Scientia Horticulturae 157 (2013) 113–118

Contents lists available at SciVerse ScienceDirect

Scientia Horticulturae

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

An integrated genetic map of pineapple (*Ananas comosus* (L.) Merr.)

Nelson de Sousa, Jorge Carlier, Tatiana Santo, José Leitão*

BioFIG, FCT, Universidade do Algarve, Campus de Gambelas, E. 8, 8005-139, Faro, Portugal

ARTICLE INFO

Article history:
Received 3 January 2013
Received in revised form 9 April 2013
Accepted 12 April 2013

Keywords:
Pineapple
Ananas
Genetic map
Bromeliaceae
DNA markers

ABSTRACT

The inclusion of multiple new markers, in particular 41 sequence specific markers, resulted in drastically improved version of a previously published F1-based genetic map of pineapple (*Ananas comosus*). The integration of the new version of this map with a more recently published F2-based map resulted in a map that assembles 741 loci: 739 DNA markers (25 SSR, 12 EST_SSR, 22 SCAR, 8 CAPS, 20 ISSR, 109 RAPD, and 543 AFLP), one isozyme (PGM) locus and the morphological trait 'piping', in 28 integrated linkage groups, spanning 2113 centimorgans (cM) and covering approximately 86% of the genome. Four small F1-based linkage groups and 5 small F2-based linkage groups assembling more than two markers, totalling 292 cM, remained not integrated. The present integrated genetic map is expected to be a helpful tool in genomic studies on pineapple and other *Bromeliaceae* genera and species.

1. Introduction

Pineapple (*Ananas comosus* (L.) Merr.), with a global production of 18 million tons in 2007 ranks third, after banana and citrus, among the world most important tropical fruit crops (Coppens d'Eeckenbrugge et al., 2011)

Consumed mostly as fresh fruit, pineapple is also worldwide commercialized as canned slices, chunks, juice and juice concentrate. Bromelain, used as a meat-tenderiser and nutraceutical, and fiber are other two important derivates from pineapple, a species which importance in the ornamental plant market is rapidly increasing (Coppens d'Eeckenbrugge et al., 2011).

The pineapple taxonomy was recently revised and simplified by Coppens d'Eeckenbrugge and Leal (2003) who downgraded the two genera, *Ananas* and *Pseudoananas*, and the seven species of the Smith and Downs (1979) classification, to two species: *A. comosus* ($2n = 50$) and *Ananas macrodontes* ($2n = 100$). According to this new classification *A. comosus* is subdivided in five botanical varieties: var. *comosus*, var. *ananassoides*, var. *erectifolius*, var. *parguazensis* and var. *bracteatus*, which include the former (Smith and Downs, 1979) diploid species. *A. comosus* var. *comosus* is the domesticated form with largest fruits and the most widely cultivated and commercialized pineapple. *A. comosus* var. *bracteatus* is cultivated as living fences and for fiber production, and used in traditional medicine (Coppens d'Eeckenbrugge et al., 2011).

Although slowly, the genomics and transcriptomics of pineapple is developing and so far among the approximately 1300 genomic sequences uploaded to the main genome data bases (www.ncbi.nlm.nih.gov) mostly (1163 sequences) belong to var. *comosus* followed by var. *bracteatus* (105 sequences). Mostly uploaded by Moyle et al. (2005) and by Ong et al. (2012) the totality of the uploaded 5941 pineapple expressed sequence tags (ESTs) are from *var. comosus* which contributes with most (183 out of 199) of the so far unveiled protein sequences.

The first genetic maps of pineapple were constructed using 46 F1 plants (Carlier et al., 2004). The map of the female parent *A. comosus* var. *comosus* (cv. Rondon, clone BR50) gathered 156 molecular markers in 30 linkage groups covering 31.6% of the genome length, whereas the map of the male parent *A. comosus* var. *bracteatus* (Branco do mato, clone Br20) included 335 molecular markers assembled in 50 linkage groups that covered 57.2% of the estimated length of the genome of this botanical variety (Carlier et al., 2004). These two F1-based maps were later integrated with markers which segregation was analysed among the subsequent F2 population, given rise to a new map with 46 linkage groups covering 62% of the estimated average length of the parental genomes (Carlier et al., 2006). Updated versions of this provisional integrated map were later included in review articles on pineapple genomics (Carlier et al., 2007; Botella and Smith, 2008).

Recently, we constructed a genetic map exclusively based on a F2 population assembling 492 DNA markers (57 RAPD, 22 ISSR, 348 AFLP, 20 SSR,12 EST-SSR, 25 SCARs, 8 CAPS) and the coding for spineless leaves locus "piping" into 40 linkage groups that cover approximately 80% of the estimated average genome length (Carlier et al., 2012).

Here we report the construction of drastically improved version of the F1 map (Electronic Supplementary Material) and its integration with the F2-based map which resulted in the assemblage

* Corresponding author. Tel.: +351 289 800939; fax: +351 289 818419.
E-mail address: jleitao@ualg.pt (J. Leitão).

0304-4238/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.scienta.2013.04.018

of an integrated genetic map gathering 741 markers in 28 linkage groups which cover 86% of the pineapple genome. Four small F1-based and 5 F2-based linkage groups with more than 2 markers remained non-integrated.

2. Materials and methods

2.1. Plant materials

The F1-based map (Electronic Supplementary Material) was reconstructed using 46 F1 plants originated from the cross between *A. comosus* var. *comosus* (cv. Rondon, clone BR 50) and *A. comosus* var. *bracteatus* (Branco do mato, clone BR 20), utilizing the pseudo-testcross strategy (Grattapaglia and Sederoff, 1994; Carlier et al., 2004). The F2-based genetic map (Carlier et al., 2012) was constructed using 135 F2 plants derived from the selfing of a single F1 plant.

2.2. DNA marker analyses

Procedures for DNA extraction, DNA-markers analysis, conversion of randomly amplified markers into Sequence Tagged Site (STS) markers and identification and retrieving of SSR ad EST-SSR markers from public genome data bases (www.ncbi.nlm.nih.gov) were as previously described in Farinhó et al. (2007) and Carlier et al. (2012).

2.3. Labeling of markers

Markers with dominant inheritance are identified by the suffixes Ac or Ab according to their parental origin, var. *comosus* or var. *bracteatus*, respectively. These suffixes are absent in co-dominant markers. Markers heterozygous in both parents, and segregating 3:1 among the F1-population, are suffixed with a "cb".

RAPD markers are identified by the letters OP (Operon Technologies) followed by the reference of the kit of primers, the primer number and the estimated size of the marker in base pairs, e.g. OPAB09_1200_Ac stands for the 1200-bp marker inherited from var. *comosus* amplified by the primer 09 of the kit AB.

ISSR markers are denoted by this acronym followed by the number ascribed to the primer (Farinhó et al., 2004) followed by the estimated size of the marker, e.g. ISSR23_550_Ab for the 550-bp marker derived from var. *bracteatus* amplified by the ISSR primer 23.

AFLP markers are identified by the two or three last nucleotides of the two primers (pEcoRI.pMseI) in each combination followed by the length of the marker in bp; e.g. AGC.CAG_243_Ac identifies the 243-bp AFLP marker inherited from var. *comosus*, amplified by primers ending, respectively, with the random nucleotide sequences AGC and CAG.

SSR and EST-SSR markers are denoted, by the respective acronyms followed by the identification code of the respective accession in the NCBI database, e.g. EST_SSR_CO731816.

SCAR markers are identified by the prefix "Sc" before the designation of the respective original marker, e.g. Sc_ISSR03_662_Ac stands for the dominant SCAR marker derived from marker ISSR03_662_Ac.

CAPS markers are denoted by the letters CAP followed by the names of the marker it originates from and used restriction enzyme, e.g. CAP_OPD01_615/MspI identifies the sequence specific marker originated from the RAPD marker OPD01_615 which polymorphism is revealed by the restriction enzyme *Msp*I.

Additional information regarding marker identification and labeling is provided in the text, figures and tables.

Table 1
Sequence specific markers only present in the F1 map.

Sequence-specific markers	Homozygous vs heterozygous		Alleles (B*–C*)
	B*	C*	
SSR_OPC09_387	Het	NotA	ab_??
SSR_AJ845041	Het	NotA	ab_??
SSR_AJ845058	Het	NotA	ab_??
SSR_AJ845065	Het	Het	ab_cd
SSR_AJ845068	Het	NotA	ab_??
SSR_AJ845074	Het	Hom	ab_aa
SSR_AJ845075	Het	NotA	ab_??
EST_SSR_CO731755	Het	Hom	ab_aa

B* and C* – var. *bracteatus* and var. *comosus*, respectively. Het – heterozygous; Hom – homozygous; NotA – not amplified.

2.4. Map construction and genome length estimation

The assemblage and the graphical representation of the linkage groups were performed using the JoinMap 3.0 program (Van Ooijen and Voorrips, 2001) set for the Kosambi mapping function.

Except for 14 F1-based linkage groups established for a LOD between 3 and 4, all the remaining F1-groups and all F2-based linkage groups were established with a minimum LOD of 4.0. The order of the loci within the integrated (F1 + F2) groups was established for a minimum LOD of 2.0, except for linkage groups: Int_LG3, Int_LG5, Int_LG11, Int_LG25, which order was determined for a LOD between 1 and 2. Markers which segregation was slightly ($\chi^2_{\alpha=0.05} \leq \chi^2 < \chi^2_{\alpha=0.01}$), more pronouncedly ($\chi^2_{\alpha=0.01} \leq \chi^2 < \chi^2_{\alpha=0.01}$) or strongly distorted ($\chi^2 \geq \chi^2_{\alpha=0.01}$) in regard to the expected Mendelian segregation ratios were labelled with one, two or three asterisks, respectively.

The genome lengths of the parental genotypes were estimated according to the method 3 of Chakravarti et al. (1991).

3. Results

The new F1-based map (Electronic Supplementary Material) is constituted by 42 linkage groups, ranging from 12.9 cM to 103.6 cM with an average size of 44 cM, which span over 1850 cM.

Nevertheless, though the new F1-map includes 19 markers common to both botanical varieties (suffixed with *cb*) and 14 co-dominant sequence specific markers, 13 linkage groups still are constituted by markers arising only from var. *bracteatus* and 10 groups by markers only from var. *comosus*. However, the improvement of the F1-based map becomes evident if we recall that the previous version was much more fragmented with 80 linkage groups, 50 constituted by markers of var. *bracteatus* and 30 groups by markers of var. *comosus*, among which 26 groups were simple pairs of markers. For the construction of the new version of the F1-map largely contributed the integration of 41 STS markers (10 SCAR, 3 CAPS, 21 SSR, 7 EST-SST) out of 61 markers of this type analysed among this population (Table 1. Electronic Supplementary Material).

Eight of the sequence specific markers included in the F1-map were mapped uniquely in this map (Table 1). The other 33 sequence specific markers were already included in the previously constructed F2-based map (Carlier et al., 2012) and together with an additional set of 128 markers: 109 AFLP, 18 RAPD and the "piping" locus common to both maps, they were used to integrate the F1-based and F2-based linkage groups. These markers are identified with the symbol "□" (Fig. 1).

The new integrated map gathers 741 loci: 739 DNA markers (25 SSR, 12 EST_SSR, 22 SCAR, 8 CAPS, 20 ISSR, 109 RAPD, 543 AFLP), one isozyme (PGM) locus and the morphological trait 'piping', in 28 linkage groups constituted by markers from the two merging maps

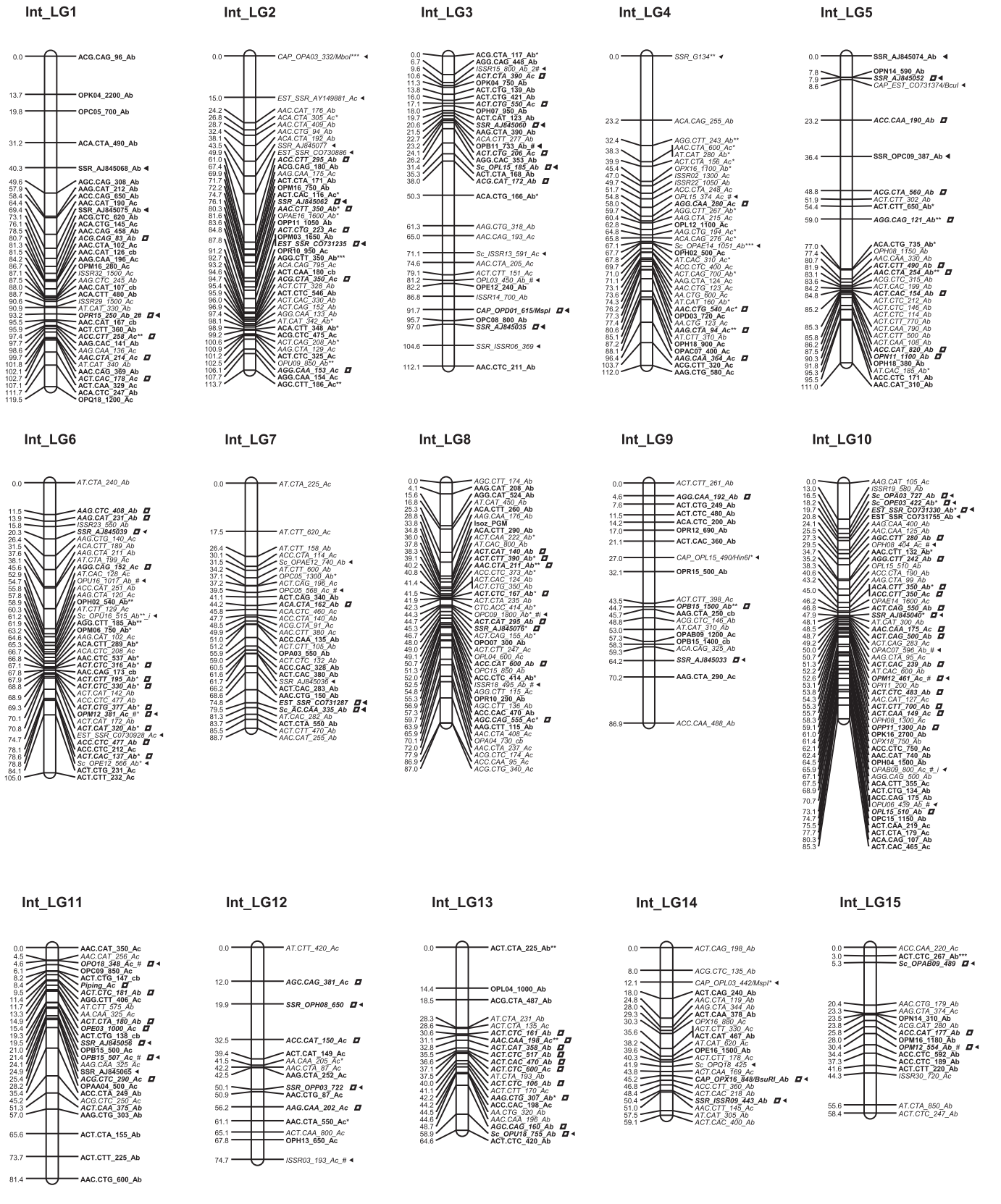

Fig. 1. The integrated genetic map of pineapple (*Ananas comosus*). Linkage groups are numbered according to size. Markers analyzed among the F1 population are in bold. Markers analyzed in the F2 population are italicized. Markers mapped in both populations, used to integrate the F1 and F2 linkage groups, are simultaneously in bold and italic and labelled with a □. Markers which sequence is known are labeled with a ◄.

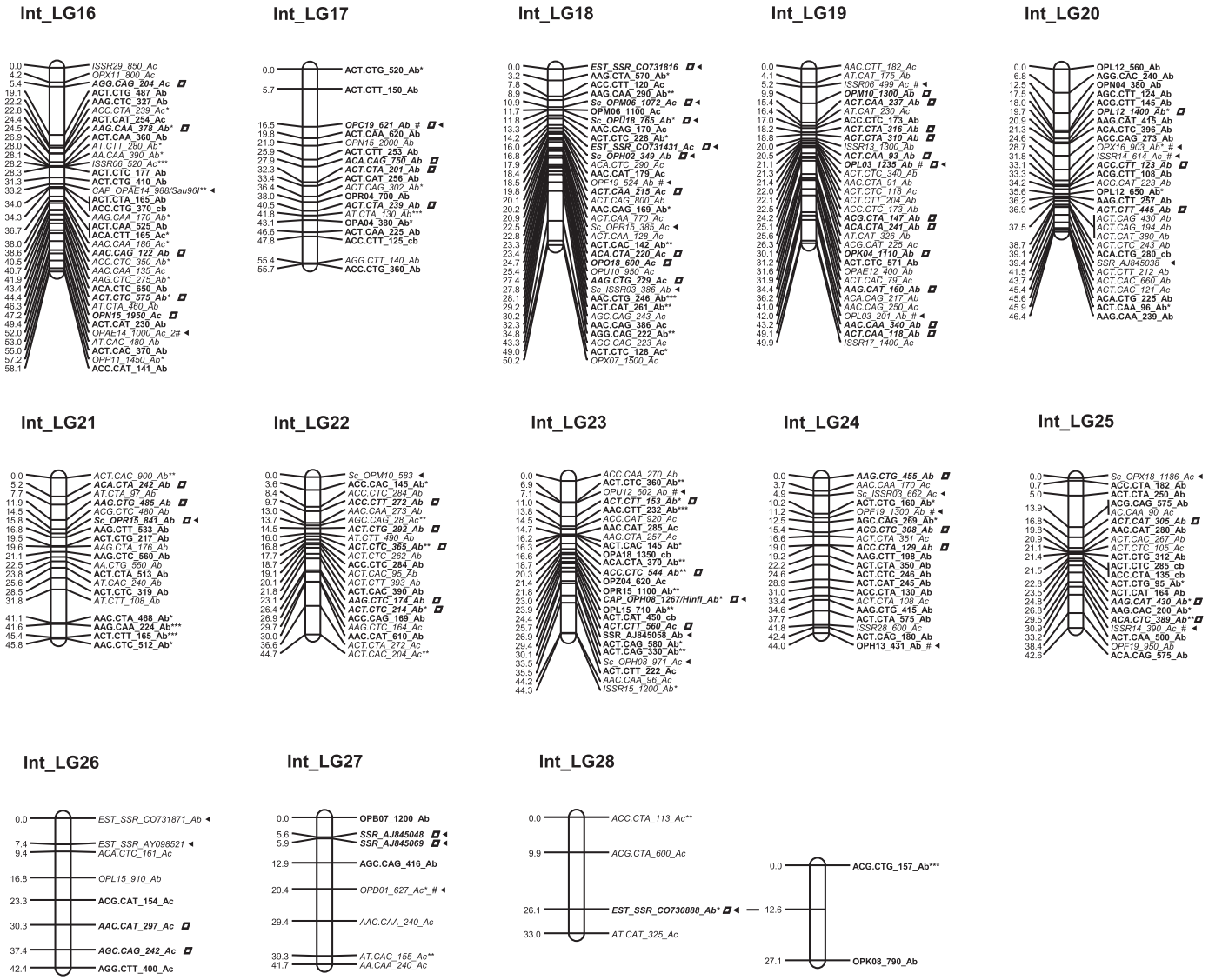

Fig. 1. (*Continued*).

(Table 2, Fig. 1). Nine small linkage groups with more than two markers from both maps remained not integrated and are shown in Fig. 2.

In addition to SSR and EST_SSR markers, which sequences have been retrieved from public databases (www.ncbi.nlm.nih.gov/) the integrated map includes sequence specific SSR, SCAR and CAPS markers derived from sequenced randomly amplified (RAPD, AFLP, ISSR) markers (Figs. 1 and 2; Tables 1 and 2). In 34 cases the specific markers derived of the sequenced RAPD and ISSR markers were monomorphic between the progenitors and could not be re-mapped as STS markers. These markers are labelled with the symbol "#" in Figs. 1 and 2 and the respective sequences (www.ncbi.nlm.nih.gov) can be used for genetic mapping using other mapping populations or for other purposes in genetic or genomic studies.

The F1- and F2-groups that have originated the new integrated groups are discriminated in Table 3. The integration resulted in few main alterations in the previously established groups: (a) the former linkage F2-group 4 was split in two smaller groups (4a and 4b); (b) three pairs of small F2-groups 28/38, 21/22, and 15/36 fused in 3 larger linkage groups; and c) multiple small F1-groups were included (e.g., groups 22, 35, 36 or groups 12, 25, 40) into large integrated groups (Table 3).

Table 2
Loci in the integrated map.

Type	No.	Observations
Morphological	1	Piping
SSR[a]	19	All sequenced
EST_SSR[a]	12	All sequenced
SSR[b]	1	Sequenced
SSR_RAPD	3	All sequenced
SSR_ISSR	2	All sequenced
Sc_RAPD	18	All sequenced
Sc_ISSR	3	All sequenced
Sc_AFLP	1	Sequenced
CAPS_EST	1	Sequenced
CAPS_RAPD	7	All sequenced
Isoz_PGM	1	
AFLP	543	
RAPD	109	24 sequenced (#)
		2 sequenced (2#)
		2 sequenced (#_i)
ISSR	20	5 sequenced (#)
		1 sequenced (2#)
Total	741	

– Sequenced but monomorphic as STS marker; 2# – with two different sequences; #_i – incomplete sequence.

[a] Sequences retrieved from public databases.

[b] Other origin.

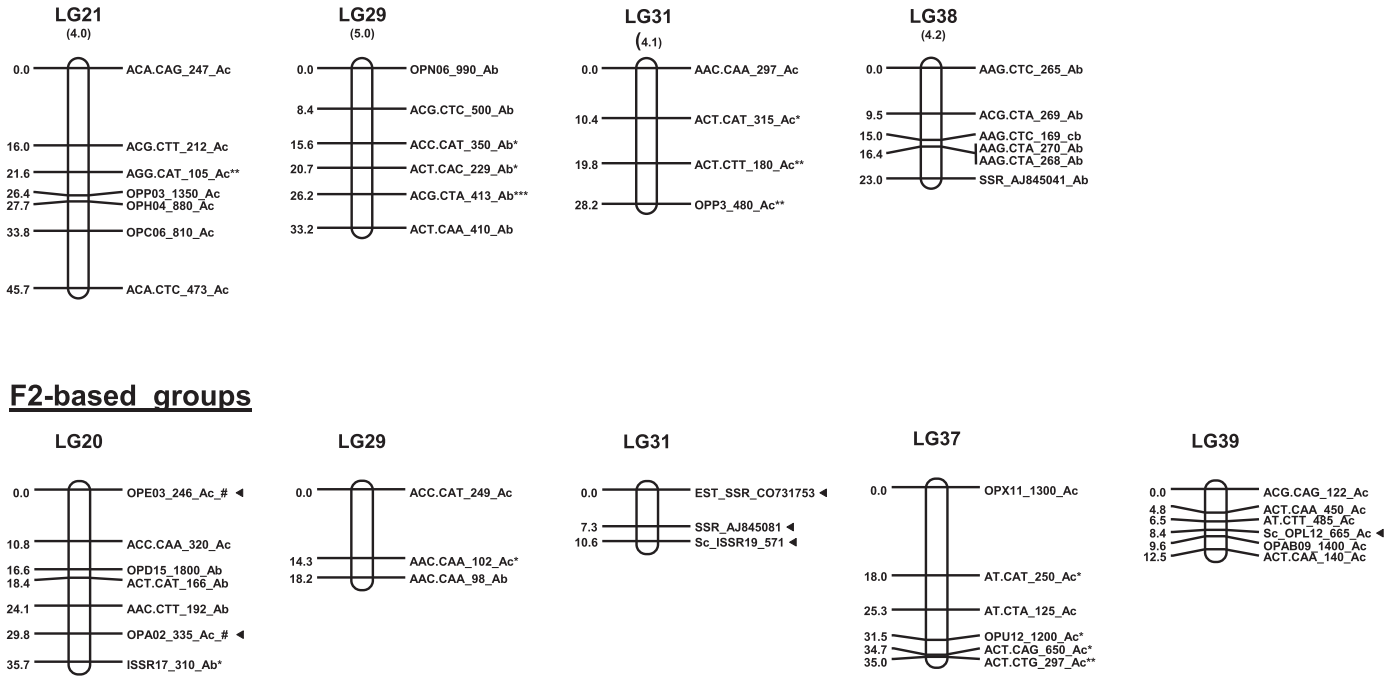

Fig. 2. The F1- and F2-linkage groups assembling more than two markers that remained non-integrated.

The sum of the lengths of the 28 integrated (F1 + F2) groups totalise 1958 centimorgans (cM), a value that increases to 2113 cM when the average distance between markers (2.76 cM) is added to both ends of all linkage groups. The genome length of the *A. comosus* var. *comosus* and *A. comosus* var. *bracteatus* genotypes used as progenitors of the mapping populations in this study, were previously estimated, respectively, as 2814 and 2126 cM to which correspond a species average genome length of 2470 cM (Carlier et al., 2012). Having into consideration all these values the present integrated map is estimated to cover 86% of the averaged pineapple (*A. comosus*) genome. The non integrated F1- and F2-groups span over 130.1 cM and 112 cM, respectively.

Table 3
Origin of the integrated linkage groups (new map).

Integrated Map (Int_LGs)[a]	F1-Map (LGs)[a]	F2-Map (LGs)[a]
Int_LG01	6	27
Int_LG02	3	1
Int_LG03	22, 35, 36	2
Int_LG04	11	3
Int_LG05	4, 37	7
Int_LG06	10, 34, 39	8
Int_LG07	17	5
Int_LG08	1	25
Int_LG09	9	15, 36
Int_LG10	12, 25, 40	6
Int_LG11	2	28, 38
Int_LG12	5	21, 22
Int_LG13	13, 41	24
Int_LG14	15	9
Int_LG15	24	10
Int_LG16	16, 23	11
Int_LG17	18	4(b)
Int_LG18	19, 30	17
Int_LG19	20	12
Int_LG20	7	26
Int_LG21	14	34
Int_LG22	28	4(a)
Int_LG23	27	13
Int_LG24	8	14
Int_LG25	26	16
Int_LG26	32	18
Int_LG27	42	19
Int_LG28	33	23

[a] Int_LGs – integrated linkage groups.

4. Discussion

During the last few years we have published some provisional pineapple genetic maps integrating the first constructed F1-based map (Carlier et al., 2004) with a successively growing F2-based map.

The first of these provisional maps gathered 574 markers in 46 linkage groups (Carlier et al., 2006), while a second version included in Carlier et al. (2007) integrated 661 loci in 36 linkage groups. A new version, assembled for inclusion in the review article on pineapple published by Botella and Smith (2008), was still very fragmented but with 32 major linkage groups integrating markers from both populations.

Recently we have reconstructed the F1-map including 41 sequence specific markers (SSR, EST-SSR, SCAR and CAPS) and several other new markers, most of them previously analysed among the F2 population (Electronic Supplementary Material). The integration of this drastically improved F1-based map with the previously published F2-based map (Carlier et al., 2012) allowed the construction of the present integrated map constituted by 28 linkage groups, a number still not agreeing completely, but much closer, to the $n = 25$ chromosomes of the species.

For the robustness of the new integrated map testifies the fact that markers used to join the two maps are assembled in the same way and in the same order both in the F1- and in the F2-based map.

Although relatively high (17.6% in the F1- and 14.2% in the F2-based map), the number of markers with distorted segregation is in the range of the commonly found in intraspecific crosses of multiple other plant species (Jenczewski et al., 1997).

In the integrated map some linkage groups still are constituted uniquely (Int_LG21) or almost exclusively (Int_LG5, Int_LG17 and

Int_LG20) by markers from var. *bracteatus* and, at much less extent, by markers from var. *comosus* (e.g. Int_LG12).

We can hypothesize that the assemblage of linkage groups exclusively, or almost exclusively, constituted by markers arising from one of the botanical varieties (progenitors), is caused by the existence of genomic regions of var. *bracteatus* with no homologous counterpart in var. *comosus* (and vice versa) which tend to be inherited as blocks with very low internal recombination. A similar trend is observed with regards to markers with skewed segregation: in the F1-map three relatively small linkage groups (27, 30 and 39) are mostly constituted by distorted markers arising from var. *bracteatus* while in the F2-based map (Carlier et al., 2012) three linkage groups (8, 11 and 25) assemble 21 distorted markers arising from this botanical variety.

Nevertheless, these and other issues regarding the highly fragmented pineapple genomes organized in large number of very small chromosomes, from $2n = 50$ in *A. comosus* to $2n = 100$ in *A. macrodontes* (Gitaí et al., 2005), are expected to be completely elucidated by genome sequencing which, with every passing day, becomes technically easier to perform and economically more affordable to accomplish.

The genomic studies on pineapple (Ananas) are expected to be significant far beyond this genus and to be particularly relevant for the other *Bromeliaceae* genera and species. This has been already demonstrated by Wöhrmann and Weising (2011) who have amplified eighteen pineapple EST-SSR markers through a wide range of genera and species from six subfamilies of the *Bromeliaceae*. The present integrated map is expected to be a useful tool for marker assisted selection, positional cloning of genes of interest and physical mapping in Ananas and, simultaneously, a useful instrument for comparative genomic studies within the *Bromeliaceae* family.

Acknowledgments

This research was supported by the project PTDC/AGR-GPL/77398/2006: “Construction of an integrated genetic map of Pineapple”, funded by the Fundação para a Ciência e a Tecnologia (FCT), Portugal. Nelson Horta Sousa and Tatiana Espírito Santo were recipients of Research FCT fellowships awarded within the framework of the same project. Jorge Carlier is the recipient of the post-doctoral grant SFRH/BPD/41714/2007 awarded by the FCT.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.scienta.2013.04.018.

References

Botella, J.R., Smith, M.K., 2008. Genomics of pineapple, crowning the king of tropical fruits. In: Moore, P.R., Ming, R. (Eds.), Genomics of Tropical Crop Plants, Plant Genetics and Genomics Series, vol. 1. Springer, New York, pp. 441–451.

Carlier, J.D., Reis, A., Duval, M.F., Coppens d'Eeckenbrugge, G., Leitão, J.M., 2004. Genetic maps of RAPD, AFLP and ISSR markers in *Ananas bracteatus* and *Ananas comosus* using the pseudo-testcross strategy. Plant Breed. 123, 186–192.

Carlier, J.D., Nancheva, D., Coppens d'Eeckenbrugge, G., Leitão, J.M., 2006. Genetic mapping of DNA markers in pineapple. Acta Hortic. 702, 79–86.

Carlier, J.D., Coppens d'Eeckenbrugge, G., Leitão, J.M., 2007. Pineapple. In: Kole, C. (Ed.), Genomic Mapping and Molecular Breeding in Plants, vol. 4, Fruits and Nuts. Springer, Berlin, pp. 331–342.

Carlier, J.D., Sousa, N.H., Santo, T.E., Coppens d'Eeckenbrugge, G., Leitão, J.M., 2012. A genetic map of pineapple (*Ananas comosus* (L.) Merr.) including SCAR, CAPS, SSR and EST-SSR markers. Mol. Breed. 29, 245–260.

Chakravarti, A., Lasher, L.A., Reefer, J.E., 1991. A maximum likelihood method for estimating genome length using genetic linkage data. Genetics 128, 175–182.

Coppens d'Eeckenbrugge, G., Leal, F., 2003. Morphology anatomy and taxonomy. In: Bartholomew, D.P., Paull, R.E., Rohrbach, K.G. (Eds.), The Pineapple: Botany, Production and Uses. CABI, Wallingford, Oxford, UK, pp. 13–32.

Coppens d'Eeckenbrugge, G., Sanewski, G.M., Smith, M.K., Duval, M.-F., Leal, F., 2011. Ananas. In: Kole, C. (Ed.), Wild Crop Relatives: Genomic and Breeding Resources, Tropical and Subtropical Fruits. Springer, Berlin, pp. 21–41.

Farinhó, M., Coelho, P., Carlier, J., Svetleva, D., Monteiro, A., Leitão, J., 2004. Mapping of a locus for adult plant resistance to downy mildew in broccoli (*Brassica oleracea* convar. italica). Theor. Appl. Genet. 109, 1392–1398.

Farinhó, M., Coelho, P., Monteiro, A., Leitão, J., 2007. SCAR and CAPS markers flanking the *Brassica oleracea* L. Pp523 downy mildew resistance locus demarcate a genomic region syntenic to the top arm end of *Arabidopsis thaliana* L. chromosome 1. Euphytica 157, 215–221.

Gitaí, J., Horres, R., Benko-Iseppon, A.M., 2005. Chromosomal features and evolution of *Bromeliaceae*. Plant Syst. Evol. 253, 65–80.

Grattapaglia, D., Sederoff, R., 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudotestcross: mapping strategy and RAPD markers. Genetics 137, 1121–1137.

Jenczewski, E., Gherardi, M., Bonnin, I., Prosperi, J.M., Olivieri, I., Huguet, T., 1997. Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). Theor. Appl. Genet. 94, 682–691.

Moyle, R.L., Crowe, M.L., Ripi-Koia, J., Fairbairn, D.J., Botella, J.R., 2005. PineappleDB: an online pineapple bioinformatics resource. BMC Plant Biol. 5, 21 http://www.biomed central.com/1471-2229/5/21

Ong, W.D., Voo, C.L., Kumar, S.V., 2012. Development of ESTs and data mining of pineapple EST-SSRs. Mol. Biol. Rep. 39 (5), 5889–5896.

Smith, L.B., Downs, R.J., 1979. Bromelioideae (*Bromeliaceae*) Flora Neotropica. New York Botanical Garden, New York, NY, USA.

Wöhrmann, T., Weising, K., 2011. In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across *Bromeliaceae*. Theor. Appl. Genet. 123, 635–647.

Van Ooijen, J.W., Voorrips, R.E., 2001. JoinMap Version 3.0 Software for the Calculation of Genetic Linkage Maps. Plant Research International, Wageningen.