Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Current status and future perspectives on the evolution of *cis*-regulatory elements in plants

Alan E. Yocca[1,2] and Patrick P. Edger[2,3]

Abstract

Cis-regulatory elements (CREs) are short stretches (~5–15 base pairs) of DNA capable of being bound by a transcription factor and influencing the expression of nearby genes. These regions are of great interest to anyone studying the relationship between phenotype and genotype as these sequences often dictate genes' spatio-temporal expression. Indeed, several associative signals between genotype and phenotype are known to lie outside of protein-coding regions. Therefore, a key to understand evolutionary biology requires their characterization in current and future genome assemblies. In this review, we cover some recent examples of how CRE variation contributes to phenotypic evolution, discuss evidence for the selective pressures experienced by non-coding regions of the genome, and consider several studies on accessible chromatin regions in plants and what they can tell us about CREs. Finally, we discuss how current advances in sequencing technologies will improve our knowledge of CRE variation.

Addresses

[1] Department of Plant Biology, Michigan State University, East Lansing, MI, 48824, USA

[2] Department of Horticulture, Michigan State University, East Lansing, MI, 48824, USA

[3] Genetics and Genome Sciences Program, Michigan State University, East Lansing, MI, 48824, USA

Corresponding author: Edger, Patrick P (edgerpat@msu.edu)

Current Opinion in Plant Biology 2022, **65**:102139

This review comes from a themed issue on **Genome studies and molecular genetics (2022)**

Edited by **Eunyoung Chae** and **Daniel Koenig**

For a complete overview see the Issue and the Editorial

Available online 24 November 2021

https://doi.org/10.1016/j.pbi.2021.102139

Keywords

Cis-regulatory evolution, Genome evolution.

Evolution of CREs

Cis-regulatory element (CRE) variation is critical to phenotypic evolution in all organisms. Broadly, regulatory regions are enriched for loci associated with phenotypic variation, for example, in maize [1,2]. We will note a few recent examples in pineapple (*Ananas comosus*) and rice (*Oryza sativa*), though multiple others exist [3,4]. The pineapple genome revealed shifts in temporal gene expression enabled the transition from C3 to crassulacean acid metabolism (CAM) photosynthesis [5]. In addition, enrichment for circadian clock associated CREs upstream CAM genes implicates CRE changes driving the transition to CAM photosynthesis in pineapple. Recently, Wu et al. [6] were able to increase the grain number per panicle in rice through the engineering of a naturally occurring CRE. They uncovered the CRE controlling the gene *OsREM20* varied between the two main groups of rice, *indica* and *japonica*. Phylogenetic analyses revealed the alleles were likely inherited independently from *O. rufipogon* and were likely artificially selected for during rice domestication. Importantly, the majority allele in *japonica* rice, when introgressed into the *indica* background increased the grain number per panicle. These examples highlight the contribution of CRE variation to observed phenotypic variation (see Table 1).

The diversity of non-protein–coding DNA can be studied in the context of a pan-genome. A pan-genome refers to the total complement of genetic variation present within a specific taxonomic group (e.g. species-level). Recent surveys of genetic variation in plants uncover a largely variable amount of genetic material absent in the reference; shared protein-coding regions between genomes ranges between 20 and 80% [7]. Compared to protein-coding regions, the amount of intraspecific variation of non-protein–coding regions in plant genomes remains relatively poorly understood. We will describe findings from recent studies regarding CRE variation and selection.

Depending on the class of element, there may be greater selective pressure on non-coding than protein-coding sequences. Williamson et al. [8] investigated selection across the genome of *Capsella grandiflora*. They found stronger signals of selection in protein-coding than non-protein-coding regions. However, selection signals in conserved noncoding sequences (CNSs) approached but did not quite reach that of 0-fold nonsynonymous protein-coding sites (those that will alter amino acids with any substitution). This suggests selection on the non-protein–coding genome may reflect patterns close to protein-coding regions. Obviously, substitutions affecting amino acids will have more direct and measurable effects on a molecular level, and

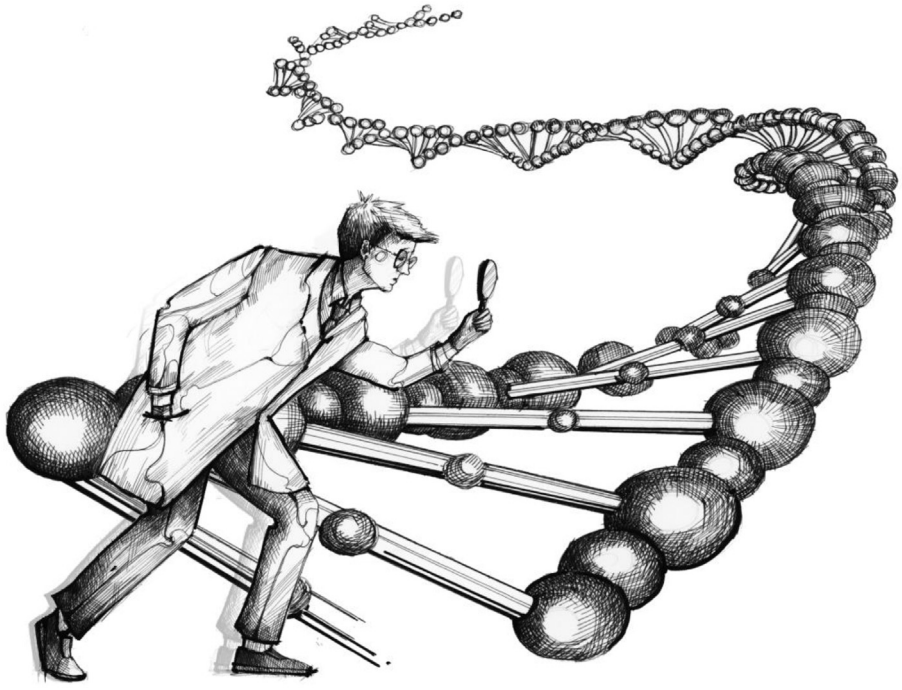


Table 1

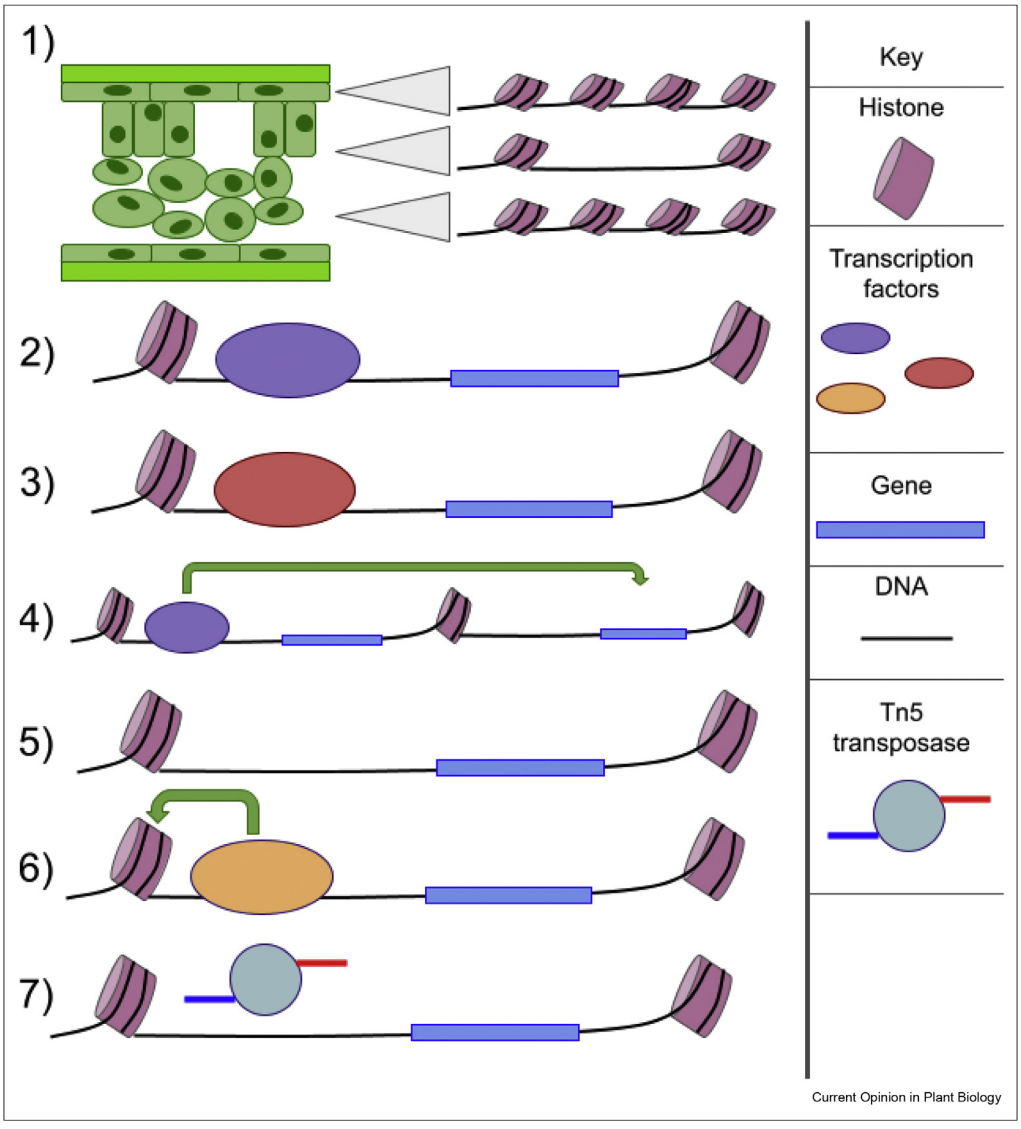
A glossary of terms used throughout this manuscript. Artwork by Juliana Žamoit depicting a scientist searching for *Cis-regulatory elements* throughout the genome.

Term	Abbreviation	Definition
Cis-regulatory element	CRE	A short, often 12–15 base-pair, segment of DNA to which transcription factors bind and influence the expression of nearby genes.
Transcription factor binding site	TFBS	The specific motif to which a transcription factor binds within a *cis*-regulatory element.
Assay for transposase accessibility followed by sequencing	ATAC-seq	A sequencing technique described by Buenostro et al. and Lu et al. to detect areas of accessible chromatin.
Accessible chromatin regions	ACR	Regions of DNA that are uncondensed to the point to allow transcription factor binding.
Conserved noncoding sequences	CNS	Regions of non-protein–coding DNA that exhibit sequence conservation above random expectation, implying some form of selection retains these sequences over millions of years of evolution.

transcription factor binding sites (TFBSs) are somewhat degenerative in nature, occasionally withstanding multiple substitutions without affecting binding. However, the integrity of transcriptional coordination must experience selection. Indeed, we found CNSs exhibit presence–absence variation at rates lower than many protein-coding regions [9]. Perhaps there exist non-protein–coding regions experiencing stronger selection than most protein-coding regions.

The strength of selection on noncoding regions depends on sequence context. A recent study investigated mutation probability genome-wide [10]. They found not all genomic regions experience equal mutation rates. Beyond the selection experienced by some genomic sites for their contribution to fitness, perhaps these critical sites additionally are shielded from mutations by their genomic context as described by Monroe et al. [10]. Counter to that thought, circumstances might

Figure 1



Modeling the lack of correlation between chromatin accessibility and gene expression. (1) Plant leaf cross-section displaying accessibility is different between cell types at the same locus. (2) A transcription factor (TF) in purple is bound yet is only poised and not actively driving transcription of its proximate gene in blue. (3) The TF binding in red represses transcription of the proximate gene in blue. (4) The TF binding in purple affects a distal gene. (5) The region is accessible, yet no TF is bound. (6) The region is accessible because a pioneer transcription factor in orange is remodeling the chromatin and not driving the expression of the proximate gene. (7) The region is accessible, yet the Tn5 transposase has not inserted adapters into this region.

incentivize higher mutation rates in transcriptional regulating regions relative to protein-coding regions because of limited negative pleiotropy [11,12]. For example, modification of a CRE may exclude a gene product in a single Spatio-temporal context, whereas a protein-coding modification will impact all contexts.

CREs are ripe for exploring the evolutionary landscape in this regard. As with nearly everything in this world, these theories can coexist. They aren't mutually exclusive, and life certainly exhibits a distribution of all these outcomes. Surely some CREs experience stronger selection than most protein-coding regions, while others experience non-lethal mutations. As with protein-coding genes, the contribution of a DNA segment to fitness determines its position within a selection curve. Indeed, a recent study by Joly-Lopez et al. inferred selection strength genome-wide in rice and created these selection curves (Figure 1a [4]). This distribution of selection strength was varied across non-coding space. Importantly, there

were non-coding regions with selection values higher than some protein-coding regions. What is important is identifying where in the spectrum a CRE of interest lies and what that sequence contributes to cellular functions and ultimately to an organism's phenotype. One might spend an entire career disentangling the non-protein–coding genome's mysteries, and the year 2021 is a good year to start.

CNS as a proxy for *cis*-regulatory elements

The alignment of genomes from closely related species has been used to identify conserved and putatively functional non-protein–coding regions, including CNSs, across diverse angiosperm lineages [13–17]. Consistently, CNSs exhibit great overlap with functional genomics data and signatures of TFBSs. However, the overlap is incomplete as CNSs encode numerous functions beyond *cis* transcriptional regulation, such as long non-coding RNA [13,17,18]. Two studies have addressed how variation in CNS within a population affects gene expression [9,16]. They both find an association between CNS absence and decreased gene expression. This suggests CNSs more often encode transcriptional activators compared to repressors. In the future, it will be interesting to discover the ratio of encoded activators to repressors, although transcriptional regulation does not divide neatly into these binary categories.

Accessible chromatin signature of *cis*-regulatory elements

Short TFBSs (5-6bp) can randomly occur over 100,000 times in a genome the size of maize (~3Gbp) and even 10,000 times in a genome the size of *Arabidopsis thaliana* (~130Mbp). Therefore, functional genomics data are critical to validate a putative TFBS. Several functional genomics methods have been developed to generate this evidence. For brevity, we will address assay for transposase accessible chromatin followed by sequencing (ATAC-seq [19,20]) though others certainly exist [21–23]. ATAC-seq leverages an active Tn5 transposase that preferentially inserts adapters into regions of accessible chromatin. Sequencing from these adapters enables a quantitative readout of chromatin accessibility. Analysis of sequence enrichment relative to naked DNA enables one to define accessible chromatin regions (ACRs) genome-wide. ACRs are indicative of transcription factor binding [24,25].

Tissue-specific ACRs do not always positively correlate with the expression of adjacent genes [26]. We model several reasons why accessibility does not perfectly correlate with gene expression in Figure 1. (1) Tissue homogenization drowns out cell-type–specific associations. (2) Transcription factors (TFs) are bound in the accessible region yet are only poised and not actively driving transcription of its proximate gene. (3) The TF binding to the ACR represses transcription. (4) The TF binding to the ACR regulates a more distal gene. (5) The ACR represents a binding site open for business yet is not actively conducting business. (6) The region is accessible because a pioneer TF in orange is remodeling the chromatin and not driving the expression of the proximate gene. Pioneer TFs bind heterochromatin and alter chromatin architecture, or recruit remodelers rather than directly driving the expression of proximate genes. Pioneer TFs are known to generate ACRs [27–29]. (7) The region is accessible, yet the Tn5 transposase has not inserted adapters into this region. Insertion bias and Tn5 transposase inefficiency are key considerations in ATAC-seq analyses [30–32].

Often, discordance between ACRs and gene expression is hypothesized to result from tissue and/or cell-type homogenization when profiling ACRs. We think exploring cell-type–specific relationships will improve the observed correlation between ACRs and gene expression. Mahar et al. [26] explored accessible chromatin regions both across plant species and *A. thaliana* hair and non-hair root epidermal cells. They found no correlation between ACR count upstream genes and that gene's expression. Between root cell types, >50% of the ACRs were shared. Importantly, there were strong associations between cell-type–specific expression and cell-type–specific ACRs. However, there were also several examples of reverse associations, which may represent repressor binding. Alexandre et al. [33] investigated ACRs between different A. thaliana ecotypes. A key finding was an increase in ACRs near conditionally expressed genes. The authors hypothesized these ACRs likely represent poised TFs that are bound and await stimulus to drive expression. A separate study by Tannenbaum et al. [34] pooled root and leaf tissues in A. thaliana for ATAC-seq and RNA-seq. Though specific correlations were not investigated, they found <50% of tissue-specific ACR were located near a tissue-specific differentially expressed gene.

The best attempts to address the correlation between ACRs and gene expression are recent studies by Marand et al. [35] and Farmer et al. [36]. Through single-cell RNA-seq and ATAC-seq, Marand et al. [35] found the Spearman's correlation coefficient between ACRs and gene expression range from 0.52 to 0.69 depending on the cell type, compared to no correlation, or <50% concordance from bulked studies. Perhaps simultaneous ATAC and RNA sequencing could improve this correlation. Importantly, this study also found several accessible regions in silenced genes. By overlaying H3K27me3 information, they hypothesized these accessible regions were necessary for the recruitment of

silencing machinery, specifically the polycomb repressive complex, which recruits H3K27me3. This highlights the complex relationship between ACRs and gene expression. Farmer et al. [36] also integrated single-cell RNA-seq and ATAC-seq. Although calculating Kendall's tau rank correlation and not Spearman's correlation coefficient, they did report positive correlations between gene expression and chromatin accessibility within cell-type clusters in agreement with results from Marand et al. [35].

The evolutionary conservation of ACRs has also been explored recently by Lu et al. [37] across thirteen angiosperms. With a specific focus on distal (>2 kb from the closest gene) ACRs (dACRs), their analyses provide exceptional insight into the evolution of chromatin accessibility. A companion publication demonstrated these dACRs often contain functional *cis*-regulatory elements [38]. Importantly, the distance between dACRs and genes increases with genome size, yet not linearly. This distance increase appears to involve transposable elements. Furthermore, there was strong enrichment for transposable elements and species-specific dACRs, further implicating mobile elements in the rewiring of transcriptional networks. At this point, the data are clear; transposable elements are major contributors to evolutionary novelty, at least partially through rewiring of transcriptional networks [39,40].

The case for studying ACRs

How does current knowledge of the relationship between ACRs and CREs motivate future researchers to profile ACRs in newly sequenced genomes? What do researchers hope to gain by sequencing new genomes? Regardless of the goal, there are numerous benefits to accessible chromatin profiling. Quantitative accessible chromatin information provides a genome-wide profile of candidates for CREs and phenotypically critical loci. Targeted genome editing must assess more than just the protein-coding regions of the genome to design resilient crops in the future. Indeed, recently targeted promoter editing in tomato (*Solanum lycopersicum*) successfully manipulated fruit locule number [41]. However, accessible chromatin profiling in plants is still largely hindered by the presence of cell walls. More efficient ACR profiling methods could revolutionize future research. We need to strive beyond studies that present new genomes that provide 'insights into' without additional functional data or evolutionary comparisons.

Long-read sequencing will improve CRE studies

Sequence technology improvements continue to advance our ability to characterize non-protein—coding regions of the genome. Current long-read sequencing technologies offered by Pacific Biosciences single-molecule real-time sequencing (PacBio SMRT) and Oxford Nanopore Technologies (ONT) boast read length, N50 values longer than many whole-genome assemblies generated a decade ago [42]. Several genome assemblies leveraging long-read sequencing now produce ungapped sequences spanning telomere to telomere [43,44]. However, we must closely monitor the base pair accuracy of these long reads when studying CREs, as TF-binding might be sensitive to only a few base pair substitutions [45]. PacBio SMRT circular consensus sequencing's (CCS) strategy currently confers a distinct advantage in base pair accuracy over ONT [45]. Because base pair errors are random, repeatedly sequencing the same molecule, the CCS strategy increases accuracy up to 99%. Currently, ONT allows for only a single pass across molecules, leaving the accuracy low around 85%. For a thorough comparison of the two technologies, see the study by Amarasinghe et al. [46]. Sequencing read accuracy are likely to continue to improve with future advancements in technologies.

Long-read sequencing will improve our understanding of plant genome organization. Primarily, the resolution of repetitive regions (often in noncoding regions) will improve. Even though long-read genome assemblies are increasing in frequency, population studies currently use short-read technologies. This introduces reference bias that may be mitigated with long-read sequencing. For example, short-read sequencing may suggest a CRE exists as reads will align to the reference CRE. However, the CRE may exist in a repetitive region or an alternate, non-reference location.

To characterize this population-level CRE variation, we propose shallow long-read sequencing. Long-read sequencing currently is cost prohibitive for this application but successfully resolves large-scale variants that short-read sequences cannot. A potential strategy to reduce cost is targeted long-read sequencing. Recently, software developed for the Oxford Nanopore sequencing platform can target specific sequences by first reading through a molecule, and if it does not match the desired sequence, will expel the molecule from the sequencing pore [47]. Sequence accuracy must greatly increase for this application in the future, however. Currently, PacBio CCS platform boasts the greatest accuracy and read length trade-off. A sequence capture approach, as commonly used in phylogenomics [48], could be combined with PacBio to target specific regions of the genome to investigate the evolution of specific CREs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Michigan State University AgBioResearch, USDA-NIFA HATCH 1009804, NSF 2029959, and USDA 2020-67013-30870.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Tu X, Mejía-Guerra MK, Valdes Franco JA, Tzeng D, Chu P-Y, Shen W, Wei Y, Dai X, Li P, Buckler ES, *et al.*: **Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors**. *Nat Commun* 2020, **11**:5089.

2. Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES: **Association mapping across numerous traits reveals patterns of functional variation in maize**. *PLoS Genet* 2014, **10**, e1004845.

3. Groen SC, Ćalić I, Joly-Lopez Z, Platts AE, Choi JY, Natividad M, Dorph K, Mauck 3rd WM, Bracken B, Cabral CLU, *et al.*: **The strength and pattern of natural selection on gene expression in rice**. *Nature* 2020, **578**:572–576.

4. ** Joly-Lopez Z, Platts AE, Gulko B, Choi JY, Groen SC, Zhong X, Siepel A, Purugganan MD: **An inferred fitness consequence map of the rice genome**. *Native Plants* 2020, **6**:119–130.
This study inferred selection across the entire genome of *Oryza sativa*. It provides a direct quantitative comparison of selective pressure across protein-coding and noncoding regions

5. Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang M-L, Chen J, Biggers E, *et al.*: **The pineapple genome and the evolution of CAM photosynthesis**. *Nat Genet* 2015, **47**:1435–1442.

6. Wu X, Liang Y, Gao H, Wang J, Zhao Y, Hua L, Yuan Y, Wang A, Zhang X, Liu J, *et al.*: **Enhancing rice grain production by manipulating the naturally evolved cis-regulatory element-containing inverted repeat sequence of OsREM20**. *Mol Plant* 2021, **14**:997–1011.

7. Bayer PE, Golicz AA, Scheben A, Batley J, Edwards D: **Plant pan-genomes are the new reference**. *Native Plants* 2020, **6**:914–920.

8. Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI: **Evidence for widespread positive and negative selection in coding and conserved noncoding regions of Capsella grandiflora**. *PLoS Genet* 2014, **10**, e1004622.

9. Yocca AE, Lu Z, Schmitz RJ, Freeling M, Edger PP: **Evolution of conserved noncoding sequences in Arabidopsis thaliana**. *Mol Biol Evol* 2021, **38**:2692–2703.

10. Monroe JG, Grey Monroe J, Srikant T, Carbonell-Bejerano P, Exposito-Alonso M, Weng M-L, Rutter MT, Fenster CB, Weigel D: Mutation bias shapes gene evolution in Arabidopsis thaliana. [date unknown], doi:10.1101/2020.06.17.156752.

11. Carroll SB: **Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution**. *Cell* 2008, **134**:25–36.

12. Stern DL, Orgogozo V: **The loci OF evolution: How predictable is genetic evolution?** *Evolution* 2008, **62**:2155–2177.

13. Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM, *et al.*: **An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions**. *Nat Genet* 2013, **45**:891–898.

14. Burgess D, Freeling M: **The most deeply conserved noncoding sequences in plants serve similar functions to those in vertebrates despite large differences in evolutionary rates**. *Plant Cell* 2014, **26**:946–961.

15. Van de Velde J, Heyndrickx KS, Vandepoele K: **Inference of transcriptional networks in Arabidopsis through conserved noncoding sequence analysis**. *Plant Cell* 2014, **26**:2729–2745.

16. Song B, Buckler ES, Wang H, Wu Y, Rees E, Kellogg EA, Gates DJ, Khaipho-Burch M, Bradbury PJ, Ross-Ibarra J, *et al.*: **Conserved noncoding sequences provide insights into regulatory sequence and loss of gene expression in maize**. *Genome Res* 2021, https://doi.org/10.1101/gr.266528.120.

17. Lai X, Behera S, Liang Z, Lu Y, Deogun JS, Schnable JC, Stag-Cns: **An order-aware conserved noncoding sequences discovery tool for arbitrary numbers of species**. *Mol Plant* 2017, **10**:990–999.

18. Van de Velde J, Van Bel M, Vaneechoutte D, Vandepoele K: **A collection of conserved noncoding sequences to study gene regulation in flowering plants**. *Plant Physiol* 2016, **171**:2586–2598.

19. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ: **ATAC-seq: a method for assaying chromatin accessibility genome-wide**. *Curr Protoc Mol Biol* 2015, **109**. 21.29.1–21.29.9.

20. Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ: **Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes**. *Nucleic Acids Res* 2017, **45**:e41.

21. Zhao H, Zhang W, Zhang T, Lin Y, Hu Y, Fang C, Jiang J: **Genome-wide MNase hypersensitivity assay unveils distinct classes of open chromatin associated with H3K27me3 and DNA methylation in Arabidopsis thaliana**. *Genome Biol* 2020, **21**:24.

22. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE: **High-resolution mapping and characterization of open chromatin across the genome**. *Cell* 2008, **132**:311–322.

23. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD: **FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin**. *Genome Res* 2007, **17**:877–885.

24. Gross DS, Garrard WT: **Nuclease hypersensitive sites in chromatin**. *Annu Rev Biochem* 1988, **57**:159–197.

25. Henikoff S: **Nucleosome destabilization in the epigenetic regulation of gene expression**. *Nat Rev Genet* 2008, **9**:15–26.

26. Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, Zumstein K, Woodhouse M, Bubb K, Dorrity MW, *et al.*: **Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules**. *Plant Cell* 2018, **30**:15–36.

27. Takaku M, Grimm SA, Shimbo T, Perera L, Menafra R, Stunnenberg HG, Archer TK, Machida S, Kurumizaka H, Wade PA: **GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler**. *Genome Biol* 2016, **17**.

28. Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK: **Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape**. *Nat Biotechnol* 2014, **32**:171–178.

29. Yamaguchi N: **LEAFY, a pioneer transcription factor in plants: a mini-review**. *Front Plant Sci* 2021, **12**:701406.

30. Meyer CA, Liu XS: **Identifying and mitigating bias in next-generation sequencing methods for chromatin biology**. *Nat Rev Genet* 2014, **15**:709–721.

31. Li Z, Schulz MH, Zenke M, Costa IG: Identification of transcription factor binding sites using ATAC-seq. [date unknown], doi: 10.1101/362863.

32. Karabacak Calviello A, Hirsekorn A, Wurmus R, Yusuf D, Ohler U: **Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling**. *Genome Biol* 2019, **20**:42.

33. Alexandre CM, Urton JR, Jean-Baptiste K, Huddleston J, Dorrity MW, Cuperus JT, Sullivan AM, Bemm F, Jolic D, Arsovski AA, *et al.*: **Complex relationships between chromatin accessibility, sequence divergence, and gene expression in Arabidopsis thaliana**. *Mol Biol Evol* 2018, **35**:837–854.

34. Tannenbaum M, Sarusi-Portuguez A, Krispil R, Schwartz M, Loza O, Benichou JIC, Mosquna A, Hakim O: **Regulatory chromatin landscape in roots uncovered by coupling INTACT and ATAC-seq**. *Plant Methods* 2018, **14**:113.

35. ** Marand AP, Chen Z, Gallavotti A, Schmitz RJ: **A cis-regulatory atlas in maize at single-cell resolution**. *Cell* 2021, **184**: 3041–3055. e21.

This study applied single-cell RNA and ATAC sequencing in maize to investigate regulatory principles dictating ACR to expression dynamics at the finest resolution to date

36. Farmer A, Thibivilliers S, Ryu KH, Schiefelbein J, Libault M: **Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis roots at the single-cell level**. *Mol Plant* 2021, **14**: 372–383.

37. * Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ: **The prevalence, evolution and chromatin signatures of plant regulatory elements**. *Native Plants* 2019, **5**:1250–1259.

An investigation of common regulatory dynamics in an evolutionary framework across thirteen angiosperms. The authors present strong evidence for the involvement of transposable elements and CRE evolution.

38. * Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, *et al.*: **Widespread long-range cis-regulatory elements in the maize genome**. *Native Plants* 2019, **5**:1237–1249.

The authors document ubiquitous long-range cis-regulatory interactions in maize. These interactions had previously been considered rare exceptions.

39. Judd J, Sanderson H, Feschotte C: **Evolution of mouse circadian enhancers from transposable elements**. *Genome Biol* 2021, **22**:193.

40. Zhao H, Zhang W, Chen L, Wang L, Marand AP, Wu Y, Jiang J: **Proliferation of regulatory DNA elements derived from transposable elements in the maize genome**. *Plant Physiol* 2018, **176**:2789–2803.

41. Wang X, Aguirre L, Rodríguez-Leal D, Hendelman A, Benoit M, Lippman ZB: **Dissecting cis-regulatory control of quantitative trait variation in a plant stem cell circuit**. *Native Plants* 2021, **7**: 419–427.

42. Payne A, Holmes N, Rakyan V, Loose M: **BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files**. *Bioinformatics* 2019, **35**:2193–2198.

43. * Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, *et al.*: **The complete sequence of a human genome**. *bioRxiv* 2021, https://doi.org/10.1101/2021.05.26.445798.

The Telomere 2 Telomere consortium releases a landmark achievement, having completely sequenced the human genome without gaps two decades after the human genome project.

44. Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang W-C, Zhang L, Zhang X, Tang R, *et al.*: **The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication**. *Nat Genet* 2019, **51**:865–876.

45. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, *et al.*: **Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome**. *Nat Biotechnol* 2019, **37**:1155–1162.

46. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q: **Opportunities and challenges in long-read sequencing data analysis**. *Genome Biol* 2020, **21**:30.

47. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M: **Readfish enables targeted nanopore sequencing of gigabase-sized genomes**. *Nat Biotechnol* 2021, **39**:442–450.

48. Andermann T, Torres Jiménez MF, Matos-Maraví P, Batista R, Blanco-Pastor JL, Gustafsson ALS, Kistler L, Liberal IM, Oxelman B, Bacon CD, *et al.*: **A guide to carrying out a phylogenomic target sequence capture project**. *Front Genet* 2019, **10**:1407.