Creation and Genomic Analysis of Irradiation Hybrids in *Populus*

Matthew Zinkgraf,[1,2] Kathy Haiby,[3] Meric C. Lieberman,[5] Luca Comai,[4,5] Isabelle M. Henry,[4,5] and Andrew Groover[1,4]

[1]U.S. Forest Service, Pacific Southwest Research Station, Davis, California
[2]Department of Computer Science, University of California Davis, California
[3]Greenwood Resources, Portland, Oregon
[4]Department of Plant Biology, University of California Davis, California
[5]Genome Center, University of California Davis, California

Establishing efficient functional genomic systems for creating and characterizing genetic variation in forest trees is challenging. Here we describe protocols for creating novel gene-dosage variation in *Populus* through gamma-irradiation of pollen, followed by genomic analysis to identify chromosomal regions that have been deleted or inserted in each F1 individual. Irradiated pollen is used in a controlled, interspecific cross to create F1 progeny that carry deletions and insertions of chromosomal regions. These insertions and deletions result in novel changes in gene dosage that in turn affect both qualitative and quantitative phenotypic traits. The protocols described here outline the processes involved in optimizing irradiation levels and performing controlled crosses, sowing seed and propagating seedlings, and genome-wide sequence-based analysis of deletions and insertions in the F1 progeny. The same approach could be applied to other vegetatively propagated species. © 2016 by John Wiley & Sons, Inc.

Keywords: forest trees • gene dosage • genomics • mutagenesis • *Populus*

How to cite this article:
Zinkgraf, M., Haiby, K., Lieberman, M.C., Comai, L., Henry, I.M., and Groover, A. 2016. Creation and genomic analysis of irradiation hybrids in *Populus*. *Curr. Protoc. Plant Biol.* 1:431-450.
doi: 10.1002/cppb.20025

INTRODUCTION

In contrast to most domesticated annual crops, long-lived perennial plants such as forest trees present additional challenges that require novel methods to create and characterize genetic and phenotypic variation. For example, the genus *Populus* has many advantages for breeding compared to other tree species, traditional breeding is still a time-consuming and laborious process. *Populus* species vary in many commercial traits, but typically require several years to reach sexual maturity, suffer from inbreeding depression, and require significant field space to grow mature trees for breeding and progeny testing. Like other forest trees, *Populus* is largely undomesticated and, while significant genetic variation is displayed within and among species, it has not been screened for useful spontaneous mutations or taken through generations of selection for agronomic properties, as have traditional domesticated crops. One challenge is to greatly accelerate the process of creating and characterizing new cultivars with novel properties to meet specialized needs for industrial (e.g., biofuels production) or ecological (e.g., restoration of polluted sites) applications.

Gene dosage describes the effective relative copy number of genes and is believed to be central to heterosis as well as to quantitative trait variation. Gene dosage relationships can be altered naturally by a variety of processes, including through copy number variation,

Volume 1

Current Protocols in Plant Biology 1:431-450, June 2016
Published online June 2016 in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cppb.20025

ploidy changes, the presence of null alleles, or—in interspecific hybrids—by failure of protein complexes to properly form or function due to divergence. Gene dosage can also be readily altered through mutagenesis, as described here.

Gene dosage also plays a role in *Populus* ecology and breeding. For example, *Populus* species such as *P. tremuloides* frequently display triploidy under natural conditions (Mock et al., 2012). For industry, interspecific hybrids can display significant heterosis, and can be deployed as clones using vegetative cuttings (Stettler et al., 1996). Importantly, vegetative propagation allows for the utilization of clones with reduced fertility or even sterile clones, including aneuploid or triploid clones. *Populus* species are divided among three subsections, and the success of interspecific hybridization varies among species within and, in some cases, between subsections (Eckenwalder, 2001). Although different breeding schemes have been employed for *Populus* (Riemenschneider et al., 2001), identification of superior interspecific F1 hybrid clones and their direct deployment as vegetative cuttings has been a mainstay for breeding success. Additionally, dosage variation in the form of triploid clones has been useful in some commercial applications, but triploids occur sporadically, and ploidy and other forms of dosage variation are not systematically assayed or exploited in poplar breeding programs.

In the following protocols, the aim is to alter gene dosage relationships in an F1 family of *Populus*, and then identify and characterize novel genotypic and phenotypic variation. We chose to perform these experiments in an interspecific hybrid population of *P.* × *canadensis* (*P. deltoides* × *P. nigra*) that is important to industry, produces large numbers of seed, and which facilitates identification of single nucleotide polymorphisms between the parents. However, the same approach could be used for other inter or intraspecific crosses of *Populus* or potentially other species that can be vegetatively propagated. Pollen from the male parent is subjected to gamma irradiation to cause deletions and insertions of chromosomal segments, and is then used to fertilize a receptive female (Basic Protocols 1). Next, female catkins are bagged, capsules allowed to dehisce, and seed is collected, cleaned, and sown (Basic Protocols 2). Resulting seedlings are then subjected to a genomic analysis using low-pass Illumina sequencing to produce high-resolution sequence-based karyotypes that reveal features including ploidy level and the location of deletions and insertions for each individual (Basic Protocols 3 and 4). Genotypes of interest can be clonally propagated as cuttings for additional study.

BASIC PROTOCOL 1

POLLEN IRRADIATION AND PERFORMING CONTROLLED CROSSES

Performing successful *Populus* crosses first requires a collection of healthy, fertile branches from the desired species and genotypes. Interspecific crosses have the potential advantages of displaying heterosis and providing large numbers of single nucleotide polymorphisms useful in genomic analysis, as described in Basic Protocol 4. However, it should be cautioned that many crosses are infertile or are successful in only one direction. For example, *P. deltoides* × *P. nigra* cross is viable, whereas *P. nigra* × *P. deltoides* is infertile.

Populus is dioecious, establishes floral buds in the growing season prior to overwintering, and flowers in the spring shortly before vegetative buds flush. Thus, trees must be evaluated for sexual maturity and for gender in the spring of the year prior to performing crosses. Floral buds tend to form basal to vegetative buds. Male catkins carry many flowers that produce copious pollen, characteristic of these wind-pollinated trees. Female catkins carry several or many capsules, each with a few or many seeds, depending on the species and success of the cross.

Fertile branches are typically high in the tree and are often collected by aerial lifts or even by shooting them from the tree with a rifle. Branches may be kept in cold storage for

at least several days prior to rooting or forcing. Male branches are placed as cuttings in water and forced in permissive conditions, and pollen is collected and dried. Pollen can then be gamma irradiated prior to being used in crosses. Female branches must first be rooted in soil in a cold room, using a false-bottom bucket method. Only when sufficient roots have formed can female branches be moved to breeding enclosures, temperature forced, and then pollinated.

Materials

Fertile male and female branches from individuals to be crossed
Dip-N'Grow (*http://www.dipngrow.com/*)
Potting soil (Sunshine Mix 4; *http://sunshineadvanced.com/*)
FloraLife FS1107 (*http://www.floralife.com/*)

Five-gallon buckets, to hold male branches during pollen collection process
Glassine bags, 6 × 3.5 × 13-in. (Brown Paper Goods, Waukegan, IL)
Desiccator and desiccant
False-bottom five-gallon buckets for rooting female branches (preferably opaque, to allow visualization of root development)
Temperature probes (placed inside each female bucket to monitor temperatures)
Pail heaters (Grainger, cat. no. SDH5120-3; *http://www.grainger.com/*)
Trash bags
Pollen-tight breeding enclosures
Scintillation vials (Fisher Scientific, cat. no. 03-227-4)
Gamma-irradiation source
Pollen applicators or atomizers

Identify species and collect branches

1. Determine and source appropriate breeding germ plasm.

   It is highly advisable to work with an experienced breeder to select and acquire appropriate germ plasm for your project. Selecting species combinations and genotypes already known to produce good seed set is paramount to success.

2. Collect fertile branches from male and female trees to be crossed.

   Branches should be collected in winter while trees are dormant, but after chilling requirements have been met. This will vary depending on location and species, but generally mid-winter is a good time. Because fertile branches tend to be high off the ground, an aerial lift or marksman are commonly employed for collections. Branches can be shipped, but minimize the amount of time they are exposed to warmer temperatures. Branches can be stored for several days at 4°C prior to rooting or forcing.

Force male branches and collect pollen

3. Make fresh basal cuts of male branches, and place them in a bucket of water.
4. Allow to develop at 16°C day and 5°C night temperatures.
5. Every three days make a fresh basal cut and change the water.
6. Monitor floral buds daily for swelling, and once male catkins emerge, attach glassine bags to the branch to hold the catkin and catch pollen.
7. Once catkins have shed pollen, empty contents into a pollen drying box for 24 to 36 h, or until pollen is no longer sticky or clumpy, then gather pollen on a piece of paper and place in scintillation vial.

   Populus produce copious pollen. One branch should provide several milliliters of pollen.

Avoiding cross-contamination of pollen can be a major challenge in some experiments. The safest approach is to sequentially force each male genotype to be utilized.

8. Transfer pollen to a scintillation vial (uncapped), and desiccate at 15°C in a desiccator until ready for use.

Root and force female branches

9. Make a fresh basal cut on each female branch, dip in rooting compound (1 part Dip-N′Grow concentrate to 10 parts distilled water) for 3 to 5 sec.

10. Place each branch in a false-bottom five-gallon bucket (Fig. 1A), with branches passing through holes in the plexiglass false bottom into the water compartment below.

11. Place premoistened soil in the top compartment to within ~2 in. of the top of the bucket, moderately packing the soil to remove any air pockets.

12. Carefully cover braches with a plastic bag (e.g., a trash bag) to prevent branches and buds from desiccating.

13. Fill the bottom compartment with deionized water supplemented with 30 cc of Floralife, leaving an ~2 cm air gap between the top of the water and soil. Change the water through the PVC tube (Fig. 1A; W) as needed to reduce bacterial and fungal growth.

    Floralife helps to inhibit microbial growth that would otherwise occlude the water-conducting vessel elements.

14. Keep the buckets in 4°C coolers until sufficient roots have formed.

    Significant root development is required to support the development of catkins and leaves after buds break. Root systems should be developed within 2 to 6 weeks, depending on the species, and roots should be visible through the opaque bucket.

    P. nigra and poplars of subsection tacamahaca typically root well without hormones or soil heating. But with hard to root species, including P. deltoides, the soil must be warmed to ~23°C while the buckets are in cold storage by wrapping with an electric pail heater (Grainger). Use temperature probes to monitor soil temperature. To avoid anoxia, soil should be moist but not wet. Alternatively, some breeders will use grafting to maintain female genotypes for breeding.

15. After sufficient root development, move females to be crossed to a pollen-tight enclosure (see Fig. 1B, for example) in a greenhouse or growth room environment at 15° to 18°C days with nighttime cooling.

16. Monitor daily for swelling of floral buds, and twice weekly, drain the water from the lower compartment, flush fresh water through the PVC tube (Fig. 1A; W), and refill water compartment with deionized water and 30 cc of Floralife, leaving a ~2-cm air gap.

    Expect buds to break and catkins to be receptive as early as 2 to 5 days after moving to permissive temperatures. Figure 1C shows the development of a female bud. Catkins will emerge prior to vegetative buds breaking (Fig. 1D). Receptive catkins will be well expanded, bracts excised, and the stigmatic surface will appear moist.

    The window for receptivity is only a few days, so monitor carefully.

Irradiate pollen and perform crosses

17. One to two days before breeding, transfer desiccated pollen in a scintillation vial into a calibrated gamma-irradiation source, and irradiate to the desired exposure.

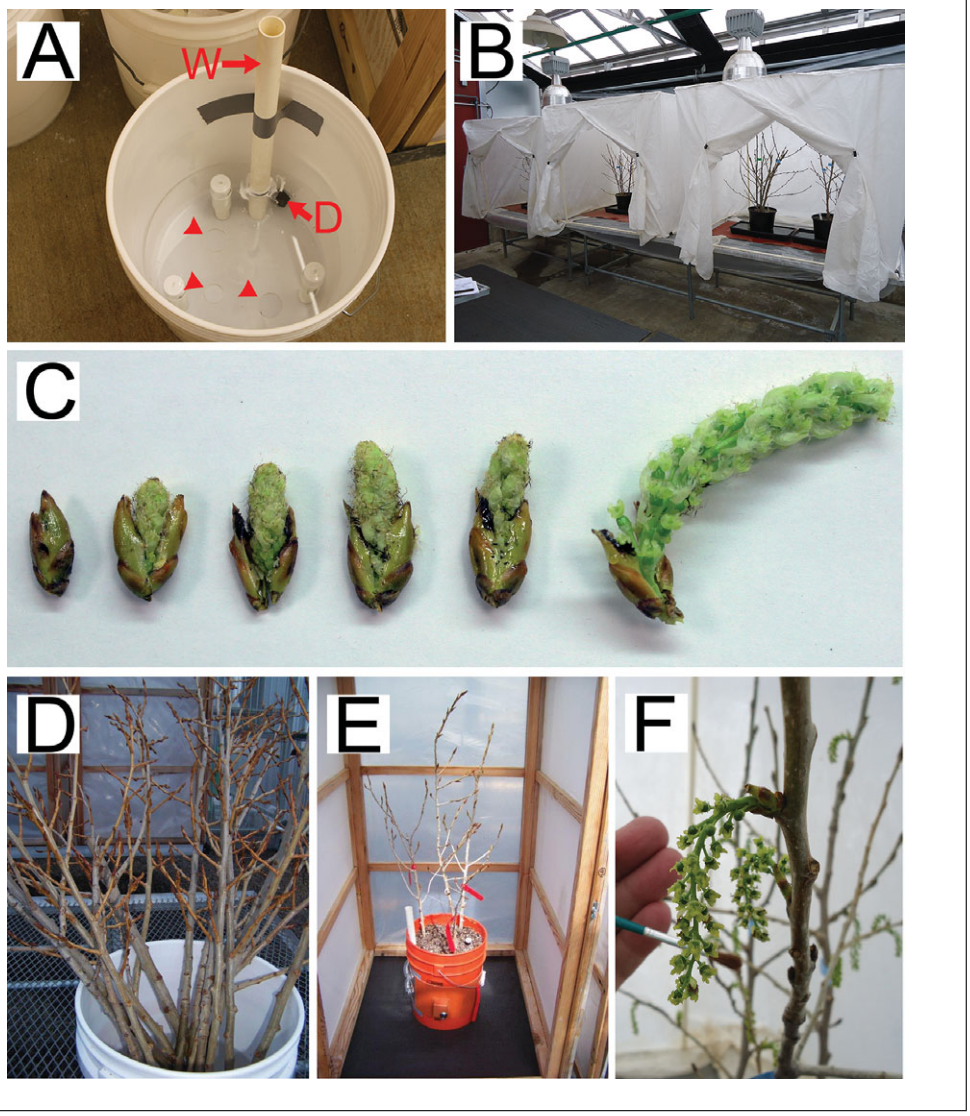

Figure 1 Controlled pollination methods for *Populus*. (**A**) False-bottom bucket for rooting female branches. An opaque five-gallon bucket is modified by adding a Plexiglas insert containing holes (arrows) to allow individual branch ends to pass through the Plexiglas. In practice, branches are placed through the holes, the bucket is filled with soil above the Plexiglass, and the compartment below the Plexiglas is filled with water. It is crucial to leave an air gap between the top of the water and the Plexiglas to prevent the soil from getting too wet. To change the water, a drain hole and plug (labeled with a red D) were inserted in the side of the bucket, near the bottom, to drain the water, and a PVC pipe was inserted through the false bottom to fill the bottom compartment with fresh water. (**B**) Example of pollen-tight breeding enclosures. (**C**) Developmental series of catkin development. (**D**) Male *P. nigra* branches in a water-filled bucket being temperature forced. (**E**) Female *P. deltoides* being temperature forced in a breeding enclosure (door removed) after rooting in a cold room. (**F**) Receptive *P. deltoides* catkins being hand pollinated.

The optimal irradiation level must be determined empirically to meet your experimental goals. For each level to be tested, desiccated pollen should be aliquoted to glass scintillation vials and introduced into the gamma irradiation source for the correct time to achieve the desired exposure. Note that moisture content could have a dramatic effect on the effective irradiation dose.

For desiccated P. deltoides × P. nigra, we found 100 gray (Gy) to be an optimal level, while 150 or 200 gray resulted in catkin abortion and 50 gray gave much lower frequency of deletions.

Include a non-irradiated control and enough samples to determine a dose-response curve. For our experiments, a range of 50 to 200 gray was used, with 100 gray being optimal. The optimal range is rather narrow, with 50 gray giving much fewer insertions and deletions, and 125 gray and above giving inviable crosses characterized by catkin abortion.

18. Rehydrate the pollen by transferring it to an open vial and placing in a closed container with a small volume of water in the bottom for 15 to 30 min.

    The rehydrated pollen will be somewhat sticky and clumpy compared to desiccated pollen, and ready for use.

19. Apply pollen to female catkin, dip a paintbrush into the pollen to pick it up, place the brush in front of a female catkin, and use a puff of air to dust the female catkin. Alternatively use an aspirator to apply pollen.

20. After pollinating each catkin, close the pollen-tight breeding enclosure.

    After closing the pollen-tight enclosure, air-borne pollen can be effectively reduced from the breeding room by fine misting with water. To avoid contamination, rinse hands, and clean equipment prior to moving on to the next pollination with different pollen.

BASIC PROTOCOL 2

COLLECTING, CLEANING, AND SOWING SEED

Populus produce seed within capsules, which will dehisce as they ripen to release the wind-dispersed seed. Seed are small and characterized by copious seed hairs (cotton) that carry the seed aloft for dispersal. *Populus* seed loses viability rapidly, and thus seed must be collected, cleaned, and stored below freezing without delay. Typically seed is easier to work with if "cleaned" by removing the cotton. Care must be taken to avoid cross contamination of seed across experiments. Seedling establishment requires moist conditions, and care should be taken not to over water to avoid fungus gnat larva outbreaks.

Materials *(also see Basic Protocol 1)*

Fertilized female branches (from Basic Protocol 1)
Potting soil (Sunshine Mix #4)
Subdue MAXX Fungicide (Syngenta)

CAUTION: *Mefenoxam is the active ingredient. Follow manufacturer's instructions for application, hazards, required personal protective equipment, and disposal.*

Gnatrol biological larvicide (active ingredient *Bacillus thuringiensis)*

Glassine bags, 6 × 3 ½ × 13-in. (Brown Paper Goods, Waukegan, IL)
Soil sieves
Yellow Cone-tainers and leach racks (SC10; *https://www.stuewe.com/products/rayleach.php*)
Misting bench

Collect seeds

1. Move fertilized females from pollination enclosures to a greenhouse or growth room.

   Wait a few days to a week after pollination before moving female branches out of the pollination enclosures. At this point, females should no longer be receptive but will benefit from slightly warmer conditions and better airflow, which will help the catkins develop to maturity. At this point, there should be vigorous vegetative growth as well. If insufficient roots were formed or pollination was unsuccessful, the catkins will shrivel and abscise.

   At this point and for the remainder of catkin development, change water by draining from the bottom of the false-bottom buckets, flush using the PVC pipe, and refill with deionized

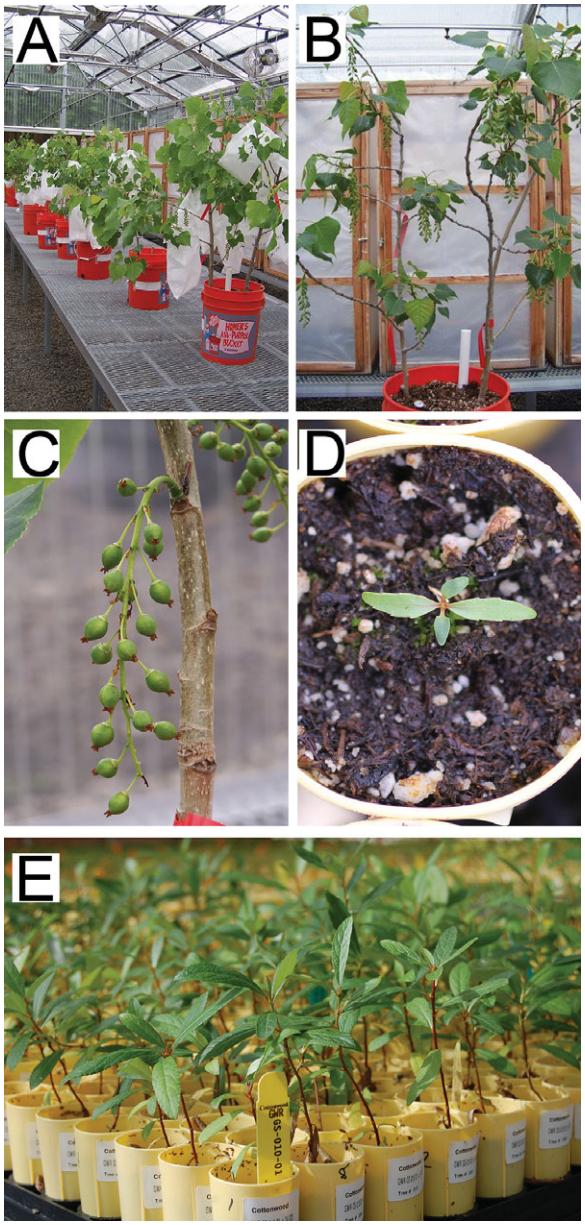

Figure 2 Seed collection and germination of *Populus*. (**A**) *Populus* female branches that have been bagged to collect developing seed. (**B**) *P. deltoides* female carrying numerous developing catkins. (**C**) *P. deltoides* female catkin showing developing capsules. (**D**) Newly germinated *P. deltoides* × *nigra* seedling under mist. (**E**) First-year growth of *Populus* under greenhouse conditions.

water and Floralife. The soil compartment must be kept moist to help sustain healthy plants.

2. When the capsules begin to crack open and seeds begin to dehisce, attach glassine bags to the branches using masking tape to contain the catkin(s).

   Examples of maturing capsules and bagged capsules are shown in Figure 2A-C.

   The catkins should be sealed in the bag. Do not mist plants or allow the catkins to get moist during the dehiscence period. Place in a greenhouse location protected from strong air currents to prevent any unwanted seed dispersal and losses.

Handle the catkins with care. If they prematurely detach from the stems, the seed will not mature.

3. Allow each catkin to dehisce and begin to release seed, remove the glassine bag containing the catkin from the plant to harvest the seed. Bags can be stored in a cooler, and placed in −2° to −4°C storage for a couple of weeks.

Clean and sow seed

4. Clean seed in small lots by passing through a soil sieve or similar fine-meshed screen. Rub the seed and cotton lightly across the screen to allow seed to drop to a waiting sheet of paper, leaving the cotton and other larger debris on the screen. Alternatively, devices can be constructed to allow compressed air to separate the seed from cotton.

   Sow seed soon after collecting to obtain optimal viability.

5. Sow seed directly on top of soil, misting periodically through the day, just enough to keep the top of the soil moist, as shown in Figure 2D.

   Small Cone-tainers work well for this purpose (Fig. 2D) and can support the seedling for the first months of growth (Fig. 2E) before being stepped up into larger tubes.

6. Ensure that the soil remains moist throughout seed germination and seedling establishment.

   In general, poplars thrive in moist but well-drained soil. Pretreatment of soil with fungicide prior to sowing can prevent unwanted fungus (e.g., Subdue MAXX Fungicide).

   Overwatering will also increase the habitat for fungus gnat infestation, and their larvae will kill new seedlings overnight by girdling the plant just below soil level. The larvae are transparent and difficult to see. "Gnatrol" is the only commercial product known to effectively destroy the larva.

   Depending on your greenhouse environment, intermittent misting may be required to establish the seedlings.

7. Apply periodic fertilizer and pesticides as needed to support healthy, vigorous growth. Water stress can make the seedlings more susceptible to insect infestations.

BASIC PROTOCOL 3

DETECTION OF LARGE-SCALE INSERTION AND DELETION MUTATIONS

Each seedling can be evaluated by genomic analysis to identify large-scale insertions and deletions (indels) using low-pass Illumina sequencing followed by bioinformatics analysis to align high quality sequencing reads to a reference genome and identify regions exhibiting altered coverage. For each tree, genomic DNA can be extracted from fresh, frozen, or desiccated leaf tissue using a protocol appropriate to the species and tissue at hand. Indexed genomic sequencing libraries can be produced from each DNA sample using commercially available reagents following the manufacturer's recommended procedures. Next, libraries are pooled (e.g., the experiments here pooled ∼96 libraries per Illumina HiSeq 2000 lane) and sequenced. Read coverage varies across genomic regions, depending on a number of factors including the presence of repeated regions, PCR, and sequencing bias. Raw read coverage is therefore not a reliable measure of copy number. Instead, normalized coverage ratios between a sample and a control more reliably identify indel mutation events. In short, high-quality sequencing reads are aligned to a reference sequence, the genome is partitioned into non-overlapping "bins" or regions of a predefined size, and coverage information across these bins is then compared between samples to detect reliable dosage variations (Figs. 3 and 4).

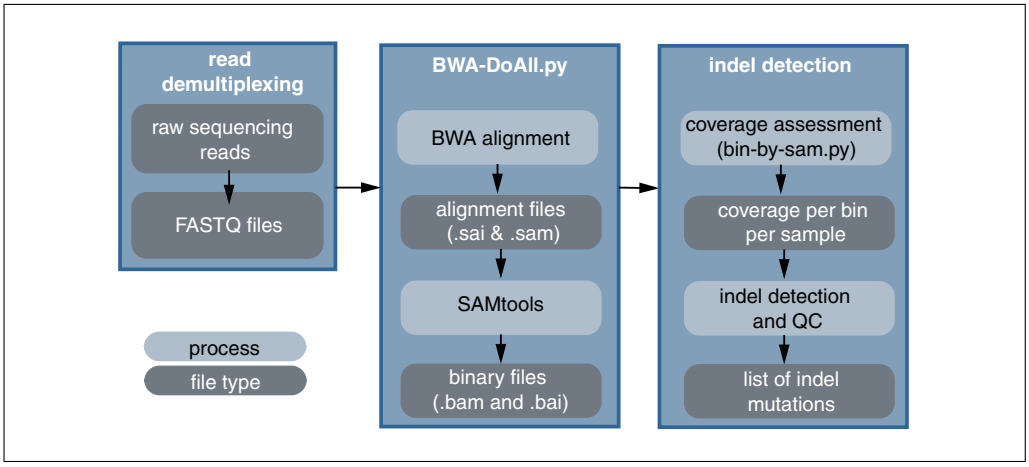

Figure 3 Steps involved in the detection of large-scale deletions and insertions. Alignment files are generated from FASTQ files using BWA and SAMtools. Read coverage across the genome is compiled by `bin-by-sam.py`.

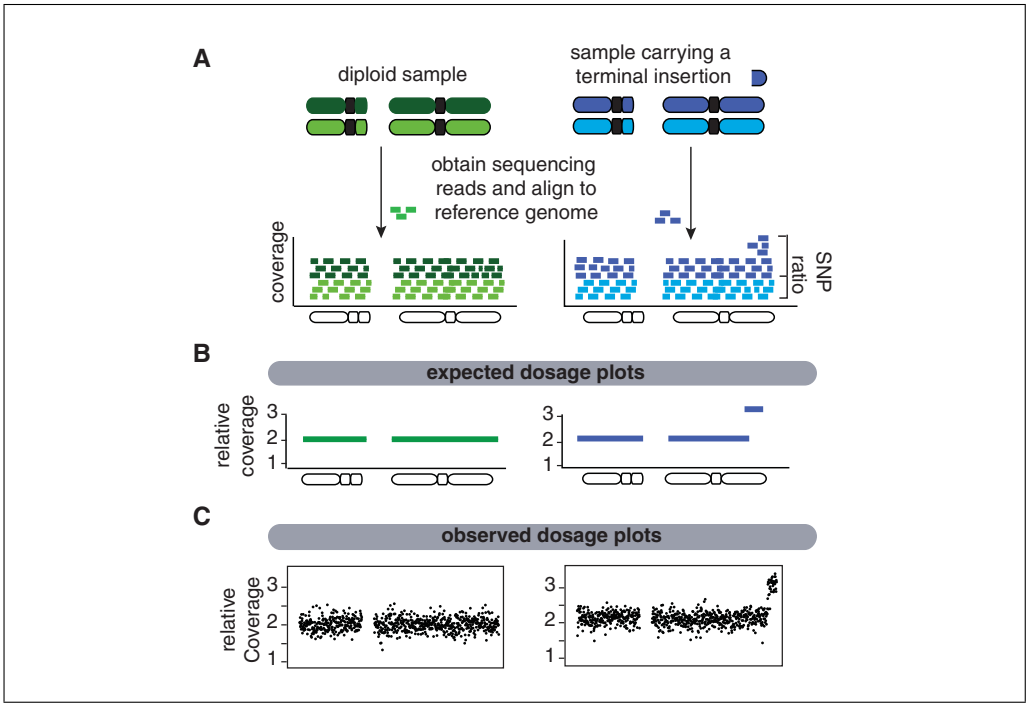

Figure 4 Expected and observed dosage plots. (**A**) Sequencing reads are obtained and aligned to the reference genome sequence for two samples, one diploid and the other carrying an insertion of the terminal end of the second chromosome. (**B**) Expected relative read coverage is directly proportional to sequence copy number. (**C**) Observed relative read coverage also depends on a variety of other factors, resulting in relative coverage values oscillating around the expected values. Each dot on the observed dosage plots represents the relative coverage value for one 100-kb non-overlapping bin. Data shown is representative of the type of dosage plots obtained from analysis of interspecific hybrids of poplar carrying indel mutations following gamma irradiation of pollen grains (Henry et al., 2015).

Materials

Software

Linux or MacOS, Unix-based computer system
Python2.6 or Python2.7
Allprep (if necessary)
BWA version 0.5.7 (*http://bio-bwa.sourceforge.net/*)
SAMtools version 1.18-dev or 1.19 (*http://SAMtools.sourceforge.net/*)

Table 1 `BWA-DoAll.py` Parameters

Parameter	Description
`-d STR`	Path to the database FASTA file to be used for alignment.
`-t INT`	Number of threads to use for alignment (Default = 1).
`-O`	To disable the program's unique read selection module (Optional). For more information consult BWA-DoAll documentation.
`-b STR`	Path to the directory containing the `BWA` executable.
`-a STR`	Path to the directory containing the `SAMtools` executable.
`-s STR`	Path to the folder containing `BWA-DoAll.py`.

BWA-DoAll (*http://comailab.genomecenter.ucdavis.edu/index.php/Bwa-doall*)
Bin-by-sam (*http://comailab.genomecenter.ucdavis.edu/index.php/Bin-by-sam*)

Data

High-quality demultiplexed FASTQ read sequence files (".fq" files).
Reference genomic file in FASTA format

Align reads

1. Prepare the work environment.

   This must be a Unix-based environment with the noted software installed. Please refer to individual software documentation to install programs.

2. Obtain the starting FASTQ read files and place them into a working directory containing no other file.

   These should be high-quality demultiplexed FASTQ (.fq) files. Sequences should have been processed for quality and adapter contamination, which is often handled by the sequencing facility. If FASTQ file preprocessing is needed, a program such as Allprep (https://github.com/Comai-Lab/allprep) can be used. File names should contain only alphanumeric characters.

3. Prepare the genomic reference file.

   The genomic reference sequence file should be in FASTA (.fa) format and in a different directory than the working directory. Note that for successful reference indexing (needed, for example, by the BWA aligner), the FASTA file must be in the correct FASTA format, with the reference sequence broken into many lines of up to 120 characters, not one line per sequence.

4. Align reads to the reference genome sequence using BWA and SAMtools to create files that can be easily parsed into coverage data.

   Many aligners are available. A standard aligner that creates many useful file types is BWA followed by SAMtools (Li and Durbin, 2009; Li et al., 2009). SAMtools output results in a .sam format that provides map location and quality score for each read and each sample.

   One option is to run the program BWA-DoAll in the directory containing only the sample FASTQ files. BWA-DoAll generates SAM, SAI, sorted BAM, and sorted BAI output files in corresponding result directories for each type of output (Fig. 3).

   Parameters available for `bwa-doall.py` *are summarized in Table 1.*

   The following example Python command line will run BWA-DoAll, using the reference sequence file `/genomes/organism/ref.fa` *with 8 threads and all executables that are installed in respectively named folders in the directory* `/software`*:*

```
python2.6 /pathtoBWA-DoAll/bwa-doall.py -d
/genomes/organism/ref.fa -O -t 8 -b /software/bwa/ -a
/software/samtools/ -s /software/BWA-DoAll/
```

Table 2 `Bin-by-sam.py` Parameters

Parameter	Description	
`-c STR`	Name of the control .sam file.	If no control file is specified, the normalized coverage for each sample is calculated as the mean percent reads in a bin divided by the mean percent reads of all libraries for that bin. If a control library is specified, the normalized coverage is calculated by dividing each mean percent read value by the corresponding value for the control sample. To ease downstream visualization, relative coverage values are multiplied by the background ploidy (`-p` parameter below).
`-o STR`	Output file name	A result file with this name will be created with information on the collected bins, and a second file with the name `readcounts-`<selected-name> will be generated with read counts and mapped counts per sample.
`-s INT`	Binsize (bps)	This parameter allows you to set the size of the bin. Only one bin size can be used per run, but it may be helpful to run several different sizes to change detection resolution. A small size will improve resolution but increase noise, whereas a large bin size will reduce noise but decrease the ability to detect smaller indel mutations.
`-b`		This option inserts blank lines between each reference-sequence bin set in the result file. This is for downstream analyses allowing the reference sequences to be more easily visualized graphically. This is not recommended if the reference sequences contain a high number of small references as opposed to a few large assembled chromosomes.
`-p INT`	Ploidy (number)	This integer scalar is used to indicate the background ploidy level. The default is 2 (diploid), this is used as a multiplier in the relative coverage calculation.
`-r STR`	Name of reference remove file.	This option will accept a text file containing reference names to remove from the output file. For more information consult `bin-by-sam.py` documentation (Optional).

BWA-DoAll outputs a .sam, .sai, .sorted.bam and .sorted.bam.bai file for each input .fq file with all files of the same data type placed together in a directory named accordingly. The BAM directory can later be used for pileup generation and parsing (see Basic Protocol 4), and the SAM directory will be used for step 5 of this protocol. The SAM and BAM files can be also used with many well-documented tools for calling single-nucleotide polymorphisms (Li et al., 2009).

Analyze dosage

5. Analyze dosage by running the program `bin-by-sam.py`, which is run on the directory of SAM files created by BWA-DoAll or any other software producing SAM alignment files of the correct format. The current working directory should be changed to the `/sam` directory created by BWA-DoAll in step 4, or in any other directory containing `.sam` files.

   This script divides the genome into non-overlapping consecutive regions or bins of a specified size, and for each sample, it records how many reads falls into each bin. The raw coverage values are then normalized either to a control size or to the average of all samples analyzed (see below). The resulting normalized read coverage values can be used to detect indel mutations (step 6).

   The parameters available for `bin-by-sam.py` *are summarized in Table 2.*

Dosage analysis of samples L1, L2, and L3 without specifying a control sample:

Chromosome	Start	End	L1	L2	L3	L1/NA	L2/NA	L3/NA
Chr01	1	1000000	1438	1023	1325	2.27	1.62	2.10

Dosage analysis using sample L3 as a control:

Chromosome	Start	End	L1	L2	L3	L1/L3	L2/L3	L3/L3
Chr01	1	1000000	1438	1023	1325	2.17	1.54	2.0

Figure 5 Example outputs of dosage analysis using the Bin-by-sam script. (**upper**) Example header and output of `bin-by-sam.py` run on three samples (named L1, L2, and L3) without specifying a control sample. (**lower**) Same example using sample L3 as a control.

Example command line:

```
python2.6 /pathToBinBySam/bin-by-sam.py -o run1-100k-
bins.txt -s 100000 -b
```

When completed, two new result files will have been created in the current working directory. One file (`readcounts-output.txt`*) will contain both the total number of reads per sample and the reads per megabase counted (i.e., normalized) per sample used in the analysis. The other file will contain the full table of read counts and normalized reads for each bin and for each sample, with the following columns (see Fig. 5 for an example output):*

1. Chromosome (or reference sequence) name

2. Start position of the bin (bps).

3. End position of the bin (bps)

Next N columns: Number of reads per bin, one column per sample (N samples).

Last N columns: Normalized reads counts, one column per sample (N samples).

6. Visually identify large-scale insertion and deletions using the coverage information obtained in step 5.

This step can be performed using a variety of methods, either computational or manual. We typically visualize coverage plots using JMP (SAS Institute, Inc.) and detect indel mutations by careful examination and identification of outliers using the dosage plots produced by the "overlay plot" function (Fig. 4C). Overall, relative coverage values for a diploid genome are expected to oscillate around 2.0. However, many factors contribute to variation around this value, including differential mapping efficiency, quality of the reference sequence, number of reads obtained, and random sampling of sequencing reads. The actual coverage values may be more or less noisy (Fig. 4).

The most important aspect of indel detection is to detect instances where consecutive bins exhibit the same coverage. We typically require a minimum of three consecutive bins with altered coverage in order to call an indel mutation. Below are additional general considerations to keep in mind in order to select high quality indel mutations:

(a) The output produced using bin-by-sam.py only informs us about the dosage of each chromosome or chromosomal piece relative to each other, within a sample. It is therefore very well suited to detect variation within a genome, including chromosome copy number or the presence of large deletions or insertions. On the other hand, it cannot detect ploidy variation—i.e., variation in the number of full genomic copies—such as triploidy or tetraploidy. Ploidy variation can be detected using parental allelic bias (see Basic Protocol 4) instead.

(b) Some regions of the genome exhibit extreme coverage values. For example, repeated regions can exhibit extremely high coverage. It is sometimes useful to exclude these regions from the analysis at this stage. This can be done pre-alignment, by using a RepeatMasked reference sequence for alignment, or post-alignment by excluding chromosome bins that exhibit higher than expected coverage (coverage outliers).

(c) If the reference genome sequence contains organellar sequences, these should be excluded from the coverage calculations (`-r` option in `bin-by-sam.py`), as the amount of organellar DNA will vary greatly depending on the type of tissue or the DNA extraction method used.

(d) Mutations are expected to be independent in all samples. When possible, it is important to compare indel mutations detected from multiple samples and discard mutations identified in more than one sample (unless the pedigree of the samples is such that they could have inherited the same mutation). Specific regions can exhibit coverage variation that is not due to copy number variation, especially if the reference genome sequence used is not exactly the same as the samples.

(e) Sequence coverage will dictate the bin size that can be used. As a rule of thumb, we select a bin size that allows for an average of 100 reads per bin for all samples. Analyzing multiple bin sizes can be useful to confirm indel mutations.

GENOTYPING OF MUTATIONS

BASIC PROTOCOL 4

Once insertion and deletion mutations have been identified, it can be useful to evaluate the origin of the mutation, i.e., which parent contributed the indel. This information can be used, for example, to determine the frequency of mutations contributed by the non-irradiated parent, the identification of F1 ploidy variants, and which parent contributed an unreduced gamete in the case of triploid F1 individuals. These analyses are performed using the same sequencing read data as those used for mutation detection (Basic Protocol 3). The pipeline described here starts from the output files of Basic Protocol 3 and provides the steps necessary to evaluate the ratio of genotypes within each indel by calculating the ratio of parental allele contributions.

The steps described in this protocol are summarized in Figure 6. In short, SAMtools (Li et al., 2009) `mpileup` function is used to generate a multiple library pileup file that is then parsed into a more readable format, containing information about which alleles are found at each position in the reference genome and for each of the sample alignment files present in the directory. A parental allelic ratio is then assigned to each

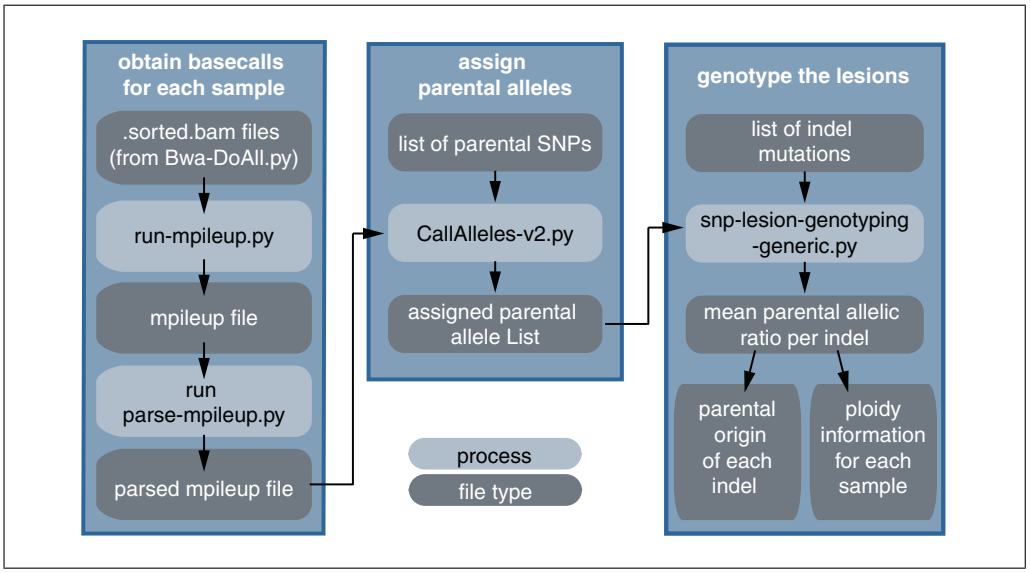

Figure 6 Pipeline to determine parental contribution. Generated sorted BAM files are combined and converted using mpileup to a multiple alignment file, which is then parsed. The parsed multiple alignment file is used to collect allele data for each sample at pre-determined parental SNP positions. Regions of interest and allelic information are combined to determine the allelic ratio of each indel, as well as the overall allelic ratio of each sample. This data is used to infer parental contributions to each indel and the ploidy of the samples.

position previously determined to be polymorphic between the two parental genotypes. Finally, for each sample, parental ratio values for each position are averaged across each indel mutation and across the rest of the genome and these mean values are compared to each other. Genome-wide allelic information is also used to infer the ploidy of each sample.

Materials

Software

Linux or MacOS, Unix-based computer system
Python2.6 or Python2.7
Mpileup generation and parsing tools
(*http://comailab.genomecenter.ucdavis.edu/index.php/Mpileup*)
CallAllelesAB (*http://comailab.genomecenter.ucdavis.edu/index.php/Call_Alleles*)
indel-allelic-ratio.py
(*http://comailab.genomecenter.ucdavis.edu/index.php/Indel_allelic_ratio*)

Data

SAMtools sorted BAM files (`.sorted.bam`; see Basic Protocol 3)
Reference genome file in FASTA format (`.fa`)
List of parental SNP alleles
List of indels (identified in Basic Protocol 3)

Generate a multiple library pileup file

1. Prepare the work environment.

   This must be a Unix-based environment with the noted software installed. Please refer to individual software documentation to install programs.

2. Place the alignment (`.sorted.bam`) files into a working directory containing no other files.

   The sorted BAM files were previously generated using SAMtools in Basic Protocol 3. The file names should end in "`_aln.sorted.bam`*" and should have been previously processed for quality and adapter contamination (see Basic Protocol 3). File names should contain only alphanumeric characters.*

3. Execute the program `run-mpileup.py` to create an mpileup file with read coverage information for every sample.

   The parameters available for `run-mpileup.py` *are summarized in Table 3.*

   The following is an example command line to execute `run-mpileup.py` *with 8 threads, using the reference file* `/genomes/organism/ref.fa`*, the output* `result_mpileup.txt`*, and all executables installed into respectively named folders in a* `/software/` *directory:*

```
python2.6 /path-to-run-mpileup/run-mpileup.py -r
/genomes/organism/ref.fa -o result_mpileup.txt -s
/software/samtools/ -t 8
```

Table 3 `run-mpileup.py` Parameters

Parameter	Description
`-r STR`	Path to the database FASTA file that was used for alignment.
`-t INT`	Number of threads to use for mpileup generation (Default = 1).
`-o STR`	A result mpileup file with this name will be created.
`-s STR`	Path to the directory containing the `SAMtools` executable.

Table 4 `mpileup-parser.py` Parameters

Parameter	Description
`-f STR`	Path to the mpileup output file generated in step 3.
`-t INT`	Number of threads to use for mpileup generation (Default = 1).

4. Parse the mpileup file using `mpileup-parser.py` to create a more compact and more readable file. The parsed mpileup file will be named `parsed-result_mpileup.txt`.

   The parameters available for `mpileup-parser.py` *are summarized in Table 4.*

   The following is an example command line to run `mpileup-parser.py` *with 8 threads, using the multiple alignment file* `result_mpileup.txt`*:*

   ```
   python2.6 /path-to-mpileup-parser/mpileup-parser.py -f
   result_mpileup.txt -t 8
   ```

Analyze parental genotypes of indels

5. Prepare the parental single-nucleotide polymorphism (SNP) file.

   To call parental alleles, a parental SNP file must be available. This file must be a tab-delimited text file with the following columns (see Fig. 7A for example):

   1. Chromosome (or reference sequence) name. This name must match the name present in the alignment FASTA reference exactly.

   2. The bp position within the chromosome or reference.

   3. The reference sequence base nucleotide at that position.

   4. The genotype of parent A at that position.

   5. The genotype of parent B at that position.

   Multiple methods can be used to create such a file, either from existing SNP files or from the parental sequencing reads. However, the selection of robust SNPs between the parental genotypes is beyond the scope of this protocol. Only positions that are homozygous in both parents but different between the two parents should be used.

6. Assign alleles by running the program `CallAllelesAB.py` with the parental SNP file.

   CallAlleles uses the following logic to call parental alleles:

   a. If only one allele type is present, and it is one of the parental alleles, that parental allele is assigned.

   b. If two alleles are present, but the less frequent allele represents less then 10% of the calls, the more frequent parental allele is assigned.

   c. If two alleles are present and both represent >10% of the calls, and both alleles correspond to the two parental alleles, the allele is called as a Het.

   d. In all other bi-allelic cases with each >10%, the genotype called is "na"

   The parameters available for `CallAllelesAB.py` *are summarized in Table 5.*

   The following is an example command line to run `CallAllelesAB.py` *using the parental SNP file* `/data/p-SNPS.txt`*, the parsed pileup* `/data/parsed-result_mpileup.txt`*, and the output file* `output.txt`*:*

   ```
   python2.6 /path/to/CallAllelesAB.py -s input-
   parental-SNPFile.txt -m create-parsed_mpileup.txt -o
   output_file_name.txt
   ```

A SNP input format:

Chr	Pos	Ref	A	B
Chr01	8445	T	A	T
Chr01	8556	T	T	G
Chr01	8678	A	G	A
Chr01	8680	G	G	*
Chr01	8722	C	C	T

B Indel input format:

Sample	Chrom	Start	End
Sample1	Chr01	47100001	49300000
Sample1	Chr02	8200001	13800000
Sample2	Chr08	100001	12800000
Sample3	Chr02	1	6700000
Sample4	Chr08	18200001	18700000

C Parent A allele ratio:

Sample	Chrom	Start	End	SNP_Indel	%AIndel	SNPsGenome	%AGenome
L161	Chr08	161001	1946546	796	40	129131	51
L161	Chr05	157001	2589004	2128	95	129131	51
L161	Chr02	189001	2020000	334	38	129131	51
L217a	Chr14	30001	840000	3175	35	162111	51
L217a	Chr01	471001	4930000	515	91	162111	51
L7b	Chr15	1	540000	7280	35	479185	50
L7b	Chr15	70001	1528577	1391	31	479185	50

Figure 7 Examples of input file formats and output of parental allele ratios. (**A**) Format of the parental SNP input data. (**B**) Format of the indel input data. (**C**) Example of the output from the genotype analysis showing the parental contribution to each indel.

Table 5 `CallAllelesAB.py` Parameters

Parameter	Description
`-s STR`	Path to the input parental SNP file created in step 6.
`-m STR`	Path to the parsed pileup file created in step 5.
`-o STR`	This is the output file name.

The output file has the following format:

Columns 1 through 5 are only present once. Columns 6 though 11 will be present for each sample in the `parsed-mpileup.txt` *file.*

1. Chromosome (or reference sequence) name.

2. Position in reference sequence.

3. Chromosome or reference sequence nucleotide base.

Table 6 `indel-allelic-ratio.py` Parameters

Parameter	Description
`-a STR`	Path to the output file generated by `CallAlleles-v2.py` in step 7.
`-l STR`	Path to the region of interest file created in step 8.
`-o STR`	This is the output file name.

4. Parental A allele from the SNP file.

5. Parental B allele from the SNP file.

6. SNP Type: This will be '1' if homozygous, and '2' if heterozygous.

7. SNP1, the most common allele.

8. SNP2, the second most common allele.

9. Total coverage at that position.

10. The percent parental allele A: This will be '1' if homozygous A, '0' if homozygous B, and '0.5' if heterozygous.

11. The coverage of the parental A allele.

7. Prepare the file of indel mutations for genotyping.

This information derives from the analysis of the output of Basic Protocol 3. This file should be a tab-delimited text file with the following columns (see Fig. 7B, as example of the file format):

1. Sample Name: This name must be exactly the same name as those used in the `parsed-mpileup.txt` *file. All samples with indel mutations should be represented in the* `parsed-mpileup.txt` *file.*

2. Chromosome (or reference sequence) name.

3. Start of indel region (bps).

4. End of indel region (bps).

8. Genotype the indels by running the `indel-allelic-ratio.py` script.

This script combines the indel information obtained in Basic Protocol 3 with the allelic calls obtained in step 7 of this protocol to calculate the mean allelic ratio across each indel mutation and across the rest of the genome, for each sample. The script outputs the following: the number of SNP alleles available in each indel, the percentage of alleles from parent A within each indel, the number of SNP alleles present outside of all indels for that sample, and the corresponding percentage of alleles from parent A.

The parameters available for `indel-allelic-ratio.py` *are summarized in Table 6.*

The following is an example command line to run `indel-allelic-ratio.py` *using the allele file* `AB-alleles.txt`*, indel file* `AB-indels.txt`*, and the output file* `AB-output-geno.txt`*:*

```
python2.6 /path-to-allelic-ratio-script/indel-
allelic-ratio.py -a AB-alleles.txt -l AB-indels.txt -o
AB-output-geno.txt
```

The output file `AB-output-geno.txt` *is a tab-delimited text file with one line per indel mutation (see Fig. 7C for example output). The columns are defined as follows:*

1-4: Correspond to columns 1 through 4 of the input indel file.

Table 7 Expected Percentage of Parent A Alleles Depending on Sample Ploidy and Dosage Variation

Ploidy	Insertion from parent A	Deletion from parent A	Insertion from parent B	Deletion from parent B	Rest of the genome
AB diploid	67	0	33	100	50
ABB triploid	50	0	25	50	33
AAB triploid	75	50	50	100	67

5: Total number of called SNPs in the sample that lie within the indel.

6: Percent SNPs within the indel mutation that derive from parent A.

7: Total number of called SNPs in the sample that lie outside of indel mutations.

8: Percent SNPs outside of all indel mutations that derive from parent A.

If a sample contains multiple indel mutations, the values in the last two columns will be same for each line of that sample.

The output file `AB-output-geno.txt` *can be used to evaluate the parental origin of each indel mutation as well as the ploidy of each sample, as described in step 9.*

9. Analyze parental allele data.

The output of step 8 can serve two purposes. First, it can be used to investigate the parental contribution to each indel. For example, in a diploid individual, an insertion originating from parent A is expected to exhibit a mean percentage of 67% A allele (2A:1B). On the other hand, if the insertion originates from the B parent, a percent A allele close to 33% (1A:2B) is expected instead. Second, it can be used to determine the ploidy of each individual. For a diploid individual, the part of the genome that is not included in indels is expected to display a percent A allele close to 50% (1A:1B). If that is not the case, the background ploidy can be assessed. For example, a triploid with two genomic copies from genome A will display a background percentage of A allele close to 66% instead. Similar expectations can be calculated for all possible combinations of ploidy and dosage variation, as described in Table 7.

An example distribution of % parent A alleles in diploid and triploid samples carrying indel mutations can be found in Figure 4 of Henry et al. (2015). Panel B of that figure also exemplifies the relationship between dosage variation as detected by Basic Protocol 3 and indel genotyping as described above.

COMMENTARY

Background Information

Tree breeding has typically been focused on manipulating "natural" genetic variation. For the dioecious species of *Populus*, this has included both intra- and inter-specific crosses of promising genotypes, followed by progeny testing to evaluate F1 families and individual F1 clones. This approach has been highly productive and generated superior clones, but has its limitations with regards to research and improvement of poplar.

The protocols described here create novel genetic and phenotypic variation beyond what would normally be observed in an F1 family. This variation is of course a source of potentially novel cultivars for industrial plantings. But this novel genetic variation also allows the functional characterization of genes and genetic pathways that may show little natural variation, and thus be transparent to traditional approaches—e.g., quantitative trait loci mapping or association genetics studies.

Importantly, effective genomic approaches that can assay the function of a significant portion of a genome have been largely lacking in trees. The approach described here pairs the creation of novel genetic and phenotypic variation with precision genotyping of the causative deletions and additions of chromosomal regions, and is capable of covering the poplar genome with insertions and deletions with a reasonable number of clones. We expect this will be a useful tool for poplar and other clonally propagated species for both research and

cultivar production. The approach also sheds light on the potential range of phenotypic variation that can be produced by a species or hybrid.

Poplar breeding is both an art and a science. Breeding methods for poplar are relatively well established (Stanton and Villar, 1996), but do vary among workers and species combinations. Good information is also available for the fecundity and performance of many poplar hybrids (Dickmann and Isebrand, 2001; Eckenwalder, 2001), as well as for different breeding strategies (Bissof and Gullberg, 1996; Riemenschneider et al., 2001). However, the successful breeder must be adaptable and understand how to make adjustments to the breeding process to ensure success.

Critical Parameters

The breeding of forest trees presents logistical challenges for short-term research. Forest trees, including poplars, flower once per year, so it is thus critical to organize all aspects of breeding well in advance. It is advisable, whenever possible, to utilize genotypes that are known to perform well together in crosses. Variation in annual weather can affect the timing of spring bud burst and other factors, and should be considered in determining when to collect fertile branches. Wherever possible, create redundancy by using multiple branches per genotype and using multiple genotypes to help ensure success.

It is imperative that female branches be well rooted prior to temperature forcing and crossing. Without adequate roots to support growth, catkins may emerge but will abscise. Monitor rooting of the female branches by looking for roots as they reach through the soil to the edge of the transparent five gallon false-bottom buckets. The soil must be kept moist but not saturated with water in the rooting bucket. If the soil is overly wet, anaerobic conditions and decay will effectively halt root development.

Determining when females are receptive for pollination is not difficult, but two applications of pollen, 2 to 3 days apart, will help ensure success and full setting of seed. If pollination is not successful or if the pollen is not viable, catkins will typically abort. Poplar seed rapidly loses viability, and should be sown soon after collection.

For genomic analysis, multiple control progeny from crosses between parents without irradiation are necessary to experimentally establish the presence or absence of preexisting insertions or deletions and the rate of spontaneous insertion or deletion. Sequence coverage can be adjusted to allow for the detection of smaller insertions or deletions, or to more precisely map the junctions of insertions and deletions. Likewise, sequence coverage and bin size can be adjusted to accommodate mapping of reads from more distantly related species (e.g., aspens) to the current *P. trichocarpa* reference genome.

Troubleshooting

If a cross fails, remove the female branch from its rooting bucket and check for root development. If there are few or poorly developed roots, this is the likely cause of failure. If roots developed well, inviable pollen or mistimed application of pollen would be likely causes. If one has multiple false-bottom rooting buckets for female branches, one can temperature force them in succession and potentially correct mistakes in subsequent crosses.

For seedling development, carefully monitor for signs of wilting, and adjust mist and water accordingly. Poplar seed are very small and the early stages of seedling development are sensitive to temperature and desiccation. Monitor carefully for disease and insect pests (e.g., fungal gnat larvae). Embryo rescue can be helpful for some more difficult crosses (Stanton and Villar, 1996).

Anticipated Results

In our example (*P. deltoides* × *nigra*), a single female branch with several catkins has the potential to produce hundreds of seedlings. By taking a conservative approach and collecting and pollinating multiple female branches, the successful application of this protocol should produce many hundreds or seedlings that together will saturate the genome with insertions and deletions.

In our cross of *P. deltoides* with gamma-irradiated *P. nigra* pollen, we observed ~8% triploid individuals in the F1 population. Most of the triploid individuals originated from a single cross, while other crosses with the same parents or the same irradiation levels generated much lower percentages of triploid progeny. This suggests that environmental factors might affect 2*N* gamete production. Triploid individuals carried a higher number of indel mutations per individual than their diploid siblings.

Time Considerations

Genotypes and individual trees for crossing should be identified the fall prior to performing crosses. Fertile branches should be collected in

mid to late winter. Rooting of females and performing crosses may take up to two months. Seed collection, sowing, germination, and initial seedling establishment will typically occur in spring to early summer. Robust seedlings with extensive root systems should be well established by the end of fall. Seedlings can be allowed to go dormant or forced to continue growth in controlled growth rooms. In the second year of growth, seedlings can be vegetatively propagated for experiments.

Acknowledgements

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0007183. We thank Annie Mix and Courtney Canning for assistance in propagation of poplar. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Literature Cited

Bissof, S. and Gullberg, U. 1996. Poplar breeding and selection strategies. National Research Council of Canada, Ottawa, Ontario, Canada.

Dickmann, D.I. and Isebrand, J.G. 2001. Poplar clones: An introduction and caution. National Research Council of Canada, Ottawa, Ontario, Canada.

Eckenwalder, J.E. 2001. Keys to species and main crosses. National Research Council of Canada, Ottawa, Ontario, Canada.

Henry, I.M., Zinkgraf, M.S., Groover, A.T., and Comai, L. 2015. A system for dosage-based functional genomics in poplar. *Plant Cell* 27:2370-2383 doi: 10.1105/tpc.15.00349.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760. doi: 10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079. doi: 10.1093/bioinformatics/btp352.

Mock, K.E., Callahan, C.M., Islam-Faridi, M.N., Shaw, J.D., Rai, H.S., Sanderson, S.C., Rowe, C.A., Ryel, R.J., Madritch, M.D., Gardner, R.S., and Wolf, P.G. 2012. Widespread triploidy in Western North American aspen (*Populus tremuloides*). *PLoS One* 7:e48406. doi: 10.1371/journal.pone.0048406.

Riemenschneider, D.E., Stanton, B.J., Vallee, G., and Perinet, P. 2001. Poplar breeding strategies. National Research Council of Canada, Ottawa, Ontario, Canada.

Stanton, B and Villar, M. 1996. Controlled reproduction of *Populus*. *In* Biology of *Populus* and its implications for management and conservation. Stettler, Bradshaw, Heilman, and Hinckley, Eds. NRC Research Press, Ottawa, Ontario, Canada.

Stettler, R.F., Zsuffa, L., and Wu, R. 1996. The role of hybridization in the genetic manipulation of *Populus*. NRC Press, Ottawa, Ontario, Canada.

Internet Resources

https://www.python.org/
Python language website.

http://samtools.sourceforge.net/
SAMtools software website.

http://bio-bwa.sourceforge.net/
BWA software website.

http://comailab.genomecenter.ucdavis.edu/index.php/Bwa-doall
BWA-DoAll software website.

http://comailab.genomecenter.ucdavis.edu/index.php/Bin-by-sam
Bin-by-sam software website.

http://comailab.genomecenter.ucdavis.edu/index.php/Mpileup
Mpileup generation and parsing tools website.

http://comailab.genomecenter.ucdavis.edu/index.php/Call_Alleles
CallAllelesAB software website.

http://comailab.genomecenter.ucdavis.edu/index.php/Indel_allelic_ratio
Indel-allelic-ratio software website.