



*Yan Su, Dehuo Hu, Huiquan Zheng**

Detection of SNPs based on DNA specific-locus amplified fragment sequencing in Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook)

Received: 16 February 2016; Accepted: 20 June 2016

Abstract: Compared to angiosperms, conifers represent more complex genomes with larger giga-genome size. To detect large-scale single nucleotide polymorphisms (SNPs), whole genome sequencing of a conifer population is still unaffordable. In this work, we report the use of DNA specific-locus amplified fragment sequencing (SLAF-seq) for large-scale SNP detection in Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook), an ecological and economic important conifer in China. SLAF libraries of 18 parent clones of a Chinese fir 2.5 generation seed orchard were sequenced and a total of 117,924 SLAFs were developed. We detected 147,376 SNPs from these SLAFs; 146,231 of them represented simple nucleotide change in A/G, C/T, A/C, A/T, C/G or G/T. The most frequent SNPs occurred in C/T (34.3%), while the majority of SNPs (68.2%) belonged to transition events (A/G and C/T). Notably, all the sequenced samples had high portion (78.2–80.9%) of common SNPs indicating that the Chinese fir genomes tended to change its nucleotides at common loci. 48,406 informative SNPs were then successfully utilized to genotype the tested samples ($n = 18$) followed by a phylogenetic tree to clarify their genetic relationship. Furthermore, a set of very high linkage disequilibrium (0.51–1.00) were identified from these informative SNPs. In brief, our work demonstrated that SLAF-seq is an alternative and cost-effectively high-throughput approach for large-scale SNP exploitation in Chinese fir. While the obtained SNPs offer useful marker resource for further genetic and genomic studies and will be helpful for Chinese fir breeding programs.

Keywords: Conifer, SLAF, SNP, genotyping and linkage disequilibrium

Addresses: Y. Su, College of Forestry and Landscape Architecture, South China Agricultural University, 510642 Guangzhou, People's Republic of China

Y. Su, D. Hu, H. Zheng, Guangdong Provincial Key Laboratory of Bio-control for the Forest Disease and Pest, Guangdong Academy of Forestry, 510520 Guangzhou, People's Republic of China, e-mail: zhenghq@sinogaf.cn

*Corresponding author

Introduction

Single nucleotide polymorphism (SNP) is the most abundant form of variation within a species at DNA level (Mammadov et al., 2012). It has been ex-

tensively used for plant genetic and genomic studies and the breeding programs. Owing to rapid development of high-throughput sequencing technologies (e.g. next generation sequencing, NGS) and sophisticated computational methods, genome-wide mining

of SNPs currently seems to be more accessible than ever before, especially for the angiosperms with simple genomes that could be decoded by whole genome sequencing (WGS) (Mammadov et al., 2012; Kumar et al., 2012; He et al., 2014). In contrast to angiosperms, conifers represent more complex genomes that have genome sizes ranging from 18 Gbp to over 35 Gbp and accumulate very long introns and terminal repeat retrotransposons (Murray et al., 2012; Prunier et al., 2016). With NGS, several ecological and economic important conifers, such as Norway spruce (*Picea abies* (L.) Karst.) (Nystedt et al., 2013), white spruce (*P. glauca* (Moench) Voss) (Birol et al., 2013) and loblolly pine (*Pinus taeda* L.) (Neale et al., 2014), have been sequenced for their giga-genomes. However, whole genome sequencing of a conifer population is still unaffordable because of its dramatic cost and analytical effort (Du et al., 2012; De La Torre et al., 2014). Thus, alternative strategies for large-scale SNP detection in conifer are highly required.

Restriction site associated DNA sequencing (RAD-seq) represents a feasible and flexible strategy for population SNP detection in plants (Miller et al., 2007; Du et al., 2012; Sun et al., 2015). Quite simply, the genome's complexity could be reduced by a represented reduced representation library (RRL) that enriched numbers of RAD loci (or restriction enzyme-based specific genomic fragments); in-parallel RRL sequencing of sample DNAs allowed genome-wide SNP detection for researchers. Recently, an enhanced version of RAD-seq called as specific-locus amplified fragment sequencing (SLAF-seq) was suggested by Sun et al. (2013). This version has several attractive features: (1) no requirement of the reference genome sequences, (2) pre-designed RRL schemes avoiding large set of repetitive sequences, (3) double barcode system couple with NGS technique amenable for high-throughput and accurate SNP detection in large populations, and (4) cost-effective for researchers. Application of SLAF-seq for *de novo* SNP discovery, genotyping and mapping has been demonstrated in a variety of plant species including rice (*Oryza Sativa* L.) (Xu et al., 2015), soybean (*Glycine max* (L.) Merr.) (Han et al., 2015), mei (*Prunus mume* Sieb. et Zucc) (Zhang et al., 2015), tree peony (*Paeonia* Sect. *Moutan*) (Cai et al., 2015), and tea plant (*Camellia sinensis*) (Ma et al., 2015).

Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook, $2n = 2x = 22$ chromosomes) is the most important conifer in southern China due to its great afforestation values in timbers and ecological contributions (Zheng et al., 2015). Its estimated genome size was approximately 26.9 Gbp (Murray et al., 2012), close to $200\times$, $53\times$ and $12\times$ the genomes of *Arabidopsis*, *Populus* and *Zea mays* L., respectively. Similar to some advanced tree breeding programs, improvement of Chinese fir recently tends to utilize DNA marker to

decipher genetic polymorphisms aiming to improve the accuracy of selection and to shorten the breeding cycles. The applied marker systems included RAPD (random amplified polymorphic DNA), AFLP (amplified fragment length polymorphism), ISSR (inter simple sequence repeat), SSR (simple sequence repeat) and SRAP (sequence-related amplified polymorphism) (Zheng et al., 2015), and they were consistent PCR-based and medium-throughput methods (Mammadov et al., 2012). To capture more variation, high-throughput genome-wide SNP detection in Chinese fir is required. In this work, we aim to: (1) test the efficacy and efficiency of SLAF-seq for SNP detection in Chinese fir, (2) utilize the high-throughput SNP genotyping to elucidate the genetic distance and relationship of 18 Chinese fir parent clones in a commercial 2.5 generation seed orchard, and (3) unravel notable SNP linkage disequilibrium in this species.

Materials and methods

The study was carried out with 18 parent clones (SP01–SP18) of a Chinese fir 2.5 generation seed orchard in Lechang, Guangdong, China. The parents included seven from Guangdong (hybrids from local breeding program: SP02, SP04, SP05, SP08, SP15, SP16 and SP18), seven from Guangxi (SP01, SP03, SP07, SP09, SP10, SP11, SP14), two from Fujian (SP12 and SP13), one from Hunan (hybrid: SP06) and one from Guizhou (SP17). These clones represented different genotypes that had no pedigree relationship with each other, except for SP01 with SP09 and SP02 with SP14. They were selected for both gain and diversity in a Second-Cycle Chinese fir Breeding Program of Guangdong with following criteria: (1) the offspring should have a positive realized gain ($\geq 5\%$) in height, diameter and stem volume compared to that of the first generation seed orchard, (2) medium to high numbers of male and female strobili, (3) flowering synchronization of male and female strobili, and (4) different origin as far as possible.

Total genomic DNA was extracted from each clone leaves with a DNAsure Plant Kit (TIANGEN, Beijing, China), and evaluated by agarose gel electrophoresis and spectrophotometric analysis. The qualified DNA was then subjected to the specific-locus amplified fragment sequencing (SLAF-seq) as described by Sun et al. (2013) with some modifications: (1) restriction enzyme and sizes prediction using *P. abies* genomic sequence (<http://congenie.org/>) as training data, and this was because *P. abies* represented a high-coverage draft genome assembly that could be used as a reference in conifer, (2) SLAF library construction targeted to *EcoR* V (5'-GAT/ATC-3') specific loci, (3) pair-end sequencing (2×80 bp)

of the 314–344 bp SLAF fragments using Illumina High-seq™ 2500 system (Illumina, San Diego, CA, USA), and (4) read depth $\geq 10\times$.

SLAF-seq reads were grouped and genotyped using the methods as described by Sun et al. (2013). In brief, pair-end reads with clear index information and over 90% identity were grouped together and defined as a SLAF. Single nucleotide polymorphisms were identified by population ($n = 18$) SLAF aligning using BLAST-like alignment tool (BLAT) (Kent 2002). The BLAT was also used to analyze the sequence similarity of the SLAFs to current known Chinese fir mRNA/genomic DNA sequence in NCBI (<http://www.ncbi.nlm.nih.gov/>). SNP-based genetic distance analyses were performed with a Maximum Composite Likelihood model (Tamura et al., 2004). All ambiguous positions were removed for each sequence pair. There were a total of 48,406 positions (SNPs) in the final dataset. Phylogenetic tree construction was conducted in MEGA5 (Tamura et al., 2011). Linkage disequilibrium analyses between SNPs were carried out using Haploview 4.2 (Barrett et al., 2005) with an r^2 cutoff of 0.10.

Results and discussion

To improve SLAF-seq efficiency, we utilized conifer *P. abies* genomic sequence as a training data to predict appropriate restriction enzyme suitable for RRL construction in Chinese fir. During this process three criteria were considered (Sun et al., 2013): (1) the number of SLAFs must be more than 50,000 which could satisfy basis genetic analysis needs, (2) genome-wide distribution of SLAFs, and (3) avoiding repeated sequences as much as possible. Briefly, this pre-design experiment *in silico* simulated a series

of digestion assays, providing a reference number and size of the restriction fragments generated from different endonucleases, and therefore being useful for endonuclease selection regarding the above criteria. According to *in silico* analysis result *EcoR V* was thought to be suitable for genome digestion during SLAF preparation in Chinese fir. The tested DNA samples ($n = 18$; SP01–SP18) were then subjected to the *EcoR V*-specific SLAF-seq assays. In parallel sequencing indicated that at least 23.76M high-quality reads (each read approximately 80-bp in length; average read depth $13.85\times$) could be obtained from the present 18-plex libraries with a high-quality bases ratio (Q30) ranging from 92.6 to 95.0% (mean = 94.1%) (Table 1). The GC (guanine-cytosine) content of the reads was about 37.0%, similar to that prevailing in Norway spruce (37.9%) (Nystedt et al., 2013) and loblolly pine (38.2%) (Neale et al., 2014), and seemed to be higher than that (35.0%) for chloroplast (Zhu et al., 2016). Based on these genome-wide reads, a total of 117,924 SLAFs were developed; 49.6% of them (58,437) present genetic variations (single nucleotide polymorphism (SNP) and insertion-deletions). Merely 2.3% of the SLAFs (2,733) harbored repeat sequences. In general, conifer genomes comprised large amount of repeat sequences (approximately 58.0%) (Prunier et al., 2016). In these repeat regions, the proportion of false discovery and false negative SNPs would be rather high (Du et al., 2012). Exclusion of repeat sequence as much as possible definitely promotes the SNP discovery accuracy in Chinese fir.

The SLAFs were then subjected to a BLAT query against known Chinese fir mRNA/genomic DNA sequence in NCBI (<http://www.ncbi.nlm.nih.gov/>) (up to Dec 31, 2015; supplementary table). 3,579 hits were observed, covering 2.4% of the SLAFs (2,879) corresponding to 604 mRNAs/unigenes, 6 microsatellite and 4 SRAP markers respectively. 381 SLAFs may be misleading as they had more than one target hits. Of the matching SLAFs, 1,282 harbored SNPs, representing the polymorphic SLAFs with annotations; notably, some of them targeted to the wood quality associated sequence, e.g. expansin-like protein mRNA, SRAP marker Me19/Em22-557, Me10/Em20-1629 and Me2/Em26-469. However, most of the SLAFs (97.6%) targeted to unknown genome regions and remained elusive.

Here, we concentrated on the single nucleotide variations and collectively, detected 147,376 SNPs in the tested samples ($n = 18$). Out Of these SNPs, 146,231 represented simple nucleotide change in A/G, C/T, A/C, A/T, C/G or G/T; 68.2% of these (99,694) were transitions (A/G and C/T), while the other (46,537) were transversions (A/C, A/T, C/G and G/T). The most frequent SNPs occurred in C/T (34.3%) and A/T change represented the least type

Table 1. Paired-end SLAF-seq statistics in this study

Feature	Value
Number of reads	23,760,000
Average Q30 percentage	94.1%
Average GC percentage	37.0%
Total number of SLAFs	117,924
Average depth	13.85
Number of polymorphic SLAFs	58,437
Number of no polymorphic SLAFs	56,754
Number of SLAFs with repeats	2,733
Number of SNPs (MAF>0.05)	147,376
Transition	
A/G	49,588 (33.9%)
C/T	50,106 (34.3%)
Transversion	
A/C	19,083 (13.0%)
A/T	6,351 (4.3%)
C/G	6,655 (4.6%)
G/T	14,448 (9.9%)

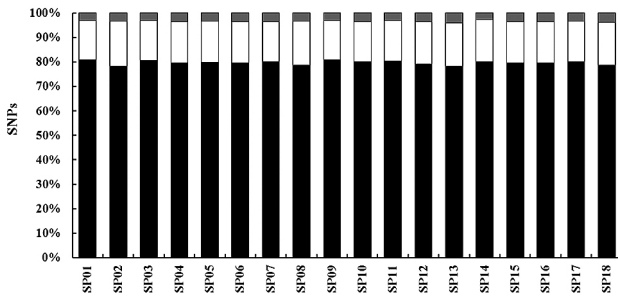
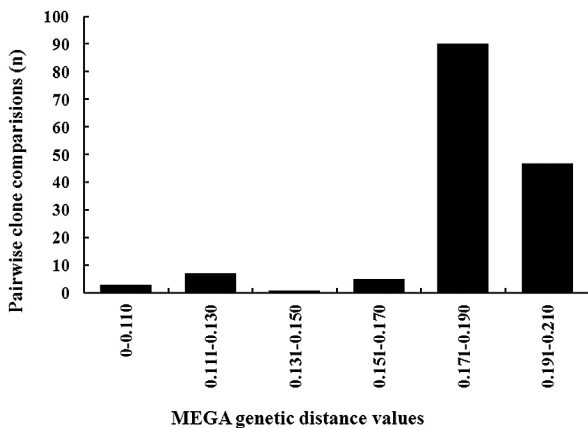


Fig. 1. Composition of SNP types in 18 Chinese fir parent clones. SNPs were grouped into three categories: rare SNPs found in less than 6 clones (in grey), medium frequency SNPs found in between 7–12 clones (in white) and common SNPs (in black) found in more than 12 clones

of variations (4.3%). Overall, the transition:transversion ratio (Ts/Tv) was 2.14, agreeing with common observations that nucleotide transitions are favored several-fold over transversions in molecular evolution (Stoltzfus & Norris, 2015). According to the population ($n = 18$) SNP frequencies, we grouped SNPs into three categories: rare SNPs (found in less than 6 clones), medium frequency SNPs (found in between 7–12 clones), and common SNPs (found in more than 12 clones). As shown in Fig. 1, all the clones have high portion (78.2–80.9%) of common SNPs indicating that the Chinese fir genomes tended to change its nucleotides at common loci; merely 2.6–4.1% and 16.0–18.5% of the SNPs were rare and medium frequency SNPs respectively. Kim and Tai (2013) also employed a RRL sequencing method (RESCAN) to simultaneously identify and genotype SNPs in rice, but observed a limited common SNPs ratio (approximately 4.5–35.0%) in cultivar ($n = 45$).

A set of informative SNPs ($n = 48,406$) were then used to estimate the genetic distance and relationship of the present 18 Chinese fir parent clone (Fig. 2). The results indicated that the genetic distance varied from 0.006 (SP01:SP09) to 0.210 (SP01:SP03) with an overall average distance of 0.180. 58.8% of the pairs of clones fell the genetic distance between 0.171 and 0.190, while only three comparisons (SP01:SP09, SP02:SP14, and SP02:SP08) present the values lower than 0.110. The extremely close kinship of SP01 with SP09 (0.006), and SP02 with SP14 (0.007) furthered our understanding of relationship of the clones in addition to their traditional pedigree relationship. It should be bared in mind that the range of the genetic distances of tested clones seemed to be rather narrow (0.006–0.210), and this was not expected because these clones were carefully selected not only for their gains but also their diversity in origins (different geographical genetic origins from southern China). Genome-wide SNP-based dendrogram analysis definitely provided an efficient and accuracy tool to address the genetic relationship of the germplasms and it will help us to select more divergent parents to improve the diversity of the seed orchard, as well as to better protect and manage them in the future. For conifer SNP genotyping, whole genome sequencing remained unattainable, but flexible methods such as population RNA-seq, EST re-sequencing approach and specific SNP array seemed to be rather feasible (Uchiyama et al., 2012; Howe et al., 2013; Liu et al., 2014; Plomion et al., 2014; Heer et al., 2016). In this study, we demonstrated that SLAF-seq is also an efficient method for the population SNP genotyping in conifer (Chinese fir).

A



B

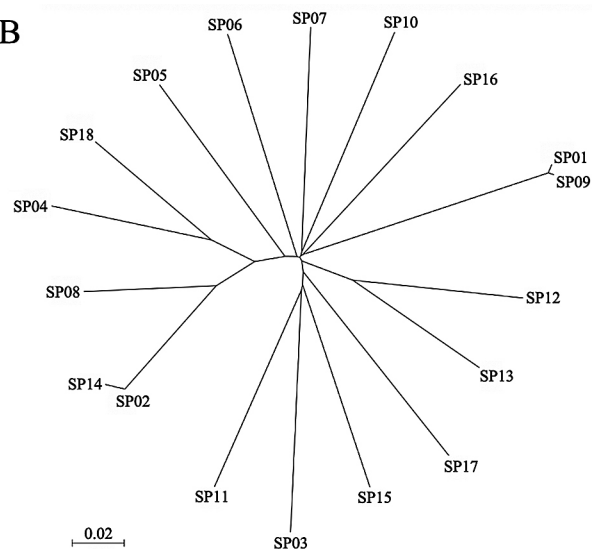


Fig. 2. Phylogenetic analysis of the 18 Chinese fir parent clones (SP01–SP18) based on 48,406 informative SNPs from SLAF-seq. (A) Distribution of pairwise MEGA genetic values for the parent clones; (B) Neighbor-joining tree showing genetic relationships among the parent clones based on MEGA genetic distance

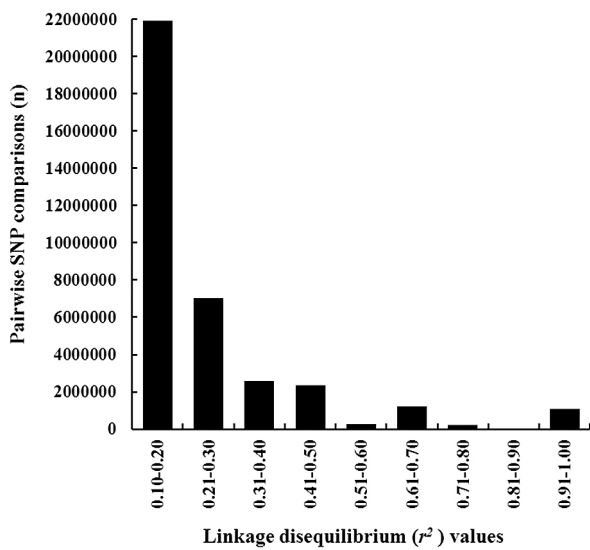


Fig. 3. Distribution of different level of linkage disequilibrium (r^2) calculated by pairwise comparison of 48,406 single nucleotide polymorphisms (SNPs) in Chinese fir. Only pairwise SNP comparisons with $r^2 \geq 0.10$ are included

In the next step, we aimed to track the possible SNP linkage disequilibrium (LD) in Chinese fir. A collection of 36,547,629 pairwise comparisons with relatively high LD ($r^2 \geq 0.10$) were found among the above mentioned 48,406 SNPs (Fig. 3). Majority of the LD estimates (92.6%) presented a r^2 lower than 0.50 (0.10–0.50); only 7.4% (2,703,496) displayed very high LD (0.51–1.00). In terms of other conifer mean r^2 values (gene-based SNP pairwise comparisons) ranging from 0.19 (white spruce) to 0.38 (Douglas-fir (*Pseudotsuga menziesii*)) (Pavy et al., 2012), 13.9% of the Chinese fir SNP LD showed substantial and above value (r^2 : 0.41–1.00). Although their genetic distances are still unknown, SNP loci with high LD may locate on the same chromosome (Liu et al., 2014).

Conclusions

The results obtained by SLAF-seq analysis of the collection ($n = 18$) of Chinese fir parent clones in this study suggested that (1) SLAF-seq is an alternative and cost-effectively high-throughput approach for large-scale SNP exploitation in conifer (Chinese fir), and (2) at least of 147,376 SNPs could be detected from the *EcoR* V-based DNA SLAF-seq, offering useful marker resource for further genetic and genomic studies (e.g. genotyping, genetic diversity analysis, population structure characterization and association study). Notably, the identified LD value allowed us to better understand the SNP LD patterns of Chinese fir, and will be helpful for further breeding programs.

Acknowledgments

This research was supported by the Special Plan for the Cultivation of High-level Talents of Guangdong (No. 2014TQ01N140), the Special Fund for Forest Scientific Research in the Public Welfare (No. 201404127) and the National Natural Sciences Foundation of China (No. 31200506). We thank Biomarker Technologies Co Ltd (Beijing, China) for assistance in sequencing.

References

- Barrett JC, Fry B, Maller J & Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.
- Biról I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MM, Keeling CI, Brand D, Vandervalk BP, Kirk H, Pandoh P, Moore RA, Zhao Y, Mungall AJ, Jaquish B, Yanchuk A, Ritland C, Boyle B, Bousquet J, Ritland K, MacKay J, Bohlmann J & Jones SJM (2013) Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 29: 1492–1497.
- Cai C, Cheng FY, Wu J, Zhong Y & Liu G (2015) The first high-density genetic map construction in tree peony (*Paeonia* Sect. *Moutan*) using genotyping by specific-locus amplified fragment sequencing. *PLoS One* 10: e0128584.
- De La Torre AR, Biról I, Bousquet J, Ingvarsson PK, Jansson S, Jones SJM, Keeling CI, MacKay J, Nilsson O, Ritland K, Street N, Yanchuk A, Zerbe P & Bohlmann J (2014) Insights into conifer giga-genomes. *Plant Physiology* 166: 1724–1732.
- Du Y, Jiang H, Chen Y, Li C, Zhao M, Wu J, Qiu Y, Li Q & Zhang X (2012) Comprehensive evaluation of SNP identification with the restriction enzyme-based reduced representation library (RRL) method. *BMC Genomics* 13: 77.
- Han Y, Zhao X, Cao G, Wang Y, Li Y, Liu D, Teng W, Zhang Z, Li D, Qiu L, Zheng H & Li W (2015) Genetic characteristics of soybean resistance to HG type 0 and HG type 1.2.3.5.7 of the cyst nematode analyzed by genome-wide association mapping. *BMC Genomics* 16: 598.
- He J, Zhao X, Laroche A, Lu ZX, Liu H & Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* 5: 484.
- Heer K, Ullrich KK, Liepelt S, Rensing SA, Zhou J, Ziegenhagen B & Lars Opgenoorth L (2016) Detection of SNPs based on transcriptome sequencing in Norway spruce (*Picea abies* (L.) Karst).

- Conservation Genetics Resources 8: 105–107. doi:10.1007/s12686-016-0520-4.
- Howe GT, Yu J, Knaus B, Cronn R, Kolpak S, Dolan P, Lorenz WW & Dean JFD (2013) A SNP resource for Douglas-fir: *de novo* transcriptome assembly and SNP detection and validation. *BMC Genomics* 14: 137.
- Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Research* 12(4): 656–664.
- Kim SI & Tai TH (2013) Identification of SNPs in closely related *temperate japonica* rice cultivars using restriction enzyme-phased sequencing. *PLoS One* 8: e60176.
- Kumar S, Banks TW & Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. *International Journal of Plant Genomics* 2012: 831460.
- Liu JJ, Snieszko RA, Sturrock RN & Chen H (2014) Western white pine SNP discovery and high-throughput genotyping for breeding and conservation applications. *BMC Plant Biology* 14: 380.
- Ma JQ, Huang L, Ma CL, Jin JQ, Li CF, Wang RK, Zheng HK, Yao MZ & Chen L (2015) Large-scale SNP discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). *PLoS One* 10: e0128798.
- Mammadov J, Aggarwal R, Buyyarapu R & Kumpatla S (2012) SNP markers and their impact on plant breeding. *International Journal of Plant Genomics* 2012: 728398.
- Miller MR, Dunham JP, Amores A, Cresko WA & Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240–248.
- Murray BG, Leitch IJ & Bennett MD (2012) Gymnosperm DNA C-values database (release 5.0, Dec 2012).
- Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D, Crepeau MW, Cardeno C, Koriabine M, Holtz-Morris AE, Liechty JD, Martinez-Garcia PJ, Vasquez-Gross HA, Lin BY, Zieve JJ, Dougherty WM, Fuentes-Soriano S, Wu LS, Gilbert D, Marcias G, Roberts M, Holt C, Yandell M, Davis JM, Smith KE, Dean JFD, Lorenz WW, Whetten RW, Sederoff R, Wheeler N, McGuire PE, Main D, Loopstra CA, Mockaitis K, de Jong PJ, Yorke JA, Salzberg SL & Langly CH (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15: R59.
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Källner M, Luthman J, Lysholm F, Nittylä T, Olson A, Rilakovic N, Ritland C, Rosselló JA, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J & Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 497: 579–584.
- Pavy N, Namroud MC, Gagnon F, Isabel N & Bousquet J (2012) The heterogeneous levels of linkage disequilibrium in white spruce genes and comparative analysis with other conifers. *Heredity (Edinb)* 108: 273–284.
- Plomion C, Chancerel E, Endelman J, Lamy JB, Mandrou E, Lesur I, Ehrenmann F, Isik F, Bink MCAM, van Heerwaarden J & Bouffier L (2014) Genome-wide distribution of genetic diversity and linkage disequilibrium in a mass-selected population of maritime pine. *BMC Genomics* 15: 171.
- Prunier J, Verta JP & MacKay JJ (2016) Conifer genomics and adaptation: at the crossroads of genetic diversity and genome function. *New Phytologist* 209: 44–62.
- Stoltzfus A & Norris RW (2015) On the causes of evolutionary transition: transversion bias. *Molecular Biology and Evolution* 33: 595–602.
- Sun R, Chang Y, Yang F, Wang Y, Li H, Zhao Y, Chen D, Wu T, Zhang X & Han Z (2015) A dense SNP genetic map constructed using restriction site-associated DNA sequencing enables detection of QTLs controlling apple fruit quality. *BMC Genomics* 16: 747.
- Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X & Zheng H (2013) SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 8: e58700.
- Tamura K, Nei M & Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences USA* 101: 11030–11035.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.
- Uchiyama K, Ujino-Ihara T, Ueno S, Taguchi Y, Futamura N, Shinohara K & Tsumura Y (2012) Single nucleotide polymorphisms in *Cryptomeria japonica*: their discovery and validation for genome map-

- ping and diversity studies. *Tree Genetics and Genomes* 8: 1213–1222.
- Xu F, Sun X, Chen Y, Huang Y, Tong C & Bao J (2015) Rapid identification of major QTLs associated with rice grain weight and their utilization. *PLoS One* 10: e0122206.
- Zhang J, Zhang Q, Cheng T, Yang W, Pan H, Zhong J, Huang L & Liu E (2015) High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* Sieb. et Zucc). *DNA Research* 22: 183–191.
- Zheng H, Duan H, Hu D, Li Y & Hao Y (2015) Genotypic variation of *Cunninghamia lanceolata* revealed by phenotypic traits and SRAP markers. *Dendrobiology* 74: 85–94.
- Zhu W, Liu T, Liu C, Zhou F, Lai XE, Hu D, Chen J & Huang S (2016) The complete chloroplast genome sequence of *Cunninghamia lanceolata*. *Mitochondrial DNA* 5: 1–2.