Methods

Microsatellite primer resource for *Populus* developed from the mapped sequence scaffolds of the Nisqually-1 genome

T. M. Yin¹, X. Y. Zhang¹, L. E. Gunter¹, S. X. Li², S. D. Wullschleger¹, M. R. Huang² and G. A. Tuskan¹

¹Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6422, USA; ²The Key Lab of Forest Genetics and Gene Engineering, Nanjing Forestry University, Nanjing, China

Summary

Author for correspondence: Gerald A. Tuskan Tel: +1 (865) 576-8141 Fax: +1 (865) 576-9939 Email: gtk@ornl.gov

Received: 22 July 2008 Accepted: 9 September 2008

New Phytologist (2009) **181**: 498–503 **doi**: 10.1111/j.1469-8137.2008.02663.x

Key words: allelic variability, amplification rate, microsatellites, *Populus*, simple sequence repeat (SSR) primers.

• In this study, 148 428 simple sequence repeat (SSR) primer pairs were designed from the unambiguously mapped sequence scaffolds of the Nisqually-1 genome. The physical position of the priming sites were identified along each of the 19 *Populus* chromosomes, and it was specified whether the priming sequences belong to intronic, intergenic, exonic or UTR regions.

• A subset of 150 SSR loci were amplified and a high amplification success rate (72%) was obtained in *P. tremuloides*, which belongs to a divergent subgenus of *Populus* relative to Nisqually-1. PCR reactions showed that the amplification success rate of exonic primer pairs was much higher than that of the intronic/intergenic primer pairs.

• Applying ANOVA and regression analyses to the flanking sequences of microsatellites, the repeat lengths, the GC contents of the repeats, the repeat motif numbers, the repeat motif length and the base composition of the repeat motif, it was determined that only the base composition of the repeat motif and the repeat motif length significantly affect the microsatellite variability in *P. tremuloides* samples.

• The SSR primer resource developed in this study provides a database for selecting highly transferable SSR markers with known physical position in the *Populus* genome and provides a comprehensive genetic tool to extend the genome sequence of Nisqually-1 to genetic studies in different *Populus* species.

Introduction

The genus *Populus* possesses many characteristics that are conducive to functional genomic studies and as such it has been widely accepted as a model system in tree genomic research (Wullschleger *et al.*, 2002). Under the efforts of numerous scientists worldwide, the genome of a black cottonwood (*Populus trichocarpa* Torr. & Gray ex Brayshaw), clone 383–2499, 'Nisqually-1', has been sequenced and publicly released (Tuskan *et al.*, 2006). It is the first sequence of a woody perennial plant. However, the applicability of the Nisqually-1 genome sequence to studies of alternate *Populus* genotypes and species remains undetermined.

Microsatellites or simple sequence repeats (SSRs) have been shown to be among the most powerful genetic markers for aligning the genome of different species (Yin *et al.*, 2004), genetic fingerprinting (Schlotterer, 2001), linkage analysis (Dib *et al.*, 1996), population genetics (Wyman *et al.*, 2003) and clonal fidelity (Rajora & Rahman, 2004). Earlier studies suggest that SSRs are potentially transferable across genera of *Salicaceae* (Tuskan *et al.*, 2004; Hanley *et al.*, 2006). Moreover, the *Populus* genome project revealed that the chromosomal structure in modern *Populus* arose from an ancient wholegenome duplication event known as 'salicoid' duplication (Tuskan *et al.*, 2006) and our recent comparative mapping study demonstrated that genomes of alternate *Populus* species maintained the basic genome structure after salicoid duplication (Yin *et al.*, 2008). Therefore, it may be feasible to use SSRs to build a platform to study all *Populus* taxa as a macrogenetic system and to validate genetic findings across different *Populus* species.

To date, SSR primers for *Populus* have been designed from sequences that were randomly selected based on either library enrichment or shotgun sequencing strategies from various Populus species (Tuskan et al., 2004). Many of these primers' sequences show low to no homology to the genome sequence of Nisqually-1 and thus no reliable physical position can be deduced for these loci, impairing their utility and application. As a resource for the international Populus community, we developed primers that amplified microsatellites consisting of repetitive motifs of 2-5 bp from the unambiguously mapped sequence scaffolds of the Nisqually-1 genome. Our primary objective was to create a publicly available comprehensive genetic resource for Populus; our secondary objective was to test the utility and allelic variability of these SSR loci in P. tremuloides (Michx.), a member of a divergent subgenus within Populus.

Materials and Methods

Sputnik program coding with C language (C. Abajian, University of Washington, USA) was used to search DNA sequence files in Fasta format for microsatellite repeats. SSR primers were subsequently designed by Primer 3 (http:// frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). SSR repeat and primer distributions, which are classified into intronic, intergenic, exonic, or UTR regions, were derived from a Fortran coding program created by the authors based on the comparison of the physical locations of SSRs and genes annotated at the US Department of Energy's Joint Genome Institute homepage (http://genome.jgi-psf.org/Poptr1_1/ Poptr1_1.home.html). SSR density was calculated as the number of SSR loci within a moving 2 Mb window divided by the effective A, T, G, C reads within the same region.

To validate SSR interspecific utility, 100 SSR primer pairs in intronic/intergenic genomic space and 50 in exonic genomic space were selected from across the Nisqually-1 genome to amplify *P. tremuloides* template DNA in order to evaluate amplification success rate and allelic variability. Initially, a single P. tremuloides genotype was used to measure amplification success rate, and then 10 alternate P. tremuloides genotypes were used to test allelic variability. PCR reactions were performed as described by Tuskan et al. (2004) and Yin et al. (2004). The electrophoresis conditions were controlled by the default module for microsatellite genotyping on the ABI 3730xl (Applied Biosystems, Foster City, CA, USA). Regression and ANOVA analyses were performed to test the influence of different factors on the allelic variability. An Anderson-Darling test was used to test for normality of SSR distribution at significance levels $P \le 0.05$ and $P \le 0.01$.

In order to compare the amplification success rates across members of different subgenus in *Salicaceae*, we recorded PCR amplification success from 100 randomly selected SSR primer pairs for two *P. trichocarpa*, *P. deltoides* and *P. fremontii* genotypes and a single genotype for all other species within the genus.

Results

A total of 148 428 SSR primer pairs were designed from the unambiguously mapped sequence scaffolds of the Nisqually-1 genome. The complete information for all SSR primers is listed in the Supporting Information, Table S1. The principal metrics in this table include a description of the SSR sequence, the physical position of SSR, the melting temperature (T_m) of each primer, the GC content of the primer sequences, and the expected PCR product sizes in Nisqually-1. The visual representation of each SSR primer position per chromosome is shown in Fig. S1.

An Anderson-Darling test for normality across the whole genome shows that the SSR numbers (mean = 524, SD = 55, adjusted $A^2 = 0.445$) in each 2 Mb window were normally distributed at $\alpha = 0.05$ level (critical $A^2 = 0.752$). There were, however, four windows that had SSRs that exceeded the expected number and three windows that had fewer than expected numbers of SSRs (Table S2). Overall these results indicate that there are no large physical gaps among the SSR priming sites within the genome. Thus this primer resource will facilitate the generation of evenly distributed SSR markers across the *Populus* genome.

On average, SSRs occurred approximately every 2.5 kb within the Populus genome. At the subchromosomal level, SSR location varied by genic region, with 85.4% found in intergenic regions, 10.7% in introns, 2.7% in exons and 1.2% in UTR. Interestingly, the frequency of SSRs within exons varied by chromosome, from zero SSRs in exons on chromosome V to 765 on chromosome I, averaging 316 SSRs within exons per chromosome across the genome (Fig. 1). Based on the even distribution assumption, chromosome V would be expected to contain 242 microsatellite repeats in exonic regions. Furthermore, chromosome V shares large duplicated segments with chromosomes II, III and VII (Tuskan et al., 2006) and exons in paralogous genes found on these homologous chromosomes do contain SSRs. Thus, for undetermined reasons, it appears that the loss of exonic SSRs is unique to chromosome V and occurred after the salicoid duplication event.

Single sequence repeat amplification success rate across different *Populus* and *Salix* species showed that SSR primer amplification rates were higher among taxa closely related to *P. trichocarpa*, including species of *Leucoides*, *Aigeiros* and *Tacamahaca* subgenera, and lower in *P. tremuloides* (a member of *Leuce*), and lowest among members of *Turanga* section (Fig. 2).

Among the 150 SSR primer pairs tested in a single *P. tremuloides* genotype, 103 produced amplified product (Table S3). The amplification test confirmed that 63% of SSRs with priming sites in intronic/intergenic regions were successfully amplified; by contrast, amplification success rate was 80% for primers designed from exonic sequences (Table S3). We then tested 57 primer pairs, including 25 with priming sites in exonic or UTR sequences and 32 with priming sites in



Fig. 1 The distribution of simple sequence repeats (SSRs) in transcribed regions per chromosome (labeled as linkage groups, LG) in the *Populus* genome. Note: no SSRs were found in transcribed regions on chromosome V. Based on the even distribution assumption, chromosome V would be expected to contain 242 microsatellite repeats. The first number underneath the linkage group designation is the number of SSRs located in transcribed regions on each chromosome followed by the percentage of all SSRs located on each chromosome.

intronic/intergenic sequences, among 10 randomly selected *P. tremuloides* genotypes to determine allelic variability among primer pairs. The average allele number revealed by the exonic SSR loci was higher than that obtained per intronic/intergenic SSR loci (4.25 vs 3.25 alleles); however, ANOVA analysis indicated that this difference was not significant (P > 0.05). Therefore, it appears that the location of the priming sites significantly influences amplification rate but not allelic variability.

Based on 100 selected primer pairs, allelic variability across members of the genus *Populus* did not significantly vary with SSR repeat length, GC content of the repeats or repeat motif numbers (F = 0.23, F = 0.94 and F = 0.502, respectively, critical F = 3.84 at $P \le 0.05$). However, when we analyzed the allelic variability by repeat motif length, a significant negative correlation was detected such that the average allele number decreased from 4.29, 2.91 and 2.00 in di-, tri- and tetranucleotide repeats, respectively. Therefore, of the tested parameters, only repeat motif length significantly affected the SSR allelic variability among members of the genus.

We also compared the variability of SSR with repeat motifs of [AAT]/[TTA], [AC]/[TG], [AT]/[TA] and [AG]/[TC]. Among these repeat motifs, the [AG]/[TC] motif results in the highest polymorphism rate; [AAT]/[TTA] yields the lowest. Significant differences in allelic variability were detected among $[AAT]_n/[TTA]_n vs [AG]_n/[TC]_n (\alpha \le 0.01) and [AC]_n/[TG]_n vs [AG]_n/[TC]_n (\alpha \le 0.05)$. Thus, the base composition of the SSR repeat motifs significantly affected allelic variability among the SSR primer pairs tested in this study.

Discussion

Our study demonstrates that the microsatellite markers derived from a single clone of *Populus trichocarpa*, Nisqually-1, have relatively high amplification rate in *P. tremuloides*, a member of the *Leuce* subgenera. The genus *Populus* contains six subgenera, including *Abaso*, *Leuce*, *Leucoides*, *Aigeiros*, *Turanga* and *Tacamahaca* (Eckenwalder, 1996; Shi *et al.*, 2001), of which the subgenera *Leuce* is more dissimilar to Nisqually-1 (a member of *Tacamahaca*), than are members of any other subgenus. The amplification success rates were higher in members of all other subgenera, expect for *P. euphratica*, which is the sole representative of the *Turanga* subgenera (Fig. 2). However, the phylogenetic position of *Turanga* is controversial. According to Eckenwalder's consensus cladogram (1996), *Turanga* is the most distant from *Tacamahaca* among all the





Fig. 2 Amplification rates of simple sequence repeat (SSR) primers across different *Populus* and *Salix* species. These estimates are based on 100 randomly chosen microsatellite primers. The x-axis indicates the species and the y-axis indicates the amplification rate. Red bars, species that belong to section *Tacamahaca*; purple bars, species that belong to section *Aigeiros*; blue bars, species that belong to section *Leucoides*; green bars, species that belong to section *Leuce*; light blue bar, species that belong to section *Turanga*; and pink bars, *Salix* species. The appearance of subgenus from left to right is ordered according to the increment of their phylogenetic divergence from *Tacamahaca*. For primers and origins of these samples, refer to Tuskan *et al.* (2004) and Yin *et al.* (2008).

subgenera of *Populus*. Nonetheless, we achieved moderate amplification success rate in *P. euphratica* in the *Turanga* subgenus and in members of *Salix*. In general, the amplification success rates in different species are positively correlated with their phylogenetic divergence from Nisqually-1.

It should be noted that the tested primers were randomly selected from an overall list of potential primer sequences in the Nisqually-1 genome. Data in Fig. 2 indicate that the amplification success rate was less than 100% for *P. trichocarpa*. These results suggest that the 72% transferability estimated for *P. tremuloides* and all other tested species is probably an underestimation. Our results indicate that the amplification rate in *P. trichocarpa* is high (*c.* 99%), and therefore the underestimation should be minor.

It is universally recognized that coding sequences are better conserved than noncoding sequences. In this study, we verified that amplification success rates of SSR primers were dramatically influenced by their priming site position (i.e. exon vs intron). Our experimental data demonstrated that primers located in exonic regions of Nisqually-1 have significantly higher amplification rates in a genetically divergent *Populus* species than primers designed from Nisqually-1 noncoding space. Therefore, the exonic primers would be especially useful in supplying a common language for the *Populus* community to communicate and validate findings among different *Populus* studies. In contrast to the amplification rate, the subchromosomal locations of the priming sites did not significantly affect the allelic variability of microsatellites. Evidence of the influence of flanking sequences on microsatellite variability from any organism is limited and inconclusive. Glenn *et al.* (1996) detected significant influence of the flanking sequence in alligator; however, no such effects were detected in studies by Balloux *et al.* (1998) on shrews or by Bachtrog *et al.* (2000) on *Drosophila.* In the only comparative SSR study in plants, Gao & Xu (2008) found that mutation rates of microsatellites did not significantly differ among motifs of di-, tri- and tetra-nucleotide repeats in four subspecies of cultivated rice *O. sativa* and its three relatives, *O. rufipogon, O. glaberrima* and *O. officinalis.*

Despite the wide occurrence of microsatellites, the basis of their variability is still not well understood. In this study, we found that the SSR variability in *P. tremuloides* was significantly influenced by the repeat motif length and the base compositions of the repeat motif, but not by the number of repeat units. Consistent with our findings, Bachtrog *et al.* (2000) reported that the base composition of repeat motif significantly influenced microsatellite mutation rates. However, whereas the repeat number has been observed to be positively correlated with microsatellite variability in a variety of organisms (Jin et al., 1996; Wierdl et al., 1997; Schlotterer et al., 1998; Schug et al., 1998), this trend was not apparent in our study, and it is reasonable to speculate that the repeat number of SSRs is not conserved among highly diverged species. Although we might expect SSRs with longer repeat lengths to reveal higher polymorphism among genotypes, based on our study, this expectation is not warranted. In support of this, the findings among different studies for the influence of the repeat motif length on microsatellite variability are not consistent. There is a trend for a higher mutation rate for SSRs with dinucleotide repeat motifs than SSRs with longer repeat motifs (Chakraborty et al., 1997; Kruglyak et al., 1998; Schug et al., 1998), which is consistent with our results. By contrast, Weber & Wong (1993) observed higher mutation rates for tetranucleotide repeats than for dinucleotide repeats in humans.

The SSR markers developed in this study provide a comprehensive genetic resource that can be used to link findings from other *Populus* studies to the sequenced genome. These primers also represent a valuable resource for the selection of genetic markers for studying population structure, genetic vs geographic variation, and sequence-dependent evolution in *Populus*. The prospective markers can also be used to explore the distribution of recombination hotspots in the *Populus* genome and promote sharing of associated data, such as QTL position validation across unrelated pedigrees. The reported SSRs are especially useful for generating lists of candidate genes occurring in QTL intervals. Moreover, the SSR primers developed in this study can be used to selectively target regions of the whole genome for efficiently closing gaps in the genetic map.

Although the parameters we supplied cannot measure SSR polymorphism per se, our survey of the influence of different factors on microsatellite variability provides a reference for selecting SSR loci that potentially yield greater allelic variability. Our study confirmed that SSRs with priming sites in exons had greater utility across species than those with priming sites in introns and intergenic regions. Without such a resource, researchers would have to randomly test several hundred primer pairs to obtain a marker in a target region. This study provides practical information for selecting SSR primers and can thus reduce the time and cost associated with the development of highly transferable, highly polymorphic markers for Populus that can be applied to genetic mapping efforts, allelic variation discovery studies, and molecular breeding efforts related to fiber and energy production, carbon sequestration and bioremediation.

Acknowledgements

We thank M. Schuster at the University of Tennessee for establishing the web resources, Dr J. Armento in Oak Ridge for his comments and editing for this manuscript, and R. M. Tuskan for perspectives on objectives and procedures. Special thanks go to the editor and anonymous reviewers for their help in formulating the final revision. Funding for this research was provided by the Educational Department of China (NCET-04-0516), the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL) and the US Department of Energy, Office of Science, Biological and Environmental Research Carbon Sequestration Program and Bioenergy Science Center. ORNL is managed by UT-Battelle, LLC for the US Department of Energy under contract no. DE-AC05-00OR22725.

References

- Bachtrog D, Agis M, Imhof M, Schlotterer C. 2000. Microsatellite variability differs between dinucleotide repeat motifs: evidence from Drosophila melanogaster. Molecular Biology and Evolution 17: 1277–1285.
- Balloux F, Ecoffey E, Fumagalli L, Goudet J, Wyttenbach A, Hausser J. 1998. Microsatellite conservation, polymorphism, and GC content in shrews of the genus *Sorex* (*Insectivora, Mammalia*). *Molecular Biology and Evolution* 15: 473–475.
- Chakraborty R, Kimmel M, Stivers D, Davison L, Deka R. 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences, USA* 94: 1041–1046.
- Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Kazan J, Seboun E *et al.* 1996. A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* 380: 152–154.
- Eckenwalder JE. 1996. Systematics and evolution of *Populus*. In: Stettler RF, Bradshaw HD, Heilman PE, Hinckley TM, eds. *Biology of* Populus *and its implications for management and conservation*. Ottawa, Canada: NRC Research Press, 7–32.
- Gao LZ, Xu HY. 2008. Comparisons of mutation rate variation at genome-wide microsatellites: evolutionary insights from two cultivated rice and their wild relatives. *BMC Evolutionary Biology* 8: 11. doi:10.1186/ 1471-2148-8-11.
- Glenn TC, Stephan W, Dessauer HC, Braun MJ. 1996. Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. *Molecular Biology and Evolution* 13: 1151–1154.
- Hanley SJ, Mallott MD, Karp A. 2006. Alignment of a *Salix* linkage map to the *Populus* genomic sequence reveals macrosynteny between willow and *Populus* genomes. *Tree Genetics and Genomes* **3**: 35–48.
- Jin L, Macaubas C, Hallmayer J, Kimura A, Mignot E. 1996. Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proceedings of the National Academy of Sciences, USA* 93: 15285–15288.
- Kruglyak S, Durrett RT, Schug MD, Aquadro CF. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences, USA* 95: 10774–10778.
- Rajora O, Rahman M. 2004. Microsatellite DNA and RAPD fingerprinting, identification and genetic relationships of hybrid poplar (*Populus* × *canadensis*) cultivars. *Theoretical and Applied Genetics* **106**: 470–477.
- Schlotterer C. 2001. Genealogical inference of closely related species based on microsatellites. *Genetic Research* 78: 209–212.
- Schlotterer C, Ritter R, Harr B, Brem G. 1998. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution* 15: 1269–1274.
- Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF. 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution* 15: 1751–1760.

- Shi QL, Zhuge Q, Huang MR, Wang MX. 2001. Phylogenetic relationship of *Populus* sections by ITS sequence analysis. *Acta Botanica Sinica* 43: 323–325.
- Tuskan GA, DiFazio SP, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray ex Brayshaw). *Science* 313: 1596–1604.
- Tuskan GA, Gunter LE, Yang ZM, Yin TM, Sewell MM, DiFazio SP. 2004. Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa. Canadian Journal of Forest Research* 34: 5–93.
- Weber JL, Wong C. 1993. Mutation of human short tandem repeats. *Human Molecular Genetics* 2: 1123–1128.
- Wierdl M, Dominska M, Petes TD. 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* 146: 769–779.
- Wullschleger SD, Jansson S, Taylor G. 2002. Genomics and forest biology: *Populus* emerges as the perennial favorite. *Plant Cell* 14: 2651–2655.
- Wyman J, Bruneau A, Tremblay MF. 2003. Microsatellite analysis of genetic diversity in four populations of *Populus tremuloides* in Quebec. *Canadian Journal of Botany* 81: 367.
- Yin T, DiFazio SP, Gunter LE, Riemenschneider D, Tuskan GA. 2004. Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theoretical and Applied Genetics* 109: 451–463.
- Yin TM, DiFazio SP, Gunter LE, Zhang XY, Swell MM, Woolbright S, Allan GJ, Kelleher CT, Douglas CJ, Tuskan GA. 2008. Genome structure and primitive sex chromosome revealed in *Populus. Genome Research* 18: 422–430.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Distribution of microsatellites, including the SSR primers along each *Populus* chromosome.

Table S1 Database of SSR primer sequences and parameters

 developed from the mapped sequence scaffolds on each *Populus*

 chromosome

Table S2 The distribution test of SSR densities in a 2 Mbsliding window along each *Populus* chromosome

Table S3 PCR amplification results and primer pair information used in testing SSR amplification success rate and SSR allelic variability in *P. tremuloides*

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



About New Phytologist

- New Phytologist is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at **www.newphytologist.org**.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *Early View* – our average submission to decision time is just 29 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £139 in Europe/\$259 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office (newphytol@lancaster.ac.uk; tel +44 1524 594691) or, for a local contact in North America, the US Office (newphytol@ornl.gov; tel +1 865 576 5261).