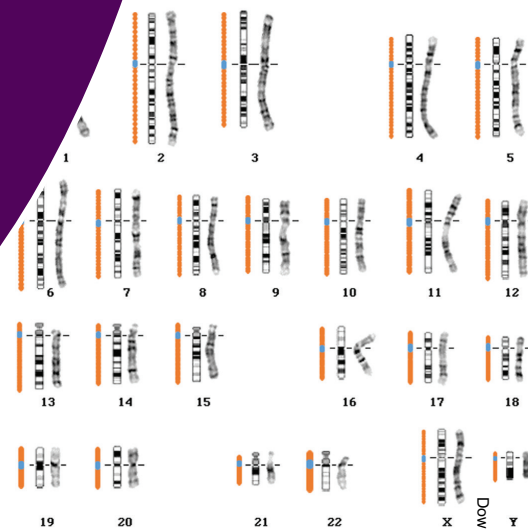


Aaron Gálvez Salido, Rafael Navajas-Pérez

**ABSTRACT**

Idiograms are graphical representations in which chromosomes appear ordered by their relative size, homology groups, and the presence of cytogenetic landmarks such as centromeres, secondary constrictions, and banding patterns. They have become crucial elements in cytogenetic research. Here, we describe a simple protocol for spreadsheet software to draw idiograms and map any type of sequence onto them. These graphs can be easily used to present research results and can become the basis to raise and resolve questions related to different biological issues. The methodology is inexpensive, easy to follow, and allows customization to meet a wide range of educational scenarios focused in genetics and evolution.

Key Words: DNA; protocol; physical map; idiogram; karyotype; evolution.

○ Introduction**The Evolution of Cytogenetic Studies**

It is commonly acknowledged that chromosomes were first observed in plants and animals by Carl Nagéli (1842) and Walther Flemming (1879). During the following decades, the field of cytogenetics experienced a great advance, and at the beginning of the 20th century, the behavior of chromosomes during mitosis and meiosis had been already described. However, the early cytogenetic studies were strictly restricted to morphological features of the chromosomes, and milestones such as the human chromosome number, $2n = 46$, and the first human karyotype were not achieved until the middle of the 20th century (Tjio & Levan, 1956).

From the 1970s onward, chromosome identification methods improved, and banding techniques, *in situ* hybridization,

and immunofluorescent location of proteins were fully developed (Qaisar et al., 2017), making karyotyping a recurring element in current research studies that require chromosome counting, measurement of chromosome features (such as number, size, and shape), and the position of certain structures (centromeres, secondary constrictions, and repeat DNA clusters) (Heslop-Harrison & Schwarzacher, 2011). Also, chromosome analysis techniques are important for detecting structural variation and chromosome mutations. Structural chromosomal abnormalities are related to genetic disease and include deletions (e.g., Jacobsen and cri-du-chat syndromes), duplications (e.g., Charcot-Marie-Tooth disease) and rearrangements (e.g., some types of cancer, such as Philadelphia chromosome, caused by a reciprocal translocation) (OMIM, 2022; Weckselblatt & Rudd, 2015).

The ability to determine the position of DNA sequences on chromosomes, the integration of molecular and cytogenetic data, and the necessity to standardize the system for referring to sequences, have opened the field of chromosome mapping (O'Connor, 2008).

Chromosome mapping is defined as the identification of the location of DNA sequences on a chromosome and the estimation of the relative distance between them.

Chromosome Mapping & Physical Maps

The International System for Cytogenetic Nomenclature compiles the standards of chromosome mapping. This is based on descriptors that refer unequivocally to a specific region of a given chromosome. The short arm of the chromosome is denominated *p* (named after the French word for small, *petit*) and the long arm, *q* (or queue arm); *p* and *q* arms are defined by the relative position of the centromere. Each arm of the chromosome can be divided into regions, and each region

is assigned a number following the direction centromere-telomere ($p1, p2, p3, \dots$). Banding and other staining procedures help define

Physical mapping can help students visualize the position of the main chromosome structures and understand its application in taxonomic and evolutionary studies.

the different regions, even at a higher level, which is expressed by adding an additional digit to the number of the region (e.g., Prader-Willi syndrome is due to absence of paternally expressed imprinted genes at 15q11.2–q13; Cassidy et al., 2012) (O'Connor, 2008).

Nowadays, the enormous amount of genomic sequences generated by Next Generation Sequencing and the increasing number of assembled genomes available have brought into play fine physical mapping. Through coordinates, it is possible to identify the position of any particular sequence in a chromosome or assembled DNA sequence with a level of precision in the base pair (bp) range. Chromosome number, followed by the start and stop positions of the sequence to be mapped, and the reference to the assembly version are sufficient to locate a sequence on a chromosome (e.g., in assembly GRCh38, the coordinates for a human SRY gene are chrY:2786854–2787681). This is the basis of the navigation tools as the UCSC Genome Browser, among others (Karolchik & Kent, 2009).

○ Motivation

Significant efforts have been made to automate cytogenetic data collection and interpretation by the development of different types of software. There are a number of programs devoted to drawing idiograms and mapping sequences. Most of them are restricted to the available sequenced genomes, such as the NCBI Genome Decoration Page (<https://www.ncbi.nlm.nih.gov/genome/tools/gdp>), require programming skills such as RIdeogram (Hao et al., 2020), or rely on genetic-map data or digital image sources such as Karyo-Measure (Mahmoudi & Mirzaghaderi, 2021). Here we describe Excel Mapper, an easy-to-follow protocol to draw idiograms and map any type of sequence. It is based on spreadsheet software functions, making programming skills unnecessary. Source data can come from any csv (comma-separated values) file with chromosome data and sequence coordinates. This information can be extracted from different databases or generated ad hoc by the user. Output can be customized using format options.

○ Methodology

Excel Mapper is based on scatter charts in which the x-axis contains the chromosome number and the y-axis contains the corresponding sequence positions to graph. Chromosome positions are divided into intervals using Sturges's Rule (Sturges, 1926), and then the number of bps each interval will contain (range) is calculated.

Protocol 1 (see Supp. P1a, available with the online version of this article) and Protocol 2 (Supp. P2a) contain the step-by-step instructions on how to draw idiograms and map sequences, respectively. Example raw data are provided in csv format, including chromosome number and their size for Protocol 1 (Supp. P1b) and sequence coordinates (chromosome number, start and stop positions) for Protocol 2 (Supp. P2b). Solved idiograms and maps are also provided in the csv files.

A typical Excel Mapper experiment starts with the graphing of an idiogram and continues by mapping different types of sequences (e.g., centromeres and telomeres). To differentiate them, it is advisable to map different sequences separately by using different datasets (see for example Figure 1).

The protocols were developed using Microsoft Excel but are replicable in free and/or open software. For more info, see the tutorial at https://youtu.be/6WhEGcS2_SU.

○ Excel Mapper for Classrooms

Protocols for chromosome banding, probe hybridizations, or image processing can be time-consuming and nonachievable during a regular biology course. We propose two activities based on Excel Mapper, which, while instructive, can partially substitute some wet lab activities:

Activity 1: The construction of a human idiogram.

Students will be trained on chromosome classification, identification of main morphological types based on the position of the centromeres, and chromosome nomenclature. Students will understand the importance of chromosome classification in the description and identification of living organisms.

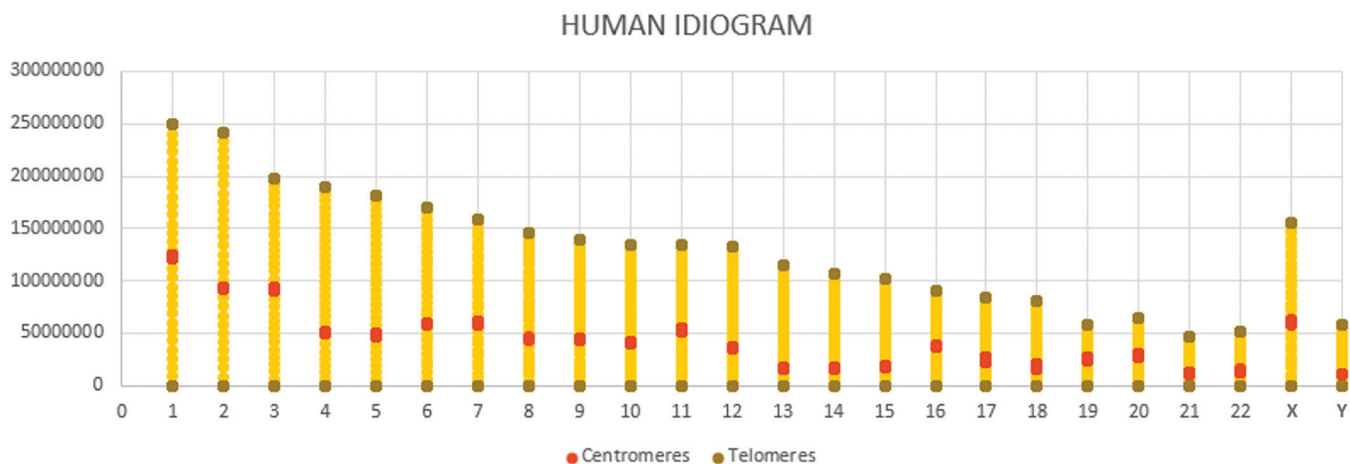


Figure 1. Excel Mapper output representing a human idiogram with centromeric and telomeric regions mapped.

Activity 2: A comparative chromosome analysis between humans and apes. Fusion of ancestral ape chromosomes left in human chromosome 2 distinctive relic telomeric sequences and a vestigial centromere (Azzalin et al., 2001; Miga, 2017). This can be traced by comparing chromosome number and mapping sequences in different closely related primate genomes and by comparing these data with the human genome. Students will understand the importance of cytogenetic data in evolutionary studies.

All class materials, including a PowerPoint presentation, a class handout, and csv with raw data, are provided for both Activity 1 (Supp. A1a–A1c) and Activity 2 (Supp. A2a–A2c).

○ Conclusion

Excel Mapper is a method for drawing idiograms using spreadsheet software and mapping different types of sequences using their genomic coordinates. It is inexpensive, easy to follow, and highly customizable. Physical mapping can help students visualize the position of the main chromosome structures and understand its application in taxonomic and evolutionary studies. As a result, we believe that it can significantly enrich evolution lessons at different educational levels.

○ Supplementary Material

Methodology

- P1a: Protocol 1 for drawing a basic idiogram
- P1b: Sample spreadsheet for protocol 1
- P2a: Protocol 2 for mapping sequences onto an idiogram
- P2b: Sample spreadsheet for protocol 2

Activity 1: Human Idiogram Construction

- A1a: Presentation
- A1b: Handout
- A1c: Dataset

Activity 2: Human-Ape Chromosome Evolution

- A2a: Presentation
- A2b: Handout
- A2c: Dataset
- Video tutorial: https://youtu.be/6WhEGcS2_SU

○ Acknowledgements

The authors thank the anonymous reviewers for their helpful comments. This work was supported by the University of Granada (PID-16-04).

References

- Azzalin, C.M., Nergadze, S.G. & Giulotto, E. (2001). Human intrachromosomal telomeric-like repeats: Sequence organization and mechanisms of origin. *Chromosoma*, *110*, 75–82.
- Cassidy, S.B., Schwartz, S., Miller, J.L. & Driscoll, D.J. (2012). Prader-Willi syndrome. *Genetics in Medicine*, *14*, 10–26.
- Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., et al. (2020). RIdiogram: Drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Computer Science*, *6*, e251.
- Heslop-Harrison, J. & Schwarzacher, T. (2011). Organisation of the plant genome in chromosomes. *Plant Journal*, *66*, 18–33.
- Karolchik, D. & Kent, W.J. (2009). The UCSC Genome Browser. *Current Protocols in Bioinformatics*, *1.4.1*.
- Mahmoudi, S. & Mirzaghaderi, G. (2021). Tools for drawing informative idiograms. *bioRxiv*. <https://doi.org/10.1101/2021.09.29.459870>.
- Miga, K.H. (2017). Chromosome-specific centromere sequences provide an estimate of the ancestral chromosome 2 fusion event in hominin genomes. *Journal of Heredity*, *1*, 45–52.
- O'Connor, C. (2008). Chromosome mapping: Idiograms. *Nature Education*, *1*(1), 107.
- Online Mendelian Inheritance in Man (OMIM). (2022). An online catalog of human genes and genetic disorders. Johns Hopkins University. <https://omim.org>.
- Qaisar, U., Tayyeb, A. & Bhat, T.A. (2017). Techniques of chromosomal studies. In T. Bhat & A. Wani (Eds.), *Chromosome Structure and Aberrations*. Springer.
- Sturges, H.A. (1926). The choice of a class interval. *Journal of the American Statistical Association*, *21*(153), 65–66. <https://doi.org/10.1080/01621459.1926.10502161>.
- Tjio, J.H. & Levan, A. (1956). The chromosome number of man. *Hereditas: Genetiskt Arkiv*, *42*(1–2), 1–6.
- Weckselblatt, B. & Rudd, M.K. (2015). Human structural variation: Mechanisms of chromosome rearrangements. *Trends in Genetics*, *31*(10), 587–99. <https://doi.org/10.1016/j.tig.2015.05.010>.

AARON GÁLVEZ SALIDO is a PhD student in the Departamento de Genética at the Universidad de Granada, Campus de Fuentenueva s/n 18071, Granada, Spain. RAFAEL NAVAJAS-PÉREZ (rnavaajas@ugr.es) is a professor in the Departamento de Genética at the Universidad de Granada, Campus de Fuentenueva s/n 18071, Granada, Spain.