Genome Resequencing in *Populus*: Revealing Large-Scale Genome Variation and Implications on Specialized-Trait Genomics

Wellington Muchero, Jessy Labbé, Priya Ranjan, Stephen DiFazio, and Gerald A. Tuskan

Abstract To date, *Populus* ranks among a few plant species with complete genome sequences and other highly developed genomic resources. With the first reference genome among all tree species, *Populus* has been adopted as a suitable model organism for genomic studies in trees. However, far from being just a model species, *Populus* is a key renewable resource that plays a significant role in providing raw materials for the biofuel and pulp and paper industries. Therefore, aside from leading frontiers of basic tree molecular biology and ecological research, *Populus* leads frontiers in addressing global economic challenges related to fuel and fiber production. The latter fact suggests that research aimed at improving quality and quantity of *Populus* as a raw material will likely drive the pursuit of more targeted and deeper research in order to unlock the economic potential tied in biological processes that drive this tree species. Advances in genome sequence-driven technologies, such as resequencing individual genotypes, which in turn facilitates large scale SNP discovery and identification of large scale polymorphisms are key determinants of future success in these initiatives. In this treatise we discuss implications of genome sequence-enable technologies on *Populus* genomic and genetic studies of complex and specialized traits.

W. Muchero (✉) • J. Labbé • G.A. Tuskan
Biosciences Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA
e-mail: mucherow@ornl.gov

P. Ranjan
University of Tennessee, Knoxville, TN 37916, USA

S. DiFazio
West Virginia University, Morgantown, VA 26506, USA

T. Fenning (ed.), *Challenges and Opportunities for the World's Forests in the 21st Century*, Forestry Sciences 81, DOI 10.1007/978-94-007-7076-8_25,

1 Introduction

The genus *Populus* is an economically important tree crop widely grown as feedstock for lignocellulosic biofuels and pulp and paper products, in part, for its rapid growth (Tuskan 1998; Yang et al. 2009) and ability to thrive on economically marginal lands that are not suitable for food crop production (Tuskan and Walsh 2001). Aside from its key importance as an industrial feedstock, *Populus* is also a biological model system for perennial tree crops due to its relatively compact genome, high level of interspecies diversity and ease of experimental manipulation compared to other tree genera (Tuskan et al. 2006). Given its central role in both economic and scientific realms, substantial investment has been made in developing genomic and genetic resources to facilitate experimental enquiry into the biology of adaptive traits in *Populus*. These resources include (1) the first reference genome of any tree species (Tuskan et al. 2006), over 900 resequenced *P. trichocarpa*, a repository of more than 48 million high-quality single nucleotide polymorphism (SNPs), Illumina Infinium SNP arrays with 34,131 and 5,390 SNP probes (Slavov et al. 2012; Geraldes et al. 2013), over 2,200 SSR markers with known positions on the genome assembly (Muchero et al. 2013). These cumulative resources have been used in wide-ranging studies ranging from population genetics studies to candidate gene identification using quantitative trait loci (QTL) mapping. The afore-mentioned 5,390 SNP array was recently applied to genotype three mapping pedigrees, 52–124, 545 and 54B, resulting in successful construction of higher density genetic maps with 3,500, 1,200 and 530 SNP markers, respectively. All together, these resources have been brought to bear on fundamental questions of *Populus* genome characteristics such as patterns of recombination across chromosomes, linkage disequilibrium and *P. trichocarpa* population structure along a Pacific Northwest cline (Slavov et al. 2012). In addition, the genomic resources and understanding of genome attributes are being used to identify genetic determinants as well as putative causal mutations for traits such as cell wall characteristics, biomass production, tolerance to heavy metals (Induri et al. 2012) and phenological traits using QTL and association mapping approaches.

The on-going resequencing initiative targeted at completion of 1,100 genomes is well underway with the fascinating possibility of saturating nucleotide polymorphisms discovery. In this regard, *Populus* is keeping pace with the *Arabidopsis* model system, which represents the most mature genomic resources of any plant species, where at least 80 ecotypes have been resequenced in the on-going 1001 genomes project (Cao et al. 2011). At present, over 900 *Populus* genotypes have been successfully resequenced. This collection of genomes will revolutionize approaches in identifying causal polymorphism on scales beyond individual SNPs (Abraham et al. 2012). Still, large-scale structural variation such as insertions/deletions (INDELs) and whole-gene deletions are not readily apparent based

high-throughput genotyping technologies and hence, are typically under-represented in genetic mapping studies. Sequenced-based studies targeted at associating INDELs and whole-gene deletion polymorphisms to relevant economically important traits should, therefore, benefit greatly from current advances in *Populus* genomics.

It is our objective, in remainder of this chapter, to illustrate this potential by exploring the genomic distribution of INDELS using receptor kinases, which exhibit disproportionately higher levels in INDEL polymorphisms compared to other gene classes, as a proof-of-concept. In addition, we explore SNP marker coverage and segregation patterns in regions harboring this class of genes in genetic maps as a means of illustrating the poor representation resulting from technical challenges presented by INDELs in high-throughput genotyping assays.

2 Beyond Single Nucleotide Polymorphism

As stated above, one of the areas of *Populus* genomics that should receive significant benefit from whole-genome resequencing will be the identification of large-scale polymorphisms such as small INDELs and whole-gene deletions. Due to the fact that the majority of genotyping assays are designed to characterize difference in allelic forms that are expected to conform to expected segregation patterns, instances where genome segments are missing often result in "anomalous" segregation that are typically excluded from further analysis. In on-going analyses, we observed that this is especially true for genomic regions that are typified by long stretches of tandem repeats that in turn exhibit frequent events of insertions and or deletions. Receptor kinases possess these hallmark features and therefore, provide good targets for exploring limitations of current genotyping technologies in identification and mapping of INDELs. We demonstrate that in light of these limitations, the genetic basis of traits mediated by these polymorphisms may not be fully accounted for. Additionally, receptor kinases have been implicated in a variety of economically important biological process that require highly specific recognition between individual cells such as pollen recognition in self incompatibility (Iwano and Takayama 2012), between organisms as innate immune response to pathogen attack (Schwessinger and Ronald 2012), and/or in mycorrhizal formation (Kereszt and Kondorosi 2011). Specifically, D-mannose lectins are known to facilitate host recognition of bacterial, fungal, viral and insect pests or symbionts in plants based on highly specific binding of mannose moieties on microbial cell walls (Schwessinger and Ronald 2012). With abundant evidence linking receptor kinases such as lectins to innate immunity in both plants and animals, exploration of polymorphisms occurring among these key genes and their potential effects on *Populus*-microbe or *Populus*-insect interactions hold significant promise in unraveling the molecular genetics behind such processes.

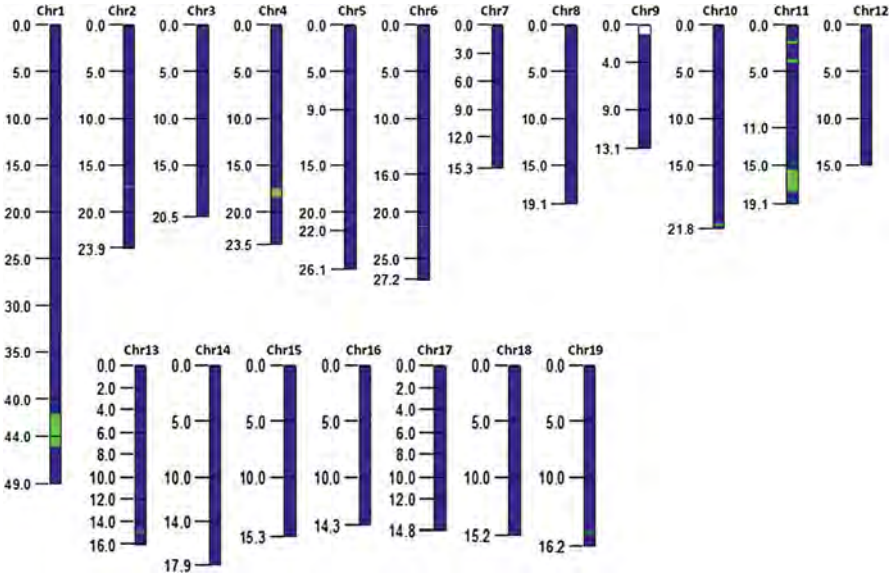

Fig. 1 Genomic distribution of D-mannose lectin paralogs (*green*) in the *Populus* genome

To that end, we analyzed the distribution of 96 paralogous D-mannose lectins within the *Populus* genome. This exercise revealed their presence on chromosomes I, II, IV, VI, X, XI, XIII and XIX with chromosomes I and XI accounting for 78 (81 %) of the D-mannose lectin genes in large tandem repeats (Fig. 1).

3 INDELs Based on Resequenced Genotypes

Based on the above observations, we focused attention on tandem repeats on chromosome I and XI in assaying structural variation in 45 *P. trichocarpa*, 2 *P. deltoides* and 1 *P. tremuloides* resequenced genomes. Variation among lectins within *P. trichocarpa* species revealed patterns of small to large-scale deletions involving intergenic, promoter and genic regions (Fig. 2). Notably, there was evidence of apparent whole-gene deletions when *P. trichocarpa* was compared to other *Populus* species, albeit based on limited alternate species genomes. In such cases, only SNP polymorphisms were found in within-*trichocarpa* comparisons, suggesting high levels of conservation of these genes in this species whereas no apparent homologs were found in either *P. deltoides* or *P. tremuloides*. These observations are in line with *Populus*-microbe interactions that exhibit within-species and between-species specificity. We hypothesize that significant advances in understanding these highly specific interaction can be made by focusing on the types of polymorphisms that occur at the species level.

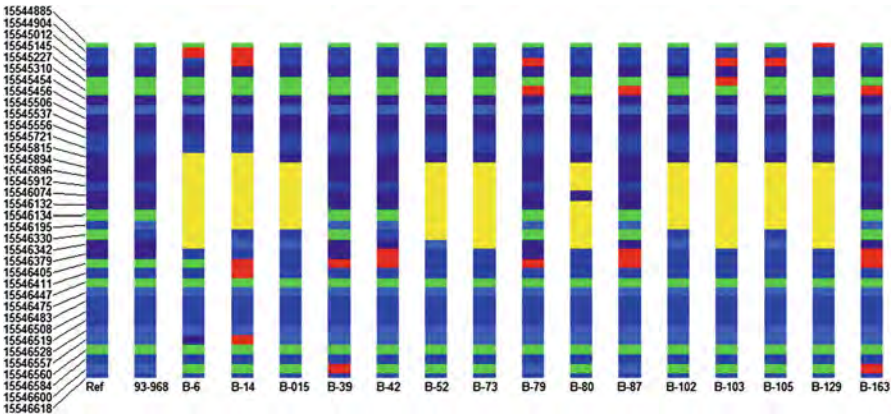

Fig. 2 INDEL polymorphisms (*yellow*) up to 564 bp in a region harboring D-mannose lectins on chromosome XI. Y-axis represents physical positions and the X-axis represents re-sequenced genotypes

4 Coverage in Genetic Maps

Tagu et al. (2001) reported significant segregation in *Populus* colonization by the symbiont *Laccaria bicolor* in F_1 progeny from a *P. trichocarpa* x *P. deltoides* interspecific cross. Further, Labbé et al. (2011) suggested preferential colonization of *P. trichocarpa* compared to *P. deltoides* and mapped QTL associated with percent colonization in the F_1 pedigree family 54B. Interestingly, a major QTL appeared to co-locate with regions harboring receptor kinases on chromosome XI described above. Recognizing that most *Populus* genetic maps did not include deletion polymorphisms, we were interested in exploring how alternate markers would be distributed in that region. We analyzed three recently completed SNPs maps for the hybrid *Populus* families 52–124 (pseudo-backcross), 545 (F_1) and 54B (F_1) for marker coverage in these specific regions. We observed large gaps in all three maps with the 52–124 having the shortest gap of 584 KB in linkage group intervals corresponding to the genomic region harboring the receptor kinase tandem repeat. Figure 3 illustrates these gaps on chromosome XI for F_1 maps 545 and 54B, respectively. Interestingly, the apparent presence of INDELs was independently validated based on SNP genotyping with probes targeting SNPs in chromosome I and XI intervals. SNP segregation violating expected segregation patterns were observed in the F_1 progeny (Fig. 4c and d). In this regard, both parental genotypes appear to be homozygous (red and blue dots), as such, all F_1 progeny would be expected to be heterozygous (maroon dots in the center). However, both heterozygous and homozygous forms are observed for the progeny in the SNP assay. This segregation pattern could be explained if one parent had A- genotype, - representing a deletion; and the other parent having the BB genotype. The SNP assay recognizes a single signal ‘A’ hence the appearance of ‘A-’as being homozygous. Similarly, progeny

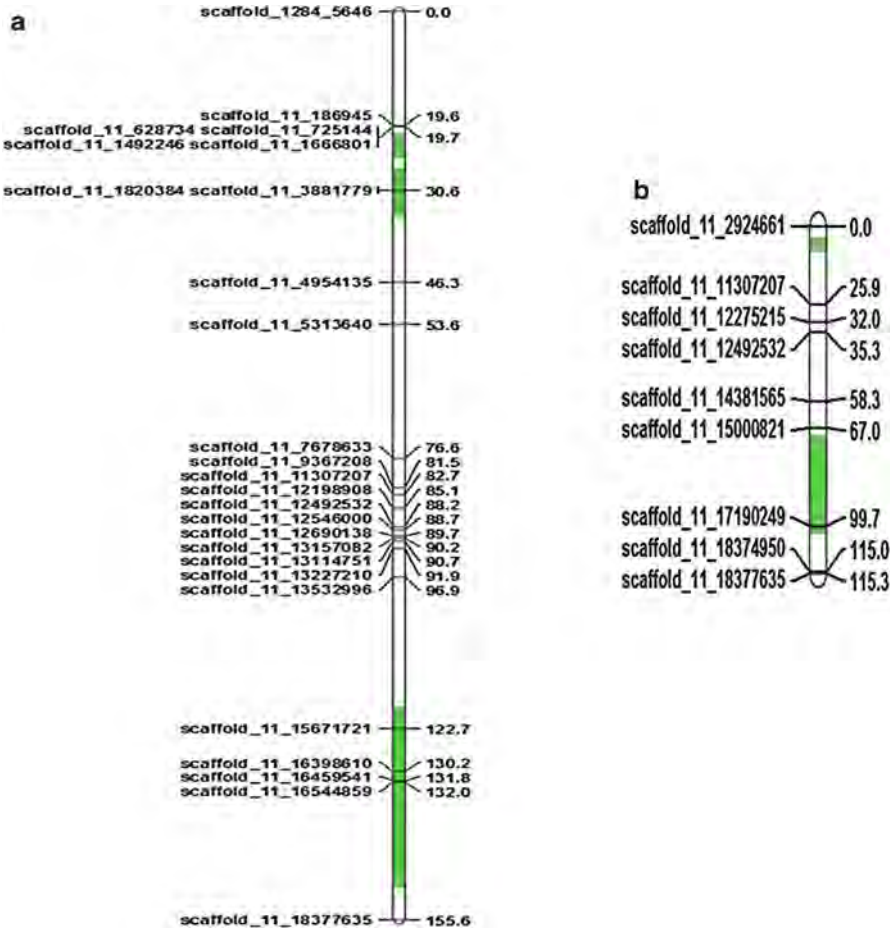

Fig. 3 Marker coverage in genomic intervals (*green*) harboring D-mannose receptor kinases in F_1 genetic maps (**a**) Family 545 and (**b**) Family 54B

that inherit the chromosome with the deletion will have genotype 'B-', which appears in the SNP assay to be homozygous BB, given the detection of only the B allele signal. Without prior knowledge of INDELs, these results would be treated as erroneous and typically excluded from map creation leading to gaps evident in the genetics maps illustrated in Fig. 3.

These observations highlight the challenges of identifying genomic regions that harbor genetic determinants of plant-microbe or plant-insect interactions, since resistance genes typically share the hallmark feature of occurring in tandem repeats with INDELs as the predominant form of functional polymorphism. As such, in work aimed at understanding plant-microbe interaction, inclusion of INDELs in genetic maps is paramount and its importance cannot be over-emphasized. Since the

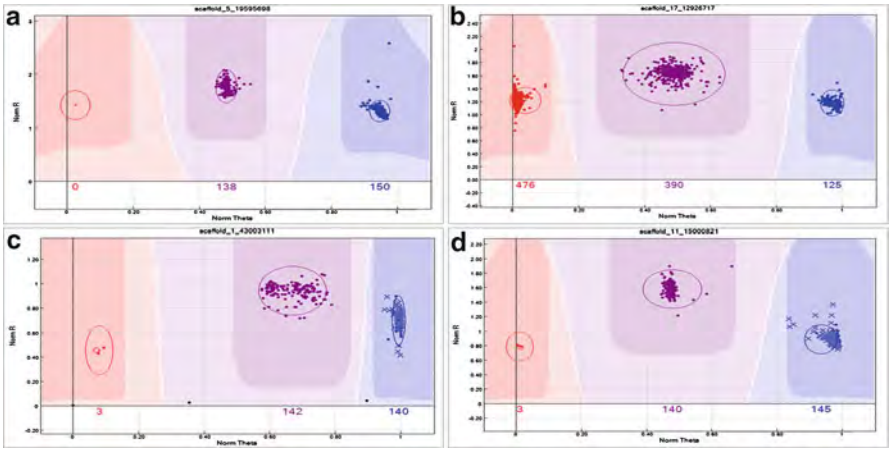

Fig. 4 Genome studio graphical representation of a SNP segregating normally in (**a**) a *Populus* pedigree and (**b**) a naturally varying collection of *P. trichocarpa*, unexpected SNP segregation in a F_1 *Populus* pedigree corresponding the region harboring D-mannose lectins on (**c**) chromosome I and (**d**) chromosome XI. GenomeStudio software recognizes progeny with unexpected genotypes (blue *x* instead of *dots*) based on pedigree information. Three samples on the *left* in panels c and d represent the replicate parent samples with the A-genotype

present discussion only considered D-mannose lectins, we believe this lack of adequate coverage is characteristic of other gene classes that have been shown to exhibit similar patterns of distribution and polymorphism content (Rodgers-Melnick et al. 2012).

5 *Populus* Phenomics

With the unprecedented amount of genotypic information becoming available for *Populus* through the resequencing initiative, the key limitation now remains access to high-quality phenotypic data. In an effort to bridge that gap, an international collaboration resulted in setting up of four common gardens with replicated 1,100 *P. trichocarpa* genotypes established to represent the genetic diversity in *P. trichocarpa* along the Pacific Northwest cline (Slavov et al. 2012). As a step toward association studies that include SNP and INDEL polymorphisms, we have established the presence of sufficient genetic variation in key traits such as susceptibility to key pathogens *Taphrina* spp. and *Venturia* spp. in common garden setting (Fig. 5). In addition, beneficial rhizosphere microbial interactions are being explored within the context of a U.S. Department of Energy project entitled Plant Microbe Interfaces at the Oak Ridge National Laboratory. This project seeks, in part, to establish the role that *Populus* genotype plays in shaping

Fig. 5 Disease symptoms showing evidence of genetic segregation in *P. trichocarpa* common gardens in the Pacific Northwest. (**a**) Leaf blistering caused by *Taphrina populina*, (**b**) shepherd's crook and (**c**) leaf necrosis caused by *Venturia*

microbial community composition in the rhizosphere. This work will undoubtedly benefit from the revelation of large structural polymorphisms in addition to individual SNPs as has been described above.

Acknowledgements This work was supported, in part, by the BioEnergy Science Center and by the Office of Science, Biological and Environmental Research (BER), as part of the Plant Microbe Interfaces Scientific Focus Area at Oak Ridge National Laboratory. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Dept. of Energy under contract DE-AC05-00OR22725. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding received for this study.

References

Abraham P, Adams R, Giannone R, Kalluri U, Ranjan P, Erickson B, Shah M, Tuskan G, Hettich R (2012) Defining the boundaries and characterizing the landscape of functional genome expression in vascular tissues of *Populus* using shotgun proteomics. J Proteome Res 11(1):449–460

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S et al (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet 43:956–963

Geraldes A, DiFazio S, Slavov GT, Priya R, Muchero W et al (2013) A 34K SNP genotyping array for *Populus trichocarpa* (black cottonwood): design, application to the study of natural populations and transferability to other *Populus*. Molecular Ecology Resources 13(2):306–323

Induri BR, Ellis DR, Slavov GT, Yin T, Zhang X, Muchero W, Tuskan GA, DiFazio SP (2012) Identification of Quantitative Trait Loci (QTL) and candidate genes for cadmium tolerance in *Populus*. Tree Physiol 32(5):626–638

Iwano M, Takayama S (2012) Self/non-self discrimination in angiosperm self-incompatibility. Curr Opin Plant Biol 15:78–83

Kereszt A, Kondorosi E (2011) Unlocking the door to invasion. Science 331:865–866
Labbé J, Jorge V, Kohler A, Vion P, Marçais B, Bastien C, Tuskan GA, Martin F, Le Tacon F (2011) Identification of quantitative trait loci affecting ectomycorrhizal symbiosis in an interspecific F1 poplar cross and differential expression of genes in ectomycorrhizas of the two parents: *Populus deltoides* and *Populus trichocarpa*. Tree Genet Genom 7:617–627
Muchero W, Sewell MM, Ranjan P, Gunter LE, Tschaplinski TJ, Yin T, Tuskan GA (2013) Genome anchored QTLs for biomass productivity in Hybrid *Populus* grown under contrasting environments. PLoS ONE 8(1):e54468
Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, Brunner AM, DiFazio SP (2012) Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. Genome Res 22:95–105
Schwessinger B, Ronald PC (2012) Plant innate immunity: perception of conserved microbial signatures. Annu Rev Plant Biol 63:451–482
Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W et al (2012) Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. New Phytologist 196(3):713–725. doi:10.1111/j.1469-8137.2012.04258.x
Tagu D, Rampant PF, Lapeyrie F, Frey-Klett P, Vion P, Villar M (2001) Variation in the ability to form ectomycorrhizas in the F1 of an interspecific poplar (*Populus* spp.) cross. Mycorrhiza 10:237–240
Tuskan GA (1998) Short-rotation forestry: what we know and what we need to know. Biomass Bioenergy 14(4):307–315
Tuskan GA, Walsh M (2001) Short-rotation woody crop systems, atmospheric carbon dioxide and carbon management. For Chron 77:259–264
Tuskan GA et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr & Gray). Science 313:1596–1604
Yang X, Kalluri UC, DiFazio SP, Wullschleger SD, Tschaplinski TJ, Cheng Z-M, Tuskan GA (2009) Poplar genomics: state of the science. Crit Rev Plant Sci 28:285–308