

Representing SSR Molecular Marker Profiles Using Concepts from Andean Khipus

Genebanks increasingly use molecular markers for routine characterization of ex-situ collections and farmer managed diversity. The International Potato Center presently uses a SSR marker-kit to create molecular profiles for potato accessions. We identified a need for a compact graphical representation that allows comparative presentation of molecular diversity and accession characteristics - thereby permitting biologists and collection curators to have a simple means to interpret molecular data. Inspired by the ancient Andean qipus we devised a graph that standardizes representation while leaving room for updates of the marker kit and the collection of accessions. The molecular khipu permits combining and annotating a set of SSR loci with allele frequency and allele size distribution information. The design is flexible to incorporate updates on genetic diversity information. Graphical means facilitate reading of allele diversity information. As a compact graphical view it facilitates information storage and exchange. The SSR khipu will be useful to other genebanks and breeders. Software to create graphics in single or batch mode is available as R package 'khipu'.

1 AUTHORS AND AFFILIATIONS

Authors:

Reinhard Simon[1], Pablo Carhuapoma[1], Vilma Hualla[1], Marc Ghislain[1], Felipe de Mendiburu[1, 2], Rene Gomez[1], Cinthya Zorrilla[1,2], William Roca[1], Stef de Haan[1], Merideth Bonierbale[1]

[1] International Potato Center (CIP)International Potato Center (CIP)

[2] Universidad Agraria La Molina (UNALM)

INTRODUCTION

Genebanks increasingly use molecular markers for routine characterization of ex-situ collections and farmer managed diversity. CIPs genebank presently uses a SSR marker-kit [8] to assist cultivar or genotype identification through molecular profiling. Applications include the comparison of ex-situ and in-situ collections to assess effectiveness and orient conservation strategies [7] or to document provenance and attribution as in an in-situ catalog [6]. A primary motivation came from Andean potato growing communities that are in a working relationship with CIP: farmers called for support to aid in registry and identification of landraces and protect native varieties against 'biopiracy'. As a visual aid to compare SSR marker profiles between accessions in these contexts we identified the need for a compact graphical presentation similar to a 'bar-code'.

A first set of criteria included:

- Amenable to standardization using bioinformatics tools
- compact presentation (e.g. a chart of maximum 2.5cm x 2cm width by height)
- all SSR marker information for a given land-race
- Convey the individuality of a genotype in comparison to diversity in a group
- Recognize the contribution of Andean farmers to the development of the potato crop

To our knowledge no tool exists that would have allowed us to produce such compact graphs based on a set of informative SSR marker. In the context of the production of the first in-situ catalog ([6] and [9]) of Andean potato landraces it seemed worthwhile to consider the use or adaptation of the traditional Andean information communication tool - the khipu (see Figure 1).

Khipus consisted of a set of chords organized as a set of pendants hanging from a backbone chord. They used colors and knots to store a wide variety of information from tribute statistics (see Figure 2) to state history [2] to sins [4]. Numbers apparently were represented as groups of knots and in a top-down order from 1000s to 100s to 10s and ones [16]. Currently, khipus are not any more read by local people but some are still used for ceremonial purposes [17]. The interpretation and usage of khipus is still not fully understood [16], so we did not aim to replicate a historically accurate way of coding SSR marker information. We rather used the khipu as an inspiration to design our own version. In this paper we summarize the process, design elements and evolution, implementation, use and reception. Details on usage can be obtained from the tutorial [13] available together with the software.

METHODS

Process

The first idea of using the khipu concept was refined by matching the properties of a set of molecular markers against the generic properties of a khipu (see details below). Then, in a first round, design criteria were consolidated and several draft designs tested on real data. Subsequently, the molecular khipu idea was presented to the general public through posters at CIP's genebank foyer, to scientists and breeders through journals [14], at conferences ([10] and [12]), and through web sites ([5]) as well as to farmers ([6] and [9]) to solicit feedback over the course of several years. Recently, the molecular khipu idea was consolidated in a publicly available open source software ([11] and [13]).

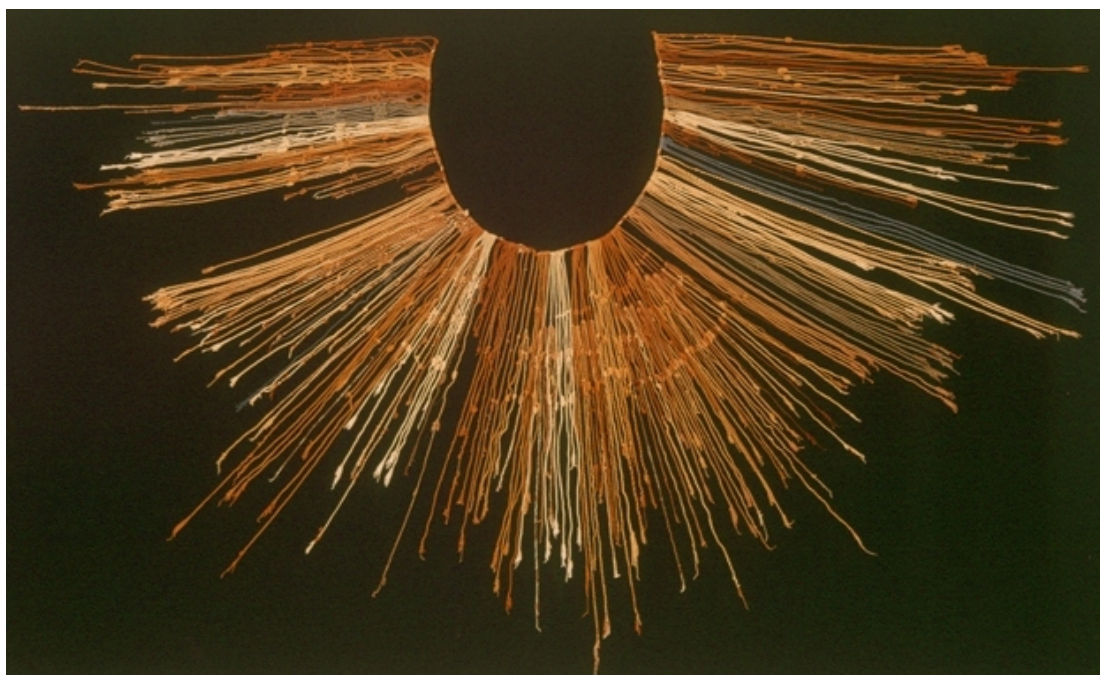


Figure 1. An example of a traditional khipu. Source: wikipedia ([1])

Design principle

The basic design principle is described in more detail in the legend of Figure 3. In summary, SSR marker data are generated in a molecular laboratory applying each marker separately to a batch of genotypes. In the case of the molecular marker kit for potato [8] also the set of SSR markers is defined. Therefore, the SSR data can be simply re-organized to form a molecular profile or molecular khipu. The principal equivalencies used in Figure 3 to turn the profile into a khipu are listed in Table 1. An important assumption is that only single-copy SSR markers are used. A first prototype using real data is shown in Figure 4.

Table 1. Comparison of Traditional and Molecular khipu Concepts

Traditional	Molecular
Set of data	Set of SSR marker
Chord	Marker or Locus
Knot position	Allele size
Knot color or knot size	Allele frequencies

Design refinements

The next principal idea was to include comparative elements into each molecular khipu: how does this genotype compare to a reference genepool? To this end, we used in the second prototype the size of an elliptic dot or 'node'. Color was used to indicate to which chromosome a marker belongs as well as to indicate the range of allele sizes for each marker. See Figure 5. This prototype also tried to more closely visualize the string structure of a physical original khipu.

However, eventually some of these latter elements were removed following general design principles of favoring simplicity and avoiding to depend only on color to convey meaning [15].

Software implementation and availability

A first version was implemented using the programming language PHP; subsequently, the algorithm was transferred to the language R and ultimately organized as a reusable package along with a tutorial. The package can be found at: <http://cran.r-project.org/web/packages/quipu/index.html>.

RESULTS

Khipu features

The current version of the molecular khipu or SSR marker graph is displayed and described in detail in Figure 6. In short, the standard graph has been enriched in terms of annotation and interpretation while the more playful graphical elements were removed. Many features can be customized as described in more detail in [13].

The more compact version is shown and described in Figure 7. Briefly, it remains close to prototype I in Figure 4 with a few enhancements regarding allele frequencies.

2 DISCUSSION

Related statistical graphs

Perhaps a similar graph could be constructed based on a set of boxplots each showing a summary of a molecular marker with superimposed dots for each allele of an individual genotype. This would have the advantage of showing more statistical information. However, for compact figures in a catalog setting it would probably appear overly congested. Another related graph is the 'graphical genotype' chart as proposed by [18]. This latter chart also shows the molecular marker pattern across a set of genotypes but with an emphasis on explicitly comparing genotypes side-by-side whereas the molecular khipu puts an emphasis on an individual genotype. Both charts can complement each other.

2.1 Reception

The tool has been useful for registry of molecular passports and as a communication tool for genebanks both in-situ [6] and ex-situ as well as for breeding materials [5]: particularly, it added value to local Peruvian potato landraces, served as an example of participatory knowledge generation, and assisted in the creation of printed community potato inventories or catalogs. It was also conceived as a tool to orient conservation efforts and as an educational tool for local farmers communities and schools. However, as documented in [9] this has only been partially fulfilled: while the scientists interviewed were overall comfortable with the molecular chart, farmers in general did not 'understand' it. A different strategy would be required if the molecular khipu were to be used in dialogue on diversity with the diverse communities originally targeted.

2.2 Acknowledgements

We are grateful for farmers communities in Huancavelica and in Cuzco ("Asociacion del Parque de la papa") for providing primary materials and data. Additional support came from the NGO ANDES.

REFERENCES

- [1] Ableiter, C. (2007). Inca Quipu.
- [2] Ascher, M. (1983). The Logical-Numerical System of Inca Quipus. *IEEE Annals of the History of Computing*, 5(3):268–278.
- [3] Ayala, Poma de, G. (1612). Quipucamayoc.
- [4] Beyersdorff, M. (2005). Writing Without Words/Words Without Writing: The Culture of the Khipu. *Latin American Research Review*, 40(3):294–311.
- [5] CIP, I. P. C. (2013). Catalogue of CIP Advanced Clones.
- [6] de Haan, S., editor (2006). *Catalogo de Variedades de Papa Nativa de Huancavelica - Peru*. International Potato Center (CIP) and Federation of Farmer Communities of the Department of Huancavelica (FEDECCH), Lima.
- [7] de Haan, S., Núñez, J., Bonierbale, M., Ghislain, M., and van der Maesen, J. (2013). A Simple Sequence Repeat (SSR) Marker Comparison of a Large In- and Ex-situ Potato Landrace Cultivar Collection from Peru Reaffirms the Complementary Nature of both Conservation Strategies. *Diversity*, 5(3):505–521.
- [8] Ghislain, M., Núñez, J., Rosario Herrera, M., Pignataro, J., Guzman, F., Bonierbale, M., and Spooner, D. M. (2008). Robust and highly informative microsatellite-based genetic identity kit for potato. *Molecular Breeding*, 23(3):377–388.

- [9] Scurrah, M., De Haan, S., and Winge, T. (2013). Cataloguing potato varieties and traditional knowledge from the Andean highlands of Huancavelica, Peru. In Andersen, R. and Winge, T., editors, *Realising Farmers Rights to Crop Genetics Resources*, pages 65–79. Routledge, London and New York, 1 edition.
- [10] Simon, R. (2006). Lightning Talk: Visualizing SSR data from a Genbank.
- [11] Simon, R., Carhuapoma, P., Hualla, V., de Haan, S., Ghislain, M., Nunez, J., Gomez, R., de Mendiburu, F., Roca, W., and Bonierbale, M. (2014). quipu: Summary charts of micro satellite profiles for a set of biological samples.
- [12] Simon, R., de Haan, S., Ghislain, M., Nunez, J., Zorilla, C., Gomez, R., Mendiburu, F., Bonierbale, M., and Roca, W. (2007). Standardizing Visualization Of Molecular Marker Sets In A Genebank Using Concepts From Ancient Peruvian Khipus. In *Plant and Animal Genomics*, San Diego. Plant and Animal Genomics Conference.
- [13] Simon, R. and Núñez, J. (2014). Tutorial: quipu package.
- [14] Stapleton, P. (2006). Tying the genome up in knots. *GeneFlow*, page 51.
- [15] Tufte, E. (2001). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, 2nd edition.
- [16] Urton, G. (2012). MATHEMATICS AND ACCOUNTING IN THE ANDES BEFORE AND AFTER THE SPANISH CONQUEST. In Mukhopadhyay, S. and Wolff-Michael Roth, editors, *Alternative Forms of Knowing (in) Mathematics: Celebrations of Diversity of Mathematical Practices*, chapter 1, pages 17–32. Sense Publishers, Rotterdam - Boston - Taipei.
- [17] Wikipedia (2014). Quipu.
- [18] Young, N. D. and Tanksley, S. D. (1989). Restriction fragment length polymorphism maps and the concept of graphical genotypes. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 77(1):95–101.



Figure 2. An example of a khipucamayoc (khipu keepers) managing apparently statistical information. Source: [3]. In the lower left corner apparently a simple counting system is shown based on pebbles or seeds to collect data. The cumulative results are summarized in a khipu.

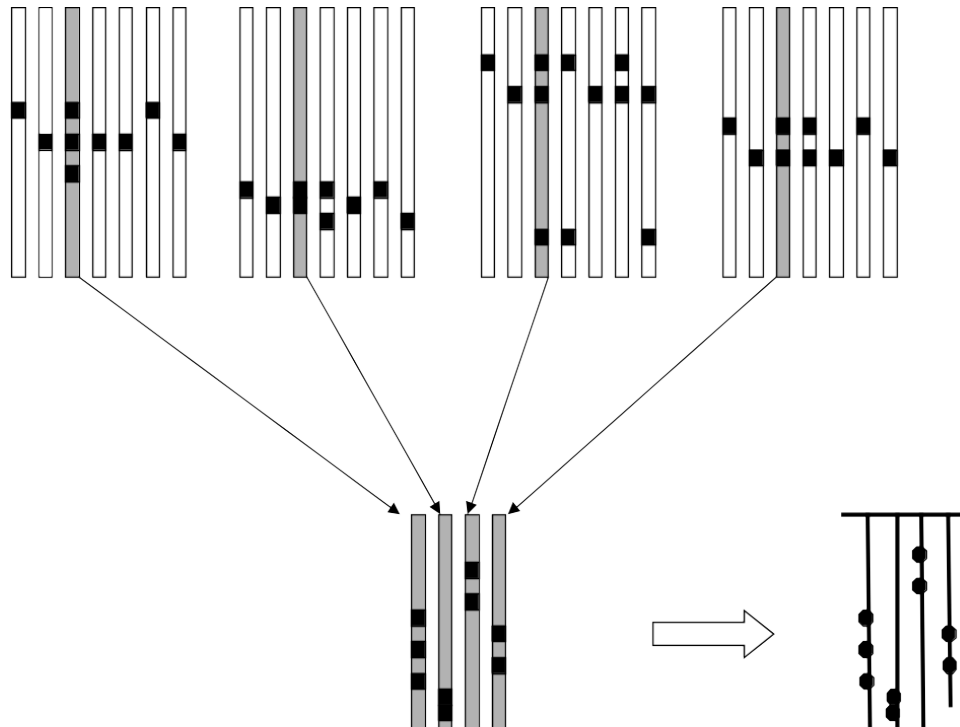


Figure 3. How the khipu is constructed. In the upper part of the figure schematic representations of four SSR marker gels are drawn. Each vertical 'bar' represents the lane in a gel where a SSR marker has been visualized. Differently located circles correspond to different alleles of a SSR marker. The higher up the circle, the more base pairs it has. For each genotype there may be more than one allele per lane up to the number of chromosome copies (ploidy number); only single-copy SSR markers are considered. Typically, for each gel a set of genotypes is characterized with the same SSR marker. Therefore, in order to compile data for a profile of a genotype we will need to simply extract the lane or SSR marker information from the original gel. E.g. for genotype number 3 (marked in dark gray) we can virtually re-assemble the lanes from each original as indicated. With three simple graphical means this is transformed into a 'molecular khipu': a) the reduction of the lane to a symbolic lane, b) the addition of a top line to indicate that these markers comprise part of a set (now each lane represents a distinct gel or marker), and c) the representation of dyed fragments as dots. As a technical note: In a gel the fragments would separate according to an inverse logarithmic relationship. A direct linear scale is used in the molecular khipu for simplicity.

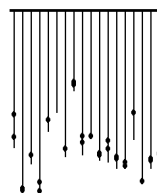


Figure 4. A first prototype of the molecular khipu. Each pendant or vertical line represents a SSR marker locus. Superimposed dots represent alleles of a given size. The concept of a set is symbolized by the top horizontal backbone line. Annotations, scales, and legend are omitted to enable more minute images for display in a catalog. However, close by alleles overlap and may not be easily noticed.

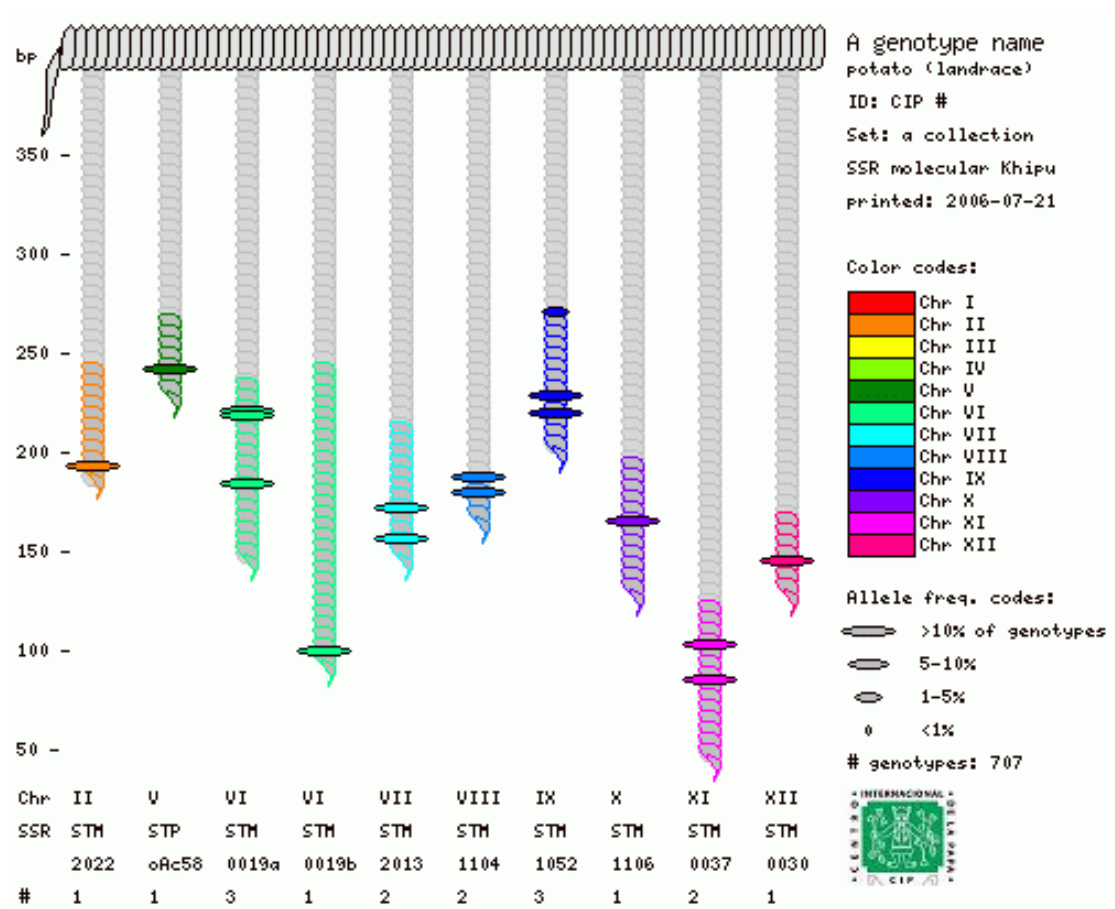


Figure 5. A second prototype of the molecular khipu. This prototype has the same basic features as prototype I. Pendant lines represent marker loci, dots or ellipses alleles of a certain size and the horizontal top backbone highlights the set concept for this data. The most striking contrasts are the use of colors to distinguish the different chromosomes to which each marker belongs and the use of size variation of each allele according to its frequency. Both axes are labeled. The right hand side contains information about the population from which the allele frequencies were derived. Further embellishments include structuring the lines in a chord-like manner.

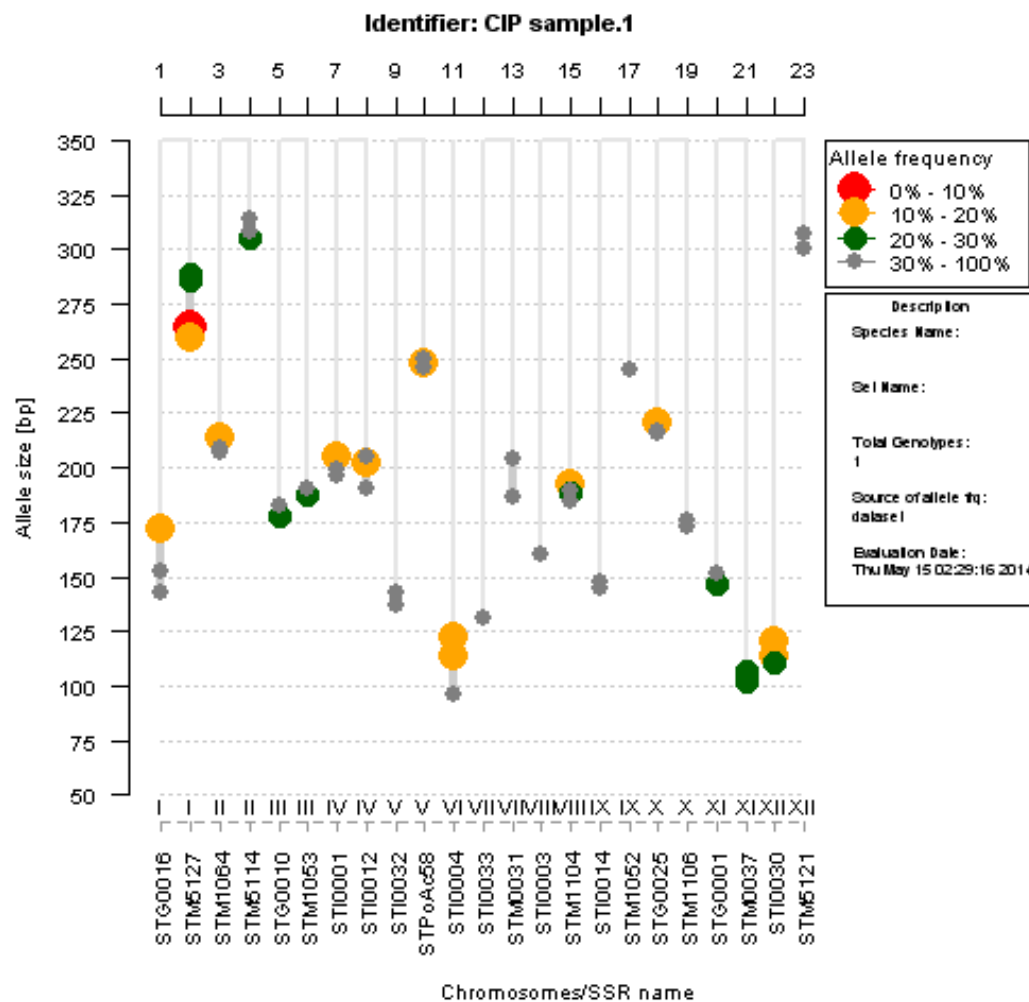


Figure 6. The current version of the molecular khipu as created by the corresponding R package 'khipu'. The top 'backbone' line has been modified (compared to the prototype versions) to convey additional information: that is the running number of the SSR marker locus. The pendant lines represent an individual marker locus. Upon each locus the range of the allele sizes in the reference population is superimposed as a thicker gray line. Alleles are sorted by decreasing dot radius to minimize occultation. Loci or 'lines' belonging to the same chromosome are joined by an upper thicker gray line. Both dot size and dot color can be configured to show allele frequencies in up to four classes. Marker loci are ordered from left to right by chromosome or if that information is missing by order in the given table. Along the y-axis the base pair counts are given in increasing order from bottom to top in a custom range which defaults to values from CIPs experience. The legend on the right hand has three groups. The top group shows the chosen color and size symbols corresponding to the allele frequency classes. Class breaks can also be modified. The second box displays additional information for the underlying population. Lastly, an optional logo image can be inserted into the lower right part.

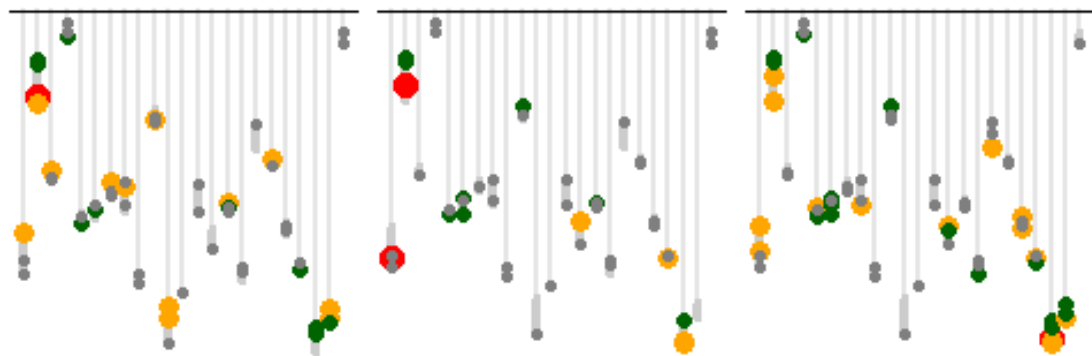


Figure 7. The current compact version of the molecular khipu as created by the correspondent R package 'khipu'. The khipu concept of a set is marked by the upper black horizontal line. Marker loci are symbolized by pendant chords or vertical lines. This y-axis also represents allele sizes increasing from bottom to top. They are in a light gray while allele base pair range is superimposed by a thicker and darker gray line fragment. On top of this, alleles are shown as dots and colored according to a classification scheme. The color scheme should be supplied in a separate way as this format is meant for compact display in a catalog. For the same reason axes units and legends are omitted. This figure shows three molecular khipus of distinct genotypes to illustrate the use of the compact chart in a comparative setting.