PITRO50K: A MULTISPECIES GENOTYPING ARRAY FOR TROPICAL PINES

Colin Jackson[1], Nanette Christie[2], Melissa Reynolds[2], Christopher Marais[2], Madison Caballero[3], Erik Visser[2], Sanushka Naidoo[2], Dominic Kain[4], Ross W. Whetten[1], Fikret Isik[1], Jill Wegrzyn[3], Gary Hodge[1], Alexander A. Myburg[2], Juan Jose Acosta[1]

[1] Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, NC, USA (cajacks7@ncsu.edu; jjacosta@ncsu.edu);
[2] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, 0002, South Africa;
[3] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA;
[4]HQPlantations Pty Ltd, North Lakes, LD, 4509, Australia

Using a combination of reduced representation sequencing methods, we performed targeted SNP discovery towards the development of a genome wide, multi-species genotyping array for tropical and subtropical pine species. Pooled RNA-seq data for five species of tropical pines, originating from pathogen challenge experiments, were used to identify transcript based SNP markers. Additionally, target capture sequencing was performed for six species, utilizing pooled DNA from 81 provenances that represented the natural range of the species across Mexico and Central America. Variant calling yielded a total of 1.8 million candidate SNP probesets, 1.3 million and 563K from RNA-seq and target capture respectively. In total, 300K RNA-seq and 120K target capture derived probesets were evaluated on a 420K screening array through the genotyping of 576 trees from eight species, representing the original 81 provenances and commercial breeding material. Evaluation of the screening array probesets resulted in the selection of 50K SNPs for inclusion on the commercial array. These markers included 20K polymorphic SNPs for *P. tecunumanii*, *P. patula*, *P. caribaea*, and *P. oocarpa*, 15K for *P. maximinoi* and *P. greggii*, 13K for *P. elliottii*, and 8K for *P. pseudostrobus*. Of the 50K markers, 75% are polymorphic in two or more species. The Pitro50K genotyping array represents the first high throughput and affordable genomic tool for these species of pines and their hybrids which represent the majority of tropical and subtropical pine plantations globally. Potential applications for the array include population genomics, genetic mapping, molecular breeding, species and hybrid identification, and genetic resource management.