



*R*Ideogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms

Zhaodong Hao^{1,2}, Dekang Lv³, Ying Ge³, Jisen Shi¹, Dolf Weijers², Guangchuang Yu⁴ and Jinhui Chen¹

¹Key Laboratory of Forest Genetics & Biotechnology of Ministry of Education, Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, Jiangsu, China

²Laboratory of Biochemistry, Wageningen University, Wageningen, Haarlem, Netherlands

³Institute of Cancer Stem Cell, Dalian Medical University, Dalian, Liaoning, China

⁴Institute of Bioinformatics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China

ABSTRACT

Background. Owing to the rapid advances in DNA sequencing technologies, whole genome from more and more species are becoming available at increasing pace. For whole-genome analysis, idiograms provide a very popular, intuitive and effective way to map and visualize the genome-wide information, such as GC content, gene and repeat density, DNA methylation distribution, genomic synteny, etc. However, most available software programs and web servers are available only for a few model species, such as human, mouse and fly, or have limited application scenarios. As more and more non-model species are sequenced with chromosome-level assembly being available, tools that can generate idiograms for a broad range of species and be capable of visualizing more data types are needed to help better understanding fundamental genome characteristics.

Results. The R package *R*Ideogram allows users to build high-quality idiograms of any species of interest. It can map continuous and discrete genome-wide data on the idiograms and visualize them in a heat map and track labels, respectively.

Conclusion. The visualization of genome-wide data mapping and comparison allow users to quickly establish a clear impression of the chromosomal distribution pattern, thus making *R*Ideogram a useful tool for any researchers working with omics.

Subjects Bioinformatics, Data Science, Graphics, Visual Analytics

Keywords Genome, Chromosome, Idiogram, R package, Data visualization

INTRODUCTION

Recently, with the development of sequencing technologies, especially rapid advances in third generation sequencing including Pacific Biosciences (*Eid et al., 2009*) and Oxford Nanopore Technologies (*Laver et al., 2015*), BioNano genome mapping (*Cao et al., 2014*) and high-throughput chromatin conformation capture sequencing (*Dekker et al., 2002*), more and more species have their genomes sequenced or updated to the chromosome level (*Jiao & Schneeberger, 2017; Phillippy, 2017*). After the chromosome-level genome completion, an overview of some genome characteristics can help to better understand a

Submitted 29 August 2019

Accepted 19 December 2019

Published 20 January 2020

Corresponding authors

Guangchuang Yu, gcyu1@smu.edu.cn

Jinhui Chen, chenjh@njfu.edu.cn

Academic editor

Sebastian Ventura

Additional Information and
Declarations can be found on
page 7

DOI 10.7717/peerj-cs.251

© Copyright
2020 Hao et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

species genome, such as gene and transposon distribution across the sunflower genome (Badouin et al., 2017).

An idiogram, also known as a karyotype, is defined as the phenotypic appearance of chromosomes in the nucleus of an eukaryotic cell and has been widely used to visualize the genome-wide data since the first web server, *Idiographica*, came online in 2007 (Kin & Ono, 2007). There are dozens of tools have been developed for circular genome visualization with a Perl language-based tool *Circos* being the most used one (Krzywinski et al., 2009; Parveen, Khurana & Kumar, 2019). In contrast, there are not many alternatives for non-circular plots of whole genome information on idiograms. Although few R packages, like *GenomeGraphs* (Durinck et al., 2009), *ggbio* (Yin, Cook & Lawrence, 2012), *IdeoViz* (Pai & Ren, 2014), *chromPlot* (Orostica & Verdugo, 2016) and *chromDraw* (Janecka & Lysak, 2016), and JavaScript libraries, like *Ideogram.js* (Weitz et al., 2017) and *karyotypeSVG* (Prlic, 2017), have been developed for non-circular genome visualization, they are either limited in several species and data visualization types or lacking the ample customization. Recently, two R packages, *karyoploteR* (Gel & Serra, 2017) and *chromoMap* (Anand, 2019), with strengthened capacities have been developed.

However, one function that all these non-circular plots fail to achieve, as *Circos* does, is to visualize the relationship between two or more species using Bezier curves on idiograms. This function is very useful and allows to interpret genome-wide relationships more intuitively, especially in the visualization of whole genome duplication. Indeed, *Circos* is usually used to show syntenic blocks both in inter- and intraspecies genome comparisons using Bezier curves (Hu et al., 2019; Wang et al., 2019). Thus, there is a lack of a R package for non-circular genome visualization and allowing to visualize genome-wide relationships between two or more species using Bezier curves on idiograms.

Scalable Vector Graphics (SVG) is a language for describing two-dimensional graphics applications and images. SVG graphics is defined in an eXtensible Markup Language (XML) text file which means that one can easily use any text editor or drawing software to create and edit SVG graphics. Most R graphics packages are built on two graphics systems, the traditional graphics system and the grid graphics system. Here, we developed an R package (*RIdeogram*) to draw high-quality idiograms without species limitations, that allows to visualize and map whole-genome information on the idiograms based on the SVG language. Besides, *RIdeogram* can also be used to show the genome synteny with Bezier curves linking the syntenic blocks on idiograms.

DESCRIPTION

The package *RIdeogram* is written in R (R Core Team, 2018), one of the most popular programming languages widely used in statistical computing, data analytics and graphics. However, this new R graphics package is not built based on any existing graphics systems. We use the R environment to read the custom input files and calculate the drawing element positions in a coordinate system. Then, we use R to write all element information into a text file following the XML format which are used to define graphics by the SVG language. A list of the currently implemented commands is given in Table 1. In general, there are three

Table 1 Functions contained in the package *RIdeogram*.

Function name	Description
GFFex	Extract information from a GFF3 format genome annotation file
ideogram	Map and visualize the genome-wide data on the ideograms
convertSVG	Convert the output file from the SVG format to the format users chose
svg2tiff	Convert the output file from the SVG format to the TIFF format
svg2pdf	Convert the output file from the SVG format to the PDF format
svg2jpg	Convert the output file from the SVG format to the JPG format
svg2png	Convert the output file from the SVG format to the PNG format

main functions, *GFFex*, *ideogram* and *convertSVG* implemented in the package *RIdeogram*. Users can use the function *data* to load the example data or the basic R function *read.table* to load the custom data from local files. The function *GFFex* can be used to extract the information from a GFF3 format genome annotation file. Then, the function *ideogram* can be used to compute the information for all drawing elements based on the input files and generate a A4-sized SVG file containing a vector graphic which can be conveniently viewed and modified using the software Adobe Illustrator or Inkscape. Alternatively, users can also use the function *convertSVG* to convert this SVG file into an adjustable image format (pdf, png, tiff, or jpg) with a user-defined resolution according to the practical requirements.

In general, there are two types of data, i.e., continuous and discrete data. For mapping and visualizing, *RIdeogram* considers the continuous data, such as gene density across the whole genome in 1-Mb windows, as overlaid features and maps them on the ideograms with dark/light colors representing high/low values. For the other data type that are scattered throughout the whole genome, such as the chromosomal distribution of members in one gene family, *RIdeogram* can add track labels next to the ideograms with three shapes (box, circle and triangle) available to represent different characteristics of these members, such as the subclade that one gene member belongs to. Users can also combine the shapes and colors to represent more than three distinct characteristic types. Furthermore, users can also map the continuous data as a heatmap, a line or area chart along the ideograms. In addition, *RIdeogram* also provides functions for the visualization of dual and ternary genome synteny using Bezier curves on the ideograms.

RIdeogram is available through CRAN (<https://cran.r-project.org/web/packages/RIdeogram/>) and is developed on GitHub (<https://github.com/TickingClock1992/RIdeogram>). Further extensions in development and fixes can be seen in the issue listing page on the package's GitHub page. The new function that we are planning to implement in next version include, but are not limited to, developing more types of data visualization along the ideograms, visualizing genome synteny for more species and enlarging the

user-specified genome regions to display detailed characteristics, as we gather more from users.

EXAMPLES

Our first example use the data contained in this package. After the completion of genome sequencing, assembly and annotation, *RIdeogram* can be used to give some idea of how genes are distributed across the whole genome. The example data contained numbers of protein-coding genes calculated in 1-Mb windows which can be considered as continues data and positions of 500 random selected non-coding RNAs, including ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and microRNAs (miRNAs), which can be considered as discrete data. *RIdeogram* maps the gene density information on the idiograms as overlaid features in a heat map and adds track labels next to the idiograms with green boxes, purple circles and orange triangles representing rRNAs, tRNAs and miRNAs, respectively (Fig. 1). Obviously, inter- and intra-chromosomal gene distributions are non-uniform. For instance, the chromosomal regions adjacent to the centromeres are gene-poor in chromosome 1, 9 and 16 while those are gene-rich in chromosome 11, 14 and 17. This function can be applied to many different situations, such as single nucleotide polymorphism (SNP) density and candidate markers (Fig. S1 & Data S1, original data see *Li et al., 2019*), DNA methylation dynamics and potential activated genes (Fig. S2 & Data S2, original data see *Huang et al., 2019*) and transcription factor (TF) binding sites and candidate target genes (Fig. S3 & Data S3, original data see *Shamimuzzaman & Vodkin, 2013*).

Besides visualizing some specific genome characteristics across the whole genome at the chromosome level as showed in Fig. 1, *RIdeogram* can also be used to compare two relevant genome features, such as gene and repeat density, which will provide some important implications for better understanding the relevance of chromosomal distribution patterns of these two features. The example data implemented in this package also contained the information of long terminal repeat (LTR) distribution across the human genome. Since the transposable elements have been suggested to have a potential detrimental effect on gene expression (*Hollister & Gaut, 2009*), the distributions of gene and LTR are supposed to be opposite across the whole genome as a result of natural selection. As expect, the region that has a relatively high gene content usually has a relatively low LTR density and vice versa (Fig. S4), indicating that LTR seems to avoid inserting in the regions with a high gene content in the genome. This similar phenomenon was also observed in the sunflower genome explained using two idiogram graphics, one showing the gene distribution and the other showing the LTR distribution (*Badouin et al., 2017*). Using *RIdeogram*, users can integrate these two graphics into one, much easier for researchers to interpret and readers to understand. Apart from the differences, this function can also be used to show the similarities, like the similar genetic diversity patterns across the whole genome between two geographical groups of the same species, in different label types (Fig. S5 & Data S4, Fig. S6, original data see *Chen et al., 2019*).

In addition, *RIdeogram* can also be used to show syntenic comparisons between two or three genomes. As shown in Fig. 2, the syntenic blocks between each pair of species,

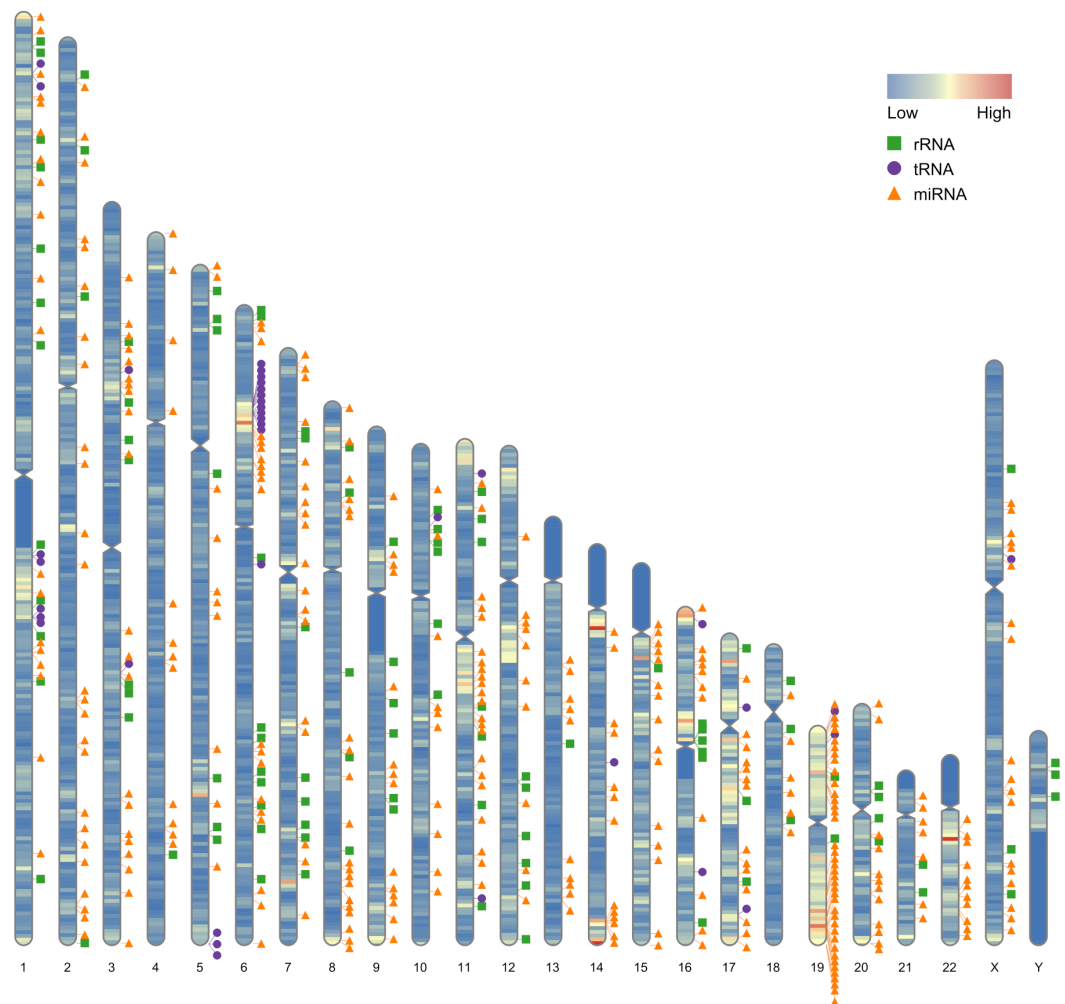


Figure 1 Gene distribution across the whole human genome. The overlaid heatmap shows the gene density and the tack labels refer to 500 random selected RNAs consisted of rRNAs (green boxes), tRNA (purple circles) and miRNA (orange triangles) locus across the human genome. Annotation information was downloaded from the GENCODE website (<https://www.genencodegenes.org>).

Full-size DOI: 10.7717/peerjcs.251/fig-1

which were identified using MCScan (*Tang et al., 2008*), were plotted. Particularly, a typical ancestral region in the basal angiosperm *Amborella* can be tracked to up to two regions in *Liriodendron* and to up to three regions in grape. Based on the fact that no lineage-specific polyploidy event has been found in *Amborella* and a whole-genome triplication has been detected in grape, it is reasonable to assume a single *Liriodendron* lineage-specific whole genome duplication event (*Chen et al., 2019*). Furthermore, *RIdeogram* allows to visualize a dual genome comparison, such as the genome synteny between human and mouse (*Fig. S7* and *Data S5*). Compared to autosomes, the syntenic blocks between human and mouse X chromosomes occupy almost the entirety of each X chromosome, suggesting a highly conserved syntenic relationship of the X chromosome within the eutherian mammalian lineage (*Ross et al., 2005*).

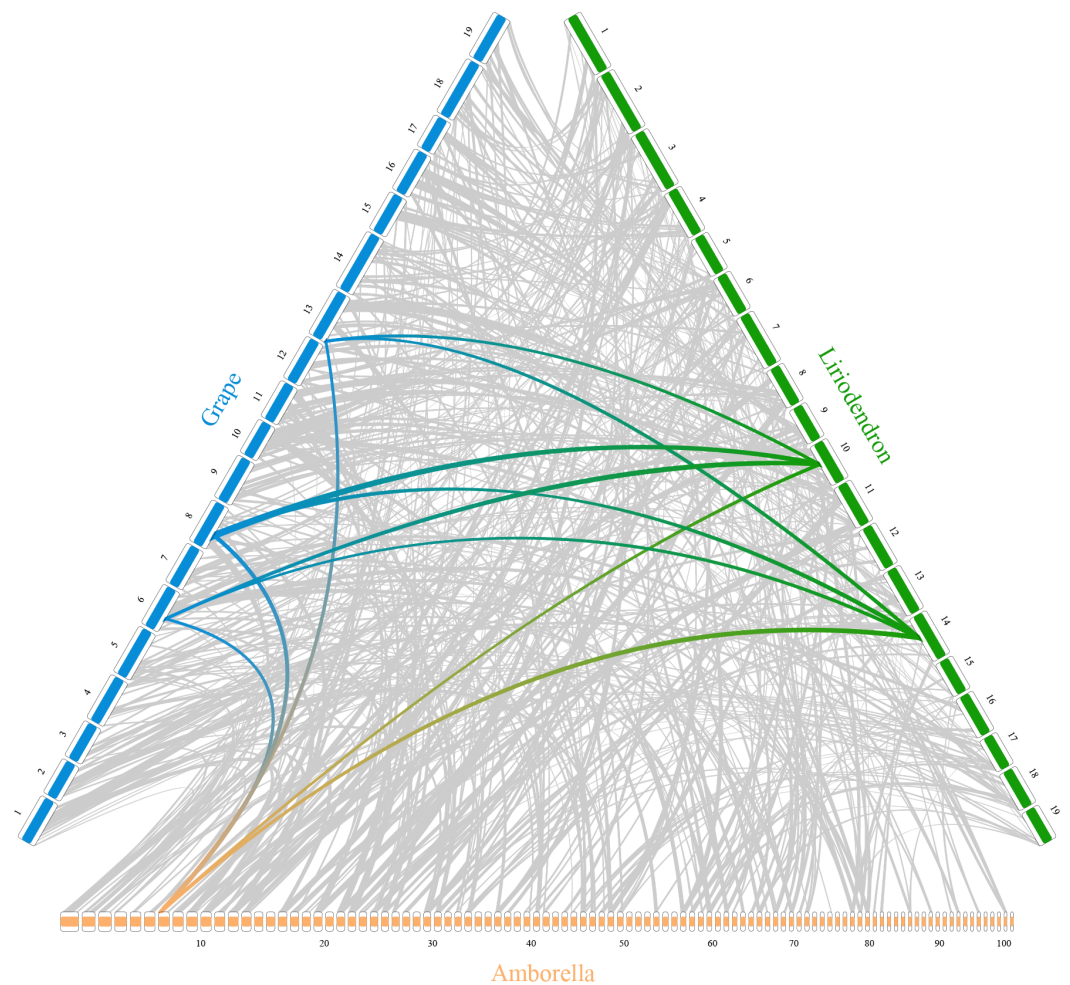


Figure 2 Syntenic comparison of three plant genomes. Genome syntenicity patterns show that a typical ancestral region in the basal angiosperm *Amborella* can be tracked up to two regions in *Liriodendron* and up to three regions in grape. Gray wedges in the background highlight major syntenic blocks spanning more than 30 genes between the genomes (highlighted by one syntenic set shown in colored).

Full-size  DOI: [10.7717/peerjcs.251/fig-2](https://doi.org/10.7717/peerjcs.251/fig-2)

CONCLUSION

The *RIdeogram* package provides an efficient and effective way to build ideograms with no species limitations and map genome-wide information on the ideograms for better visualizing and understanding the chromosomal distribution patterns of some particular genomic features. Meanwhile, this package can be also used to visualize syntenic analysis between genomes. Additionally, it is user-friendly and accessible for biologists without extensive computer programming expertise. Finally, *RIdeogram* can generate two types of images, a vector graphic or a bitmap file, both in high-quality and meeting conventional requirements for direct use in presentations or journal publications.

ACKNOWLEDGEMENTS

We thank Dr. Zhongjuan Zhang for her comments on the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Key Research and Development Plan of Jiangsu Province (BE2017376), the Foundation of Jiangsu Forestry Bureau (LYKJ[2017]42), the Qinglan Project of Jiangsu Province and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Key Research and Development Plan of Jiangsu Province: BE2017376.

Foundation of Jiangsu Forestry Bureau: LYKJ[2017]42.

Qinglan Project of Jiangsu Province.

Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Zhaodong Hao conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Dekang Lv performed the experiments, performed the computation work, authored or reviewed drafts of the paper, and approved the final draft.
- Ying Ge performed the experiments, performed the computation work, authored or reviewed drafts of the paper, typeset the code, and approved the final draft.
- Jisen Shi and Dolf Weijers performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Guangchuang Yu and Jinhui Chen conceived and designed the experiments, performed the computation work, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Data and codes are available at GitHub: <https://github.com/TickingClock1992/RIdeogram>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.251#supplemental-information>.

REFERENCES

- Anand L. 2019. chromoMap: interactive visualization and mapping of chromosomes. *bioRxiv* DOI 10.1101/605600.
- Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Briere C, Owens GL, Carrere S, Mayjonade B, Legrand L, Gill N, Kane NC, Bowers JE, Hubner S, Bellec A, Berard A, Berges H, Blanchet N, Boniface MC, Brunel D, Catrice O, Chaidir N, Claudel C, Donnadiou C, Faraut T, Fievet G, Helmstetter N, King M, Knapp SJ, Lai Z, Le Paslier MC, Lippi Y, Lorenzon L, Mandel JR, Marage G, Marchand G, Marquand E, Bret-Mestries E, Morien E, Nambeesan S, Nguyen T, Pegot-Espagnet P, Pouilly N, Raftis F, Sallet E, Schiex T, Thomas J, Vandecasteele C, Vares D, Vear F, Vautrin S, Crespi M, Mangin B, Burke JM, Salse J, Munos S, Vincourt P, Rieseberg LH, Langlade NB. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546:148–152 DOI 10.1038/nature22380.
- Cao H, Hastie AR, Cao D, Lam ET, Sun Y, Huang H, Liu X, Lin L, Andrews W, Chan S, Huang S, Tong X, Requa M, Anantharaman T, Krogh A, Yang H, Cao H, Xu X. 2014. Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *Gigascience* 3:Article 34 DOI 10.1186/2047-217X-3-34.
- Chen J, Hao Z, Guang X, Zhao C, Wang P, Xue L, Zhu Q, Yang L, Sheng Y, Zhou Y, Xu H, Xie H, Long X, Zhang J, Wang Z, Shi M, Lu Y, Liu S, Guan L, Zhu Q, Yang L, Ge S, Cheng T, Laux T, Gao Q, Peng Y, Liu N, Yang S, Shi J. 2019. Liriodendron genome sheds light on angiosperm phylogeny and species-pair differentiation. *Nature Plants* 5:18–25 DOI 10.1038/s41477-018-0323-6.
- Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* 295:1306–1311 DOI 10.1126/science.1067799.
- Durinck S, Bullard J, Spellman PT, Dudoit S. 2009. GenomeGraphs: integrated genomic data visualization with R. *BMC Bioinformatics* 10:2 DOI 10.1186/1471-2105-10-2.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138 DOI 10.1126/science.1162986.

- Gel B, Serra E. 2017.** karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**:3088–3090 DOI [10.1093/bioinformatics/btx346](https://doi.org/10.1093/bioinformatics/btx346).
- Hollister JD, Gaut BS. 2009.** Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* **19**:1419–1428 DOI [10.1101/gr.091678.109](https://doi.org/10.1101/gr.091678.109).
- Hu L, Xu Z, Wang M, Fan R, Yuan D, Wu B, Wu H, Qin X, Yan L, Tan L, Sim S, Li W, Saski CA, Daniell H, Wendel JF, Lindsey K, Zhang X, Hao C, Jin S. 2019.** The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nature Communications* **10**:Article 4702 DOI [10.1038/s41467-019-12607-6](https://doi.org/10.1038/s41467-019-12607-6).
- Huang H, Liu R, Niu Q, Tang K, Zhang B, Zhang H, Chen K, Zhu JK, Lang Z. 2019.** Global increase in DNA methylation during orange fruit development and ripening. *Proceedings of the National Academy of Sciences of the United States of America* **116**:1430–1436 DOI [10.1073/pnas.1815441116](https://doi.org/10.1073/pnas.1815441116).
- Janecka J, Lysak MA. 2016.** chromDraw: an R package for visualization of linear and circular karyotypes. *Chromosome Research* **24**:217–223 DOI [10.1007/s10577-015-9513-5](https://doi.org/10.1007/s10577-015-9513-5).
- Jiao WB, Schneeberger K. 2017.** The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* **36**:64–70 DOI [10.1016/j.pbi.2017.02.002](https://doi.org/10.1016/j.pbi.2017.02.002).
- Kin T, Ono Y. 2007.** Idiographica: a general-purpose web application to build idiograms on-demand for human, mouse and rat. *Bioinformatics* **23**:2945–2946 DOI [10.1093/bioinformatics/btm455](https://doi.org/10.1093/bioinformatics/btm455).
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009.** Circos: an information aesthetic for comparative genomics. *Genome Research* **19**:1639–1645 DOI [10.1101/gr.092759.109](https://doi.org/10.1101/gr.092759.109).
- Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, Studholme DJ. 2015.** Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification* **3**:1–8 DOI [10.1016/j.bdq.2015.02.001](https://doi.org/10.1016/j.bdq.2015.02.001).
- Li X, Singh J, Qin M, Li S, Zhang X, Zhang M, Khan A, Zhang S, Wu J. 2019.** Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnology Journal* **17**:1582–1594 DOI [10.1111/pbi.13085](https://doi.org/10.1111/pbi.13085).
- Orostica KY, Verdugo RA. 2016.** chromPlot: visualization of genomic data in chromosomal context. *Bioinformatics* **32**:2366–2368 DOI [10.1093/bioinformatics/btw137](https://doi.org/10.1093/bioinformatics/btw137).
- Pai S, Ren J. 2014.** IdeoViz: plots data (continuous/discrete) along chromosomal ideogram. R package version 1.8.0.
- Parveen A, Khurana S, Kumar A. 2019.** Overview of genomic tools for circular visualization in the next-generation genomic sequencing era. *Current Genomics* **20**:90–99 DOI [10.2174/1389202920666190314092044](https://doi.org/10.2174/1389202920666190314092044).
- Phillippy AM. 2017.** New advances in sequence assembly. *Genome Research* **27**:xi–xiii DOI [10.1101/gr.223057.117](https://doi.org/10.1101/gr.223057.117).

- Prlic A.** 2017. KaryotypeSVG—SVG based ideograms of chromosomes showing cytogenetic bands. Version 0.2.0. Available at <https://github.com/andreasprlic/karyotypeSVG>.
- R Core Team.** 2018. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at <https://www.R-project.org/>.
- Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, Howe KL, Ashurst JL, Fulton RS, Sudbrak R, Wen GP, Jones MC, Hurles ME, Andrews TD, Scott CE, Searle S, Ramser J, Whittaker A, Deadman R, Carter NP, Hunt SE, Chen R, Cree A, Gunaratne P, Havlak P, Hodgson A, Metzker ML, Richards S, Scott G, Steffen D, Sodergren E, Wheeler DA, Worley KC, Ainscough R, Ambrose KD, Ansari-Lari MA, Aradhya S, Ashwell RIS, Babbage AK, Bagguley CL, Ballabio A, Banerjee R, Barker GE, Barlow KF, Barrett IP, Bates KN, Beare DM, Beasley H, Beasley O, Beck A, Bethel G, Blechschmidt K, Brady N, Bray-Allen S, Bridgeman AM, Brown AJ, Brown MJ, Bonnini D, Bruford EA, Buhay C, Burch P, Burford D, Burgess J, Burrill W, Burton J, Bye JM, Carder C, Carrel L, Chako J, Chapman JC, Chavez D, Chen E, Chen G, Chen Y, Chen ZJ, Chinault C, Ciccodicola A, Clark SY, Clarke G, Clee CM, Clegg S, Clerc-Blankenburg K, Clifford K, Copley V, Cole CG, Conquer JS, Corby N, Connor RE, David R, Davies J, Davis C, Davis J, Delgado O, DeShazo D, Dhami P, Ding Y, Dinh H, Dodsworth S, Draper H, Dugan-Rocha S, Dunham A, Dunn M, Durbin KJ, Dutta I, Eades T, Ellwood M, Emery-Cohen A, Errington H, Evans KL, Faulkner L, Francis F, Frankland J, Fraser AE, Galgoczy P, Gilbert J, Gill R, Glockner G, Gregory SG, Gribble S, Griffiths C, Grocock R, Gu YH, Gwilliam R, Hamilton C, Hart EA, Hawes A, Heath PD, Heitmann K, Hennig S, Hernandez J, Hinzmann B, Ho S, Hoffs M, Howden PJ, Huckle EJ, Hume J, Hunt PJ, Hunt AR, Isherwood J, Jacob L, Johnson D, Jones S, Jong PJ, Joseph SS, Keenan S, Kelly S, Kershaw JK, Khan Z, Kioschis P, Klages S, Knights AJ, Kosiura A, Kovar-Smith C, Laird GK, Langford C, Lawlor S, Levensha M, Lewis L, Liu W, Lloyd C, Lloyd DM, Loulseged H, Loveland JE, Lovell JD, Lozado R, Lu J, Lyne R, Ma J, Maheshwari M, Matthews LH, McDowall J, McLaren S, McMurray A, Meidl P, Meitinger T, Milne S, Miner G, Mistry SL, Morgan M, Morris S, Muller I, Mullikin JC, Nguyen N, Nordsiek G, Nyakatura G, O'Dell CN, Okwuonu G, Palmer S, Pandian R, Parker D, Parrish J, Pasternak S, Patel D, Pearce AV, Pearson DM, Pelan SE, Perez L, Porter KM, Ramsey Y, Reichwald K, Rhodes S, Ridler KA, Schlessinger D, Schueler MG, Sehra HK, Shaw-Smith C, Shen H, Sheridan EM, Shownkeen R, Skuce CD, Smith ML, Sotheran EC, Steingruber HE, Steward CA, Storey R, Swann RM, Swarbreck D, Tabor PE, Taudien S, Taylor T, Teague B, Thomas K, Thorpe A, Timms K, Tracey A, Trevanion S, Tromans AC, d'Urso M, Verduzco D, Villasana D, Waldron L, Wall M, Wang QY, Warren J, Warry GL, Wei XH, West A, Whitehead SL, Whiteley MN, Wilkinson JE, Willey DL, Williams G, Williams L, Williamson A, Williamson H, Wilming L, Woodmansey RL, Wray PW, Yen J, Zhang JK, Zhou JL, Zoghbi H, Zorilla S, Buck D, Reinhardt R, Poustka A, Rosenthal A, Lehrach H, Meindl A, Minx PJ, Hillier LW, Willard HF, Wilson RK, Waterston RH, Rice CM,**

- Vaudin M, Coulson A, Nelson DL, Weinstock G, Sulston JE, Durbin R, Hubbard T, Gibbs RA, Beck S, Rogers J, Bentley DR. 2005. The DNA sequence of the human X chromosome. *Nature* 434:325–337 DOI 10.1038/nature03440.
- Shamimuzzaman M, Vodkin L. 2013. Genome-wide identification of binding sites for NAC and YABBY transcription factors and co-regulated genes during soybean seedling development by ChIP-Seq and RNA-Seq. *BMC Genomics* 14:477 DOI 10.1186/1471-2164-14-477.
- Tang HB, Wang XY, Bowers JE, Ming R, Alam M, Paterson AH. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research* 18:1944–1954 DOI 10.1101/gr.080978.108.
- Wang M, Tu L, Yuan D, Zhu , Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, Huang H, Yan F, Ma Y, Zhang L, Liu M, You J, Yang Y, Liu Z, Huang F, Li B, Qiu P, Zhang Q, Zhu L, Jin S, Yang X, Min L, Li G, Chen LL, Zheng H, Lindsey K, Lin Z, Udall JA, Zhang X. 2019. Reference genome sequences of two cultivated allotetraploid cottons. *Gossypium hirsutum* and *Gossypium barbadense*. *Nature Genetics* 51:224–229 DOI 10.1038/s41588-018-0282-x.
- Weitz EM, Pantano L, Zhu J, Upton B, Busby B. 2017. Viewing RNA-seq data on the entire human genome. *F1000Res* 6:Article 596 DOI 10.12688/f1000research.9762.1.
- Yin TF, Cook D, Lawrence M. 2012. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biology* 13:R77 DOI 10.1186/gb-2012-13-8-r77.