Genomics 93 (2009) 282–288

Contents lists available at ScienceDirect

Genomics

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

BAC-end sequence-based SNPs and Bin mapping for rapid integration of physical and genetic maps in apple

Yuepeng Han [a,b,1], David Chagné [c,1], Ksenija Gasic [a,2], Erik H.A. Rikkerink [d], Jonathan E. Beever [e], Susan E. Gardiner [c], Schuyler S. Korban [a,*]

[a] Department of Natural Resources and Environmental Sciences, University of Illinois, 1201 W. Gregory, Urbana, IL 61801, USA
[b] Wuhan Botanical Garden of the Chinese Academy of Sciences, Moshan, Wuhan, 430074 PR China
[c] HortResearch, Palmerston North Center, Private Bag 11030, Palmerston North 4442, New Zealand
[d] HortResearch, Mt. Albert Research Center, Private Bag 92169 Auckland 1142, New Zealand
[e] Department of Animal Sciences, University of Illinois, 1201 W. Gregory, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 26 September 2008
Accepted 6 November 2008
Available online 21 December 2008

Keywords:
Apple
Single nucleotide polymorphisms
Physical map
Genetic map
Bin mapping

ABSTRACT

A genome-wide BAC physical map of the apple, *Malus*×*domestica* Borkh., has been recently developed. Here, we report on integrating the physical and genetic maps of the apple using a SNP-based approach in conjunction with bin mapping. Briefly, BAC clones located at ends of BAC contigs were selected, and sequenced at both ends. The BAC end sequences (BESs) were used to identify candidate SNPs. Subsequently, these candidate SNPs were genetically mapped using a bin mapping strategy for the purpose of mapping the physical onto the genetic map. Using this approach, 52 (23%) out of 228 BESs tested were successfully exploited to develop SNPs. These SNPs anchored 51 contigs, spanning ~37 Mb in cumulative physical length, onto 14 linkage groups. The reliability of the integration of the physical and genetic maps using this SNP-based strategy is described, and the results confirm the feasibility of this approach to construct an integrated physical and genetic maps for apple.

Introduction

Recently, we have developed a genome-wide BAC-based physical map of the apple, *Malus* × *domestica* Borkh. [1], and our integration of the physical and genetic maps is now underway. Genome-wide integrated physical and genetic maps are highly valuable resources for fine mapping and isolation of genes and quantitative trait loci (QTLs) responsible for important biological or agronomic traits, comparative genomics, and genome sequencing as well as marker-assisted selection. Usually, BAC library screening is pursued to anchor the physical map onto the genetic map of a genome. This represents a crucial and substantial component of any effort to develop an integrated physical and genetic map. BAC library screening is generally carried out using two different strategies, a PCR-based protocol [2] and filter hybridization [3]. The PCR-based screening protocol relies on the efficiency of the BAC pooling strategy. Often, two rounds of PCR screenings are performed to assign genetic markers to positive BAC clones. For the hybridization strategy, probes that correspond to well-known genetic markers are individually prepared and pooled, and then hybridized to high-density BAC colony filters. A high-density genetic map is essential for the generation of an integrated physical and genetic map using both PCR-based screening and filter hybridization approaches.

To date, approximately 1046 molecular markers have been developed for apple [4–6]. However, most of these markers (~70%) are amplified fragment length polymorphisms (AFLPs) and random amplified polymorphic DNAs (RAPDs). These two classes of markers are inefficient for integrating physical and genetic maps as they cannot be used easily to screen BAC libraries. As only about 330 microsatellite markers have been mapped in apple [4–6], development of additional genetic markers, including single nucleotide polymorphisms (SNPs) as well as simple sequence repeats (SSRs) is a major thrust in our effort to integrate the physical and genetic maps of the apple genome.

SNPs are the most abundant form of DNA sequence variants. For example, it is estimated that human and maize genomes have 1 SNP per 1 kb and 60–120 bp, respectively [7,8], making SNPs ideal markers for construction of high-density genetic maps. Due to the proliferation of expressed sequence tag (EST) databases and/or whole genome sequences, SNP analysis is widely used for building genetic maps [9–11]. For example, a dense SNP-based genetic map that utilizes whole

* Corresponding author.
E-mail address: korban@illinois.edu (S.S. Korban).
[1] Contributed equally.
[2] Present address: Department of Horticulture, Clemson University, Clemson, SC 29634 USA.

0888-7543/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2008.11.005

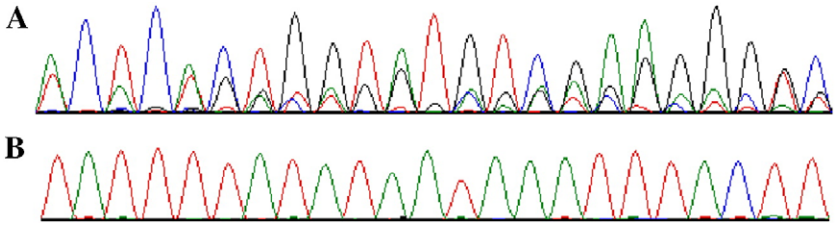

Fig. 1. Two chromatograms of sequencing of PCR products demonstrating a single band product following 1% agarose gel electrophoresis. A) Chromatogram containing frequent multiple overlapping peaks; and B) chromatogram displaying a clear sequence.

genome sequences has been recently reported for grape, *Vitis vinifera* [12]. Although SNP detection relies on bioinformatic analysis of EST and/or genomic sequence data, SNP validation is a critical phase of SNP discovery. This latter phase is both expensive and time-consuming, thus an efficient approach for establishing a dense SNP-based linkage map is highly desirable. Such an approach, based on a bin mapping strategy, has been proposed by Vision et al. [13], and this has been successfully used to map ESTs and SSRs in wheat [14], *Prunus* [15], and more recently in *Malus* [16, 17]. The bin mapping strategy relies on the use of low numbers of plants, rendering it an economical tool for SNP mapping studies. Here, we report on the development of SNPs from apple BAC-end sequences (BESs), and of mapping of these SNPs onto an existing apple genetic map using bin mapping. This genetic mapping effort simultaneously integrates the physical and genetic maps, thus rendering this SNP-based approach ideal for constructing an integrated map for apple, and will be also useful for those of other plant genomes.

Results

Candidate SNPs derived from the BAC end sequences

A total of 228 BAC end sequences were used to design primers to amplify three different apple genotypes; i.e., 'Golden Delicious', 'Co-op 17', and 'Co-op 16'. One pair of primers was designed for each BAC end sequence. Of 228 primer pairs, 53 (23%) failed to generate PCR products or amplified more than one band on 1% agarose gel electrophoresis. Products from the remaining 175 (77%) primer pairs, amplifying single band PCR products, were selected for sequencing. All sequencing results were visually checked, and two forms of chromatograms were identified: one contained multiple overlapping peaks; while, the other displayed sequence of single peaks (Fig. 1). Sequence chromatograms containing multiple overlapping peaks were not used for analysis. The remaining sequence chromatograms were aligned, and 99 sequence contigs, containing either 2 or 3 overlapping good quality DNA sequences, were generated.

Of 99 overlapping sequences, 69 contained one or more sequence variations corresponding to putative SNPs, with a total of 198 SNPs. The total length of all 99 contigs was 44,615 bp, and a single SNP could be detected every 225 bp in the apple genome. An example of a SNP identified in a contiguous sequence illustrating a putative SNP characterized by a double-peak at the level of only a single base is presented in Fig. 2. The number of SNPs in each contig sequence ranged from 1 to 16. Those A/G and C/T base substitutions made up ~65.2% of all base substitutions (Table 1). In summary, 69 (30%) of 228 BESs selected for initial primer design were successfully used to develop SNPs.

Bin mapping of candidate SNPs and anchoring of contigs

These 69 primer pairs flanking putative SNPs were used for genotyping the bin mapping set developed for a 'Royal Gala' × A689/24 population as previously described [17]. Briefly, a set of 14 highly informative individuals, representing a maximum of recombination events throughout the genome, were selected from the population used to construct a saturated genetic map. Each genotype was determined by direct sequencing. Although some sequences contained several scorable SNPs, only a single SNP was scored in the bin set for each BES. The genotype of each SNP was visually compared with the genotype of the framework markers of the bin mapping set. A total of 52 SNPs derived from BESs were mapped onto 14 linkage groups (Table 2). The newly mapped SNPs were deposited into dbSNP under accessions #102662752 to #102662801. An example of the position of these new SNPs on the genetic map is presented in Fig. 3. Candidate SNPs derived from the remaining 17 BESs could not be mapped due to either presence of monomorphic genotypes in the progeny, or difficulty in scoring of the genotype. Genetically anchored contigs were approximately 37 Mb in cumulative physical length, with an average contig size of 720 kb.

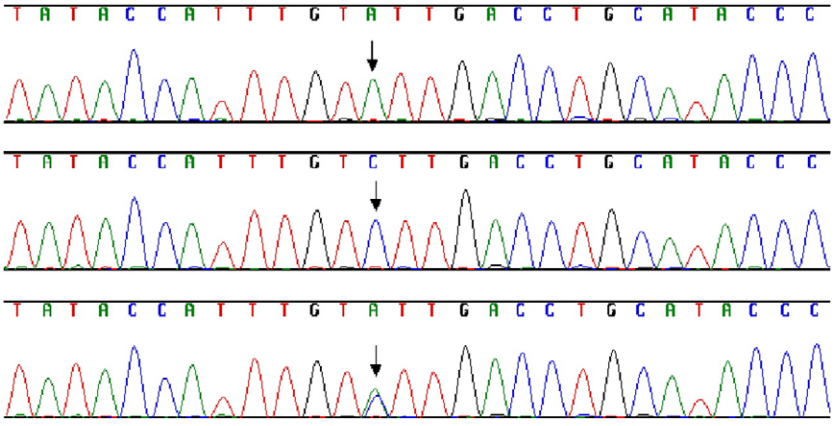

Fig. 2. An example of a SNP detected in overlapping sequences of PCR products amplified from three different apple genotypes using the same primers. Arrows indicate the A/C SNP.