Theor Appl Genet (2007) 115:237–243
DOI 10.1007/s00122-007-0559-4

ORIGINAL PAPER

Leveraging the rice genome sequence for monocot comparative and translational genomics

H. C. Lohithaswa · F. A. Feltus · H. P. Singh · C. D. Bacon · C. D. Bailey · A. H. Paterson

Received: 12 September 2006 / Accepted: 14 April 2007 / Published online: 24 May 2007

Abstract Common genome anchor points across many taxa greatly facilitate translational and comparative genomics and will improve our understanding of the Tree of Life. To add to the repertoire of genomic tools applicable to the study of monocotyledonous plants in general, we aligned *Allium* and *Musa* ESTs to *Oryza* BAC sequences and identified candidate *Allium-Oryza* and *Musa-Oryza* conserved intron-scanning primers (CISPs). A random sampling of 96 CISP primer pairs, representing loci from 11 of the 12 chromosomes in rice, were tested on seven members of the order Poales and on representatives of the Arecales, Asparagales, and Zingiberales monocot orders. The single-copy amplification success rates of *Allium* (31.3%), *Cynodon* (31.4%), *Hordeum* (30.2%), *Musa* (37.5%), *Oryza* (61.5%), *Pennisetum* (33.3%), *Sorghum* (47.9%), *Zea* (33.3%), *Triticum* (30.2%), and representatives of the palm family (32.3%) suggest that subsets of these primers will provide DNA markers suitable for comparative and translational genomics in orphan crops, as well as for applications in conservation biology, ecology, invasion biology, population biology, systematic biology, and related fields.

Communicated by E. Guiderdoni.

H. C. Lohithaswa and F. A. Feltus contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0559-4) contains supplementary material, which is available to authorized users.

H. C. Lohithaswa · F. A. Feltus · H. P. Singh · A. H. Paterson (✉)
Plant Genome Mapping Laboratory, University of Georgia,
111 Riverbend Road, Athens, GA 30602, USA
e-mail: paterson@uga.edu

H. C. Lohithaswa
University of Agricultural Sciences, Krishinagar,
Dharwad 580005, India

H. P. Singh
N.D. University of Agriculture and Technology,
Kumarganj, Faizabad 224264, Uttar Pradesh India

C. D. Bacon · C. D. Bailey
Department of Biology, New Mexico State University,
Las Cruces, NM 88003, USA

Introduction

The angiosperm class Liliopsida (monocots) is thought to have derived from a common ancestral form as determined by morphological and molecular analyses (Chase 1995; Judd et al. 2002; Davis et al. 2004). The Commelinid monocots (*sensu* Davis et al. 2004) include four orders (Arecales, Commelinales, Poales, Zingiberales) in which are found a cornucopia of the world's food resources. A significant fraction of world-wide agricultural production comes from crops in the order Poales, which includes maize (*Zea mays*), wheat (*Triticum aestivum*), rice (*Oryza sativa*), sugarcane (*Saccharum officinarum*), sorghum (*Sorghum bicolor*), barley (*Hordeum vulgare*), pineapple (*Ananas comosus*), and pearl millet (*Pennisetum glaucum*). The Zingiberales and Arecales encompass ginger (*Zingiber officinale*), banana (*Musa acuminata*), and the palms (Arecaceae), which together represent some of the most important exports from many developing countries. In addition, the Asparagales includes onion (*Allium cepa*), garlic (*Allium sativum*), *Aloe*, asparagus (*Asparagus officinalis*), *Agave*, *Iris*, leek (*Allium porrum*), vanilla (*Vanilla planifolia*) and horticultural orchids (Orchidaceae). The economic significance of these taxa clearly underlines the need for monocot genomic tools and resources, especially for those 'orphan' species for which minimal genomic data exists.

Plant biologists have capitalized on the advancement of sequencing technologies within the last decade (Paterson 2006), and the completion of the rice genome provided the first opportunity to assess the entire gene complement of a monocot (Goff et al. 2002; Yu et al. 2002; IRGSP 2005). The rice genome also serves as a framework for advances in monocot comparative biology with potential benefit for orphan crops such as pearl millet and Bermuda grass (*Cynodon dactylon*) in the Poaceae family and for crops with enormous nuclear genome content such as onion (15,797 Mbp), hexaploid wheat (15,966 Mb), and barley (4,873 Mb) (Arumuganathan and Earle 1991).

Similarly to orphan crops, many non-cultivated monocots that are important as weeds or prospective invasives are virtually unexplored at the DNA level and cannot be compared to the rice or other genomic framework, yet our knowledge of these species might benefit from comparative approaches. Thus, the vast majority of monocots lack sufficient genomic tools with which to investigate pertinent problems in conservation biology, ecology, invasion biology, population biology, and systematic biology. The development of generalized monocot genomic tools would benefit research on monocot model and non-model systems alike.

Reflecting the monophyletic nature of the Poaceae, large blocks of colinearity and synteny have been described between several grass taxa which foster inter-Poaceae genome comparison (Hulbert et al. 1990; Ahn and Tanksley 1993; Moore 1995; Paterson et al. 1995a, 2000, 2005; Devos and Gale 2000; Feuillet and Keller 2002; Bennetzen and Ma 2003). Colinearity and synteny, together with the availability of large sequence datasets from several Poaceae taxa including high coverage genomic libraries and large numbers of ESTs from multiple tissue sources, make the rice genome an attractive focal point for grass comparative genomics. However, it is unclear whether or not the rice genome is representative of the monocots as a whole, or how applicable genomic resources developed for the Poaceae will be to other monocots. By identifying and mapping genes that have remained highly conserved in DNA sequence since the radiation of monocots from their last common ancestors, this question might be investigated.

Here we utilized the "Conserved Intron Scanning Primer (CISP)" approach (Feltus et al. 2006) for developing comparative genomics resources useful across divergent monocots. A CISP pair is designed from conserved exon sequences that flank introns in order to maximize (intronic) polymorphism discovery rates within a taxon while maintaining cross taxa applicability via DNA conservation in the priming sites. This technique has been effective in both plants and animals (Aitken et al. 2004; Feltus et al. 2006; Fredslund et al. 2006). A total of 19,719 *Allium* ESTs, 15,661 *Musa* ESTs, and 2,074 *Oryza* BACs were used to identify CISPs and evaluate their suitability as pan-taxon genomic resources for both well studied models and resource poor taxa in the monocot lineage.

Materials and methods

CISP primer design

EC_oligos (Liu et al. 2004) software, which finds all identical oligonucleotides of a set length in a gene protein coding region, was used to design all possible polymerase chain reaction (PCR) primers between 19,719 *Allium* or 15,661 *Musa* ESTs, respectively, and 2,074 annotated *Oryza* BAC sequences downloaded on August 2, 2004 from NCBI (http://www.ncbi.nlm.nih.gov). Primer design criteria included: (A) PCR amplicon spans at least one intron, (B) 100% primer site identity between taxa, (C) 20 bp long primer sites, and (D) 200–2,000 bp predicted amplicon size. Primer combinations were removed if they were predicted to amplify more than one region of the rice genome (TIGR ver. 2; http://www.tigr.org). This resulted in 2,286 possible *Allium-Oryza* and 2,582 *Musa-Oryza* CISP combinations available at http://www.plantgenome.uga.edu/CISP. Most of the primer combinations overlap and represent 106 or 157 unique loci for *Allium-Oryza* and *Musa-Oryza* CISPS, respectively. Forty-eight *Allium-Oryza* and 48 *Musa-Oryza* CISP sets predicted to amplify non-redundant rice loci were synthesized (MWG Biotech, High Point, NC, USA).

Plant materials

Sampling for the PCR amplification and DNA sequencing component of the project included one genotype from rice (*Oryza sativa*, CT9993), sorghum (*Sorghum propinquum*), pearl millet (*Pennisetum glaucum*, 841B), Bermuda grass [*Cynodon transvaalensis* (2X), T574], maize (*Zea mays*, CML268), wheat (*Triticum aestivum*, M6), barley (*Hordeum vulgare*, Steptoe), banana (*Musa acuminata*), onion (*Allium cepa*), and multiple genotypes from the Arecaceae [*Ptychosperma macarthurii*, S. Zona 869 (FTG); *Metroxylon warburgii*, J. Rocal 046 (FTG); *Ravenea louvelii*, J. Roncal 48 (FTG); *Guassia maya*, A. Cuenca 28 (FTG); *Chamaedorea cataractarum*, 671084 (FTG); *Chamaedorea tuerckheimii, Chamaedorea tepejilote*, C. Bacon (UMEX); *Washingtonia filifera*].

PCR

PCR buffer conditions were the same for all primers. Reaction mixtures included 1 ng/µl genomic DNA, 0.2 mM dNTPs (Amersham), 1.25 units of Taq (Promega), 0.0626U

cloned Pfu (Stratagene), 3.0 μM forward/reverse CISP primer, 4 μM of $MgSO_4$, 3 μl 10× Cloned Pfu buffer (Stratagene) in a total reaction volume of 30 μl. PCR (MJ Research PTC−100) cycling parameters were: 94°C for 5 min followed by 94°C for 30 s, 55°C or 60°C for 45 s, 72°C for 60 s for 35 cycles, and a final extension at 72°C for 10 min. PCR products were visualized on 1.5% agarose gels stained with Ethidium Bromide. Loci were classified (0−3) according to whether they yielded no product (0); a single band (1); two bands (2); or three or more bands (3).

PCR product sequencing

Prior to sequencing, PCR products were digested with Exonuclease I/Shrimp Alkaline Phosphatase (Exo/Sap), adding 5 μl of a mixture of 1% ExoI and 10% SAP to 25 μl of PCR product, followed by a brief centrifugation then incubation at 37°C for 15 min (to react) and 80°C for 15 min (to terminate reaction). Cleaned high quality PCR products were sequenced using the ABI Big Dye 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and standard protocols (using forward primer). Finished cycle sequencing reaction products were treated with a dilute (2.2%) SDS solution, then passed through homemade Sephadex filter plates into Perkin-Elmer MicroAmp Optical 96-well reaction plates, and analyzed on an ABI 3730 automated DNA sequencer (Applied Biosystems, Foster City, CA, USA).

Sequence processing

Trace files for each locus were divided into separate projects in a genus-specific manner using the phred (http://www.phrap.org; Ewing et al. 1998) directory structure. Low quality reads (less than 75 bases at $Q > 16$) were removed using in-house Perl scripts, and the remaining reads were processed using the phredphrap script. Sequences were trimmed using Lucy2 (Li and Chou 2004) on default settings. Sequence alignments were produced by BLASTX-aligning trimmed PCR product reads with the TIGRv4 rice protein set (http://www.tigr.org) using a low-stringency cutoff ($E <= 10$). All DNA sequences have been deposited into Genbank under accession numbers: ED509338-ED509569.

Results

PCR amplification of orthologs by CISPs

A total of 19,719 *Allium* or 15,661 *Musa* ESTs each were aligned to 2,074 annotated rice BAC sequences to design a set of candidate pan-monocot CISPs (see Materials and methods). A total of 4,868 candidate CISPs (2,286 *Allium-Oryza;* 2,582 *Musa-Oryza*) passed the design criteria. However, this collection of CISPs is highly redundant due to overlapping primers predicted to amplify the same intron in the same rice genes. A non-redundant collection of representative CISPs tag 106 loci from the *Allium-Oryza* comparison and 157 loci from the *Musa-Oryza* comparison. Forty-eight CISP pairs from each set were randomly selected for synthesis.

Ninety-six CISP pairs designed from *Allium* and *Musa* ESTs (Table 1; Supplemental Table 1) were tested on seven members of the family Poaceae, one member each from the Musaceae and Alliaceae families, and several members of the Arecaceae family (Fig. 1). In *Oryza*, a high percentage (61.5%) of the primer sets amplified single bands. The single-band amplification success of these CISPs in *Sorghum* (sorghum—47.9%), *Pennisetum* (pearl millet—33.3%), *Cynodon* (Bermuda grass—31.3%), *Zea* (maize—33.3%), *Triticum* (wheat—30.2%), and *Hordeum* (barley—28.1%) provided an assessment of the degree to which these CISPs may work across members of the grass family. Amplification success rates of 37.5% for *Musa*, 31.3% for *Allium*, and 32.3% for Arecaceae, suggested many of the CISPs would work across the monocot lineage. In contrast, 124/124 CISP primers designed from grass-only EST-genome alignments (Feltus et al. 2006) failed to amplify *Allium* DNA, suggesting that the primer selection approach based on the use of distant monocot sequences is necessary for success.

Genomic distribution of CISP loci

The probable genome coverage of 96 CISPs was assessed in terms of physical distributions on the 12 rice chromosomes. An appreciable portion of the genome is sampled by this relatively small number of tested CISP sets (Fig. 2). The low coverage seen for chromosomes 11 (5 BACs) and 12 (0 BACs) is due to the paucity of ORF-annotated BACs from these chromosomes at the time the study was performed.

Sequence analysis

High quality sequences were generated from three Poaceae taxa (*Cynodon*—17 reads; *Oryza*—70 reads; *Sorghum*—29 reads) and three non-Poaceae taxa (*Allium*—30 reads; Araceae—43 reads, and *Musa*—46 reads). Sequences were examined for GC and AT content (Supplemental Table 2). The highest AT fraction was found in *Allium* (60.1% on average), whereas the lowest values were found in *Cynodon* (54.4% on average) and *Sorghum* (54.6% on average). The fact that several CISP primer sets worked in the AT-rich *Allium* genome points to the robustness of these CISP sets across genomes that vary widely in GC/AT ratios.

Table 1 PCR results of 96 *Musa-Oryza* and *Allium-Oryza* CISPs on selected monocots

PCR results	*Oryza*	*Musa*	*Allium*	*Sorghum*	*Pennisetum*	*Cynodon*	*Zea*	*Triticum*	*Hordeum*	*Arecaceae*
0 amplicon	1	27	51	28	28	54	28	33	46	55
1 amplicon	59	36	30	46	32	33	32	29	29	31
2 amplicons	28	26	12	20	19	8	23	15	19	10
3+ amplicons	8	7	3	2	17	1	13	19	2	0
Musa-derived (1 amplicon)	27	22	7	25	22	16	15	15	15	13
Allium-derived (1 amplicon)	32	14	23	21	10	14	17	14	14	18
% Success (All CISP)	61.5	37.5	31.3	47.9	33.3	34.4	33.3	30.2	30.2	32.3
% Success (*Musa*-derived)	56.3	45.8	14.6	52.1	45.8	33.3	31.3	31.3	31.3	27.1
% Success (*Allium*-derived)	66.7	29.2	47.9	43.8	20.8	29.2	35.4	29.2	29.2	37.5

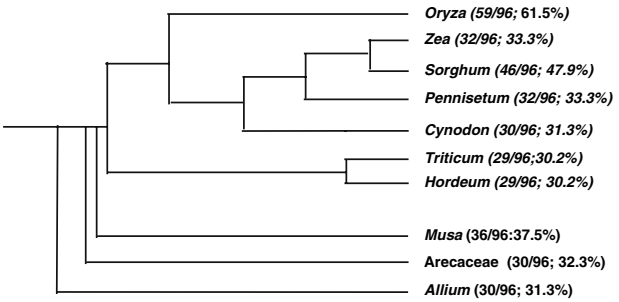

Fig. 1 Effectiveness of CISPs across the monocots. Approximate dendrogram illustrating the approximate evolutionary relationships of monocots used in the study. The values in the *parentheses* indicate successful PCR rates

In order to determine the number of loci amplified from non-Poaceae taxa that are likely to be homologous to the rice locus used in primer design, the longest, trimmed read for each CISP locus from *Allium*, Arecaceae, and *Musa* were aligned to the rice protein set using BLASTX. Using a very low stringency *E*-value ($E <= 10$) cutoff, it was determined that 56/98 (56.1%) reads matched the expected rice locus (Supplemental Table 3). *Musa* reads were more likely to hit the expected rice locus (65.2%) than *Allium* (46.9%) or Arecaceae (50.0%) sequences. For those reads that did not correspond to the expected locus, 21/96 (21.4%) did match an alternate genomic position in rice (Supplemental Table 4). Finally, 22/98 (22.4%) of the reads did not hit the rice genome even at the very low stringency threshold. These results suggest that a little more than half of the CISPs that amplified as a single PCR product provide anchor points for highly diverged monocot taxa. Furthermore, these genomic positions are specific, since they amplify a single gene and may be useful in marker development.

Discussion

The successful amplification of single-copy intron-spanning loci using the CISP primers developed with the rice genome ranged from 30–61% in Poales to 31–38% outside of the Poaceae lineage with an overall average success rate of 36.7%, suggesting that approximately one-third of these primer sets may yield genomic tags from most monocots. In many cases, these tags are suitable as genomic anchor points between the grasses and other monocots, thereby facilitating the extrapolation of information from well-characterized grass genomes across the monocot lineage.

Previous studies have shown that angiosperm genomes differ in nucleotide composition (reviewed in King 2002). For example, Kuhl et al. (2004) compared EST and genomic datasets between representatives of the Asparagales (*Allium*), Poales (*Oryza*), and Brassicales (*Arabidopsis*) and concluded that the *Allium* genome is more similar to *Arabidopsis* than to *Oryza* with respect to mean GC content and other factors. The GC composition of *Allium* was lower than the other monocots we sampled, a fact that supports previous observations that onion has a very low GC content relative to other angiosperms (Kirk et al. 1970; Stack and Comings 1979; Matassi et al. 1989). The widespread use of the CISP primers developed here will facilitate the sampling of genomes whose nucleotide content is unknown. The difference in GC content observed between *Allium* and members of Poaceae indicates that the grass genomes are not necessarily representative of all monocots and identifies the need for broader surveys to better characterize genomic variation among monocots.

The pan-monocot single-copy amplification success rates (i.e. single PCR product) of about a third, actually mask appreciably higher success rates in the target taxa. Of the 48 primers designed from *Musa-Oryza* alignment and used to amplify *Musa* DNA, 45.8% amplified a single band, 37.5% amplified multiple bands, and 16.7% did not amplify. Of the 48 primers designed from *Allium-Oryza* alignment and used to amplify *Allium* DNA, 47.9% amplified a single band, 20.8% amplified multiple bands, and 31.3% did not amplify. Thus, primers targeted to a particular taxon showed higher success rates than did, for example, *Musa*-based primers on *Allium* DNA. Moreover, in all

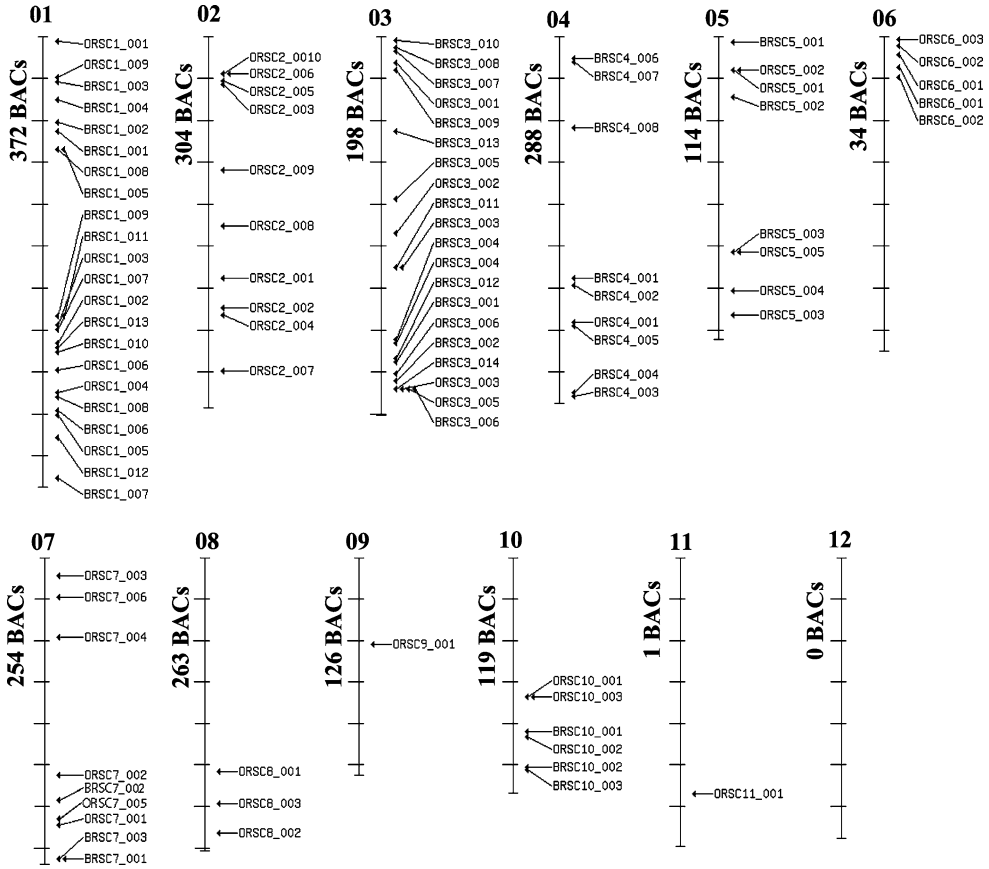

Fig. 2 Distribution of *Musa-Oryza* and *Allium-Oryza* primers on rice chromosomes. Twelve rice chromosomes are shown with the number of BACs per chromosome used in CISP primer design shown to the left of each chromosome. Each tick-mark on a chromosome represents four megabases. CISP primer sets are shown as triangles. BRSC primers were designed from *Musa-Oryza* conserved DNA segments. ORSC primers were derived from *Allium-Oryza* conserved DNA segments

genomes, many of the 'failures' are actually due to amplification of multiple loci. PCR amplifications that resulted in multiple products are presumably from specific amplifications of duplicate gene copies that maintain the conserved primer sites but differ in intron size and/or number. Therefore, the identification of discriminatory polymorphisms or the use of the excised bands as probes may shed light on the current genomic position of the duplicated genes and whether they have 'moved' from syntenic positions. This would further increase the success rate at identifying 'anchor loci' useful for genome comparisons.

With respect to the true failures, i.e. which yielded no amplification in the source genome, some otherwise-conserved loci may have undergone genomic events that interfere with PCR amplification, such as intron expansion, exon rearrangement, or exon sequence divergence. The lower than expected success rates in the design taxa, *Allium* and *Musa*, are presumably due to a combination of the above effects which could not be detected due to the lack of complete sequence information for these taxa.

While none of the primer sets amplified one and only one band in all tested samples, some sets stand out as good candidates for testing in any monocot species. Eight CISP sets amplified a single band in at least 7/10 samples (BRSC3_007, BRSC4_001, BRSC4_003, ORSC1_002, ORSC1_003, ORSC2_003, ORSC2_004, ORSC2_010). These CISP sets are likely to successfully amplify loci that can be compared across monocots and therefore represent a logical starting point for testing the CISP primers on novel monocot DNA samples. Many other primer sets also amplified single PCR product in multiple samples (<7/10), and these might also be effective tools for the study of poorly characterized genomes.

The similarities between *Oryza* and *Musa* are currently under investigation. For example, the GC distribution among putative *Musa* genes is bimodal, as also reported for *Oryza* (Carels and Bernardi 2000). Our data support the notion that *Musa* and *Oryza* are similar at the nucleotide composition level in that the AT/GC content of CISP amplicons between *Oryza* and *Musa* were almost identical. Lescot et al. (2005) compared eleven *Musa acuminata* BAC sequences with the complete genome sequence of *Oryza*, which led to the identification of several putative syntenic regions. The genomic tags generated by our approach should add to the repertoire of genomic tools needed to further elucidate the degree of synteny between *Oryza* and *Musa*.

Since CISPs were designed to amplify intron DNA which should be more polymorphic than exon DNA, the intra-taxon comparison of CISP amplicon sequences should

be useful in phylogenetic studies, marker development, and for comparative mapping across monocot families. For example, CISP primers that we developed for grasses have been successfully mapped as intron-size length polymorphisms in tef (*Eragrostis tef*; Yu et al. 2006). Mapping these loci may facilitate the development of the syntenic information that is necessary for better understanding the evolution of genes, genomes, and gene functions across plant taxa. The CISP sets for which a sampling were validated herein may serve as valuable genomic resources for orphan monocots which are less studied at the genomic level.

In addition to facilitating research focused on the improvement of orphan crops, pan-monocot tools may provide an economical means to address many additional applications (Linder and Rieseberg 2004), obviating the need to develop taxon-specific tools. For example, knowledge of taxa that represent problematic weeds, but that are closely-related to well studied crops, might benefit from comparative/translational approaches that utilize growing genomic information about weediness in botanical models (Paterson et al. 1995b; Hu et al. 2003; Westerbergh and Doebley 2004). In conservation biology, pan-taxon PCR-based tools might facilitate improved management decisions by providing non-destructive means to obtain basic information about genetic architecture of populations at low cost (DeSalle and Amato 2004). In invasion biology, such tools might be useful across a wide range of taxa not only in identifying invaders, but also to obtain molecular data providing insight into both mechanisms (Lee 2002) and management (Hufbauer 2004) of invasions. Broadly applicable and available tools are also of tremendous interest to researchers focused on ecology, population biology, and systematic biology (Kuittinen et al. 2002; Aitken et al. 2004; Linder and Rieseberg 2004).

Acknowledgments This work was supported by grants from the Rockefeller Foundation and USAID- CCGI to AHP, FAF, and HS; and NSF EF−0542228 to CDB as well as the Biotechnology Overseas Associateship of the Department of Biotechnology, Ministry of Science & Technology, Government of India to HCL.

References

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90:7980–7984

Aitken N, Smith S, Schwarz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol Ecol 13:1423–1431

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Bennetzen JL, Ma J (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 6:128–133

Carels N, Bernardi G (2000) Two classes of genes in plants. Genetics 154:1819–1825

Chase M (1995) Molecular systematics of Liliaceae. In: Rudall PJ, Cribb PJ, Cutler DF, Humphries CJ (eds) Monocotyledons: systematics and evolution. Kew Publishing, Surrey

Davis JI, Stevenson DW, Petersen G, Seberg O, Campbell LM, Freudenstein JV, Goldman DH, Hardy CR, Michelangeli FA, Simmons MP, Specht CD, Vergara-Silva F, Gandolfo M (2004) A Phylogeny of the monocots, as inferred from rbcL and atpA sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. Syst Bot 29:467–510

DeSalle R, Amato G (2004) The expansion of conservation genetics. Nat Rev Genet 5:702–712

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12:637–646

Ewing B, Hillier L, Wendl MC, Green P, Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitziel NO, Kwok PY, Gish WR, Olson SA, Bailey TL, Gribskov M, Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Feltus FA, Singh HP, Lohithaswa HC, Schulze SR, Silva TD, Paterson AH (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. Plant Physiol 140:1183–1191

Feuillet C, Keller B (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot (Lond) 89:3–10

Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L (2006) A general pipeline for the development of anchor markers for comparative genomics in plants. BMC Genomics 7:207

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100

Hu FY, Tao DY, Sacks E, Fu BY, Xu P, Li J, Yang Y, McNally K, Khush GS, Paterson AH, Li ZK (2003) Convergent evolution of perenniality in rice and sorghum. Proc Natl Acad Sci USA 100:4050–4054

Hufbauer RA (2004) Population genetics of invasions: can we link neutral markers to management? Weed Technol 18:1522–1527

Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc Natl Acad Sci USA 87:4251–4255

IRGSP (2005) The map-based sequence of the rice genome. Nature 436:793–800

Judd WS, Stevens PF, Campbell CS, Kellogg EA, Donoghue MJ (2002) Plant systematics: a phylogenetic approach. Sinauer Associates, Incorporated, Massachusetts, USA

King GJ (2002) Through a genome, darkly: comparative analysis of plant chromosomal DNA. Plant Mol Biol 48:5–20

Kirk JTO, Rees H, Evans G (1970) Base composition of nuclear DNA with the genus Allium. Heredity 25:507–512

Kuhl JC, Cheung F, Yuan Q, Martin W, Zewdie Y, McCallum J, Catanach A, Rutherford P, Sink KC, Jenderek M, Prince JP, Town CD, Havey MJ (2004) A unique set of 11,008 onion expressed sequence tags reveals expressed sequence and genomic differences between the monocot orders Asparagales and Poales. Plant Cell 16:114–125

Kuittinen H, Aguade M, Charlesworth D, Haan ADE, Lauga B, Mitchell-Olds T, Oikarinen I, Ramos-Onsins S, Stranger B, van

Tienderen P, Savolainen O (2002) Primers for 22 candidate genes for ecological adaptations in Brassicaceae. Mol Ecol Notes 2:258–262

Lee CE (2002) Evolutionary genetics of invasive species. Trends Ecol Evol 17:386–391

Lescot M, Ciampi AY, Ruiz M, Blanc G, Leebens-Mack J, Garsmeur O, D'Hont A, da Silva FR, Ronning CM, Cheung F, Haas BJ, Althoff R, Arbogast T, Hine E, Pappas G, Souza MT, Miller R, Glaszmann JC, Town CD, Piffanelli P (2005) Fresh insights into the Musa genome and its comparison with rice. In: Journées Ouvertes : Biologie, Informatique et Mathématiques, Lyon, France

Li S, Chou HH (2004) LUCY2: an interactive DNA sequence quality trimming and vector removal tool. Bioinformatics 20:2865–2866

Linder CR, Rieseberg LH (2004) Reconstructing patterns of reticulate evolution in plants. Am J Bot 91:1700–1708

Liu S, Tinker NA, Molnar SJ, Mather DE (2004) EC_oligos: automated and whole-genome primer design for exons within one or between two genomes. Bioinformatics 20:3668–3669

Matassi G, Montero LM, Salinas J, Bernardi G (1989) The isochore organization and the compositional distribution of homologous coding sequences in the nuclear genome of plants. Nucleic Acids Res 17:5273–5290

Moore G (1995) Cereal genome evolution: pastoral pursuits with 'Lego' genomes. Curr Opin Genet Dev 5:717–724

Paterson AH (2006) Leafing through the genomes of our major crop plants: strategies for capturing unique information. Nat Rev Genet 7:174–184

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming R, Wright RJ (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1540

Paterson AH, Freeling M, Sasaki T (2005) Grains of knowledge: genomics of model cereals. Genome Res 15:1643–1650

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995a) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995b) The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, Sorghum halepense (L.) Pers. Proc Natl Acad Sci USA 92:6127–6131

Stack SM, Comings DE (1979) The chromosomes and DNA of Allium cepa. Chromosoma 70:161–181

Westerbergh A, Doebley J (2004) Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. Theor Appl Genet 109:1544–1553

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Li J, Liu Z, Qi Q, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Zhao W, Li P, Chen W, Zhang Y, Hu J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Tao M, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296:79–92

Yu JK, Kantety RV, Graznak E, Benscher D, Tefera H, Sorrells ME (2006) A genetic linkage map for tef [Eragrostis tef (Zucc.) Trotter]. Theor Appl Genet 113:1093–1102