Trends in Plant Science

CellPress

Forum

Synteny detection, visualization, and its trending applications

Xi Zhang (张曦)[1,2,*] and David R. Smith[3,*]

Synteny detection analyses facilitate comparative genomics, especially when investigating gene duplications, genomic rearrangements, and ancient whole-genome duplication events. In recent years, major advancements have been seen in synteny detection and visualization. In this forum article, we explore the trending role that these tools play in gene-duplication detection, pangenome graph construction, and deep-learning-based cross-species transcript prediction.

Synteny and collinearity detection in the comparative genomics era

A key goal of comparative genomics is understanding the genetic similarity and evolutionary relationships among and within species. This includes analyzing whether homologous genes are conserved on the same chromosome (**synteny**) (see Glossary) and detecting if they are in a conserved order (**collinearity**) between distinct taxa [1]. There are different categories of homologous genes (i.e., **orthologs**, **paralogs,** and **syntelogs**). Orthologs, for example, are typically derived from speciation events, while paralogs descend from a common ancestor via duplication events. Identifying homologous genes in highly conserved order and conserved nucleotide sequences of chromosomes (**collinear blocks**) has been useful for understanding gene duplication, **genomic rearrangement**, and ancient **whole-genome duplication (WGD) events**, as well as for annotating the genomes of nonmodel species. This is especially true for plant genomes, many of which have undergone multiple WGD events, often resulting in polyploidy. Similarly, in animals, some genes (e.g., Hox genes) are highly conserved across species, emphasizing the evolutionary importance of gene collinearity [2]. Other types of **small-scale duplications (SSDs)** alongside chromosomal rearrangements, such as indels, inversions, and translocations, are significant evolutionary forces shaping gene families and genome architecture [3,4].

Given the importance of identifying collinear genes when inferring SSDs and WGD events, many bioinformatics tools apply reciprocal Basic Local Alignment Search Tool (BLAST) results to construct chains of collinear gene pairs [3]. For instance, some algorithms predict collinearity by grouping neighboring pairs from a homologous gene matrix via all-versus-all protein BLAST reciprocal searches. However, this approach can fail to identify collinearity when it arises from chromosomal rearrangements and **tandem duplications** (TDs) [4].

Consequently, various specialized bioinformatic tools have been developed to overcome these disadvantages and accurately detect gene collinearity (i.e., homologous genes in a conserved gene order) [3–5]. These tools typically employ dynamic programming algorithms to create chains of collinear gene pairs (≥5) around **anchor genes**, using a scoring matrix that rewards or penalizes the distance between anchors [4]. Currently available synteny detection tools (e.g., MCScanX [4]) can differ in various ways, including the types of input files they employ and the visualization of results (Figure 1A and Box 1).

Seeing is believing: Leveraging interactive synteny and collinearity visualization

There are many different bioinformatics tools for visualizing the results of synteny

Glossary

Anchor genes: homologous gene pairs in regions with the same gene order and orientation across different genomes, either within or between species; they are used as reference points to detect and align homologous chromosomal regions.
Collinearity: a specific type of synteny whereby homologous genes are arranged in the same order on the chromosomes of different genomes, either within or between species.
Collinear blocks: the genomic sequences/regions that contain the homologous genes shared in the regions of chromosomes with the same order and orientation; they appear in multiple genomes or between subgenomes of a polyploid; they are free from large-scale chromosomal rearrangements (e.g., inversions, transpositions, or deletions).
Collinear orthologs: they refer to genes derived from a common ancestor that appear in the same linear order on chromosomes in different species. Most people now alternatively use the term 'syntenic orthologs'.
Dispersed: it refers to a duplicate gene is translocated to a nonhomologous chromosomal region.
Genomic rearrangement: a large-scale (>50 bp to megabases) DNA change in gene content and order, which is usually caused by deletions, duplications, inversions, or translocations during cell division. However, small structural variants generally refer to smaller indels or insertions near the 50 bp threshold.
Macrosynteny: it operates at a different scale compared with synteny; it specifically refers to large-scale, chromosomal-level segments where genes are on the same chromosome, but their precise order and location may be heavily rearranged among different species or genomes.
Microsynteny: the conservation of small-scale gene content in synteny regions, which often decays more slowly than microsynteny. It is different from collinearity, which focuses on the strict, ordered arrangement of genes across larger genomic regions.
Nonanchor genes: homologous genes located in regions that do not maintain collinearity between genomes, which have undergone rearrangements or gene duplication events that play an important role in gene family expansion and genome evolution.
Orthologs: homologous genes in different species derived from a single gene in their most recent common ancestor.
Orthogroups: the collection of genes from different species that all descend from a single ancestral gene in the last common ancestor of those species.
Pangenome: the entire set of genes or DNA sequences found in all individuals of a species, combining both the core genome (shared by all) and the dispensable genome (present in some individuals but not in others).
Paralogs: homologous genes in one species derived from a duplication event. Out-paralogs arise before speciation, while in-paralogs arise after.

Proximal: it refers to a duplicate gene is located close to its parental gene on the same chromosome but is separated by a few intervening genes (usually fewer than ten).
Reticulation: also known as reticulate evolution—it refers to the origination of a lineage through the partial merging of two or more ancestral lineages. It describes a complex, network-like pattern rather than a simple branching tree of life.
Small-scale duplications (SSDs): a type of gene duplication that involves one or a few genes, which can create small evolutionary changes, facilitating functional divergence.
Syntelogs: a typical gene homology where sets of genes are derived from the same ancestral genomic region.
Synteny: the homologous genes located on the same chromosomes or in the same genomic region across different genomes, either within or between species. It should be noted that in modern genomics, the word synteny often matches collinearity to define homologous genes that are in the same order.
Syntenic orthologs: also known as collinear orthologs in modern genomics, are used to describe orthologous genes found in the same order on the chromosomes across different species. Most people now use this term to refer to collinear orthologs.
Tandem duplications: it refers to a duplicate gene is adjacent to its parental gene, which is a result of unequal chromosomal crossing over.
Transposed duplicates: a segment of DNA is copied or cut and inserted into another chromosomal region, which is generated by a DNA-based transposition event or a retrotransposition event.
Whole-genome duplication (WGD) events: a major evolutionary mechanism in which an organism acquires extra sets of chromosomes, yielding a polyploid with increased genome complexity.

analyses. Among the most popular is a 'circular plot', whereby chromosomes from two genomes are arranged in a circle and the shared conserved genomic sequences are connected with colored ribbons (e.g., Circos plot). Bar plots, dot plots, and dual synteny plots are also commonly used for visualizing synteny [3] (Figure 1B).

Each visualization tool has its limitations. For example, some do not support interactive explorations, synteny analyses of multiple genomes, or synteny relationships on multiple genome scales, such as plant genomes that have undergone WGD events. Because synteny-based comparative genomic analysis relies on human interpretation and interactive interfaces, it is important for tools to be designed to visualize multiple aspects of collinear blocks on chromosomes, including their location, size, and orientation. This can help researchers explore and interpret the consequences of **macrosynteny** or **microsynteny** [6].

Recently, a computational research group (working closely with biologists) developed two innovative and interactive tools for exploring and visualizing gene synteny: SynVisio and AccuSyn [6]. Both tools contain interactive web browsers, are fully compatible with the results from MCScanX [4], are frequently updated with new features, and together have greatly improved synteny analyses of crop genomes. SynVisio is designed for visualizing synteny in parallel plots and dot plots generated from paired genomes or stacked parallel plots or hive plots from multiple genomes. AccuSyn generates a circular plot for the genomes as well as a side-by-side plot for the collinear blocks, thus providing an overview of the genome and allowing users to explore the specific genes on the collinear blocks (Figure 1C and Box 1).

Decoding paralogs and pangenome from synteny-based analysis

User-friendly bioinformatics tools for visualizing synteny and collinearity provide an array of options and uses for downstream analyses. Synteny detection tools, for example, are commonly used for inferring ancient WGD events by looking at collinear chromosomal regions, but they are also useful for many other evolutionary analyses [4,5]. This is because anchor genes located on collinear blocks are typically under stronger purifying selection [4]. However, **nonanchor genes** often provide evidence for patterns of gene gain, loss, or transposition [5].

Take, for instance, the tool MCScanX [4]. It can perform many synteny-relevant evolutionary analyses, such as inferring modes of gene duplication (i.e., WGD, tandem, **proximal**, and **dispersed**), exploring nucleotide substitution rates (K_a/K_s) of collinear genes, detecting **collinear orthologs**, and visualizing tandem arrays and collinearity of gene families. Indeed, many of these synteny-associated analyses have already been integrated into third-party tools (Figure 1D). For example, the tools Doubletrouble [11] and DupLiCate [12] have integrated MCScanX and MCScanX-transposed, allowing them to distinguish and classify specific categories of gene duplication (e.g., WGD vs. Tandem).

Similarly, **pangenome** analyses, focusing on the entire gene set in a taxon or a clade, have been applied to distinguish genes shared by all individuals (core genome) and those unique to some (accessory genome). Recently, a synteny-based comparative pangenomic framework was established via SynPanScan[i] for investigating the genetic architecture of sexual dimorphism in kiwifruit. To identify orthologs among species, it is necessary to introduce ortholog detection tools. OrthoFinder version 2.4.0 onwards [13], for example, can use a phylogenetically informed, tree-based inference method to detect **orthogroups** across species and perform a comparative genomics analysis. GENESPACE [8] can combine the popular synteny detection tool MCScanX [4] with the orthology inference tool OrthoFinder [13] to perform a synteny-based pangenome analysis and visualize the results of multiple genomes in a riparian plot. This allows for focused detection of orthologs among collinear blocks and makes it useful for comparisons of *de novo* genome assemblies (Figure 1D). For example, DEEPSPACE v0.1[ii] and ntSynt v1.0.5[iii] support an annotation-free mode, which is more convenient for carrying out synteny-based comparative genomics.

Future perspectives

New synteny applications are continually being developed. SOI which is a robust

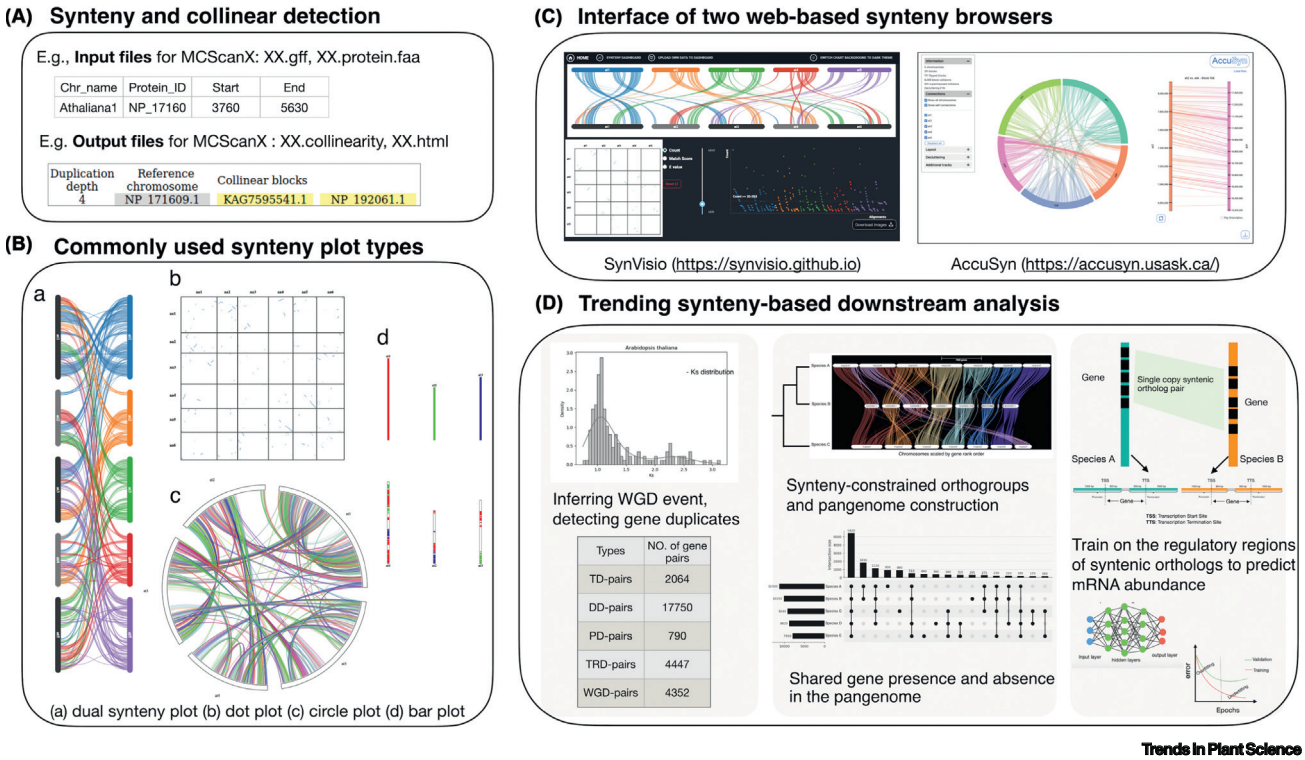

Figure 1. The overview of synteny detection and visualization and its trending applications. (A) MCScanX [4] is used as an example to demonstrate the general input and output files for synteny detection. One of the future improvements could be how to optimize different synteny and collinear detection tools to decode the increasing of genomes of the nonmodel species. (B) Some commonly used synteny plot styles including dual synteny plot, dot plot, circle plot, and bar plot are illustrated. It would be helpful if microsynteny can be easily visualized and compared within and among multiple species in the future. (C) The user-friendly synteny browsers SynVisio and AccuSyn [6] are highlighted with their working interfaces. (D) Synteny-based downstream applications are explored here with examples, such as gene duplication detection [7], pangenome graph construction [8], and deep-learning-based transcript prediction analyses [9]. DD: dispersed duplication; PD: proximal duplication; TRD: transposed duplication.

identification method of orthologous synteny [14], for instance, uses a novel approach based on the Orthology Index to detect synteny and address questions related to polyploidy, **reticulation**, and phylogenomics. Specifically, SOI might help researchers understand how accurate differentiation of **syntenic orthologs** and outparalogs contributes to the inference of phylogenetic tree reconstruction. Machine learning (ML) is also being employed for synteny detection. One example of this is Frackify[iv], which can apply the ML decision tree algorithm on MCScanX-like output files to classify genes originating from WGD into different stages of WGD events by inferring how many copies have been retained [15].

A few years ago, an ortholog-contrast deep learning (DL) model [9] was applied to different syntenic orthologs to explore the impact of cross-species transcript abundance. Not surprisingly, DL models are quickly improving and include the appearance of pre-trained DNA analysis models and large-scale generative genome language models (Figure 1D).

Box 1. Study case for synteny detection and visualization tools

MCScanX [4] can identify collinear blocks of different genomes within or between species. For MCScanX, collinear blocks are suggested to contain at least five genes (i.e., MATCH_SIZE) and the collinear genes cannot contain gaps (i.e., noncollinear genes) >25 (i.e., MAX_GAPS). Moreover, the GAP_ PENALTY value is set by default to −1 but can be adjusted to allow for more relaxed collinear gene alignments. MCScanX also has a step-by-step detection and visualization protocol [5] along with a Conda version[v] and a SnakeMake version—MCScanX_Assisatant[vi] [10], making it easier to prepare the input file and run the associated programs for nonmodel organism genomes—including those retrieved from National Center for Biotechnology Information.

SynVisio and AccuSyn [6] are two powerful web-based tools for interactive visualization of synteny at the chromosome and gene levels from single or multiple species. Some popular model species were built-in as study cases for users to explore. These synteny browsers require files from MCScanX (i.e., the '.gff' used for MCScanX running and the '.collinearity' file from MCScanX output). The optional track files in bedGraph format can be used to demonstrate synteny relationship (such as heat maps, line charts, scatter plots, or histograms).

With the rapid development of synteny detection, visualization tools, and models, the downstream analysis of synteny and collinearity will be open to more explorations.

Author contributions

The study was conceptualized by X.Z. X.Z. wrote the initial draft and performed the data analysis. D.R.S greatly contributed to the manuscript editing. All authors commented to refine the manuscript for peer review.

Acknowledgments

The authors gratefully acknowledge the funding of Discovery Grants from the Natural Sciences and Engineering Research Council of Canada. The authors appreciate the discussions on MCScanX with Dr Yupeng Wang, as well as the feedback on SynVisio and AccuSyn with Venkat Bandi and Dr Carl Gutwin.

Declaration of interests

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.

Resources

[i]https://synpanscan.atcgn.com

[ii]https://github.com/jtlovell/DEEPSPACE

[iii]https://github.com/BirolLab/ntSynt

[iv]https://gitlab.com/barker-lab/frackify

[v]https://anaconda.org/bioconda/mcscanx

[vi]https://github.com/zx0223winner/MCScanX_Assistant

[1]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada
[2]Institute for Comparative Genomics, Dalhousie University, Halifax, Nova Scotia, Canada
[3]Department of Biology, Western University, London, Ontario, Canada

*Correspondence:
xi.zhang@dal.ca (X. Zhang) and dsmit242@uwo.ca (D.R. Smith).

https://doi.org/10.1016/j.tplants.2026.05.010

References

1. Coghlan, A. *et al.* (2005) Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet.* 21, 673–682
2. Zhao, T. and Schranz, M.E. (2019) Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. *Proc. Natl. Acad. Sci. U. S. A.* 116, 2165–2174
3. Lallemand, T. *et al.* (2020) An overview of duplicated gene detection methods: why the duplication mechanism has to be accounted for in their choice. *Genes (Basel)* 11, 1046
4. Wang, Y. *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49
5. Wang, Y. *et al.* (2024) Detection of colinear blocks and synteny and evolutionary analyses based on utilization of MCScanX. *Nat. Protoc.* 19, 2206–2229
6. Bandi, V. *et al.* (2022) Visualization tools for genomic conservation. In *Plant Bioinformatics: Methods and Protocols* (Edwards, D., ed.), pp. 285–308, Springer
7. Zhang, X. *et al.* (2025) HSDSnake: a user-friendly SnakeMake pipeline for analysis of duplicate genes in eukaryotic genomes. *Bioinformatics* 41, btaf325
8. Lovell, J.T. *et al.* (2022) GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* 11, e78526
9. Washburn, J.D. *et al.* (2019) Evolutionarily informed deep learning methods for predicting relative transcript abundance from DNA sequence. *Proc. Natl. Acad. U. S. A.* 116, 5542–5549
10. Zhang, X. *et al.* (2026) Addendum: detection of colinear blocks and synteny and evolutionary analyses based on utilization of MCScanX. *Nat. Protoc.* https://doi.org/10.1038/s41596-026-01380-8
11. Almeida-Silva, F. and Van de Peer, Y. (2025) doubletrouble: an R/Bioconductor package for the identification, classification, and analysis of gene and genome duplications. *Bioinformatics* 41, btaf043
12. Natarajan, S. and Pucker, B. (2026) DupyliCate: mining, classifying, and characterizing gene duplications. *Sci. Rep.* 16 (1), 16557
13. Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238
14. Zhang, R.-G. *et al.* (2025) SOI: robust identification of orthologous synteny with the Orthology Index and broad applications in evolutionary genomics. *Nucleic Acids Res.* 53, gkaf320
15. McKibben, M.T.W. and Barker, M.S. (2023) Applying machine learning to classify the origins of gene duplications. In *Polyploidy: Methods and Protocols* (Van de Peer, Y., ed.), pp. 91–119, Springer