BMC
Proceedings

Genome-wide genotyping and SNP discovery by ultra-deep Restriction-Associated DNA (RAD) tag sequencing of pooled samples of *E. grandis* and *E. globulus*

Dario Grattapaglia[1*], Sergio de Alencar[1], Georgios Pappas[2]

From IUFRO Tree Biotechnology Conference 2011: From Genomes to Integration and Delivery
Arraial d'Ajuda, Bahia, Brazil. 26 June - 2 July 2011

Background

The availability of next generation sequencing (NGS) technologies has opened the door to new strategies of SNP discovery and genotyping. Rapid genome-wide SNP detection via deep resequencing of reduced representation libraries of restriction digested pools of genomic DNA combined with a reference genome has been successfully used for SNP discovery in microorganisms [1], plants [2]and domestic animals [3]. Taking a step further from using NGS for SNP discovery, Baird et al [1]showed that NGS of short tags derived from barcoded multiplexed genomic representations generated with restriction enzymes could be used for direct genotyping of individuals, calling this method RAD (Restriction-site associated DNA) sequencing. RAD sequencing involves cutting a genome with at least one restriction enzyme and NGS the ends of the resulting fragments. We have recently developed a first set of SNPs for high-throughput genotyping of species of *Eucalyptus*. Although SNP assay success was high, the proportion of polymorphic SNPs declined as phylogenetic distance between species increased, down to <20% when contrasting *E. grandis* and *E. globulus*, the two main worldwide commercially planted species were considered [4]. In this work we used RAD sequencing to discover polymorphic SNPs across these two species. Additionally we were interested in assessing the potential of RAD for direct genotyping-by-sequencing in *Eucalyptus*.

Methods

DNA was extracted separately from 18 unrelated individual trees of *E. grandis* and 18 of *E. globulus*. For each species three bulks of six individuals were prepared with equimolar amounts of picogreen quantified DNA. DNA samples were delivered to Floragenex who carried out the RAD reduced representation library preparation using PstI and Illumina 75 bp single-end sequencing on a GAIIx. Raw sequence data was filtered for quality and mapped onto the 11 chromosomes of the *E. grandis* reference genome available in Phytozome. SNPs in the short sequence tags were called for nucleotides with quality Q> 30 at the position and a minimum of 6X coverage.

Results and conclusions

The average sequencing depth exceeded 28X for all bulked samples, providing a minimum estimated ~5X coverage for each individual present in each bulk providing a 93.75% probability of detecting a heterozygous SNP position. With 18 individuals per species (36 chromosomes), the probability of detecting a SNP allele with a Minimum Allele Frequency (MAF) > 0.1 is > 95% [5] therefore providing good power to select informative SNPs in each species separately and even more so in both species simultaneously. RAD sequence tags may be present or absent in specific individuals depending on the presence of the PstI restriction site providing large numbers of dominant markers; SNPs detected within the aligned tags provide additional co-dominant markers (Figure 1).

Out of a total of 200,712 SNPs declared with high confidence, 42,300 were simultaneously polymorphic in

* Correspondence: dario@cenargen.embrapa.br
[1]EMBRAPA Genetic Resources and Biotechnology – Estação Parque Biológico, 70770-910 and Genomic Sciences Program - Universidade Catolica de Brazilia, 70790-160 Brazilia, DF, Brazil
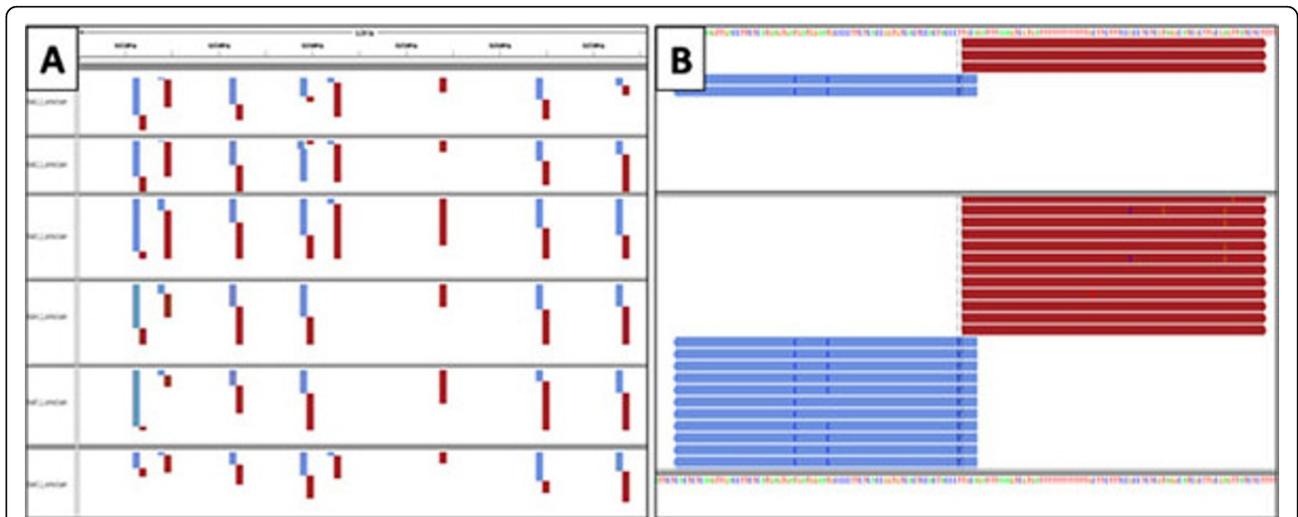Full list of author information is available at the end of the article

Figure 1 Screenshots of RAD sequencing tags mapped onto the Eucalyptus genome. (A) RAD tags generated from the six bulked samples (top to bottom) from 8 sequential PstI sites showing the variability in the presence or absence of tags across samples and in different directions (red = forward; blue=reverse); (B) close-up screenshot of SNPs detected in the RAD tags of two bulked samples.
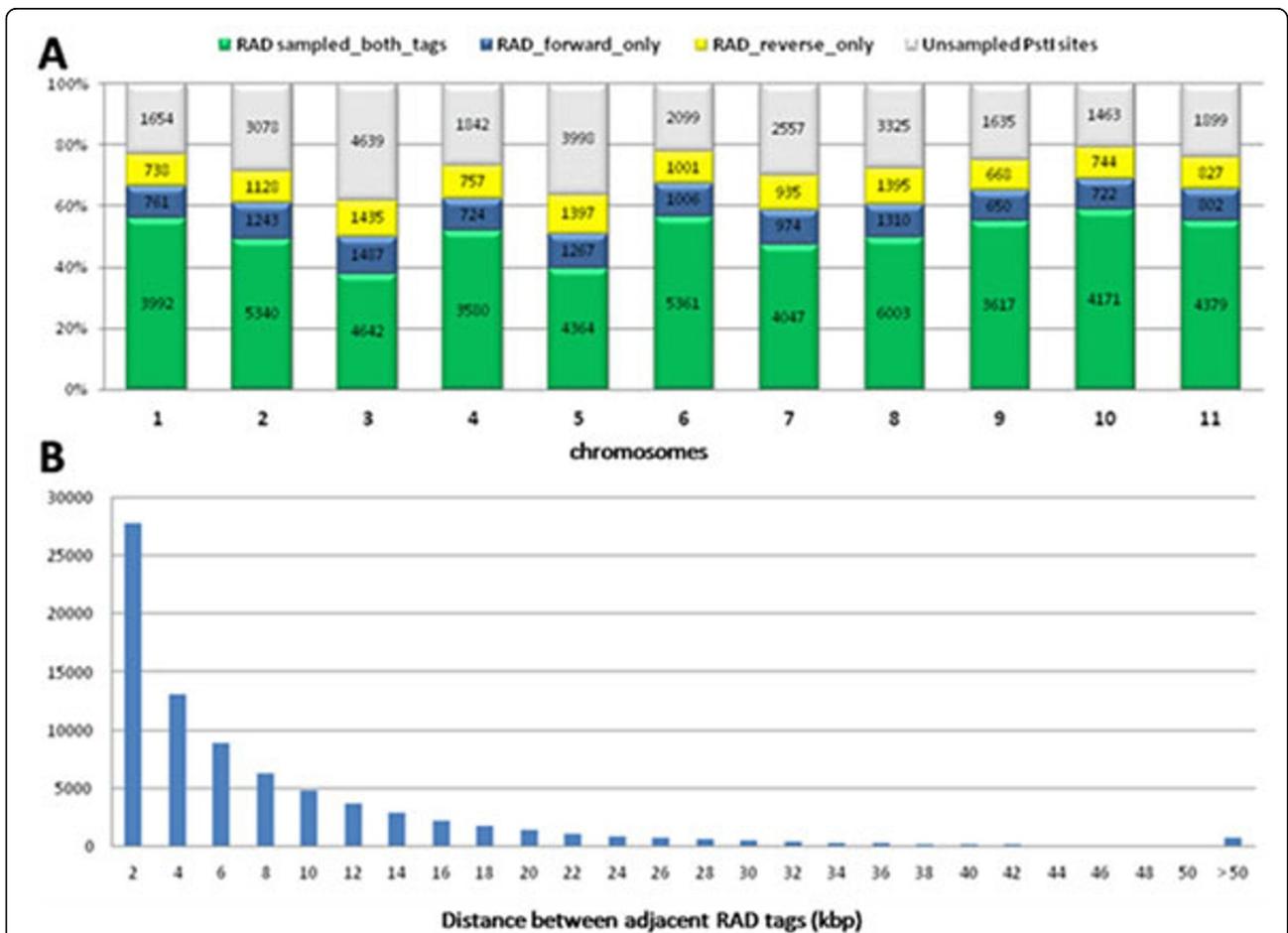

Figure 2 (A) Sampling of the PstI sites in the 11 chromosomes of *Eucalyptus* provided by RAD sequencing. Counts (into bars) and percent (Y axis) of unsampled PstI sites, RAD sequence tags generated in both directions out of the PstI site or in single directions (fwd or rev); (B) Distribution of the distances between adjacent RAD sequencing tags across the 605 Mbp of the *Eucalyptus* genome.

the two species while the remaining were fixed in one or the other. These 42,300 SNPs provide an average density of one SNP every 14 kbp in the Eucalyptus genome. These SNPs could be immediately used to select an evenly spaced set of SNPs for the development of a high density SNP genotyping platform.

Taken together, the RAD tags plus the SNPs into them provide excellent marker density for applications such as Genomic Selection [6]. Besides the RAD method, Elshire et al. [7] recently described a straightforward method of genotyping-by-sequencing. Additionally the DArT complexity reduction protocol has also been streamlined based on NGS for a number of plant genomes including *Eucalyptus* (see Sansaloni et al. this meeting). All these NGS based genotyping methods will cause a paradigm shift in our ability to carry out high density, high throughput and low-cost genotyping of large numbers of samples, unlocking incredible opportunities in forest tree genetics and breeding in the years to come.

Author details
[1]EMBRAPA Genetic Resources and Biotechnology – Estação Parque Biológico, 70770-910 and Genomic Sciences Program - Universidade Catolica de Brazilia, 70790-160 Brazilia, DF, Brazil. [2]EMBRAPA Genetic Resources and Biotechnology – Estação Parque Biológico, 70770-910, Brazilia, DF, Brazil.

References
1. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA: **Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers.** *Plos One* 2008, **3**(10).
2. Myles S, Chia JM, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D: **Rapid Genomic Characterization of the Genus Vitis.** *Plos One* 2010, **5**(1).
3. Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS: **SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries.** *Nature Methods* 2008, **5**(3):247-252.
4. Grattapaglia D, Silva-Junior O, Kirst M, Lima BM, Faria DA, Pappas GJ: **High-throughput SNP genotyping in the highly heterozygous genome of Eucalyptus: assay success, polymorphism and transferability across species.** *BMC Plant Biology* 2011, **11**:65.
5. Ott J: **Strategies for Characterizing Highly Polymorphic Markers in Human Gene-Mapping.** *American Journal of Human Genetics* 1992, **51**(2):283-290.
6. Grattapaglia D, Resende MDV: **Genomic selection in forest tree breeding.** *Tree Genetics & Genomes* 2011, **7**(2):241-255.
7. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: **A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.** *PLoS One* 2011, **6**(5):e19379.