epigenetics. *Ann. Agrarian Sci.* 16 (4), 416–421. doi:10.1016/j.aasci.2018.07.002

Bartrina, I., Otto, E., Strnad, M., Werner, T., and Schmülling, T. (2011). Cytokinin Regulates the Activity of Reproductive Meristems, Flower Organ Size, Ovule Formation, and Thus Seed Yield inArabidopsis Thaliana. *The Plant cell* 23 (1), 69–80. doi:10.1105/tpc.110.079079

Bastow, R., Mylne, J. S., Lister, C., Lippman, Z., Martienssen, R. A., and Dean, C. (2004). Vernalization Requires Epigenetic Silencing of FLC by Histone Methylation. *Nature* 427 (6970), 164–167. doi:10.1038/nature02269

Baulcombe, D. C., and Dean, C. (2014). Epigenetic Regulation in Plant Responses to the Environment. *Cold Spring Harbor Perspect. Biol.* 6, a019471. doi:10.1101/cshperspectAcknowledgments.a019471

Berger, S. (2002). Jasmonate-related Mutants of Arabidopsis as Tools for Studying Stress Signaling. *Planta* 214 (4), 497–504. doi:10.1007/s00425-001-0688-y

Berr, A., McCallum, E. J., Alioua, A., Heintz, D., Heitz, T., and Shen, W.-H. (2010). Arabidopsis Histone Methyltransferase SET DOMAIN GROUP8 Mediates Induction of the Jasmonate/Ethylene Pathway Genes in Plant Defense Response to Necrotrophic Fungi. *Plant Physiol.* 154, 1403–1414. doi:10.1104/pp.110.161497

Blüumel, M., Dally, N., and Jung, C. (2015). Flowering Time Regulation in Crops—What Did We Learn from Arabidopsis? *Curr. Opin. Biotechnol.* 32, 121–129. doi:10.1016/j.copbio.2014.11.023

Bond, D. M., Dennis, E. S., Pogson, B. J., and Finnegan, E. J. (2009). Histone Acetylation, VERNALIZATION INSENSITIVE 3 , FLOWERING LOCUS C , and the Vernalization Response. *Mol. Plant* 2 (4), 724–737. doi:10.1093/mp/ssp021

Borrdorf, O., Arcuri, D., Richards, C. L., and Pigliucci, M. (2010). Experimental Alteration of DNA Methylation Affects the Phenotypic Plasticity of Ecologically Relevant Traits in *Arabidopsis thaliana*. *Evol. Ecol* 24, 541–553.

Borg, M., Jacob, Y., Susaki, D., LeBlanc, C., Buendía, D., Axelsson, E., et al. (2020). Targeted Reprogramming of H3K27me3 Resets Epigenetic Memory in Plant Paternal Chromatin. *Nature Cell Biology* 22, 1–9. doi:10.1038/s41556-020-0515-y

Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., et al. (2017). DNA Methylation Dynamics during Early Plant Life. *Genome Biol.* 18, 179. doi:10.1186/s13059-017-1313-0

Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1991). Genetic Interactions Among floral Homeotic Genes of Arabidopsis. *Development* 112, 1–20. doi:10.1242/dev.08397210.1242/dev.112.1.1

Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (2012). The ABC Model of Flower Development: Then and Now. *Development* 139, 4095–4098. doi:10.1242/dev.083972

Boyko, A., and Kovalchuk, I. (2008). Epigenetic Control of Plant Stress Response. *Environ. Mol. Mutagen.* 49 (1), 61–72. doi:10.1002/em.20347

Brownfield, L., Hafidh, S., Durbarry, A., Khatab, H., Sidorova, A., Doerner, P., et al. (2009). ArabidopsisDUO POLLEN3 Is a Key Regulator of Male Germline Development and Embryogenesis. *The Plant cell* 21 (7), 1940–1956. doi:10.1105/tpc.109.066373

Brukhin, V., and Albertini, M. (2021). Epigenetic Modification in Plant Development and Reproduction. *Epigenomes* 5, 25. doi:10.3390/epigenomes5040025

Bucher, E., Reinders, J., and Mirouze, M. (2012). Epigenetic Control of Transposon Transcription and Mobility in Arabidopsis. *Curr. Opin. Plant Biol.* 15, 503–510. doi:10.1016/j.pbi.2012.08.006

Burgarella, C., Chantret, N., Gay, L., Prosperi, J. M., Bonhomme, M., Tiffin, P., et al. (2016). Adaptation to Climate through Flowering Phenology: a Case Study in Medicago Truncatula. *Mol. Ecol.* 25 (14), 3397–3415. doi:10.1111/mec.13683

Burn, J. E., Bagnall, D. J., Metzger, J. D., Dennis, E. S., and Peacock, W. J. (1993). DNA Methylation, Vernalization, and the Initiation of Flowering. *Proc. Natl. Acad. Sci.* 90 (1), 287–291. doi:10.1073/pnas.90.1.287

Butenko, Y., and Ohad, N. (2011). Polycomb-group Mediated Epigenetic Mechanisms through Plant Evolution. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1809, 395–406. doi:10.1016/j.bbagrm.2011.05.013

Butterbrodt, T., Thurow, C., and Gatz, C. (2006). Chromatin Immunoprecipitation Analysis of the Tobacco PR-1a- and the Truncated CaMV 35S Promoter Reveals Differences in Salicylic Acid-dependent TGA Factor Binding and Histone Acetylation. *Plant Mol. Biol.* 61 (4-5), 665–674. doi:10.1007/s11103-006-0039-2

Calarco, J. P., Borges, F., Donoghue, M. T. A., Van Ex, F., Jullien, P. E., Lopes, T., et al. (2012). Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. *Cell* 151, 194–205. doi:10.1016/j.cell.2012.09.001

Campos-Rivero, G., Osorio-Montalvo, P., Sánchez-Borges, R., Us-Camas, R., Duarte-Aké, F., and De-la-Peña, C. (2017). Plant Hormone Signaling in Flowering: An Epigenetic point of View. *J. Plant Physiol.* 214, 16–27. doi:10.1016/j.jplph.2017.03.018

Candaele, J., Demuynck, K., Mosoti, D., Beemster, G. T. S., Inzé, D., Nelissen, H., et al. (2014). Differential Methylation during maize Leaf Growth Targets Developmentally Regulated Genes. *Plant Physiol.* 164, 1350–1364. doi:10.1104/pp.113.233312

Capy, P., Gasperi, G., Biémont, C., and Bazin, C. (2000). Stress and Transposable Elements: Co-evolution or Useful Parasites? *Heredity* 85, 101–106. doi:10.1046/j.1365-2540.2000.00751.x

Carpenter, R., Martin, C., and Coen, E. S. (1987). Comparison of Genetic Behaviour of the Transposable Element Tam3 at Two Unlinked Pigment Loci in Antirrhinum majus. *Mol. Gen. Genet.* 207, 82–89. doi:10.1007/bf00331494

Castiglione, M. R., Cremonini, R., and Frediani, M. (2002). DNA Methylation Patterns on Plant Chromosomes. *Caryologia* 55 (4), 275–282. doi:10.1080/00087114.2002.10797876

Castiglione, M. R., Cremonini, R., and Frediani, M. (2002). DNA Methylation Patterns on Plant Chromosomes. *Caryologia* 55, 275–282. doi:10.1080/00087114.2002.10797876

Cavrak, V. V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L. M., Mittelsten Scheid, O., et al. (2014). How a Retrotransposon Exploits the Plant's Heat Stress Response for its Activation. *Plos Genet.* 10 (1), e1004115. doi:10.1371/journal.pgen.1004115

Chen, Q., Chen, X., Wang, Q., Zhang, F., Lou, Z., Zhang, Q., et al. (2013). Structural Basis of a Histone H3 Lysine 4 Demethylase Required for Stem Elongation in rice. *Plos Genet.* 9, e1003239. doi:10.1371/journal.pgen.1003239

Chen, T., and Li, E. (2004). Structure and Function of Eukaryotic DNA Methyltransferases. *Curr. Top. Dev Biol* 60, 55–89. doi:10.1016/S0070-2153(04)60003-2

Cheng, Y.-H., Peng, X.-Y., Yu, Y.-C., Sun, Z.-Y., and Han, L. (2019). The Effects of DNA Methylation Inhibition on Flower Development in the Dioecious Plant Salix Viminalis. *Forests* 10, 173. doi:10.3390/f10020173

Choi, J., Hyun, Y., Kang, M.-J., In Yun, H., Yun, J.-Y., Lister, C., et al. (2009). Resetting and Regulation of Flowering Locus Cexpression during Arabidopsis Reproductive Development. *The Plant J. Cell Mol. Biol.* 57 (5), 918–931. doi:10.1111/j.1365-313x.2008.03776.x

Choi, K., Zhao, X., Tock, A. J., Lambing, C., Underwood, C. J., Hardcastle, T. J., et al. (2018). Nucleosomes and DNA Methylation Shape Meiotic DSB Frequency in *Arabidopsis thaliana* Transposons and Gene Regulatory Regions. *Genome Res.* 28, 532–546. doi:10.1101/gr.225599.117

Chu, C., and Lee, T. M. (1989). The Relationship between Ethylene Biosynthesis and Chilling Tolerance in Seedlings of rice (Oryza Sativa). *Bot. Bull. Acad. Sin* 30, 263–273.

Cimini, D., Mattiuzzo, M., Torosantucci, L., and Degrassi, F. (2003). Histone Hyperacetylation in Mitosis Prevents Sister Chromatid Separation and Produces Chromosome Segregation Defects. *MBoC* 14, 3821–3833. doi:10.1091/mbc.e03-01-0860

Covarrubias, A. A., and Reyes, J. L. (2010). Post-transcriptional Gene Regulation of Salinity and Drought Responses by Plant microRNAs. *Plant, cell envi* 33 (4), 481–489. doi:10.1111/j.1365-3040.2009.02048.x

Creasey, K. M., Zhai, J., Borges, F., van Ex, F., Regulski, M., Meyers, B. C., et al. (2014). miRNAs Trigger Widespread Epigenetically Activated siRNAs from Transposons in Arabidopsis. *Nature* 508, 411–415. doi:10.1038/nature13069

Creelman, R. A., and Mullet, J. E. (1997). Biosynthesis and Action of Jasmonates in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48, 355–381. doi:10.1146/annurev.arplant.48.1.355

Creelman, R. A., and Mullet, J. E. (1995). Jasmonic Acid Distribution and Action in Plants: Regulation during Development and Response to Biotic and Abiotic Stress. *Proc. Natl. Acad. Sci.* 92 (10), 4114–4119. doi:10.1073/pnas.92.10.4114

Cubas, P., Vincent, C., and Coen, E. (1999). An Epigenetic Mutation Responsible for Natural Variation in floral Symmetry. *Nature* 401 (6749), 157–161. doi:10.1038/43657

Cui, X., Jin, P., Cui, X., Gu, L., Lu, Z., Xue, Y., et al. (2013). Control of Transposon Activity by a Histone H3K4 Demethylase in rice. *Proc. Natl. Acad. Sci.* 110, 1953–1958. doi:10.1073/pnas.1217020110

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that marks Chromosomal Polycomb Sites. *Cell* 111, 185–196. doi:10.1016/s0092-8674(02)00975-3

Daccord, N., Celton, J. M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., et al. (2017). High- Quality De Novo Assembly of the Apple Genome and

Methylome Dynamics of Early Fruit Development. *Nat. Genet.* 49, 1099–1106. doi:10.1038/ng.3886

De Lucia, F., Crevillen, P., Jones, A. M. E., Greb, T., and Dean, C. (2008). A PHD-Polycomb Repressive Complex 2 Triggers the Epigenetic Silencing of FLC during Vernalization. *Proc. Natl. Acad. Sci.* 105 (44), 16831–16836. doi:10.1073/pnas.0808687105

Deal, R. B., Topp, C. N., McKinney, E. C., and Meagher, R. B. (2007). Repression of Flowering inArabidopsisRequires Activation ofFLOWERING LOCUS CExpression by the Histone Variant H2A.Z. *Z. Plant Cell.* 19, 74–83. doi:10.1105/tpc.106.048447

Domagalska, M. A., Sarnowska, E., Nagy, F., and Davis, S. J. (2010). Genetic Analyses of Interactions Among Gibberellin, Abscisic Acid, and Brassinosteroids in the Control of Flowering Time in *Arabidopsis thaliana*. *PloS one* 5, e14012. doi:10.1371/journal.pone.0014012

Dowen, R. H., Pelizzola, M., Schmitz, R. J., Lister, R., Dowen, J. M., Nery, J. R., et al. (2012). Widespread Dynamic DNA Methylation in Response to Biotic Stress. *Proc. Natl. Acad. Sci.* 109 (32), E2183–E2191. doi:10.1073/pnas.1209329109

Dziegielewski, W., and Ziolkowski, P. A. (2021). License to Regulate: Noncoding RNA Special Agents in Plant Meiosis and Reproduction. *Front. Plant Sci.* 12, 662185. doi:10.3389/fpls.2021.662185

Eeuwens, C. J., Lord, S., Donough, C. R., Rao, V., Vallejo, G., and Nelson, S. (2002). Effects of Tissue Culture Conditions during Embryoid Multiplication on the Incidence of ``mantled'' Flowering in Clonally Propagated Oil palm. *Plant Cell, Tissue Organ. Cult.* 70, 311–323. doi:10.1023/a:1016543921508

El-Sharkawy, I., Liang, D., and Xu, K. (2015). Transcriptome Analysis of an Apple (Malus×domestica) Yellow Fruit Somatic Mutation Identifies a Gene Network Module Highly Associated with Anthocyanin and Epigenetic Regulation. *Exbotj* 66, 7359–7376. doi:10.1093/jxb/erv433

Engler, P., Weng, A., and Storb, U. (1993). Influence of CpG Methylation and Target Spacing on V(D)J Recombination in a Transgenic Substrate. *Mol. Cell Biol* 13 (1), 571–577. doi:10.1128/mcb.13.1.571-577.1993

Farmer, E. E., Alméras, E., and Krishnamurthy, V. (2003). Jasmonates and Related Oxylipins in Plant Responses to Pathogenesis and Herbivory. *Curr. Opin. Plant Biol.* 6 (4), 372–378. doi:10.1016/s1369-5266(03)00045-1

Fedoroff, N. V. (2012). Transposable Elements, Epigenetics, and Genome Evolution. *Science* 338, 758–767. doi:10.1126/science.338.6108.758

Feng, S., Jacobsen, S. E., and Reik, W. (2010). Epigenetic Reprogramming in Plant and Animal Development. *Science* 330 (6004), 622–627. doi:10.1126/science.1190614

Forestan, C., Farinati, S., Zambelli, F., Pavesi, G., Rossi, V., and Varotto, S. (2019). Epigenetic Signatures of Stress Adaptation and Flowering Regulation in Response to Extended Drought and Recovery in Zea mays. *Plant Cell Environ* 43, 55–75. doi:10.1111/pce.13660

Fu, C., Sunkar, R., Zhou, C., Shen, H., Zhang, J. Y., Matts, J., et al. (2012). Overexpression of miR156 in Switchgrass (Panicum Virgatum L.) Results in Various Morphological Alterations and Leads to Improved Biomass Production. *Plant Biotechnol. J.* 10, 443–452. doi:10.1111/j.1467-7652.2011.00677.x

Gao, L., Geng, Y., Li, B., Chen, J., and Yang, J. (2010). Genome-wide DNA Methylation Alterations of Alternanthera Philoxeroides in Natural and Manipulated Habitats: Implications for Epigenetic Regulation of Rapid Responses to Environmental Fluctuation and Phenotypic Variation. *Plant Cell Environ.* 33, 1820–1827. doi:10.1111/j.1365-3040.2010.02186.x

Gao, M.-J., Schäfer, U. A., Parkin, I. A. P., Hegedus, D. D., Lydiate, D. J., and Hannoufa, A. (2003). A Novel Protein fromBrassica Napushas a Putative KID Domain and Responds to Low Temperature. *Plant J.* 33 (6), 1073–1086. doi:10.1046/j.1365-313x.2003.01694.x

Gehring, M., Bubb, K. L., and Henikoff, S. (2009). Extensive Demethylation of Repetitive Elements during Seed Development Underlies Gene Imprinting. *Science* 324, 1447–1451. doi:10.1126/science.1171609

Geng, S., Kong, X., Song, G., Jia, M., Guan, J., Wang, F., et al. (2019). DNA Methylation Dynamics during the Interaction of Wheat Progenitor Aegilops Tauschii with the Obligate Biotrophic Fungus Blumeria Graminis F. Sp. Tritici. *New Phytol.* 221, 1023–1035. doi:10.1111/nph.15432

Ghoshal, B., and Gardiner, J. (2021). "CRISPR-dCas9-Based Targeted Manipulation of DNA Methylation in Plants," in *CRISPR-cas Methods*. Editors M. T. Islam, and K. A. Molla (Humana, New York: Springer Protocols Handbooks). doi:10.1007/978-1-0716-1657-4_5

Gourcilleau, D., Mousset, M., Latutrie, M., Marin, S., Delaunay, A., Maury, S., et al. (2019). Assessing Global DNA Methylation Changes Associated with Plasticity in Seven Highly Inbred Lines of Snapdragon Plants (Antirrhinum majus). *Genes* 10, 256. doi:10.3390/genes10040256

Grandbastien, M. (1998). Activation of Plant Retrotransposons under Stress Conditions. *Trends Plant Sci.* 3, 181–187. doi:10.1016/s1360-1385(98)01232-1

Gratani, L. (2014). Plant Phenotypic Plasticity in Response to Environmental Factors. *Adv. Bot.* 2014, 1–17. doi:10.1155/2014/208747

Haag, J. R., Ream, T. S., Marasco, M., Nicora, C. D., Norbeck, A. D., Pasa-Tolic, L., et al. (2012). *In Vitro* Transcription Activities of Pol IV, Pol V, and RDR2 Reveal Coupling of Pol IV and RDR2 for dsRNA Synthesis in Plant RNA Silencing. *Mol. Cell* 48, 811–818. doi:10.1016/j.molcel.2012.09.027

Han, M.-H., Goud, S., Song, L., and Fedoroff, N. (2004). The A*rabidopsis* Double-Stranded RNA-Binding Protein HYL1 Plays a Role in microRNA-Mediated Gene Regulation. *Proc. Natl. Acad. Sci.* 101 (4), 1093–1098. doi:10.1073/pnas.0307969100

Han, M. H., Goud, S., Song, L., and Fedoroff, N. (2019). The Arabidopsis Double-Stranded RNA-Binding Protein HYL1 Plays a Role in microRNA-Mediated Gene Regulation. *Proc. Natl. Acad. Sci. U S A.* 101, 1093–1098. doi:10.1073/pnas.0307969100

Harrison, B. J., and Fincham, J. R. S. (1964). Instability at the Pal Locus in Antirrhinum majus. *Heredity* 19, 237–258. doi:10.1038/hdy.1964.28

Hashida, S.-n., Kitamura, K., Mikami, T., and Kishima, Y. (2003). Temperature Shift Coordinately Changes the Activity and the Methylation State of Transposon Tam3 in Antirrhinum majus. *Plant Physiol.* 132 (3), 1207–1216. doi:10.1104/pp.102.017533

Hauben, M., Haesendonckx, B., Standaert, E., Van Der Kelen, K., Azmi, A., Akpo, H., et al. (200920109–20114). Energy Use Efficiency Is Characterized by an Epigenetic Component that Can Be Directed through Artificial Selection to Increase Yield. *Proc. Natl. Acad. Sci. U S A.* 106, 20109–20114. doi:10.1073/pnas.0908755106

Hauser, F., Waadt, R., and Schroeder, J. I. (2011). Evolution of Abscisic Acid Synthesis and Signaling Mechanisms. *Curr. Biol.* 21 (9), R346–R355. doi:10.1016/j.cub.2011.03.015

He, X.-J., Chen, T., and Zhu, J.-K. (2011). Regulation and Function of DNA Methylation in Plants and Animals. *Cell Res* 21, 442–465. doi:10.1038/cr.2011.23

He, Y. (2009). Control of the Transition to Flowering by Chromatin Modifications. *Mol. Plant* 2, 554–564. doi:10.1093/mp/ssp005

He, Y., Michaels, S. D., and Amasino, R. M. (2003). Regulation of Flowering Time by Histone Acetylation in Arabidopsis. *Science* 302 (5651), 1751–1754. doi:10.1126/science.1091109

Herrera, C. M., and Bazaga, P. (2012). Epigenetic Correlates of Plant Phenotypic Plasticity: DNA Methylation Differs between Prickly and Nonprickly Leaves in heterophyllousIlex aquifolium(Aquifoliaceae) Trees. *Bot. J. Linn. Soc.* 171, 441–452. doi:10.1111/boj.12007

Higo, H., Tahir, M., Takashima, K., Miura, A., Watanabe, K., Tagiri, A., et al. (2012). DDM1 (Decrease in DNA Methylation) Genes in rice (Oryza Sativa). *Mol. Genet. Genomics* 287, 785–792. doi:10.1007/s00438-012-0717-5

Hirayama, T., and Shinozaki, K. (2010). Research on Plant Abiotic Stress Responses in the post-genome Era: Past, Present and Future. *Plant J. Cell Mol. Biol.* 61 (6), 1041–1052. doi:10.1111/j.1365-313X.2010.04124.x

Horvath, D. P., Sung, S., Kim, D., Chao, W., and Anderson, J. (2010). Characterization, Expression and Function of DORMANCY ASSOCIATED MADS-BOX Genes from Leafy Spurge. *Plant Mol. Biol.* 73, 169–179. doi:10.1007/s11103-009-9596-5

Howell, M. D., Fahlgren, N., Chapman, E. J., Cumbie, J. S., Sullivan, C. M., Givan, S. A., et al. (2007). Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway inArabidopsisReveals Dependency on miRNA- and tasiRNA-Directed Targeting. *Plant Cell.* 19, 926–942. doi:10.1105/tpc.107.050062

Hsieh, D., Xia, Y., Wray, L., Qian, D., Pal, A., Dil, J. H., et al. (2009). Observation of Unconventional Quantum Spin Textures in Topological Insulators. *Science* 323, 919–922. doi:10.1126/science.1167733

Hsieh, T.-F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., et al. (2009). Genome-Wide Demethylation of Arabidopsis Endosperm. *Science* 324, 1451–1454. doi:10.1126/science.1172417

Hugouvieux, V., Kwak, J. M., and Schroeder, J. I. (2001). An mRNA Cap Binding Protein, ABH1, Modulates Early Abscisic Acid Signal Transduction in Arabidopsis. *Cell* 106 (4), 477–487. doi:10.1016/s0092-8674(01)00460-3

Ibarra, C. A., Feng, X., Schoft, V. K., Hsieh, T.-F., Uzawa, R., Rodrigues, J. A., et al. (2012). Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. *Science* 337 (6100), 1360–1364. doi:10.1126/science.1224839

Ietswaart, R., Wu, Z., and Dean, C. (2012). Flowering Time Control: Another Window to the Connection between Antisense RNA and Chromatin. *Trends Genet.* 28 (9), 445–453. doi:10.1016/j.tig.2012.06.002

Ingouff, M., Selles, B., Michaud, C., Vu, T. M., Berger, F., Schorn, A. J., et al. (2017). Live-cell Analysis of DNA Methylation during Sexual Reproduction in Arabidopsis Reveals Context and Sex-specific Dynamics Controlled by Noncanonical RdDM. *Genes Dev.* 31, 72–83. doi:10.1101/gad.289397.116

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA Pathway Prevents Transgenerational Retrotransposition in Plants Subjected to Stress. *Nature* 472 (7341), 115–119. doi:10.1038/nature09861

Ito, H. (2013). Small RNAs and Regulation of Transposons in Plants. *Genes Genet. Syst.* 88 (1), 3–7. doi:10.1266/ggs.88.3

Iwasaki, M., and Paszkowski, J. (2014). Identification of Genes Preventing Transgenerational Transmission of Stress-Induced Epigenetic States. *Proc. Natl. Acad. Sci.* 111 (23), 8547–8552. doi:10.1073/pnas.1402275111

Jacobsen, S. E., and Meyerowitz, E. M. (1997). Hypermethylated SUPERMAN Epigenetic Alleles in Arabidopsis. *Science* 277 (5329), 1100–1103. doi:10.1126/science.277.5329.1100

Jaligot, E., Adler, S., Debladis, É., Beulé, T., Richaud, F., Ilbert, P., et al. (2011). Epigenetic Imbalance and the floral Developmental Abnormality of the In Vitro-regenerated Oil palm Elaeis Guineensis. *Ann. Bot.* 108 (8), 1453–1462. doi:10.1093/aob/mcq266

Jaligot, E., Rival, A., Beulé, T., Dussert, S., and Verdeil, J.-L. (2000). Somaclonal Variation in Oil palm (Elaeis Guineensis Jacq.): the DNA Methylation Hypothesis. *Plant Cell Rep.* 19, 684–690. doi:10.1007/s002999900177

Jenuwein, T., and Allis, C. D. (2001). Translating the Histone Code. *Science* 293, 1074–1080. doi:10.1126/science.1063127

Jiang, D., Gu, X., and He, Y. (2009). Establishment of the Winter-Annual Growth Habit via FRIGIDA-Mediated Histone Methylation at FLOWERING LOCUS C in Arabidopsis. *The Plant cell* 21 (6), 1733–1746. doi:10.1105/tpc.109.067967

Jiang, D., Wang, Y., Wang, Y., and He, Y. (2008). Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb Repressive Complex 2 Components. *PLoS one* 3 (10), e3404. doi:10.1371/journal.pone.0003404

Jin, J. B., Jin, Y. H., Lee, J., Miura, K., Yoo, C. Y., Kim, W. Y., et al. (2008). The SUMO E3 Ligase, AtSIZ1, Regulates Flowering by Controlling a Salicylic Acid-Mediated floral Promotion Pathway and through Affects on FLC Chromatin Structure. *Plant J.* 53 (3), 530–540. doi:10.1111/j.1365-313X.2007.03359.x

Johns, M. A., Mottinger, J., and Freeling, M. (1985). A Low Copy Number, Copia-like Transposon in maize. *EMBO J.* 4 (5), 1093–1101. doi:10.1002/j.1460-2075.1985.tb03745.x

Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., et al. (2003). Arabidopsis MET1 Cytosine Methyltransferase Mutants. *Genetics* 163 (3), 1109–1122. doi:10.1093/genetics/163.3.1109

Kawakatsu, T., Nery, J. R., Castanon, R., and Ecker, J. R. (2017). Dynamic DNA Methylation Reconfiguration during Seed Development and Germination. *Genome Biol.* 18, 171. doi:10.1186/s13059-017-1251-x

Kellenberger, R. T., Schlüter, P. M., and Schiestl, F. P. (2016). Herbivore-induced DNA Demethylation Changes floral Signalling and Attractiveness to Pollinators in *Brassica Rapa*. *PLoS ONE* 11, e0166646. doi:10.1371/journal.pone.0166646

Kim, J.-M., To, T. K., Nishioka, T., and Seki, M. (2010). Chromatin Regulation Functions in Plant Abiotic Stress Responses. *Plant Cell Environ.* 33, 604–611. doi:10.1111/j.1365-3040.2009.02076.x

Kim, M.-S., Kondo, T., Takada, I., Youn, M.-Y., Yamamoto, Y., Takahashi, S., et al. (2009). DNA Demethylation in Hormone-Induced Transcriptional Derepression. *Nature* 461, 1007–1012. doi:10.1038/nature08456

Kinoshita, A., and Richter, R. (2020). Genetic and Molecular Basis of floral Induction in *Arabidopsis thaliana*. *J. Exp. Bot.* 71, 2490–2504.

Klee, H. J., and Giovannoni, J. J. (2011). Genetics and Control of Tomato Fruit Ripening and Quality Attributes. 942. *Annu. Rev. Genet.* 45, 41–59. doi:10.1093/jxb/eraa057

Krogan, N. T., Hogan, K., and Long, J. A. (2012). APETALA2 Negatively Regulates Multiple floral Organ Identity Genes inArabidopsisby Recruiting the Co-repressor TOPLESS and the Histone Deacetylase HDA19. *Development* 139 (22), 4180–4190. doi:10.1242/dev.085407

Kumar, S., and Mohapatra, T. (2021). Dynamics of DNA Methylation and its Functions in Plant Growth and Development. *Front. Plant Sci.* 12. doi:10.3389/fpls.2021.596236

Kundariya, H., Yang, X., Morton, K., Sanchez, R., Axtell, M. J., Hutton, S. F., et al. (2020). MSH1-induced Heritable Enhanced Growth Vigor through Grafting Is Associated with the RdDM Pathway in Plants. *Nat. Commun.* 11, 5343–5414. doi:10.1038/s41467-020-19140-x

Kwak, K. J., Kim, Y. O., and Kang, H. (2005). Characterization of Transgenic Arabidopsis Plants Overexpressing GR-RBP4 under High Salinity, Dehydration, or Cold Stress. *J. Exp. Bot.* 56, 3007–3016. doi:10.1093/jxb/eri298

Kwon, C. S., Lee, D., Choi, G., and Chung, W.-I. (2009). Histone Occupancy-dependent and -independent Removal of H3K27 Trimethylation at Cold-Responsive Genes in Arabidopsis. *Plant J. : Cel. Mol. Biol.* 60 (1), 112–121. doi:10.1111/j.1365-313x.2009.03938.x

Lang, Z., Wang, Y., Tang, K., Tang, D., Datsenka, T., Cheng, J., et al. (2017). Critical Roles of DNA Demethylation in the Activation of Ripening-Induced Genes and Inhibition of Ripening-Repressed Genes in Tomato Fruit. *Proc. Natl. Acad. Sci. USA* 114, E4511–E4519. doi:10.1073/pnas.1705233114

Lee, B.-h., Henderson, D. A., and Zhu, J.-K. (2005). TheArabidopsisCold-Responsive Transcriptome and its Regulation by ICE1. *The plant cell online* 17 (11), 3155–3175. doi:10.1105/tpc.105.035568

Lermontova, I., Sandmann, M., Mascher, M., Schmit, A.-C., and Chabouté, M.-E. (2015). Centromeric Chromatin and its Dynamics in Plants. *Plant J.* 83, 4–17. doi:10.1111/tpj.12875

Li, X., Bian, H., Song, D., Ma, S., Han, N., Wang, J., et al. (2013). Flowering Time Control in Ornamental Gloxinia (Sinningia Speciosa) by Manipulation of miR159 Expression. *Ann. Bot.* 111, 791–799. doi:10.1093/aob/mct034

Li, Y., Butenko, Y., and Grafi, G. (2005). Histone Deacetylation Is Required for Progression through Mitosis in Tobacco Cells. *Plant J.* 41, 346–352. doi:10.1111/j.1365-313X.2004.02301.x

Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S., et al. (2001). Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. *Science* 292 (5524), 2077–2080. doi:10.1126/science.1059745

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M. W., Dedhia, N., Richard McCombie, W., et al. (2004). Role of Transposable Elements in Heterochromatin and Epigenetic Control. *Nature* 430 (6998), 471–476. doi:10.1038/nature02651

Liu, C., Lu, F., Cui, X., and Cao, X. (2010). Histone Methylation in Higher Plants. *Annu. Rev. Plant Biol.* 61, 395–420. doi:10.1146/annurev.arplant.043008.091939

Liu, G., Li, W., Zheng, P., Xu, T., Chen, L., Liu, D., et al. (2012). Transcriptomic Analysis of 'Suli' Pear (Pyrus Pyrifolia white Pear Group) Buds during the Dormancy by RNA-Seq. *BMC Genomics* 13, 700. doi:10.1186/1471-2164-13-700

Liu, R., How-Kit, , A., Stammitti, L., Teyssier, E., Rolin, D., Mortain-Bertrand, A., et al. (2015). A DEMETER-like DNA Demethylase Governs Tomato Fruit Ripening. *Proc. Natl. Acad. Sci. USA* 112, 10804–10809. doi:10.1073/pnas.1503362112

López, A., Ramírez, V., García-Andrade, J., Flors, V., and Vera, P. (2011). The RNA Silencing Enzyme RNA Polymerase V Is Required for Plant Immunity. *Plos Genet.* 7, e1002434. doi:10.1371/journal.pgen.1002434

López Sánchez, A., Stassen, J. H., Furci, L., Smith, L. M., and Ton, J. (2020). The role of DNA (de)methylation in immune responsiveness of *Arabidopsis*. *Plant J.* 88, 361–374. doi:10.1111/tpj.13252

López Sánchez, A., Stassen, J. H. M., Furci, L., Smith, L. M., and Ton, J. (2016). The role of DNA (de)methylation in immune responsiveness of *Arabidopsis*. *Plant J.* 88, 361–374. doi:10.1111/tpj.13252

Loyola-Vargas, V., Ruíz-May, E., Galaz-Ávalos, R., and De-la-Peña, C. (2012). The Role of Jasmonic Acid in Root Mitochondria Disruption. *Plant Signaling Behav.* 7 (6), 611–614. doi:10.4161/psb.20070

Lloret, A., Quesada-Traver, C., and Conejero, A. (2021). Regulatory Circuits Involving Bud Dormancy Factor *PpeDAM6*. *Hortic. Res.* 8 261. doi:10.1038/s41438-021-00706-9

Lu, C., and Fedoroff, N. (2000). A Mutation in the Arabidopsis HYL1 Gene Encoding a dsRNA Binding Protein Affects Responses to Abscisic Acid, Auxin, and Cytokinin. *The Plant Cell* 12, 2351–2366. doi:10.2307/3871234

Luo, Y., Guo, Z., and Li, L. (2013). Evolutionary Conservation of microRNA Regulatory Programs in Plant Flower Development. *Developmental Biol.* 380, 133–144. doi:10.1016/j.ydbio.2013.05.009

Ma, N., Xue, J., Li, Y., Liu, X., Dai, F., Jia, W., et al. (2008). Rh-PIP2;1, a Rose Aquaporin Gene, Is Involved in Ethylene-Regulated Petal Expansion. *Plant Physiol.* 148 (2), 894–907. doi:10.1104/pp.108.120154

Madhusudhan, J. (2015). Importance of Epigenetic in Plants. *Bioengineer Biomed. Sci.* 5.

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., et al. (2006). A Naturally Occurring Epigenetic Mutation in a Gene Encoding an SBP-Box Transcription Factor Inhibits Tomato Fruit Ripening. *Nat. Genet.* 38, 948–952. doi:10.1038/ng1841

Marfil, C. F., Camadro, E. L., and Masuelli, R. W. (2009). Phenotypic Instability and Epigenetic Variability in a Diploid Potato of Hybrid Origin, Solanum Ruiz-Lealii. *BMC Plant Biol.* 9, 21. doi:10.1186/1471-2229-9-21

Martínez, C., Pons, E., Prats, G., and León, J. (2004). Salicylic Acid Regulates Flowering Time and Links Defence Responses and Reproductive Development. *Plant J.* 37, 209–217. doi:10.1046/j.1365-313x.2003.01954.x

Martinez-Pérez, M., Aparicio, F., López-Gresa, M. P., Bellés, J. M., Sánchez-Navarro, J. A., and Pallás, V. (2017). Arabidopsis m6A Demethylase Activity Modulates Viral Infection of a Plant Virus and the m6A Abundance in its Genomic RNAs. *Proc. Natl. Acad Sci.* 114, 10755–10760.

Matsoukas, I. G. (2014). Interplay between Sugar and Hormone Signaling Pathways Modulate floral Signal Transduction. *Front. Genet.* 5, 218. doi:10.3389/fgene.2014.00218

Matzke, M. A., Matzke, A. J. M., Pruss, G. J., and Vance, V. B. (2001). RNA-based Silencing Strategies in Plants. *Curr. Opin. Genet. Development* 11 (2), 221–227. doi:10.1016/s0959-437x(00)00183-0

Matzke, M. A., and Mosher, R. A. (2014). RNA-directed DNA Methylation: an Epigenetic Pathway of Increasing Complexity. *Nat. Rev. Genet.* 15 (6), 394–408. doi:10.1038/nrg3683

Matzke, M., Kanno, T., Huettel, B., Daxinger, L., and Matzke, A. J. (2007). Targets of RNA-Directed DNA Methylation. *Current Opinion in Plant Biology* 10 (5), 512–519. doi:10.1016/j.pbi.2007.06.007

Mazzitelli, L., Hancock, R. D., Haupt, S., Walker, P. G., Pont, S. D. A., McNicol, J., et al. (2007). Co-ordinated Gene Expression during Phases of Dormancy Release in Raspberry (Rubus Idaeus L.) Buds. *J. Exp. Bot.* 58, 1035–1045. doi:10.1093/jxb/erl266

McClintock, B. (1984). The Significance of Responses of the Genome to challenge. *Science* 226 (4676), 792–801. doi:10.1126/science.15739260

Meijón, M., Feito, I., Valledor, L., Rodríguez, R., and Cañal, M. J. (2011). Promotion of Flowering in Azaleas by Manipulating Photoperiod and Temperature Induces Epigenetic Alterations during floral Transition. *Physiologia plantarum* 143, 82–92. doi:10.1111/j.1399-3054.2011.01485.x

Mirouze, M., Lieberman-Lazarovich, M., Aversano, R., Bucher, E., Nicolet, J., Reinders, J., et al. (2012). Loss of DNA Methylation Affects the Recombination Landscape in Arabidopsis. *Proc. Natl. Acad. Sci.* 109, 5880–5885. doi:10.1073/pnas.1120841109

Mirouze, M., and Paszkowski, J. (2011). Epigenetic Contribution to Stress Adaptation in Plants. *Curr. Opin. Plant Biol.* 14, 267–274. doi:10.1016/j.pbi.2011.03.004

Mondal, P., Natesh, J., Penta, D., and Meeran, S. M. (2020). Progress and Promises of Epigenetic Drugs and Epigenetic Diets in Cancer Prevention and Therapy: A Clinical Update. Semin. *Cancer Biol.*, S1044. doi:10.1016/j.semcancer.2020.12.006

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C. N., Richardson, A. O., et al. (2009). Unexpected Consequences of a Sudden and Massive Transposon Amplification on rice Gene Expression. *Nature* 461, 1130–1134. doi:10.1038/nature08479

Niederhuth, C. E., and Schmitz, R. J. (2017). Putting DNA Methylation in Context: from Genomes to Gene Expression in Plants. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1860 (1), 149–156. doi:10.1016/j.bbagrm.2016.08.009

Okada, K., Ueda, J., Komaki, M. K., Bell, C. J., and Shimura, Y. (1991). Requirement of the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation. *Plant Cell* 3 (7), 677–684. doi:10.1105/tpc.3.7.677

Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vázquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., et al. (2010). Control of Female Gamete Formation by a Small RNA Pathway in Arabidopsis. *Nature* 464, 628–632. doi:10.1038/nature08828

Overmyer, K., Tuominen, H., Kettunen, R., Betz, C., Langebartels, C., Sandermann, H., Jr, et al. (2000). Ozone-sensitive Arabidopsis Rcd1 Mutant Reveals Opposite Roles for Ethylene and Jasmonate Signaling Pathways in Regulating Superoxide-dependent Cell Death. *The Plant cell* 12 (10), 1849–1862. doi:10.2307/3871197

Papareddy, R. K., Páldi, K., Paulraj, S., Kao, P., Lutzmayer, S., and Nodine, M. D. (2020). Chromatin Regulates Expression of Small RNAs to Help Maintain Transposon Methylome Homeostasis in Arabidopsis. *Genome Biol.* 21, 251. doi:10.1186/s13059-020-02163-4

Papp, I. n., Mur, L., Dalmadi, g., Dulai, S. n., and Koncz, C. (2004). A Mutation in the Cap Binding Protein 20 Gene Confers Drought. *Plant Mol. Biol.* 55, 679–686. doi:10.1007/s11103-004-1680-2

Park, K., Kim, M. Y., Vickers, M., Park, J.-S., Hyun, Y., Okamoto, T., et al. (2016). DNA Demethylation Is Initiated in the central Cells of Arabidopsis and rice. *Proc. Natl. Acad. Sci. USA* 113, 15138–15143. doi:10.1073/pnas.1619047114

Parker, A. H., Wilkinson, S. W., and Ton, J. (2021). Epigenetics: a Catalyst of Plant Immunity against Pathogens. *New Phytol.*, 2338–2429. doi:10.1111/nph.17699

Pavangadkar, K., Thomashow, M. F., and Triezenberg, S. J. (2010). Histone Dynamics and Roles of Histone Acetyltransferases during Cold-Induced Gene Regulation in Arabidopsis. *Plant Mol. Biol.* 74, 183–200. doi:10.1007/s11103-010-9665-9

Pei, H., Ma, N., Chen, J., Zheng, Y., Tian, J., Li, J., et al. (2013). Integrative Analysis of miRNA and mRNA Profiles in Response to Ethylene in Rose Petals during Flower Opening. *PLoS ONE* 8 (5), e64290. doi:10.1371/journal.pone.0064290

Pérez-Hormaeche, J., Potet, F., Beauclair, L., Le Masson, I., Courtial, B., Bouché, N., et al. (2008). Invasion of the Arabidopsis Genome by the Tobacco Retrotransposon Tnt1 Is Controlled by Reversible Transcriptional Gene Silencing. *Plant Physiol.* 147 (3), 1264–1278. doi:10.1104/pp.108.117846

Pfluger, J., and Wagner, D. (2007). Histone Modifications and Dynamic Regulation of Genome Accessibility in Plants. *Curr. Opin. Plant Biol.* 10 (6), 645–652. doi:10.1016/j.pbi.2007.07.013

Pikaard, C. S. (2006). Cell Biology of the Arabidopsis Nuclear siRNA Pathway for RNA-Directed Chromatin Modification. *Cold Spring Harbor symposia quantitative Biol.* 71, 473–480. doi:10.1101/sqb.2006.71.046

Quadrana, L., Almeida, J., Asís, R., Duffy, T., Dominguez, P. G., Bermúdez, L., et al. (2014). Natural Occurring Epialleles Determine Vitamin E Accumulation in Tomato Fruits. *Nat. Commun.* 5, 4027. doi:10.1038/ncomms5027

Ramirez-Prado, J. S., Piquerez, S. J. M., Bendahmane, A., Hirt, H., Raynaud, C., and Benhamed, M. (2018). Modify the Histone to Win the Battle: Chromatin Dynamics in Plant–Pathogen Interactions. *Front. Plant Sci.* 9–355. doi:10.3389/fpls.2018.00355

Rasia, R. M., Mateos, J., Bologna, N. G., Burdisso, P., Imbert, L., Palatnik, J. F., et al. (2010). Structure and RNA Interactions of the Plant MicroRNA Processing-Associated Protein HYL1. *Biochemistry* 49, 8237–8239. doi:10.1021/bi100672x

Reinhardt, D., Mandel, T., and Kuhlemeier, C. (2000). Auxin Regulates the Initiation and Radial Position of Plant Lateral Organs. *Plant Cell* 12 (4), 507–518. doi:10.1105/tpc.12.4.507

Regulski, M., Lu, Z., Kendall, J., Donoghue, M. T., Reinders, J., Llaca, V., et al. (2013). The Maize Methylome Influences mRNA Splice Sites and Reveals Widespread Paramutation-Like Switches Guided by Small RNA. *Genome Res.* doi:10.1101/gr.153510.112

Ríos, G., Leida, C., Conejero, A., and Badenes, M. L. (2014). Epigenetic Regulation of Bud Dormancy Events in Perennial Plants. *Front. Plant Sci.* 5, 247. doi:10.3389/fpls.2014.00247

Rodrigues, J. A., and Zilberman, D. (2015). Evolution and Function of Genomic Imprinting in Plants. *Genes Dev.* 29 (24), 2517–2531. doi:10.1101/gad.269902.115

Rohde, A., Prinsen, E., De Rycke, R., Engler, G., Van Montagu, M., and Boerjan, W. (2002). *PtABI3* Impinges on the Growth and Differentiation of Embryonic

Leaves during Bud Set in poplar. *Plant Cell* 14, 1885–1901. doi:10.1105/tpc.003186

Ruttink, T., Arend, M., Morreel, K., Storme, V., Rombauts, S., Fromm, J., et al. (2007). A Molecular Timetable for Apical Bud Formation and Dormancy Induction in poplar. *Plant Cell* 19, 2370–2390. doi:10.1105/tpc.107.052811

Saez, A., Apostolova, N., Gonzalez-Guzman, M., Gonzalez-Garcia, M. P., Nicolas, C., Lorenzo, O., et al. (2004). Gain-of-function and Loss-of-function Phenotypes of the Protein Phosphatase 2C HAB1 Reveal its Role as a Negative Regulator of Abscisic Acid Signalling. *Plant J.* 37 (3), 354–369. doi:10.1046/j.1365-313x.2003.01966.x

Saez, A., Rodrigues, A., Santiago, J., Rubio, S., and Rodriguez, P. L. (2008). HAB1-SWI3B Interaction Reveals a Link between Abscisic Acid Signaling and Putative SWI/SNF Chromatin-Remodeling Complexes in Arabidopsis. *The Plant cell* 20 (11), 2972–2988. doi:10.1105/tpc.107.056705

Sahu, P. P., Pandey, G., Sharma, N., Puranik, S., Muthamilarasan, M., and Prasad, M. (2013). Epigenetic Mechanisms of Plant Stress Responses and Adaptation. *Plant Cell Rep* 32 (8), 1151–1159. doi:10.1007/s00299-013-1462-x

Schaller, G. E., Bishopp, A., and Kieber, J. J. (2015). The Yin-Yang of Hormones: Cytokinin and Auxin Interactions in Plant Development. *Plant Cell* 27 (1), 44–63. doi:10.1105/tpc.114.133595

Schmid, M., Uhlenhaut, N. H., Godard, F., Demar, M., Bressan, R., Weigel, D., et al. (2003). Dissection of floral Induction Pathways Using Global Expression Analysis. *Development* 130, 6001–6012. doi:10.1242/dev.00842

Schmitz, R. J., Hong, L., Fitzpatrick, K. E., and Amasino, R. M. (2007). DICER-LIKE 1 and DICER-LIKE 3 Redundantly Act to Promote Flowering via Repression of FLOWERING LOCUS C in *Arabidopsis thaliana*. *Genetics* 176, 1359–1362. doi:10.1534/genetics.107.070649

Schönrock, N., Bouveret, R., Leroy, O., Borghi, L., Köhler, C., Gruissem, W., et al. (2006). Polycomb-group Proteins Repress the floral Activator AGL19 in the FLC-independent Vernalization Pathway. *Genes Dev.* 20 (12), 1667–1678. doi:10.1101/gad.377206

Seifert, F., Thiemann, A., Schrag, T. A., Rybka, D., Melchinger, A. E., Frisch, M., et al. (2018). Small RNA-Based Prediction of Hybrid Performance in maize. *BMC Genomics* 19. doi:10.1186/s12864-018-4708-8

Shafiq, S., Li, J., and Sun, Q. (2016). Functions of Plants Long Non-coding RNAs. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1859 (1), 155–162. doi:10.1016/j.bbagrm.2015.06.009

Sheldon, C. C., Hills, M. J., Lister, C., Dean, C., Dennis, E. S., and Peacock, W. J. (2008). Resetting of FLOWERING LOCUS C Expression After Epigenetic Repression by Vernalization. *Proc. Natl. Acad. Sci. U S A.* 105, 2214–2219.

Shen, J., Xie, K., and Xiong, L. (2010). Global Expression Profiling of rice microRNAs by One-Tube Stem-Loop Reverse Transcription Quantitative PCR Revealed Important Roles of microRNAs in Abiotic Stress Responses. *Mol. Genet. Genomics* 284 (6), 477–488. doi:10.1007/s00438-010-0581-0

Shen, Q., Lin, Y., Li, Y., and Wang, G. (2021). Dynamics of H3K27me3 Modification on Plant Adaptation to Environmental Cues. *Plants* 10, 1165. doi:10.3390/plants10061165

Sherman, J. D., and Talbert, L. E. (2002). Vernalization-induced Changes of the DNA Methylation Pattern in winter Wheat. *Genome* 45 (2), 253–260. doi:10.1139/g01-147

Shi, J., Dong, A., and Shen, W. H. (2015). Epigenetic Regulation of rice Flowering and Reproduction. *Front. Plant Sci.* 5, 803. doi:10.3389/fpls.2014.00803

Shikata, M., Yamaguchi, H., Sasaki, K., and Ohtsubo, N. (2012). Overexpression of Arabidopsis miR157b Induces Bushy Architecture and Delayed Phase Transition in Torenia Fournieri. *Planta* 236, 1027–1035. doi:10.1007/s00425-012-1649-3

Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., and Levy, A. A. (2015). DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in Arabidopsis. *Plant Cell* 27, 2427–2436. doi:10.1105/tpc.15.00391

Simon, J. A., and Kingston, R. E. (2013). Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Mol. Cell* 49, 808–824. doi:10.1016/j.molcel.2013.02.013

Slotkin, R. K., and Martienssen, R. (2007). Transposable Elements and the Epigenetic Regulation of the Genome. *Nat. Rev. Genet.* 8 (4), 272–285. doi:10.1038/nrg2072

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J. D., Feijó, J. A., et al. (2009). Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell* 136, 461–472. doi:10.1016/j.cell.2008.12.038

Song, Q., Zhang, T., Stelly, D. M., and Chen, Z. J. (2017). *Epigenomic and Functional Analyses Reveal Roles of Epialleles in the Loss of Photoperiod Sensitivity during Domestication of Allotetraploid Cottons*.

Song, Y., Ji, D., Li, S., Wang, P., Li, Q., and Xiang, F. (2012). The Dynamic Changes of DNA Methylation and Histone Modifications of Salt Responsive Transcription Factor Genes in Soybean. *PLoS ONE* 7, e41274. doi:10.1371/journal.pone.0041274

Soppe, W. J. J., Viñegra de la Torre, N., and Albani, M. C. (2021). The Diverse Roles of FLOWERING LOCUS C in Annual and Perennial Brassicaceae Species. *Front. Plant Sci.* 12, 627258. doi:10.3389/fpls.2021.627258

Steward, N., Ito, M., Yamaguchi, Y., Koizumi, N., and Sano, H. (2002). Periodic DNA Methylation in maize Nucleosomes and Demethylation by Environmental Stress. *J. Biol. Chem.* 277 (40), 37741–37746. doi:10.1074/jbc.M204050200

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., et al. (2014). Non-CG Methylation Patterns Shape the Epigenetic Landscape in Arabidopsis. *Nat. Struct. Mol. Biol.* 21 (1), 64–72. doi:10.1038/nsmb.2735

Sun, C., Chen, D., Fang, J., Wang, P., Deng, X., and Chu, C. (2014). Understanding the Genetic and Epigenetic Architecture in Complex Network of rice Flowering Pathways. *Protein Cell* 5 (12), 889–898. doi:10.1007/s13238-014-0068-6

Sun, Y., Fan, M., and He, Y. (2019). DNA Methylation Analysis of the *Citrullus lanatus* Response to Cucumber *GreenMottle Mosaic Virus* Infection by Whole-Genome Bisulfite Sequencing. *Genes* 10 (5), 344. doi:10.3390/genes10050344

Sung, S., and Amasino, R. M. (2004). Vernalization in *Arabidopsis thaliana* Is Mediated by the PHD finger Protein VIN3. *Nature* 427 (6970), 159–164. doi:10.1038/nature02195

Sunkar, R., and Zhu, J.-K. (2004). Novel and Stress-Regulated MicroRNAs and Other Small RNAs from Arabidopsis[W]. *The Plant cell* 16 (8), 2001–2019. doi:10.1105/tpc.104.022830

Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced Silencing by Long Antisense Transcripts of an Arabidopsis Polycomb Target. *Nature* 462 (7274), 799–802. doi:10.1038/nature08618

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., et al. (2002). Important Roles of Drought- and Cold-Inducible Genes for Galactinol Synthase in Stress Tolerance in *Arabidopsis thaliana*. *Plant J.* 29 (4), 417–426. doi:10.1046/j.0960-7412.2001.01227.x

Takeda, S., Sugimoto, K., Kakutani, T., and Hirochika, H. (2001). Linear DNA Intermediates of the Tto1 Retrotransposon in Gag Particles Accumulated in Stressed Tobacco and *Arabidopsis thaliana*. *Plant J. : Cel. Mol. Biol.* 28 (3), 307–317. doi:10.1046/j.1365-313x.2001.01151.x

Tamada, Y., Yun, J. Y., Woo, S. C., and Amasino, R. M. (2009). ARABIDOPSIS TRITHORAX-RELATED7 is Required for Methylation of Lysine 4 of Histone H3 and for Transcriptional Activation of FLOWERING LOCUS C. *The Plant cell.* 21 (10), 3257–3269. doi:10.1105/tpc.109.070060

Tanaka, N., Itoh, H., Sentoku, N., Kojima, M., Sakakibara, H., Izawa, T., et al. (2011). TheCOP1OrthologPPSRegulates the Juvenile-Adult and Vegetative-Reproductive Phase Changes in Rice. *Plant Cell* 23, 2143–2154. doi:10.1105/tpc.111.083436

Tang, X., Wang, Q., Yuan, H., and Huang, X. (2018). Chilling-induced DNA Demethylation Is Associated with the Cold Tolerance of *Hevea Brasiliensis*. *BMC Plant Biol.* 18, 70. doi:10.1186/s12870-018-1276-7

Tani, E., Polidoros, A. N., Nianiou-Obeidat, I., and Tsaftaris, A. S. (2005). DNA Methylation Patterns are Differentially Affected by Planting Density in Maize Inbreds and Their Hybrids. *Maydica* 50, 19–23.

Telias, A., Lin-Wang, K., Stevenson, D. E., Cooney, J. M., Hellens, R. P., Allan, A. C., et al. (2011). Apple Skin Patterning Is Associated with Differential Expression of MYB10. *BMC Plant Biol.* 11, 93. doi:10.1186/1471-2229-11-93

Teotia, S., and Tang, G. (2015). To Bloom or Not to Bloom: Role of MicroRNAs in Plant Flowering. *Mol. Plant* 8, 359–377. doi:10.1016/j.molp.2014.12.018

Thapa, B., and Shrestha, A. (2020). Epigenetic Mechanisms and its Role in Plant Growth and Development. *J. Plant Biochem. Physiol.* 8, 255.

Tian, L., and Chen, Z. J. (2001). Blocking Histone Deacetylation in Arabidopsis Induces Pleiotropic Effects on Plant Gene Regulation and Development. *Proc. Natl. Acad. Sci.* 98, 200–205. doi:10.1073/pnas.98.1.200

Tiwari, R., and Rajam, M. V. (2022). RNA- and miRNA-Interference to Enhance Abiotic Stress Tolerance in Plants. *J. Plant Biochem. Biotechnol.* doi:10.1007/s13562-022-00770-9

To, T. K., Kim, J.-M., Matsui, A., Kurihara, Y., Morosawa, T., Ishida, J., et al. (2011). Arabidopsis HDA6 Regulates Locus-Directed Heterochromatin Silencing in Cooperation with MET1. *Plos Genet.* 7 (4), e1002055. doi:10.1371/journal.pgen.1002055

Trap-Gentil, M.-V., Hébrard, C., Lafon-Placette, C., Delaunay, A., Hagège, D., Joseph, C., et al. (2011). Time Course and Amplitude of DNA Methylation in the Shoot Apical Meristem Are Critical Points for Bolting Induction in Sugar Beet and Bolting Tolerance between Genotypes. *J. Exp. Bot.* 62, 2585–2597. doi:10.1093/jxb/erq433

Urano, K., Kurihara, Y., Seki, M., and Shinozaki, K. (2010). 'Omics' Analyses of Regulatory Networks in Plant Abiotic Stress Responses. *Curr. Opin. Plant Biol.* 13 (2), 132–138. doi:10.1016/j.pbi.2009.12.006

van Dijk, K., Ding, Y., Malkaram, S., Riethoven, J.-J. M., Liu, R., Yang, J., et al. (2010). Dynamic Changes in Genome-wide Histone H3 Lysine 4 Methylation Patterns in Response to Dehydration Stress in *Arabidopsis thaliana*. *BMC Plant Biol.* 10, 238. doi:10.1186/1471-2229-10-238

Waddington, C. H. (2012). The Epigenotype. *Int. J. Epidemiol.* 41 (1), 10–13. doi:10.1093/ije/dyr184

Waddington, C. H. (1957). *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology*. London: George Allen & Unwin.

Wang, C., Wang, C., Xu, W., Zou, J., Qiu, Y., Kong, J., et al. (2018). Epigenetic Changes in the Regulation of *Nicotiana Tabacum* Response to Cucumber Mosaic Virus Infection and Symptom Recovery through Single-Base Resolution Methylomes. *Viruses* 10, 402. doi:10.3390/v10080402

Wang, Y., Xue, X., Zhu, J. K., and Dong, J. (2016). Demethylation of ERECTA Receptor Genes by IBM1 Histone Demethylase Affects Stomatal Development. *Development* 143, 4452–4461. doi:10.1242/dev.129932

Wang, Y., Li, L., Ye, T., Lu, Y., Chen, X., and Wu, Y. (2013). The Inhibitory Effect of ABA on floral Transition Is Mediated by ABI5 in Arabidopsis. *J. Exp. Bot.* 64 (2), 675–684. doi:10.1093/jxb/ers361

Wibowo, A., Becker, C., Marconi, G., Durr, J., Price, J., Hagmann, J., et al. (2016). Hyperosmotic Stress Memory in Arabidopsis Is Mediated by Distinct Epigenetically Labile Sites in the Genome and Is Restricted in the Male Germline by DNA Glycosylase Activity. *eLife* 5, e13546. doi:10.7554/eLife.13546

Wu, K., Zhang, L., Zhou, C., Yu, C.-W., and Chaikam, V. (2008). HDA6 Is Required for Jasmonate Response, Senescence and Flowering in Arabidopsis. *J. Exp. Bot.* 59 (2), 225–234. doi:10.1093/jxb/erm300

Wu, L., Liu, D., Wu, J., Zhang, R., Qin, Z., Liu, D., et al. (2013). Regulation of FLOWERING LOCUS T by a microRNA in Brachypodium Distachyon. *The Plant Cell* 25, 4363–4377. doi:10.1105/tpc.113.118620

Wu, M. F., Yamaguchi, N., Xiao, J., Bargmann, B., Estelle, M., Sang, Y., et al. (2015). Auxin-regulated Chromatin Switch Directs Acquisition of Flower Primordium Founder Fate. *eLife* 4, e09269. doi:10.7554/eLife.09269

Wu, R.-M., Walton, E. F., Richardson, A. C., Wood, M., Hellens, R. P., and Varkonyi-Gasic, E. (2012). Conservation and Divergence of Four Kiwifruit *SVP*-like MADS-Box Genes Suggest Distinct Roles in Kiwifruit Bud Dormancy and Flowering. *J. Exp. Bot.* 63, 797–807. doi:10.1093/jxb/err304

Xiao, D., Zhou, K., Yang, X., Yang, Y., Ma, Y., and Wang, Y. (2021). Crosstalk of DNA Methylation Triggered by Pathogen in Poplars with Different Resistances. *Front. Microbiol.* 12, 750089. doi:10.3389/fmicb.2021.750089

Xie, H., Konate, M., Sai, N., Tesfamicael, K. G., Cavagnaro, T., Gilliham, M., et al. (2017). Global DNA Methylation Patterns Can Play a Role in Defining Terroir in Grapevine (Vitis vinifera Cv. Shiraz). *Front. Plant Sci.* 8, 1860. doi:10.3389/fpls.2017.01860

Xie, K., Shen, J., Hou, X., Yao, J., Li, X., Xiao, J., et al. (2012). Gradual Increase of miR156 Regulates Temporal Expression Changes of Numerous Genes during Leaf Development in Rice. *Plant Physiol.* 158, 1382–1394. doi:10.1104/pp.111.190488

Xu, G., Lyu, J., Li, Q., Liu, H., Wang, D., Zhang, M., et al. (2020). Evolutionary and Functional Genomics of DNA Methylation in maize Domestication and Improvement. *Nat. Commun.* 11, 5539. doi:10.1038/s41467-020-19333-4

Xu, J., Chen, G., Hermanson, P. J., Xu, Q., Sun, C., Chen, W., et al. (2019). Population-level Analysis Reveals the Widespread Occurrence and Phenotypic Consequence of DNA Methylation Variation Not Tagged by Genetic Variation in maize. *Genome Biol.* 20, 243. doi:10.1186/s13059-019-1859-0

Xue, X.-Y., Zhao, B., Chao, L.-M., Chen, D.-Y., Cui, W.-R., Mao, Y.-B., et al. (2014). Interaction between Two Timing microRNAs Controls Trichome Distribution in Arabidopsis. *Plos Genet.* 10, e1004266. doi:10.1371/journal.pgen.1004266

Yaish, M. W., Colasanti, J., and Rothstein, S. J. (2011). The Role of Epigenetic Processes in Controlling Flowering Time in Plants Exposed to Stress. *J. Exp. Bot.* 62, 3727–3735. doi:10.1093/jxb/err177

Yaish, M. W. (2013). "DNA Methylation-Associated Epigenetic Changes in Stress Tolerance of Plants," in *Molecular Stress Physiology of Plants*. Editors G. Rout and A. Das (New Delhi: Springer), 427–440. doi:10.1007/978-81-322-0807-5_17

Yamagishi, K., and Kikuta, Y. (2021). Nucleoside Derivatives of 5-methylcytosine Suppress 5-Azacytidine-Induced Reactivation of a Silent Transgene in Suspension-Cultured Tobacco Cells. *Plant Biotechnol.* 38, 173–178. doi:10.5511/plantbiotechnology.20.1126b

Yamamuro, C., Miki, D., Zheng, Z., Ma, J., Wang, J., Yang, Z., et al. (2014). Overproduction of Stomatal Lineage Cells in *Arabidopsis* Mutants Defective in Active DNA Demethylation. *Nat. Commun.* 5, 4062. doi:10.1038/ncomms5062

Yamane, H., Ooka, T., Jotatsu, H., Hosaka, Y., Sasaki, R., and Tao, R. (2011). Expressional Regulation of PpDAM5 and PpDAM6, Peach (Prunus Persica) Dormancy-Associated MADS-Box Genes, by Low Temperature and Dormancy-Breaking Reagent Treatment. *J. Exp. Bot.* 62, 3481–3488. doi:10.1093/jxb/err028

Yang, X., Sanchez, R., Kundariya, H., Maher, T., Dopp, I., Schwegel, R., et al. (2020). Segregation of an MSH1 RNAi Transgene Produces Heritable Non-genetic Memory in Association with Methylome Reprogramming. *Nat. Commun.* 11, 2214–2217. doi:10.1038/s41467-020-16036-8

Yelina, N. E., Lambing, C., Hardcastle, T. J., Zhao, X., Santos, B., and Henderson, I. R. (2015). DNA Methylation Epigenetically Silences Crossover Hot Spots and Controls Chromosomal Domains of Meiotic Recombination in Arabidopsis. *Genes Dev.* 29, 2183–2202. doi:10.1101/gad.270876.115

Yu, A., Lepère, G., Jay, F., Wang, J., Bapaume, L., Yu Wang, Y., et al. (2013). Dynamics and Biological Relevance of DNA Demethylation in Arabidopsis Antibacterial Defense. *Proc. Natl. Acad. Sci.* 110, 2389–2394. doi:10.1073/pnas.1211757110

Yu, C.-W., Liu, X., Luo, M., Chen, C., Lin, X., Tian, G., et al. (2011). HISTONE DEACETYLASE6 Interacts with FLOWERING LOCUS D and Regulates Flowering in Arabidopsis. *Plant Physiol.* 156 (1), 173–184. doi:10.1104/pp.111.174417

Yuan, W., Luo, X., Li, Z., Yang, W., Wang, Y., Liu, R., et al. (2016). A *Cis* Cold Memory Element and a *Trans* Epigenome Reader Mediate Polycomb Silencing of *FLC* by Vernalization in *Arabidopsis*. *Nat. Genet.* 48, 1527–1534. doi:10.1038/ng.3712

Zemach, A., Kim, M. Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., et al. (2013). The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell* 153 (1), 193–205. doi:10.1016/j.cell.2013.02.033

Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G.-L., Walbot, V., et al. (2015). Spatiotemporally Dynamic, Cell-type-dependent Premeiotic and Meiotic phasiRNAs in maize Anthers. *Proc. Natl. Acad. Sci. USA* 112, 3146–3151. doi:10.1073/pnas.1418918112

Zhang, H., Lang, Z., and Zhu, J. K. (2018). Dynamics and Function of DNA Methylation in Plants. *Nat. Rev. Mol. Cell Biol* 19, 489–506. doi:10.1038/s41580-018-0016-z

Zhang, J.-F., Yuan, L.-J., Shao, Y., Du, W., Yan, D.-W., and Lu, Y.-T. (2008). The Disturbance of Small RNA Pathways Enhanced Abscisic Acid Response and Multiple Stress Responses in Arabidopsis. *Plant Cell Environ* 31, 562–574. doi:10.1111/j.1365-3040.2008.01786.x

Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., et al. (2011). Extensive, Clustered Parental Imprinting of Protein-Coding and Noncoding RNAs in Developing maize Endosperm. *Proc. Natl. Acad. Sci.* 108, 20042–20047. doi:10.1073/pnas.1112186108

Zhang, Y. Y., Fischer, M., Colot, V., and Bossdorf, O. (2013). Epigenetic Variation Creates Potential for Evolution of Plant Phenotypic Plasticity. *New Phytol.* 197, 314–322. doi:10.1111/nph.12010

Zhao, M., Liu, W., Xia, X., Wang, T., and Zhang, W.-H. (2014). Cold Acclimation-Induced Freezing Tolerance ofMedicago Truncatulaseedlings Is Negatively Regulated by Ethylene. *Physiol. Plantarum* 152 (1), 115–129. doi:10.1111/ppl.12161

Zhao, T., Zhan, Z., and Jiang, D. (2019). Histone Modifications and Their Regulatory Roles in Plant Development and Environmental Memory. *J. Genet. Genomics* 46, 467–476. doi:10.1016/j.jgg.2019.09.005

Zheng, X. Y., and Gehring, M. (2019). Low-Input Chromatin Profiling in Arabidopsis Endosperm Using CUT&RUN. *Plant Reprod* 32, 63–75. doi:10.1007/s00497-018-00358-1

Zhong, W., Gao, Z., Zhuang, W., Shi, T., Zhang, Z., and Ni, Z. (2013a). Genome-wide Expression Profiles of Seasonal Bud Dormancy at Four Critical Stages in Japanese Apricot. *Plant Mol. Biol.* 83, 247–264. doi:10.1007/s11103-013-0086-4

Zhong, W., Gao, Z., Zhuang, W., Shi, T., Zhang, Z., and Ni, Z. (2013). Genome-wide Expression Profiles of Seasonal Bud Dormancy at Four Critical Stages in Japanese Apricot. *Plant Mol. Biol.* 83, 247–264. doi:10.1007/s11103-013-0086-4

Zhou, C., Zhang, L., Duan, J., Miki, B., and Wu, K. (2005). HISTONE DEACETYLASE19 Is Involved in Jasmonic Acid and Ethylene Signaling of Pathogen Response in Arabidopsis. *Plant Cell* 17 (4), 1196–1204. doi:10.1105/tpc.104.028514

Zhou, D.-X. (2009). Regulatory Mechanism of Histone Epigenetic Modifications in Plants. *Epigenetics* 4 (1), 15–18. doi:10.4161/epi.4.1.7371

Zhou, S., Liu, X., Zhou, C., Zhou, Q., Zhao, Y., Li, G., et al. (2016). Cooperation between the H3K27me3 Chromatin Mark and Non-CG Methylation in Epigenetic Regulation. *Plant Physiol.* 172, 1131–1141. doi:10.1104/pp.16.01238

Zhu, H., Xia, R., Zhao, B., An, Y.-Q., Dardick, C. D., Callahan, A. M., et al. (2012). Unique Expression, Processing Regulation, and Regulatory Network of Peach (*Prunus Persica*) miRNAs. *BMC Plant Biol.* 12, 149. doi:10.1186/1471-2229-12-149

Zhu, Q.-H., Upadhyaya, N. M., Gubler, F., and Helliwell, C. A. (2009). Over-expression of miR172 Causes Loss of Spikelet Determinacy and floral Organ Abnormalities in rice (Oryza Sativa). *BMC Plant Biol.* 9, 149. doi:10.1186/1471-2229-9-149

Zong, W., Zhong, X., You, J., and Xiong, L. (2012). Genome-wide Profiling of Histone H3K4-Tri-Methylation and Gene Expression in rice under Drought Stress. *Plant Mol. Biol.* 81, 175–188. doi:10.1007/s11103-012-9990-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership