South African Journal of Botany 117 (2018) 158–168

Contents lists available at ScienceDirect

South African Journal of Botany

journal homepage: www.elsevier.com/locate/sajb

A first genetic map in the genus *Streptocarpus* generated with RAD sequencing based SNP markers☆

Y.-Y. Chen [a,b,1], K. Nishii [a,c,*,1], S. Barber [a], C. Hackett [d], C.A. Kidner [a,b], K. Gharbi [b], A.J. Nagano [e], A. Iwamoto [c], M. Möller [a,**]

[a] Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, Scotland, UK
[b] University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, Scotland, UK
[c] Tokyo Gakugei University, 4-1-1 Nukuikitamachi, Koganei, Tokyo 184-8501, Japan
[d] Biomathematics and Statistics Scotland, Invergowrie, Dundee DD2 5DA, Scotland, UK
[e] Ryukoku University, Yokotani 1-5, Seta Ohe-cho, Otsu-shi, Shiga 520-2194, Japan

ARTICLE INFO

Article history:
Received 16 October 2017
Received in revised form 18 April 2018
Accepted 8 May 2018
Available online xxxx

Keywords:
Streptocarpus rexii
Streptocarpus grandis
RAD sequencing
Genome sequence
Next generation sequencing
Genetic linkage map

ABSTRACT

Streptocarpus shows a great diversity of shoot and flower morphology, but lacks resources for genetic studies. RAD sequencing with NGS technologies provides large amounts of genotyping data quickly and cost effectively and is very useful for genetic map construction. In this study, three approaches for constructing RAD loci were combined to build the first genetic map for *Streptocarpus* (Gesneriaceae). A backcross population between *Streptocarpus grandis* and *S. rexii* was used as the mapping population. The parents are both diploid ($2n = 32$) and largely autogamous. *Streptocarpus grandis* is unifoliate, retaining the enlarged macrocotyledon as sole foliar organ which is monocarpic, and *S. rexii* is rosulate and perennial. 233 backcross plants were genotyped, and the data from 200 or 150 individuals used to generate the maps. The RAD reads were assembled using a *de novo* approach, aligned on the reference draft genome sequences of *S. rexii* using BWA or Stampy, and finally all markers were analyzed combined. In the latter, 599 mappable markers were generated which were assigned to 16 linkage groups, with an average inter-marker distance of 2.6 cM. This work provides a valuable genetic resource for further studies in the large and morphologically diverse genus *Streptocarpus*.

1. Introduction

Genetic maps are fundamental tools for exploring the genetic basis of diversity within and between species. They are the essential foundation for association mapping, to identify genome regions that are linked to biological traits of interest (Lodish et al., 2000). In model plants, classical map-based cloning techniques were key for many years of work linking functional genes to associated traits (*e.g.* Meinke et al., 2003; Oraguzie et al., 2014).

Recent advances in next generation sequencing (NGS) technologies are not only speeding up gene isolation from model plants (Austin et al., 2011), but also expanding this field to non-model organisms (Unamba et al., 2015). Several approaches employing NGS techniques have been developed to obtain genotype data for constructing genetic maps. One approach is to use NGS data, such as those obtained from transcriptomes by RNA sequencing, to generate markers which are then used to genotype the mapping population (*e.g.* Kubisiak et al., 2013; Duarte et al., 2014; Feng et al., 2016). Another is to directly generate the markers and genotype data, such as in Restriction-site Associated DNA (RAD) sequencing (Miller et al., 2007; Baird et al., 2008; Peterson et al., 2012) or Genotyping by Sequencing (GbS; Elshire et al., 2011; Russell et al., 2014). These methods do not require separate marker generation and genotyping steps, and reduces the experimental time and cost dramatically.

Markers can be assembled from RAD reads *de novo*, or generated with the aid of a reference genome (Catchen et al., 2011, 2013). The *de novo* generation of markers was seen as an advantage of RAD sequencing (Catchen et al., 2011), because in principle it can be applied without any prior genomic knowledge and is applicable to non-model organisms (Miller et al., 2007; Baird et al., 2008). Using a reference genome can improve the result, lowering the error rate and increasing the number of loci resolved (Fountain et al., 2016; Shafer et al., 2016). The reference genome is selected from the genome sequence of the same or closely related species used in RAD sequencing and helps to identify similar RAD sequences mapping to different loci in the

☆ Conflicts of interest: none.

* Correspondence to: K. Nishii, 20A Inverleith Row, Royal Botanic Garden Edinburgh, EH3 5LR Edinburgh, Scotland, United Kingdom.
** Corresponding author.
E-mail addresses: kanaenishii@gmail.com (K. Nishii), mmoeller@rbge.org.uk (M. Möller).

[1] YC and KN have equally contributed to this study.

https://doi.org/10.1016/j.sajb.2018.05.009
0254-6299/

reference genome. However, where the reference genome is very poor and does not cover the full genome, the genetic markers may not be identified from the RAD sequencing where the sequence is missing in the reference genome.

In the *de novo* approach, the genotyping error rate is strongly affected by the sequencing coverage so that it requires a higher coverage to genotype the samples correctly (Fountain et al., 2016). On the other hand, in species with only preliminary genome assemblies, the *de novo* approach might have advantages for assembling RAD markers (Wang et al., 2013). Preliminary genome assemblies potentially lack large parts of the genome. RAD sequencing evenly samples across the entire genome based on particular restriction enzyme recognition sites. Thus, RAD sequencing based *de novo* approaches can sample genome gaps and detect markers where genome sequences are not yet assembled in preliminary genomes. In this study, we performed RAD sequencing and combined *de novo* and reference-based approaches for generating a first genetic map from RAD reads in *Streptocarpus* Lindl. (Gesneriaceae, Lamiales).

The genus *Streptocarpus* currently contains 176 species and is divided into two subgenera, differing in their basic chromosome numbers, with subgenus *Streptocarpella* having $x = 15$ and subgenus *Streptocarpus* $x = 16$ chromosomes (Nishii et al., 2015). The genus includes species with highly variable vegetative and floral morphologies (Jong, 1970; Hilliard and Burtt, 1971). The vegetative forms include one-leaf plants (unifoliates), rosulates and caulescents, and the flowers are zygomorphic with at least eight distinct types (Möller and Cronk, 2001; Hughes et al., 2006; Nishii et al., 2015).

A few studies have attempted to link phenotype and genotype in *Streptocarpus*, using the candidate gene approach, usually selecting developmental genes characterized in model plant studies (*e.g.* Harrison et al., 2005; Mantegazza et al., 2009; Nishii et al., 2010; Tononi et al., 2010; Chen et al., 2017). Although these approaches provided information on molecular mechanisms and some evolutionary aspects in *Streptocarpus*, attempts to link vegetative phenotypes in *Streptocarpus*, such as unifoliateness of the one-leaf plants, and genetic loci were not successful (*e.g.* Harrison et al., 2005). A genetic map and QTL approach may be more promising in this respect.

Thus, in this study, a genetic map was calculated using a backcross population of the unifoliate *Streptocarpus grandis* N.E.Br. and the rosulate *Streptocarpus rexii* (Bowie ex Hook.) Lindl. (Fig. 1, Fig. A.1). The two species are closely related and both belong to section *Streptocarpus* in subgenus *Streptocarpus*. In this section, *S. grandis* and *S. rexii* are members of sister clades (Nishii et al., 2015). Both species are diploid with $2n = 32$ chromosomes, self-compatible and highly autogamous (Hilliard and Burtt, 1971; Hughes et al., 2005). When hybridized *S. grandis* and *S. rexii* are known to generate semi-fertile hybrids (Oehlkers, 1964).

Using this *Streptocarpus* mapping population, we applied several different methodologies of marker construction from RAD reads on the resulting maps. We performed the *de novo* approach using RAD sequencing data only, with reference-based approaches where a draft genome assembly of *S. rexii* is added. In addition, for the reference-based approach, we performed alignment with two different short read aligners, one based on the Borrows-Wheeler transform and associated data structures (BWA) (Li and Durbin, 2009), and Stampy (Lunter and Goodson, 2011). BWA uses the theory on string matching using the Burrows–Wheeler Transform (Burrows and Wheeler, 1994), and Stampy is a hash-based method digesting the target sequence to a certain size (15 mer) of the substrings, generates a hash-table, and compares those information between the target and reference (Lunter and Goodson, 2011). These two aligners were tested and the resulting markers integrated into a genetic map, together with those of the *de novo* approach (Fig. 2).

2. Materials and methods

2.1. Plant materials

Plant materials were selected from the living collection cultivated at the Royal Botanic Garden Edinburgh (RBGE). Several accessions and lineages of *Streptocarpus rexii* and *S. grandis* were used in this study (Table 1). Both species have been maintained by selfing for several generations, *S. rexii* > 8 generations, *S. grandis*^{BC} > 15 generations, and *S. grandis*^{F1} *ca.* 5 generations.

2.2. Mapping population

A backcross population between *S. rexii* and *S. grandis* was generated at RBGE (Table 1, Fig. A.1). *Streptocarpus rexii* was crossed with *S. grandis* to generate an F1 hybrid. One F1 hybrid plant was maintained by leaf cuttings and hybridized with *S. grandis* as pollen donor, for generating the backcross population. Large numbers of F2 offspring between *S. rexii* and *S. grandis* are more difficult to obtain than backcross plants because of the low pollen fertility in F1 plants (Oehlkers, 1938). A total of 233 backcross plants of (*S. rexii* × *grandis*) × *grandis* were cultivated for RAD sequencing.

2.3. Genome sequencing

Whole genome sequencing was performed using *S. rexii* (20,150, 819*A), a descendant of the parent plant of the original F1 cross (19,990,270; see Table 1). DNA was extracted using the ChargeSwitch gDNA plant kit (Thermo Fisher Scientific, Waltham, MA, USA), following the manufacturer's protocol with some modifications. In brief, leaf

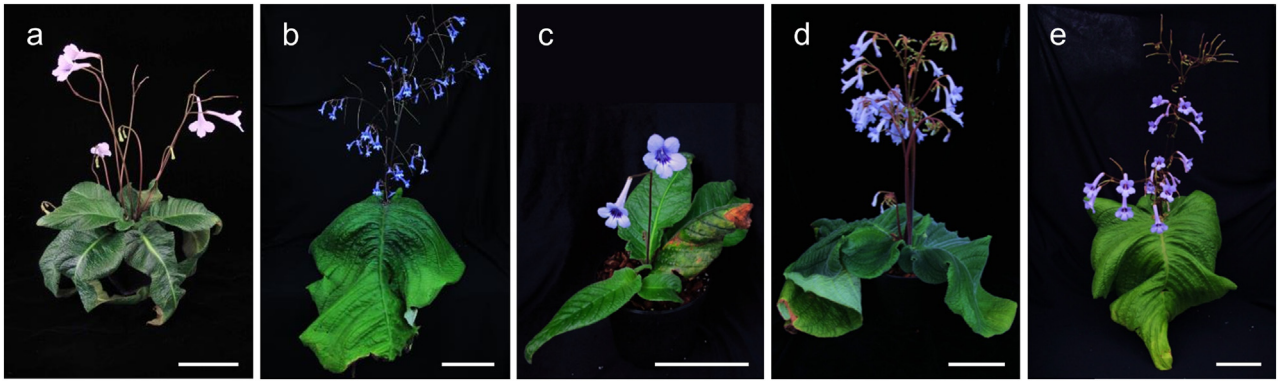

Fig. 1. Material used for this study. a *Streptocarpus rexii.* b *Streptocarpus grandis.* c *S. grandis* × *S. rexii* F1 hybrid. d,e Examples of morphological types in the backcross population used in this study. Bars = 10 cm.