CellPress

Review

Genome Mapping in Plant Comparative Genomics

Lindsay Chaney,[1] Aaron R. Sharp,[1] Carrie R. Evans,[1] and Joshua A. Udall[1,@,*]

Genome mapping produces fingerprints of DNA sequences to construct a physical map of the whole genome. It provides contiguous, long-range information that complements and, in some cases, replaces sequencing data. Recent advances in genome-mapping technology will better allow researchers to detect large (>1 kbp) structural variations between plant genomes. Some molecular and informatics complications need to be overcome for this novel technology to achieve its full utility. This technology will be useful for understanding phenotype responses due to DNA rearrangements and will yield insights into genome evolution, particularly in polyploids. In this review, we outline recent advances in genome-mapping technology, including the processes required for data collection and analysis, and applications in plant comparative genomics.

Trends

Genomic structural variations (large DNA rearrangements, such as insertions, deletions, duplications, inversions, and translocations) can lead to phenotypic differences.

Genome mapping images individual DNA molecules, fluorescently labeled at restriction enzyme recognition sites, to create an ordered barcode hundreds of kilobase-pairs long. These barcodes are used to assemble a map that spans the genome.

Due to the continuity of information in genome mapping, it is an ideal tool for use in plants that commonly contain highly repetitive genomic regions and differences in genome sizes.

Recently, advances in genome-mapping technology have made this resource more widely available. Improved optics, advanced molecular biology, informatics tools, and creative innovations have been combined to create relatively low-cost mapping tools.

Origins of Genome Mapping

Despite advances in creating new genomic tools, in some cases revisiting old approaches to scientific questions can be fruitful. This retrospective strategy brings to mind the title of a popular show tune, 'Everything old is new again' [1]. In the case of genome mapping, something old is indeed new again. For many years, cytogeneticists looked at banding patterns of condensed chromosomes and made significant deductions and contributions to our understanding of plant genome organization. In recent years, improved optics, advanced molecular biology, and creative innovations have been combined to create higher-throughput genomic tools that have roots in, and similarities to, many older cytogenetic methods. These strategies produce maps of large individual DNA molecules.

One reason why these long-molecule maps are receiving attention is because of their ability to complement genome sequencing. The relative ease of genome sequencing often overshadows its shortcomings: a puzzle with many small pieces is difficult to solve without additional, long-range information. For example, genome maps can be combined with sequence assemblies comprising numerous scaffolds and contigs, in which case they provide the necessary structure for joining contigs and improving the *de novo* assembly of plant genomes. Aside from *de novo* genome assembly, **genome mapping** (see Glossary) provides some unique research opportunities for comparative plant genomics, which were previously closed because short sequencing reads cannot detect certain large structural variations. Here, we review genome mapping, including its limitations and capacities, and explore some of its potential applications in the field of plant comparative genomics.

Comparing Plant Genomes

Comparative plant genomics examines the similarities of, and differences in, genomes between plant species. By comparing genomes of evolutionarily divergent species, we can better understand the patterns and processes that underlie plant genome evolution as well as uncover functional regions of genomes [2]. **Structural variations** are large (>1 kbp in size)

[1]Plant and Wildlife Sciences Department, Brigham Young University, Provo, UT 84602, USA

*Correspondence: jaudall@byu.edu (J.A. Udall).
@Twitter: @udall

rearrangements of DNA that include insertions, deletions, duplications (also referred to as copy number variations), inversions, and translocations (Figure 1) [3]. These genomic alterations are an important source of genetic and phenotypic diversity. For example, structural variations in plants have been associated with stress tolerance, disease resistance, domestication, increase in yields, leaf size, fruit shape, reproductive morphology, adaptation, and speciation [4–8]. Through the use of **cytogenetics**, researchers have been able to identify large chromosomal changes (e.g., translocations, aneuploidy, and loss of repeats) [9–13], yet this method is labor

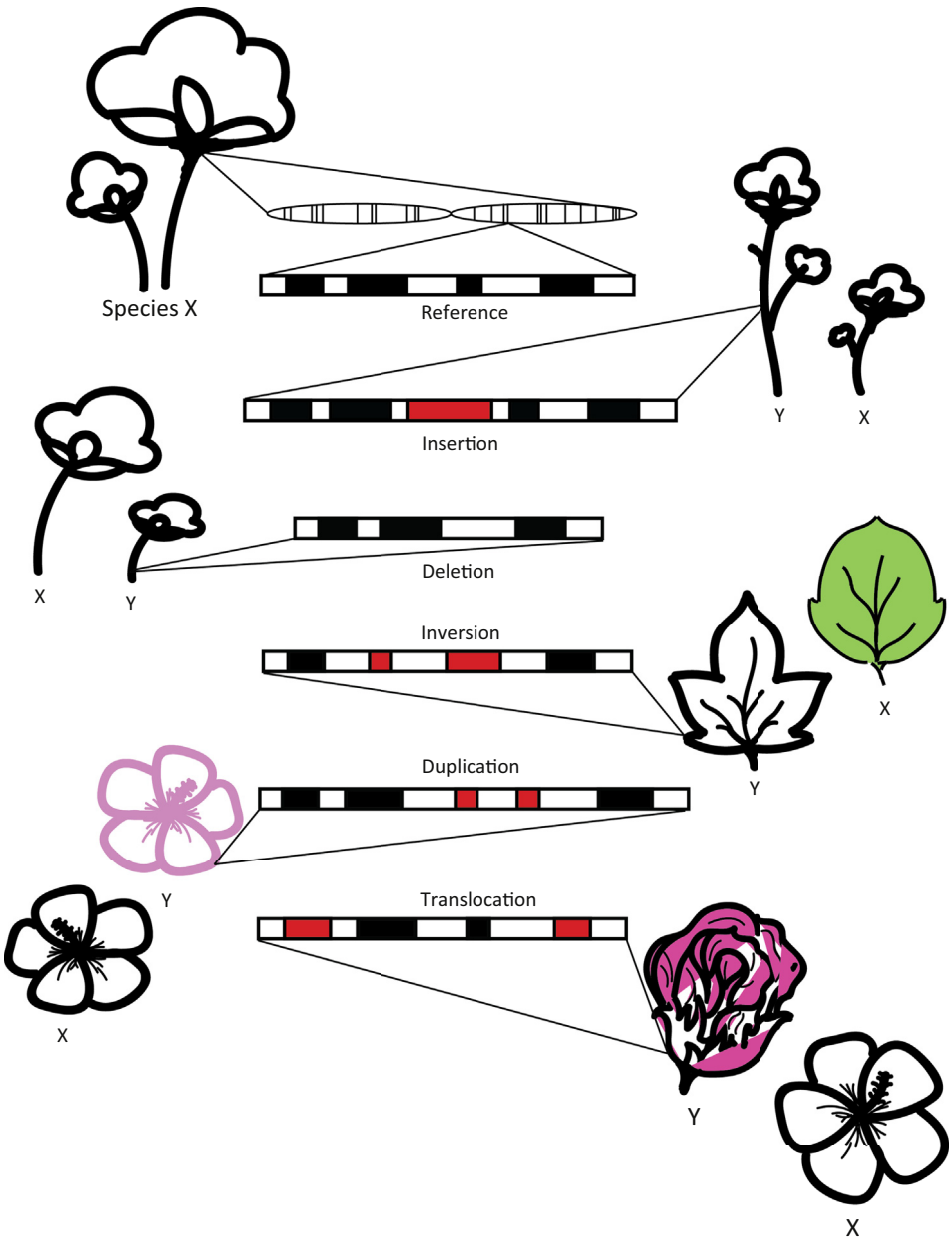

Figure 1. Conceptual Representation of Different Genomic Structural Variations to a Single Region of the Reference Genome. Structural variations are large (>1 kbp) rearrangements of DNA that frequently result in phenotypic differences. These variations include insertions, deletions, inversions, duplications, and translocations. By comparing genomes of different species, large chromosomal changes can be identified.

Glossary

Consensus genome map: a large genomic region represented by a set of contigs. Each contig comprises the total length and distribution of restriction enzyme recognition sites. A consensus map is constructed from single-molecule maps that share compatible distance patterns and, therefore, are likely to represent the same place in a genome.

Cytogenics: microscopic examination of chromosomes. One common molecular cytogenetic method is fluorescent *in situ* hybridization (FISH), where metaphase chromosomes are hybridized with a fluorescently labeled DNA probe. This method can examine a specific DNA sequence and can detect large DNA rearrangements ranging between a few mbp to 1 gbp.

Genome mapping: a variant of optical mapping commercialized as the BioNano lyrs system that uses modified restriction enzymes, fluorescent nucleotide incorporation, and automated imaging in parallel nanochannel arrays to increase the rate of data generation over traditional optical mapping.

High molecular weight (HMW) DNA: DNA molecules that are between 50 kbp and 2 mbp long. The length of the DNA molecules allows them to span large regions of the genome that are typically difficult to resolve with short read sequencing.

Molecule map: a single DNA molecule characterized by an ordered list of distances between restriction enzyme recognition sites. Also referred to as *rmaps*, these individual molecules are combined and/or assembled to construct consensus genome maps.

Optical map: a method that uses microscopic imaging to produce ordered restriction enzyme recognition site maps from a single linearized DNA molecule. Optical mapping allows detection of DNA with a resolution of 1 kbp to several mbp.

Physical map: a map of the physical distances between identifiable landmarks in DNA, generally produced using restriction enzyme digestion of bacterial artificial chromosomes (BACs). By contrast, genetic maps depict relative positions of loci based on the degree of recombination.