Zhang et al. *Zool. Res.* 2022, 43(5): 805−812
https://doi.org/10.24272/j.issn.2095-8137.2022.111

ZR | Zoological Research

Article Open Access

Coevolutionary insights between promoters and transcription factors in the plant and animal kingdoms

Jing-Song Zhang[1,#,*], Hai-Quan Wang[2,#], Jie Xia[1,3,#], Kun Sha[4], Shu-Tao He[1], Hao Dai[1], Xiao-Hu Hao[5], Yi-Wei Zhou[6], Qiu Wang[1], Ke-Ke Ding[7], Zhang-Lei Ju[1], Wen Wang[8,9,*], Luo-Nan Chen[1,10,11,*]

[1] *Key Laboratory of Systems Biology*, *Shanghai Institute of Biochemistry and Cell Biology*, *Center for Excellence in Molecular Cell Science*, *Chinese Academy of Sciences*, *Shanghai* 200031, *China*

[2] *Department of General Surgery*, *Shanghai General Hospital of Shanghai Jiao Tong University School of Medicine*, *Shanghai* 200080, *China*

[3] *College of Information Engineering*, *Zhejiang University of Technology*, *Hangzhou*, *Zhejiang* 310023, *China*

[4] *Naval Healthcare Information Center*, *Faculty of Military Health Services*, *Naval Medical University*, *Shanghai* 200433, *China*

[5] *Bioinformatics Core of Excellence Department*, *GenScript Biotech Corporation*, *Nanjing*, *Jiangsu* 211110, *China*

[6] *Waigaoqiao Free Trade Zone*, *Wuxi Biologics*, *Shanghai* 200131, *China*

[7] *Department of Cardiology*, *Tongji Hospital*, *Tongji University School of Medicine*, *Shanghai* 200065, *China*

[8] *School of Ecology and Environment*, *Northwestern Polytechnical University*, *Xi'an*, *Shaanxi* 710072, *China*

[9] *State Key Laboratory of Genetic Resources and Evolution*, *Kunming Institute of Zoology*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650223, *China*

[10] *Key Laboratory of Systems Health Science of Zhejiang Province*, *School of Life Science*, *Hangzhou Institute for Advanced Study*, *University of Chinese Academy of Sciences*, *Chinese Academy of Sciences*, *Hangzhou*, *Zhejiang* 310024, *China*

[11] *Guangdong Institute of Intelligence Science and Technology*, *Zhuhai*, *Guangdong* 519031, *China*

ABSTRACT

The divergence and continuous evolution of plants and animals contribute to ecological diversity. Promoters and transcription factors (TFs) are key determinants of gene regulation and transcription throughout life. However, the evolutionary trajectories and relationships of promoters and TFs are still poorly understood. Here, we conducted extensive analysis of large-scale multi-omics sequences in 420 animal species and 223 plant species spanning nearly a billion years of evolutionary history. Results showed that promoter GC-content and TF isoelectric points, as features/signatures that accompany long biological evolution, exhibited increasing growth in animal cells but a decreasing trend in plant cells. Furthermore, the evolutionary trajectories of promoter and TF

Received: 15 June 2022; Accepted: 18 August 2022; Online: 18 August 2022

Foundation items: This work was supported by the National Key Research and Development Program of China (2017YFA0505500 to L.N.C., 2017YFC0909502 to J.S.Z.); Strategic Priority Research Program of the Chinese Academy of Sciences (XDB38040400 to L.N.C., XDB13000000 to W.W.); National Science Foundation of China (12131020 and 31930022 to L.N.C, 61602460 to J.S.Z.); Major Key Project of PCL (PCL2021A12 to L.N.C.); Special Fund for Science and Technology Innovation Strategy of Guangdong Province (2021B0909050004 and 2021B0909060002 to L.N.C.); and Fundamental Research Funds for the Central Universities (3102019JC007 to W.W.)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: jingsong.zhang@sibcb.ac.cn; wwang@mail.kiz.ac.cn; lnchen@sibs.ac.cn

signatures in the animal kingdom provided further evidence that Mammalia as well as Aves evolved directly from the ancestor Reptilia. The strong correlation between promoter and TF signatures indicates that promoters and TFs formed antagonistic coevolution in the animal kingdom, but mutualistic coevolution in the plant kingdom. The distinct coevolutionary patterns potentially drive the plant-animal divergence, divergent evolution and ecological diversity.

Keywords: Molecular evolution; Coevolution; Promoter; Transcription factor; Plant-animal divergence

INTRODUCTION

Promoters, transcription factors (TFs), and their interactions are vital for transcriptional regulation and affect nearly all stages of the cell life cycle (Mirny, 2010). The evolution of promoters and TFs (Thomas & Chiang, 2006) in eukaryotic cells has occurred approximately 1.6–2.1 billion years (Bengtson et al., 2017; Zhu et al., 2016). However, our knowledge concerning the evolutionary trajectories and relationships of promoters and TFs remains limited. In this study, we explored promoter and TF evolution over nearly a billion years of evolutionary history by analyzing their signatures.

Promoter stability is critical for the initiation of gene transcription. The GC-content (see Materials and Methods) of promoters has a substantial effect on DNA molecular stability and gene activity because the connection between G and C bases (three hydrogen bonds, G≡C) is stronger than that between A and T bases (two hydrogen bonds, A=T), and stacking energy is more favorable for GC pairs than for AT pairs (Yakovchuk et al., 2006). GC-content is variable within a given genome (Furey & Haussler, 2003) and across organisms (Birdsell, 2002), and is well developed in biological evolution (Blanc-Mathieu et al., 2017; Clément et al., 2015; Shen et al., 2020; Su et al., 2011; Tan et al., 2011; Zahn, 2015). Therefore, GC-content is a useful signature for exploring the evolution of promoters. In addition to GC-content, the isoelectric point (pI) of proteins is a crucial physicochemical property and a major biochemical factor (Dika et al., 2015; Liu et al., 2009) affecting the structure and functions of proteins (including TFs). Thus, we explored the evolutionary trajectories and relationships of TFs and their corresponding promoters in an entire phylogeny by assessing the variation in their signatures, i.e., isoelectric points of TFs and GC-content of promoters.

The benchmark datasets included both genome and proteome sequences (from the AnimalTFDB (Hu et al., 2019), PlantTFDB (Jin et al., 2017), Ensembl (Hunt et al., 2018), EnsemblPlants (Kersey et al., 2018), and UniProt databases (The UniProt Consortium, 2019), see Supplementary Table S1) and covered almost the entire evolutionary history of the plant and animal kingdoms. We performed extensive multi-omics sequence analysis of promoter and TF signatures to evaluate their evolutionary trajectories. Results showed that the evolutionary trajectories of promoter and TF signatures shared a strikingly synchronous increase in animal cells but a synchronous decreasing trend in plant cells. These signature trajectories provide additional evidence that both Mammalia and Aves originated directly from Reptilia.

MATERIALS AND METHODS

Data acquisition and preprocessing

The genome and proteome sequence datasets were obtained from several benchmark databases (i.e., AnimalTFDB (Hu et al., 2019), PlantTFDB (Jin et al., 2017), Ensembl (Hunt et al., 2018), EnsemblPlants (Kersey et al., 2018), and UniProt (The UniProt Consortium, 2019)). Promoters are generally located upstream of the transcription start sites (TSSs) and typically contain 1 000–5 000 bases. As 2 000 bases are commonly used as a gene promoter, we selected 2 000 bases as the upstream promoter sequence.

It is inevitable that information for some sites is ambiguous due to site mutations or limitations of sequencing depth. For example, many sites are labeled N in genome sequences or X in proteome sequences (Malde, 2008). Noise from indeterminate nucleic acids and amino acids was considered in our study to increase accuracy.

GC-content

GC-content is generally the percentage of guanine or cytosine in a DNA or RNA molecule (Kudla et al., 2006; Šmarda et al., 2014; Smith, 2009). In our study, for a given promoter sequence, GC-content is the sum of the percentages of guanine and cytosine:

$$GC-content = \frac{Number\ of\ G + Number\ of\ C}{Length\ of\ sequence - Number\ of\ N} \quad (1)$$

where *N* denotes uncertain sites.

TF and protein isoelectric points

Similar to promoter data, the TF and proteome sequences also contain noise. Given protein sequence S and isoelectric point values of 20 amino acids (Supplementary Table S2) at 25 °C, the mean isoelectric point of protein S was calculated as:

$$pI = \frac{\sum_{i=1}^{|S|} I(S_i)}{|S| - |X|} \quad (2)$$

where $I(S_i)$ represents the isoelectric point of the i^{th} amino acid in sequence *S*; *X* is noise; and $|S|$ and $|X|$ are the lengths of sequences S and X, respectively.

RESULTS

Promoter and TF/TF-cofactor signatures

TFs and TF-cofactors play distinct roles in gene expression (Reiter et al., 2017; Thomas & Chiang, 2006). TFs typically bind to transcription factor binding sites (TFBSs) located in the corresponding promoters to open the double DNA strands of a gene and further control the rate of transcription of genetic information from DNA to mRNA. As intermediary proteins, TF-

cofactors are recruited by TFs to activate RNA polymerase II, thereby modulating the expression of genes. Why do they perform differential functions? The difference in spatial conformation between TF and TF-cofactor molecules is an important factor affecting their functions (Garcia et al., 2019; Frankel & Kim, 1991; Gonzalez, 2016; Liu et al., 2001). The electrical properties of amino acids are crucial physicochemical properties that shape the specific spatial conformation of proteins. Therefore, we estimated the electrical properties of TFs/TF-cofactors (97 animal species) and whole proteomes (74 animal species) in terms of isoelectric point.

Interestingly, we found that the mean isoelectric points of TFs were significantly higher than those of TF-cofactors at all phylogenetic levels of evolution (Figure 1A, total P=4.68E-71), and the combined isoelectric points of TFs and TF-cofactors were also significantly higher (total P=1.04E-18) than those of the corresponding whole proteomes (except for Arthropoda). Based on the total mean isoelectric point (=6.066 at 25 °C) of all whole proteomes, we found that amino acids with higher isoelectric points (>6.066 at 25 °C) tended to be positively charged (alkalinity), whereas those with lower isoelectric points (<6.066 at 25 °C) tended to be negatively charged (acidity) under the same physiological conditions. Thus, TFs/TF-cofactors displayed relatively stronger positive charges compared to the corresponding whole proteomes. DNA exhibits one intrinsic negative charge per base at its sugar-phosphate backbone (Fritz et al., 2002). Therefore, TF/TF-cofactor proteins with high isoelectric points may preferentially access DNA strands.

We then assessed the GC-content of promoter sequences as well as exon and coding sequences in 249 animal species. Results showed that GC-content was significantly lower in promoter sequences than in exon and coding sequences across the tree of life (Figure 1B), indicating that promoters may interact with TFs more readily. Taken together, the above findings suggest that isoelectric point and GC-content, as signatures, can represent protein family- and DNA sequence-specific physicochemical properties, respectively. Notably, TFs/TF-cofactors with higher isoelectric points are more likely to interact with promoters (TFBSs), especially those with lower GC-content. This is because TFs/TF-cofactors hold relatively stronger positive charges compared to corresponding whole proteomes, while DNA exhibits one intrinsic negative charge per base at its sugar-phosphate backbone. Furthermore, promoters with lower GC-content have lower stacking energy, which is beneficial for TF binding.

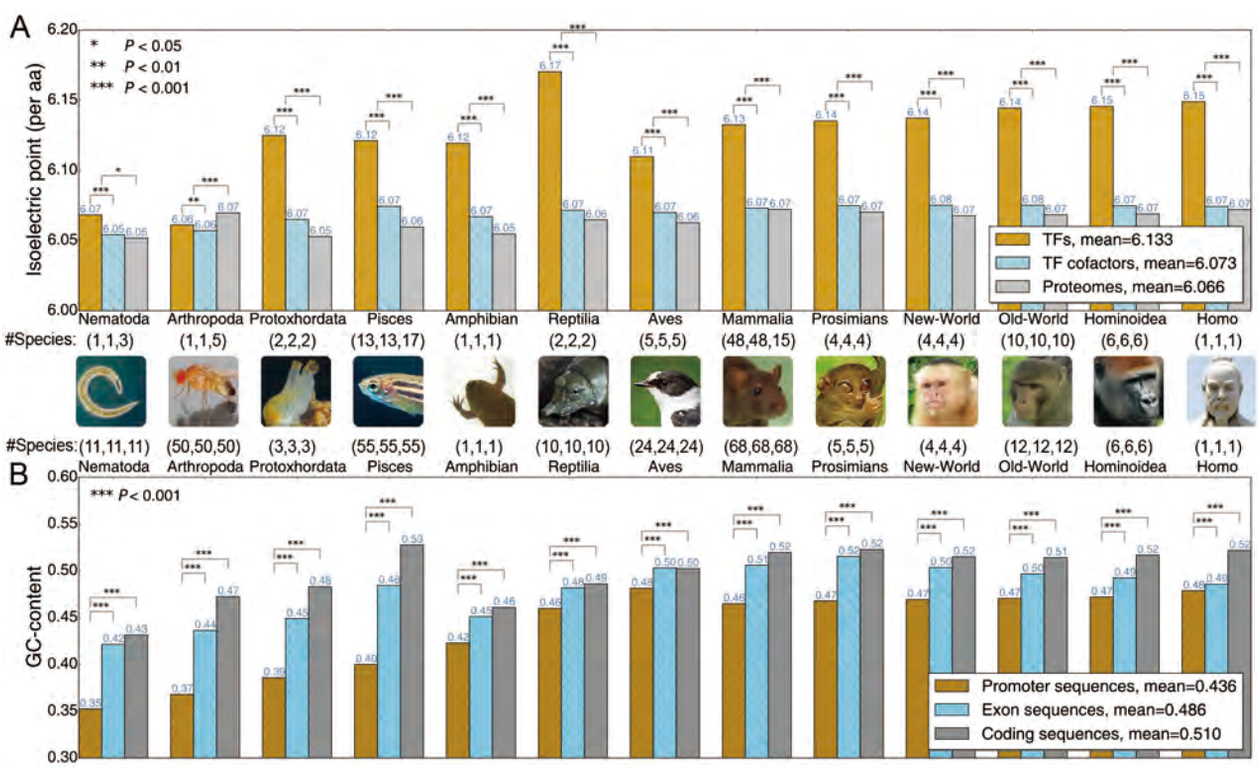

Figure 1 TF Isoelectric points and promoter GC-content as signatures

A, B: X-axis labels show species categories, which follow a phylogenetic relationship (Figure 2B). Animal logos under/over x-axis labels represent corresponding categories. Each triple like (1,1,3), or (11,11,11) element between x-axis labels and animal logos represents species number of TF/promoter sequences, TF-cofactor/exon sequences, and whole-proteome/coding sequences, respectively. A: TFs show significantly higher isoelectric points than TF-cofactors across the tree of life (Supplementary Table S3). Isoelectric points of groups of TFs and TF-cofactors are also significantly higher than those of corresponding whole proteomes (except Arthropoda). B: In 249 animal species, promoters (8 503 028 sequences) show significantly lower GC-content than that of exons (49 239 643 sequences) and coding sequences (7 311 335 sequences) across the tree of life. As signatures, isoelectric point and GC-content can characterize the specificity of protein families and DNA sequences, respectively.

Increased evolution of animal promoter and TF signatures

As signatures, GC-content and isoelectric point can represent DNA sequence and protein family specificity, respectively (Figure 1; Supplementary Table S3). Here, we first investigated the evolutionary trajectories of GC-content in the promoters of 249 animal species (Nematoda to Homo in Ensembl (Hunt et al., 2018)) grouped by the most probable (overall) evolutionary history. The phylogenetic tree is illustrated in Figure 2B using the representative logos of the animal categories. We found that promoter GC-content clearly

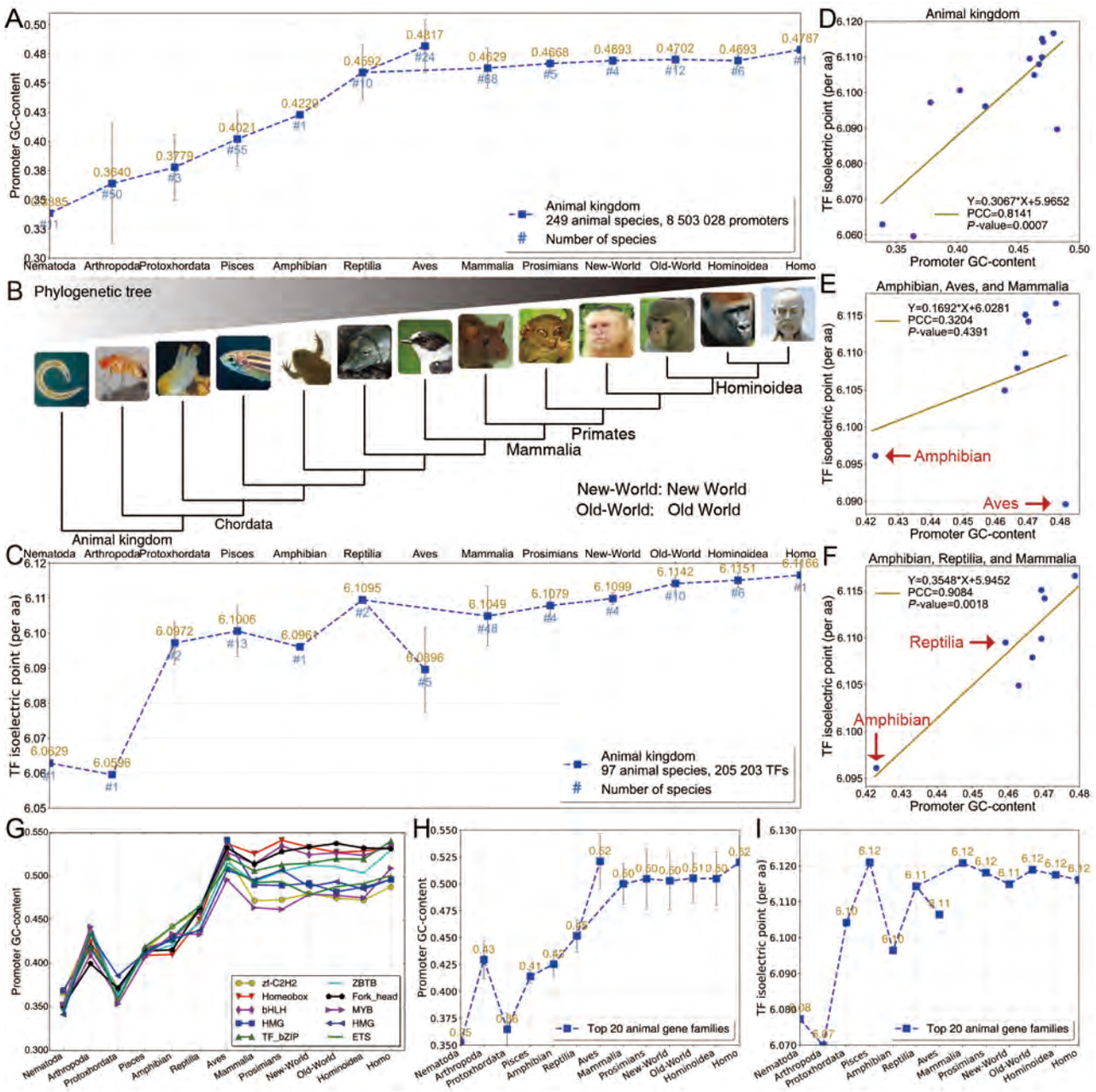

Figure 2 Trajectories and correlations of promoter GC-content and TF isoelectric points in animal evolution

A–C: A and C share the same x-axis labels, each representing an animal species category. Representative animal logos of these categories are listed in order as shown in B. A: GC-content in promoters showing increase in almost all categories (except Aves). B: Phylogenetic tree of animal categories showing (possible) evolutionary history in A and C based on Ensembl. C: Isoelectric points of TFs showing overall increase. D: Correlation between promoter GC-content and TF isoelectric points. E, F: Location of Aves (outlier) is separate from mammals and amphibians (*P*=0.3204), but points for Reptilia are very close to fitted line (PCC=0.9084). G: Promoter GC-content in top 10 gene families (Supplementary Table S4). H: Trend in promoter GC-content in top 20 gene families (Supplementary Table S4). I: Trend in TF isoelectric points in top 20 gene families (Supplementary Table S4). Correlations between promoter GC-content and TF isoelectric points show consensus at genome scale and in gene families. Overall, in evolution of animals, promoter and TF signatures showed synchronous increase and strong correlation.

increased (Figure 2A) in almost all categories (except Aves). Following convention, we use TFs hereafter to refer to TFs and TF-cofactors if not otherwise specified. We next estimated the changes in the isoelectric points of TFs during the same evolutionary process described above and found that isoelectric points showed an overall increasing trend (Figure 2C). In particular, the isoelectric points of mammals exhibited a strong monotonic trend.

Figure 2A, C showed a similar trend. Therefore, we explored the relationship between promoter GC-content and TF isoelectric points using a scatter diagram. Results showed a strong positive correlation between promoter GC-content (Figure 2A) and TF isoelectric points (Figure 2C) in animal cells, with a Pearson correlation coefficient (PCC) of 0.8141 and *P*-value of 0.0007 using two-tailed *t*-test (Figure 2D). Observation revealed that the points for Aves diverged markedly from the global trajectory in Figure 2A, C. We then analyzed the relationship between promoter GC-content and TF isoelectric points in the Amphibia, Aves, and Mammalia group. The scatter diagram (Figure 2E) showed clear separation of Aves (outlier) from Mammalia and Amphibia. Subsequently, we combined Amphibia, Reptilia, and Mammalia as a hypothetical evolutionary lineage and investigated the relationship between promoter GC-content and TF isoelectric points. Results (Figure 2F) showed that the Reptilia point was close to the fitted line (PCC=0.9084, *P*=0.0018). Regression analysis of promoter GC-content and TF isoelectric points provided additional evidence that mammals more likely evolved from reptiles than from birds, as reported in previous research (Janes et al., 2010), thus supporting the correlation between the two signatures.

We further explored the evolutionary trajectories of promoter GC-content and TF isoelectric points in both gene families and genes. Results showed that gene families and major genes exhibited similar evolutionary trends and correlations in promoter and TF signatures at the genome scale (Figure 2G–I; Supplementary Figure S1 and Tables S4, S5). Thus, promoter and TF signatures in animal cells displayed a synchronous increase with strong correlations during evolution.

Decreased evolution of plant promoter and TF signatures

We also explored the relationship between promoter and TF signatures in 223 plant species categorized by evolution. The phylogenetic tree of these plant categories is shown in Figure 3C. Due to the absence of gymnosperm data in the benchmark databases, angiosperm species were divided into three sub-groups, i.e., lower, medium, and higher angiosperms, to increase the number of evolutionary categories.

By tracking the trend of promoter GC-content in plant cells (62 plant species), we unexpectedly found an overall decrease in GC-content (Figure 3A), opposite to the trend found in animal cells. In addition, the TF isoelectric point trends in 161 plant species (Algae to Angiosperm) showed an overall decrease in the isoelectric point curve (Figure 3B), similar to the trend found for promoter GC-content (Figure 3A). Analyzing the scatter diagram (Figure 3D) between promoter GC-content and TF isoelectric points, we found a strong positive correlation (PCC=0.9357, *P*=0.0061) between all plant categories and an equally strong positive correlation (PCC=0.9357) in the Algae, Pteridophyta, and Angiosperm group, indicating that both Pteridophyta and Bryophyta may evolve directly from Algae rather than Pteridophyta evolves from Bryophyta. The evolution of TFs in gene families (Supplementary Table S6) also showed a similar isoelectric point trajectory at the genome scale (Figure 3F, G). Together, promoter GC-content and TF isoelectric point, as signatures accompanying biological evolution, exhibited a downward trend in plant cells.

Coevolution between promoters and TFs

Molecular evolution is a fundamental driver of genetic divergence, ontogenesis, and ongoing trait evolution in species (Cui et al., 2021; Yang et al., 2021). Billions of years of interactions between promoters and TFs have potentially driven their coevolution. Our results showed that paired promoter and TF signatures accompanying evolutionary processes monotonically increased in animal cells but decreased in plant cells, reflecting different evolutionary trajectories of promoters and TFs in the evolution of animals and plants. The strong correlation between promoter and TF signatures suggests that promoters and TFs formed coevolutionary relationships in plant and animal evolution.

In animal cells (Figure 4A, B), promoter GC-content clearly increased during the evolutionary process. In this case, promoter region strands tend to be harder to unwind and transcribe because: (1) GC pairs contain three hydrogen bonds while AT pairs contain only two bonds; and (2) GC-rich regions typically contribute to the base stacking of adjacent bases and therefore block interactions between promoters and TFs (Yakovchuk et al., 2006). The evolutionary increase in the isoelectric points of TFs suggests that TFs carried stronger positive electrical charges during animal evolution, thus providing more opportunities to trigger interactions with promoters, as DNA molecules exhibit an intrinsic negative charge on their double-helix backbone (Fritz et al., 2002). The opposite charges between TFs and DNA molecules increased their attraction and interactions with each other. Thus, promoters protected double-strand DNA from TF unwinding and transcription by increasing promoter GC-content. In contrast, TFs strengthened their own ability to bind to TFBSs by increasing their positive electrical properties. These findings provide potential evidence for parasitism and mutualism between promoters and TFs. Thus, the selective pressures of their physicochemical properties may have driven an evolutionary arms race between promoters and TFs, namely an antagonistic coevolutionary relationship.

In contrast to animal cells, promoter GC-content showed an overall decrease in plant cells (Figure 4C, D). This decrease may be beneficial for TF function during transcription. Interestingly, the positive electrical property of TFs was weaker, showing a similar trend as promoter GC-content. The simultaneous weakening of promoter stability and TF activity may benefit both partners, thus retaining symbiotic evolution of molecules in plant cells. Taken together, the altruistic interactions between promoters and TFs resulted in the mutualistic coevolution in plant cells.

Overall, promoters and TFs showed an antagonistic coevolutionary relationship induced by syntropic changes in

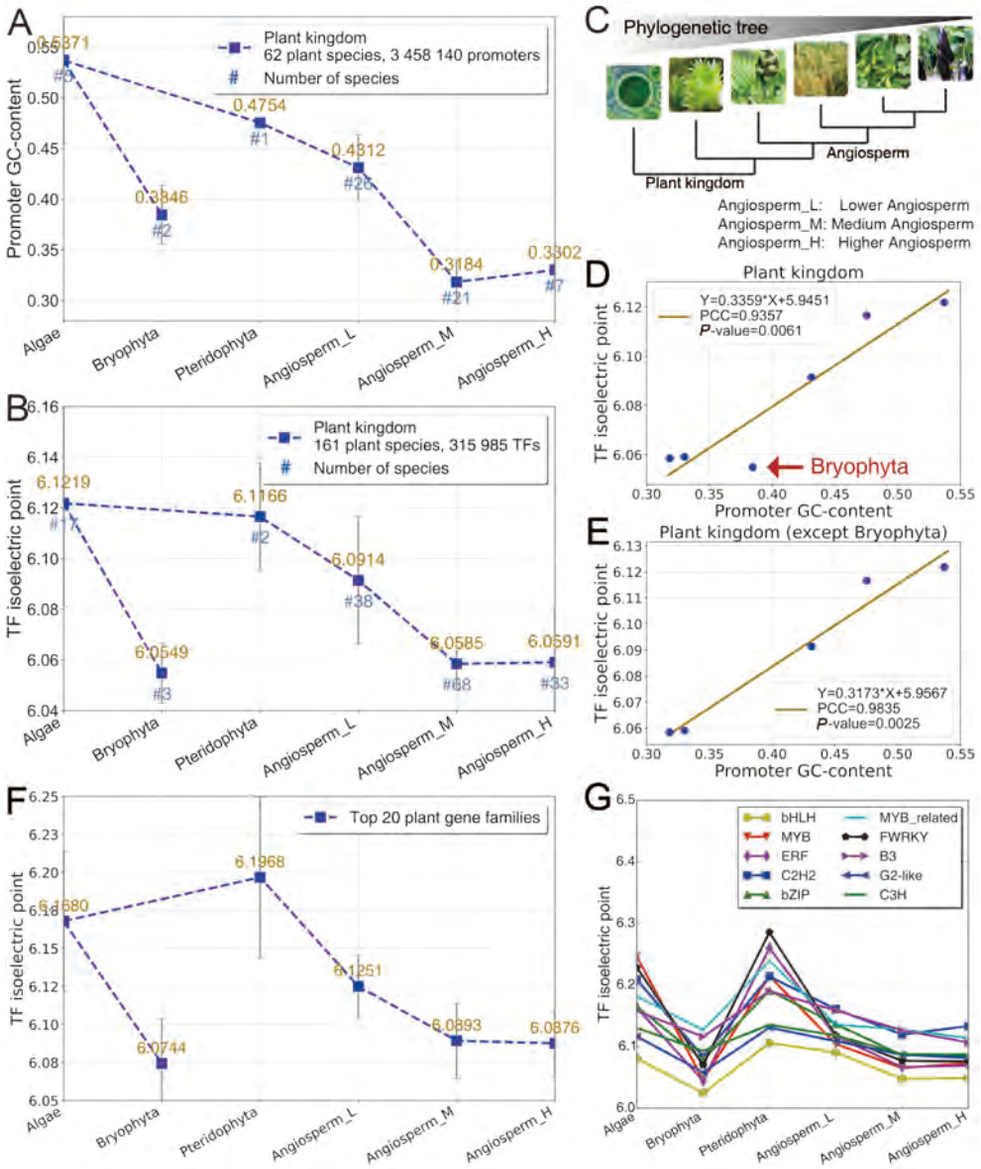

Figure 3 Trajectories and correlations of promoter GC-content and TF isoelectric points in plant evolution

A–C: X-axes share the same plant species categories, and representative logos are shown in order in C. Numbers to the right of "#" indicate number of species. Phylogenetic tree in C of plant categories shows evolutionary history based on EnsemblPlants (Kersey et al., 2018). Both promoter GC-content and TF isoelectric points decreased overall (except in Bryophyta). A and B are characterized by a similar trend. D, E: Strong positive correlation (PCC=0.9357) between promoter GC-content and TF isoelectric points in all plant categories and strong positive correlation (PCC=0.9835) in Algae, Pteridophyta, and Angiosperm group are shown, indicating that both Pteridophyta and Bryophyta may evolve directly from Algae rather than Pteridophyta evolves from Bryophyta. F: Trend in TF isoelectric points in top 20 gene families (Supplementary Table S6). G: TF isoelectric points in top 10 gene families (Supplementary Table S6). Overall, in evolution of plants, promoter and TF signatures showed synchronous changes and strong correlation.

promoter and TF signatures in animal cells, but exhibited a mutualistic coevolutionary relationship due to the altruistic features of their signatures in plant cells.

DISCUSSION

Our extensive analysis of multi-omics sequences of animal and plant species revealed several intriguing patterns. Promoter GC-content and TF isoelectric points, as signatures accompanying biological evolution, showed a continuing increase in animal cells but a decreasing trend in plant cells. The evolutionary trajectories of promoter and TF signatures in the animal kingdom provide further evidence that Mammalia as well as Aves evolved directly from a common ancestor in Reptilia. In addition, the strong correlation between promoter and TF signatures suggested that promoters and TFs formed an antagonistic coevolutionary relationship in the animal kingdom, but a mutualistic coevolutionary relationship in the

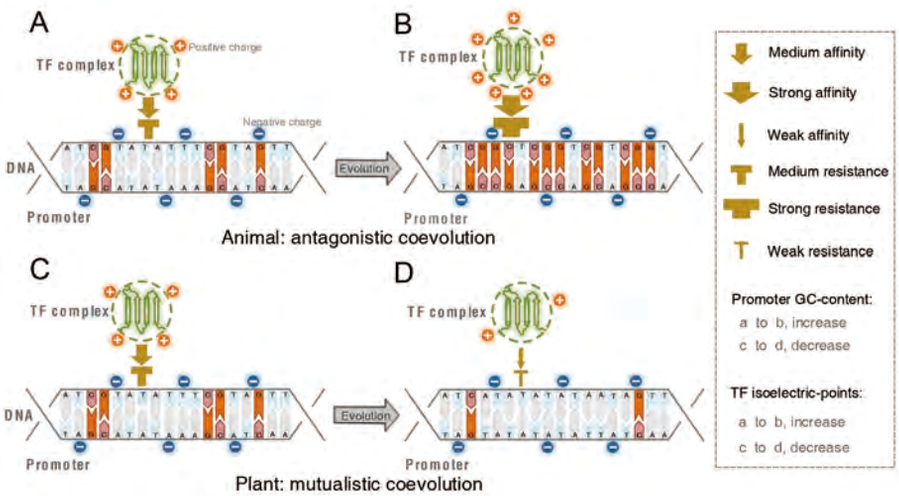

Figure 4 Coevolution between promoters and TFs (TF complexes) in plant and animal cells

A–D: Represent four different interaction statuses of promoter-TF pairs during regulation of gene expression in evolutionary processes. Negative charges around sugar-phosphate backbone of promoters are labeled. From A to B, increase in positive charges carried by TFs indicates that TFs tend to have stronger electrical property and alkalinity. Transition of A to B conveys potential antagonistic coevolution in animal cells. In contrast, decrease in positive charges from C to D indicates TFs have weaker electrical property and alkalinity. Transition of C to D conveys potential mutualistic coevolution in plant cells.

plant kingdom. Molecular adaptation (Guo et al., 2021; Peng et al., 2021) and evolution are fundamental drivers of species genetic divergence, ontogenesis, and trait evolution. Due to the vital roles of promoters and TFs in transcriptional regulation in eukaryotic cells, the distinct evolutionary trajectories and strong correlations in signatures may highlight genetic divergence between animals and plants from their common ancestor. Under natural selection, pervasive antagonistic coevolution may be a critical pattern and important driver of species diversity in the animal kingdom (~7.77 million species (Mora et al., 2011; Strain, 2011)) compared to the plant kingdom (~298 000 species (Mora et al., 2011)). These results provide a strong basis for further exploration of plant-animal evolution using conserved patterns (Zhang et al., 2015, 2016, 2020), (co-)mutations (Zhang et al., 2021b), gene regulatory networks (Dai et al., 2020), and network biomarkers (Shi et al., 2021, 2022; Zhang et al., 2021a). This study not only provides insights into the interactions between promoters and TFs, but also advances our understanding of plant-animal divergence, divergent evolution and ecological diversity.

DATA AVAILABILITY

The raw sequencing data reported in this paper were deposited in the Ensembl (http://asia.ensembl.org/index.html), EnsemblPlants (http://plants.ensembl.org/index.html), AnimalTFDB (http://bioinfo.life.hust.edu.cn/AnimalTFDB#!/), and PlantTFDB (http://planttfdb.cbi.pku.edu.cn/) websites. They are also available from the corresponding author upon reasonable request.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

L.N.C., W.W., and J.S.Z. designed the study. J.S.Z., H.Q.W., J.X., and K.S. coded programs and analyzed output data. J.S.Z. designed the phylogenetic tree. S.T.H. provided the original datasets. Y.W.Z. performed the statistical analysis. H.D. designed the isoelectric point experiments. J.S.Z. and X.H.H. designed the evolutionary mechanism figure. Q.W., Z.L.J., and K.K.D. polished the manuscript. W.W. supervised the experiments. J.S.Z. wrote the manuscript. All authors participated in result interpretation and discussion. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank Profs. Man-Yuan Long, Aaron Hsueh, Jian-Mei Guo, Tao Zeng, Zhi-Xi Su, and Fu-Yuan Zhang for useful comments on the manuscript. We also thank LetPub for linguistic assistance during the preparation of this manuscript.

REFERENCES

Bengtson S, Sallstedt T, Belivanova V, Whitehouse M. 2017. Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS Biology*, **15**(3): e2000735.

Birdsell JA. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Molecular Biology and Evolution*, **19**(7): 1181−1197.

Blanc-Mathieu R, Krasovec M, Hebrard M, Yau S, Desgranges E, Martin J, et al. 2017. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Science Advances*, **3**(7): e1700239.

Clément Y, Fustier MA, Nabholz B, Glémin S. 2015. The bimodal distribution of genic GC content is ancestral to monocot species. *Genome*

Biology and Evolution, **7**(1): 336−348.
Cui Y, Liu ZL, Li CC, Wei XM, Lin YJ, You L, et al. 2021. Role of juvenile hormone receptor Methoprene-tolerant 1 in silkworm larval brain development and domestication. *Zoological Research*, **42**(5): 637−649.
Dai H, Jin QQ, Li L, Chen LN. 2020. Reconstructing gene regulatory networks in single-cell transcriptomic data analysis. *Zoological Research*, **41**(6): 599−604.
Dika C, Duval JFL, Francius G, Perrin A, Gantzer C. 2015. Isoelectric point is an inadequate descriptor of MS2, Phi X 174 and PRD1 phages adhesion on abiotic surfaces. *Journal of Colloid and Interface Science*, **446**: 327−334.
Frankel AD, Kim PS. 1991. Modular structure of transcription factors: implications for gene regulation. *Cell*, **65**(5): 717−719.
Fritz J, Cooper EB, Gaudet S, Sorger PK, Manalis SR. 2002. Electronic detection of DNA by its intrinsic molecular charge. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(22): 14142−14146.
Furey TS, Haussler D. 2003. Integration of the cytogenetic map with the draft human genome sequence. *Human Molecular Genetics*, **12**(9): 1037−1044.
Garcia MF, Moore CD, Schulz KN, Alberto O, Donague G, Harrison MM, et al. 2019. Structural features of transcription factors associating with nucleosome binding. *Molecular Cell*, **75**(5): 921−932.e6.
Gonzalez DH. 2016. Chapter 1 - Introduction to transcription factor structure and function. *In*: Gonzalez DH. Plant Transcription Factors. Boston: Academic Press, 3–11.
Guo YT, Zhang J, Xu DM, Tang LZ, Liu Z. 2021. Phylogenomic relationships and molecular convergences to subterranean life in rodent family Spalacidae. *Zoological Research*, **42**(5): 671−674.
Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. 2019. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Research*, **47**(D1): D33−D38.
Hunt SE, McLaren W, Gil L, Thormann A, Schuilenburg H, Sheppard D, et al. 2018. Ensembl variation resources. *Database*, **2018**: bay119.
Janes DE, Organ CL, Fujita MK, Shedlock AM, Edwards SV. 2010. Genome evolution in reptilia, the sister group of mammals. *Annual Review of Genomics and Human Genetics*, **11**: 239−264.
Jin JP, Tian F, Yang DC, Meng YQ, Kong L, Luo JC, et al. 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, **45**(D1): D1040−D1045.
Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, et al. 2018. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Research*, **46**(D1): D802−D808.
Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biology*, **4**(6): e180.
Liu MH, Wu DH, Yu SC, Gao CJ. 2009. Influence of the polyacyl chloride structure on the reverse osmosis performance, surface properties and chlorine stability of the thin-film composite polyamide membranes. *Journal of Membrane Science*, **326**(1): 205−214.
Liu Q, Zhang GY, Chen SY. 2001. Structure and regulatory function of plant transcription factors. *Chinese Science Bulletin*, **46**(4): 271−278.
Malde K. 2008. The effect of sequence quality on sequence alignment. *Bioinformatics*, **24**(7): 897−900.
Mirny LA. 2010. Nucleosome-mediated cooperativity between transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(52): 22534−22539.
Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on earth and in the ocean?. *PLoS Biology*, **9**(8): e1001127.
Peng ZL, Yin BX, Ren RM, Liao YL, Cai H, Wang H. 2021. Altered metabolic state impedes limb regeneration in salamanders. *Zoological Research*, **42**(6): 772−782.
Reiter F, Wienerroither S, Stark A. 2017. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development*, **43**: 73−81.
Shen XX, Steenwyk JL, Labella AL, Opulente DA, Zhou XF, Kominek J, et al. 2020. Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Science Advances*, **6**(45): eabd0079.
Shi JF, Aihara K, Chen LN. 2021. Dynamics-based data science in biology. *National Science Review*, **8**(5): nwab029.
Shi JF, Aihara K, Li TJ, Chen LN. 2022. Energy landscape decomposition for cell differentiation with proliferation effect. *National Science Review*: nwac116
Šmarda P, Bureš P, Horová L, Leitch IJ, Mucina L, Pacini E, et al. 2014. Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(39): E4096−E4102.
Smith DR. 2009. Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Molecular Biology*, **71**(6): 627–639.
Strain D. 2011. 8.7 million: a new estimate for all the complex species on earth. *Science*, **333**(6046): 1083.
Su ZX, Huang W, Gu X. 2011. Comment on “positive selection of tyrosine loss in metazoan evolution”. *Science*, **332**(6032): 917.
Tan CSH, Schoof EM, Creixell P, Pasculescu A, Lim WA, Pawson T, et al. 2011. Response to comment on “positive selection of tyrosine loss in metazoan evolution”. *Science*, **332**(6032): 917.
The UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, **47**(D1): D506−D515.
Thomas MC, Chiang CM. 2006. The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, **41**(3): 105−178.
Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research*, **34**(2): 564−574.
Yang H, Lyu B, Yin HQ, Li SQ. 2021. Comparative transcriptomics highlights convergent evolution of energy metabolic pathways in group-living spiders. *Zoological Research*, **42**(2): 195−206.
Zahn LM. 2015. Probing plant evolution by GC content. *Science*, **347**(6220): 385−386.
Zhang CM, Zhang H, Ge J, Mi TY, Cui X, Tu FJ, et al. 2021a. Landscape dynamic network biomarker analysis reveals the tipping point of transcriptome reprogramming to prevent skin photodamage. *Journal of Molecular Cell Biology*, **13**(11): 822−833.
Zhang JS, Guo JM, Zhang M, Yu XT, Yu XQ, Guo WF, et al. 2020. Efficient mining multi-mers in a variety of biological sequences. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **17**(3): 949−958.
Zhang JS, Wang YL, Yang DY. 2015. CCSpan: mining closed contiguous sequential patterns. *Knowledge-Based Systems*, **89**: 1−13.
Zhang JS, Wang YL, Zhang C, Shi YY. 2016. Mining contiguous sequential generators in biological sequences. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **13**(5): 855−867.
Zhang JS, Zhang Y, Kang JY, Chen SY, He YQ, Han BH, et al. 2021b. Potential transmission chains of variant B. 1.1. 7 and co-mutations of SARS-CoV-2. *Cell Discovery*, **7**(1): 44.
Zhu SX, Zhu MY, Knoll AH, Yin ZJ, Zhao FC, Sun SF, et al. 2016. Decimetre-scale multicellular eukaryotes from the 1.56-billion-year-old Gaoyuzhuang Formation in North China. *Nature Communications*, **7**(1): 11500.