JOURNAL OF Evolutionary Biology

doi: 10.1111/j.1420-9101.2010.02125.x

SHORT COMMUNICATION

The molecular basis of white pericarps in African domesticated rice: novel mutations at the *Rc* gene

B. L. GROSS, F. T. STEFFEN & K. M. OLSEN

Department of Biology, Washington University in St. Louis, St. Louis, MO, USA

Keywords:

domestication;
grain colour;
Oryza glaberrima;
Oryza sativa;
parallel evolution;
proanthocyanidin.

Abstract

Repeated phenotypic evolution can occur at both the inter- and intraspecific level and is especially prominent in domesticated plants, where artificial selection has favoured the same traits in many different species and varieties. The question of whether repeated evolution reflects changes at the same or different genes in each lineage can now be addressed using the domestication and improvement genes that have been identified in a variety of crops. Here, we document the genetic basis of nonpigmented ('white') pericarps in domesticated African rice (*Oryza glaberrima*) and compare it with the known genetic basis of the same trait in domesticated Asian rice (*Oryza sativa*). In some cases, white pericarps in African rice are apparently caused by unique mutations at the *Rc* gene, which also controls pericarp colour variation in Asian rice. In one case, white pericarps appear to reflect changes at a different gene or potentially a *cis*-regulatory region.

Introduction

Domesticated plants are useful study systems for a diverse array of disciplines, and evolutionary biologists in particular have taken advantage of the abundance of genetic resources combined with the strong artificial selection applied during the domestication process to understand genetic responses to selection and the genetic basis of phenotypic diversity (Doebley *et al.*, 2006; Burke *et al.*, 2007; Burger *et al.*, 2008; Purugganan & Fuller, 2009). One question of special interest in evolutionary genetics is the genetic basis of repeated phenotypic evolution under domestication (Paterson *et al.*, 1995; Glémin & Bataillon, 2009). Repeated phenotypic evolution can occur at the interspecific level, where it is manifest as the appearance of the same trait in multiple domesticated species, and at the intraspecific level, where the same trait arises multiple times within a single crop. The advent of quantitative trait locus (QTL) mapping in the 1990s, followed by the cloning of a number of domestication genes, has made it possible to pursue the genetic basis of repeated evolution at an increasingly detailed level in several different crops.

In rice, the opportunity exists to examine the genetic basis of repeated trait evolution at both the interspecific and intraspecific level. At the interspecific level, this genus features two different crop species: Asian rice (*Oryza sativa*, derived from the Asian wild species *Oryza rufipogon*) and the independently domesticated African rice (*Oryza glaberrima*, derived from the African wild species *Oryza barthii*) (Linares, 2002; Semon *et al.*, 2005; Londo *et al.*, 2006; Caicedo *et al.*, 2007). At the intraspecific level, there have been at least two independent domestication events within Asian rice; these correspond broadly to the *japonica* rice subspecies (comprising *tropical japonica*, *temperate japonica* and *aromatic* rice varieties) and the *indica* rice subspecies (comprising *indica* and *aus* varieties) (see Sweeney & McCouch (2007) for a thorough review of rice domestication). The question of repeated intraspecific phenotypic evolution in *O. sativa* has been addressed by taking advantage of recently cloned genes (Sweeney *et al.*, 2007; Gross *et al.*, 2009; Kovach *et al.*, 2009), but the same candidate genes have rarely been leveraged to investigate the genetic basis of the evolution of similar traits in *O. glaberrima*. Here, we begin the process by exploring the genetic basis of a

Correspondence: Kenneth M. Olsen, Department of Biology, Washington University in St. Louis, St. Louis, MO, USA.
Tel.: 314 935 7013; fax: 314 935 4432; e-mail: kolsen@wustl.edu

classical domestication-associated trait, grain colour, in *O. glaberrima* as compared to *O. sativa*.

Grain colour has been the target of selection in domesticated Asian rice (as well as other domesticated grains), and the vast majority of *O. sativa* varieties have white (nonpigmented) grains, compared to the dark red grains that characterize the wild relative and other wild species (Sweeney *et al.*, 2006, 2007). The genetic basis of white grains in *O. sativa* has been pinpointed to loss-of-function mutations in the *Rc* gene, a regulatory protein in the proanthocyanidin synthesis pathway; these loss-of-function mutations prevent the development of a pigmented pericarp layer (Sweeney *et al.*, 2006). A survey of 337 white pericarp rice cultivars showed that the majority (> 97%) have a 14-bp deletion in exon 7 of the gene, resulting in a premature stop codon and a nonfunctional allele referred to as *rc*. The prevalence of the *rc* allele is because of introgression and selective breeding—the allele arose in the *japonica* variety group and spread across nearly all Asian domesticates despite their independent origins, including the *aromatic*, *indica* and *aus* varieties (Sweeney *et al.*, 2007). The same survey revealed repeated evolution of light pericarps within *O. sativa* because of independent changes at the same locus; a small number (< 3%) of the cultivars, all in the *aus* group, harboured an alternative loss-of-function allele referred to as *Rc-s*, resulting from a point mutation that causes a premature stop codon in exon 7 (Sweeney *et al.*, 2007). Thus, alleles resulting in white pericarps arose independently after each unique domestication event in *O. sativa*, although one became far more common than the other.

In contrast to Asian domesticated rice, most varieties of African domesticated rice have red pericarps and so do not differ in grain colour from their wild relative. However, collections of *O. glaberrima* from its range in western Africa do include some populations with white pericarps. In this study, we sequenced the *Rc* locus in accessions from several white pericarp populations as well as samples of the more common red pericarp African cultivars and the wild relative *O. barthii*. These sequences were compared to samples of wild, domesticated and weedy Asian rice to determine the role of the *Rc* locus in generating white pericarp African rice. This approach is predicated on the assumption that the *Rc* gene controls grain colour in African rice via the same mechanisms that it does in Asian rice, although this assumption has not yet been tested using functional genetic approaches.

We expected that there would be three possible results from the survey. The first was that the white pericarp *O. glaberrima* samples were the result of either intentional or unintentional introgression from *O. sativa*, which is also grown in western Africa and has been shown to hybridize with the native crop (Chu & Oka, 1970; Semon *et al.*, 2005). In this case, *O. glaberrima* would have an *O. sativa* allele at the locus in question, most likely the predominant *rc* allele. The second was that the *Rc* gene in white pericarp *O. glaberrima* would show no coding changes resulting in either premature stops or deleted amino acids; this would indicate that the mutations resulting in white pericarps were located in a different gene (or possibly in an undetected *cis*-regulatory region of the *Rc* gene). The third possibility was that the *Rc* locus in white pericarp *O. glaberrima* would show functionally similar but unique mutations compared to *O. sativa*, indicating that repeated phenotypic evolution occurred via changes at the same locus in the two species. Note that because several samples were used in this study, none of these possibilities were mutually exclusive; each might be true for a different sample.

Methods and materials

Samples included both newly generated DNA sequence data and previously sequenced accessions that are publicly available. Previously published sequences included 131 accessions from Gross *et al.* (2010), Genbank numbers GU261543–GU261673 and 25 accessions from Sweeney *et al.* (2007), Genbank numbers DQ885796, DQ885799–DQ885811, DQ885813–DQ885816, DQ8858 19, DQ885820, DQ885822, DQ885823 and DQ902350–DQ902352. These combined sample sets comprised 68 samples of domesticated rice (13 US cultivars and 55 Asian landraces), 57 samples of weedy *O. sativa* (red rice) from the US, 27 *O. rufipogon* samples from throughout the native range of the species (including samples labelled as the annual form *Oryza nivara*), two samples of the Latin American species *Oryza glumaepatula* and two samples of the Australasian species *Oryza meridionalis* (Table 1).

Newly sequenced samples included seven *O. glaberrima* (three with white pericarps, four with red pericarps) and two *O. barthii* (both with red pericarps); all samples were

Table 1 *Oryza* samples classified according to pericarp colour. Sample sizes are in parentheses. Details on accessions are provided in Table S1.

	Pericarp colour	
Species or variety group	Red	White
Oryza rufipogon (27)	24	3
Cultivated *Oryza sativa* variety groups (68)		
indica (18)	3	15
aus (9)	4	5
tropical japonica		
Asian cultivars (16)	3	13
US cultivars (13)	0	13
temperate japonica (7)	1	6
aromatic (5)	0	5
Weedy *O. sativa* from US (57)	56	1
Oryza barthii (2)	2	0
Oryza glaberrima (7)	4	3
Oryza glumaepatula (2)	2	0
Oryza meridionalis (2)	2	0

from the range of *O. glaberrima* cultivation in western Africa (Table 1; see Table S1 for locality details). One *O. glaberrima* sample was acquired from the USDA Germplasm Resources Information Network; all remaining *O. glaberrima* and *O. barthii* samples were acquired from the International Rice Research Institute (Table S1). All *O. glaberrima* and *O. barthii* were grown from seed in the greenhouse of Washington University in St. Louis. Plants were visually inspected for any evidence of hybridization with *O. sativa* based on species-specific ligule morphology; no plants used in the sequencing survey showed obvious morphological signs of hybridization. Fresh tissue was harvested and frozen in liquid nitrogen; DNA was extracted via a modified CTAB procedure (Gross *et al.*, 2009). Primers described in Gross *et al.* (in press) were designed to PCR amplify and sequence the coding region of *Rc* (6.4 kb) as well as the regions approximately 1.6 kb upstream and 0.8 kb downstream of the start and stop codons. PCR fragment amplification and Big Dye Terminator sequencing were performed according to conventional methods and separated on an ABI 3130 capillary sequencer at the Washington University Biology Departmental core facility. Sequencing techniques were standard and are available on request. GenBank accession numbers for newly sequenced samples are HM749735–HM749743. Sequence editing and alignment were performed using the PHRED, PHRAP and Polyphred programs (Deborah Nickerson, University of Washington) and BioLign Version 4.0.6 (Tom Hall, North Carolina State University, Raleigh, NC, USA). All *O. glaberrima* and *O. barthii* samples were homozygous, so cloning was not necessary to determine the sequences accurately.

A maximum likelihood (ML) tree was generated for all unique non-African rice sequences (as identified by Collapse 1.2 by David Posada, http://darwin.uvigo.es/) combined with all nine African rice sequences. Because the lengths of *Rc* sequences from some published GenBank accessions were shorter than the 9576-bp region sequenced for the present study, the aligned *Rc* data set for tree construction was 6675 bp. The ML tree was constructed using RAxML (Stamatakis, 2006b; Stamatakis *et al.*, 2008) *via* the CIPRES 2.2 web portal using the GTR model of molecular evolution with rate variation among sites (the only model of molecular evolution implemented by RAxML). Rate heterogeneity among sites was estimated using the GTR + CAT approximation, which is a fast computational substitution for the GTR + Γ model that can be applied to large data sets (> 50 taxa) (Stamatakis, 2006a). Bootstrap values were calculated via 100 replicates of the data set, and a consensus tree was generated using Consense from Phylip 3.68 (Felsenstein, 2005).

Results

The two nonfunctional alleles at the *Rc* locus in *O. sativa* described by Sweeney *et al.* (2006) characterize the previously sequenced data set; one is the 14-bp deletion *rc* allele and the other is the point mutation *Rc-s* allele. Neither of these mutations were found in the white pericarp *O. glaberrima* samples. Instead, two samples contained a novel point mutation predicted to result in a premature stop codon in exon 7, hereafter referred to as allele *rc-g1* (Fig. 1). This mutation is 146 bp upstream of the *Rc-s* point mutation and 201 bp upstream of the initiation of the 14-bp *rc* deletion. The third white pericarp *O. glaberrima* contained no identifiable insertions, deletions or premature stop codons, nor does it differ from the red pericarp *O. glaberrima* sequences for any predicted amino acid changes. Comparison of this sample to the most closely related red pericarp sample showed that they were identical for the sequenced region upstream of the start codon (> 1.5 kb) and for the first intron; both of these regions are potentially important as *cis*-regulatory regions in plants.

The ML tree generated in this study (Fig. 2) is broadly similar to the tree generated in Gross *et al.* (2010), which contains the same non-African samples, although there are differences in the placement of some *O. rufipogon* haplotypes and the precise relationships between some individual haplotypes within groups that are consistent across trees. However, none of these differences influence conclusions made using the previous tree, and all major groupings are identical. All *O. glaberrima* and *O. barthii* samples group together (100% bootstrap support), and are separate from all *O. sativa* and *O. rufipogon* haplotypes, consistent with the finding that even white pericarp *O. glaberrima* do not share alleles with *O. sativa*. The *O. barthii* and *O. glaberrima* samples are intermixed within the group, indicative of a close relationship between the wild and domesticated species as is seen

```
AA position in exon 7          33                                  70                                  107
Rc allele O. rufipogon         IKNYLPVSEKSSFSRWTTPEGSDDNKTMISPGTTQRMLKSILMIVPSSHCSYRGAETPESRGGKGASGTRKVGAI
rc allele O. sativa            ..................................................................CHPR*FQ
Rc-s allele O. sativa          ...............................................*..........................
red O. glaberrima              ..............................................................................
rc-g1 O. glaberrima            .*............................................................................
```

Fig. 1 Predicted amino acid sequence of positions 33–107 in *Rc* exon 7. Samples show functional *Rc* alleles (*Rc* allele of *Oryza rufipogon* and red pericarp *Oryza glaberrima*), known nonfunctional alleles in *Oryza sativa* (*rc* and *Rc-S*), and the newly described mutation in white pericarp *O. glaberrima* (*rc-g1*). Amino acids identical to the first sequence are represented by periods (.), and stop codons are represented by asterisks (*). The white pericarp *O. glaberrima* with no obvious mutation is not shown, but is identical to the red pericarp *O. glaberrima* in this region. Amino acid positions are relative to the first complete codon in exon 7, not the first codon in the *Rc* locus.

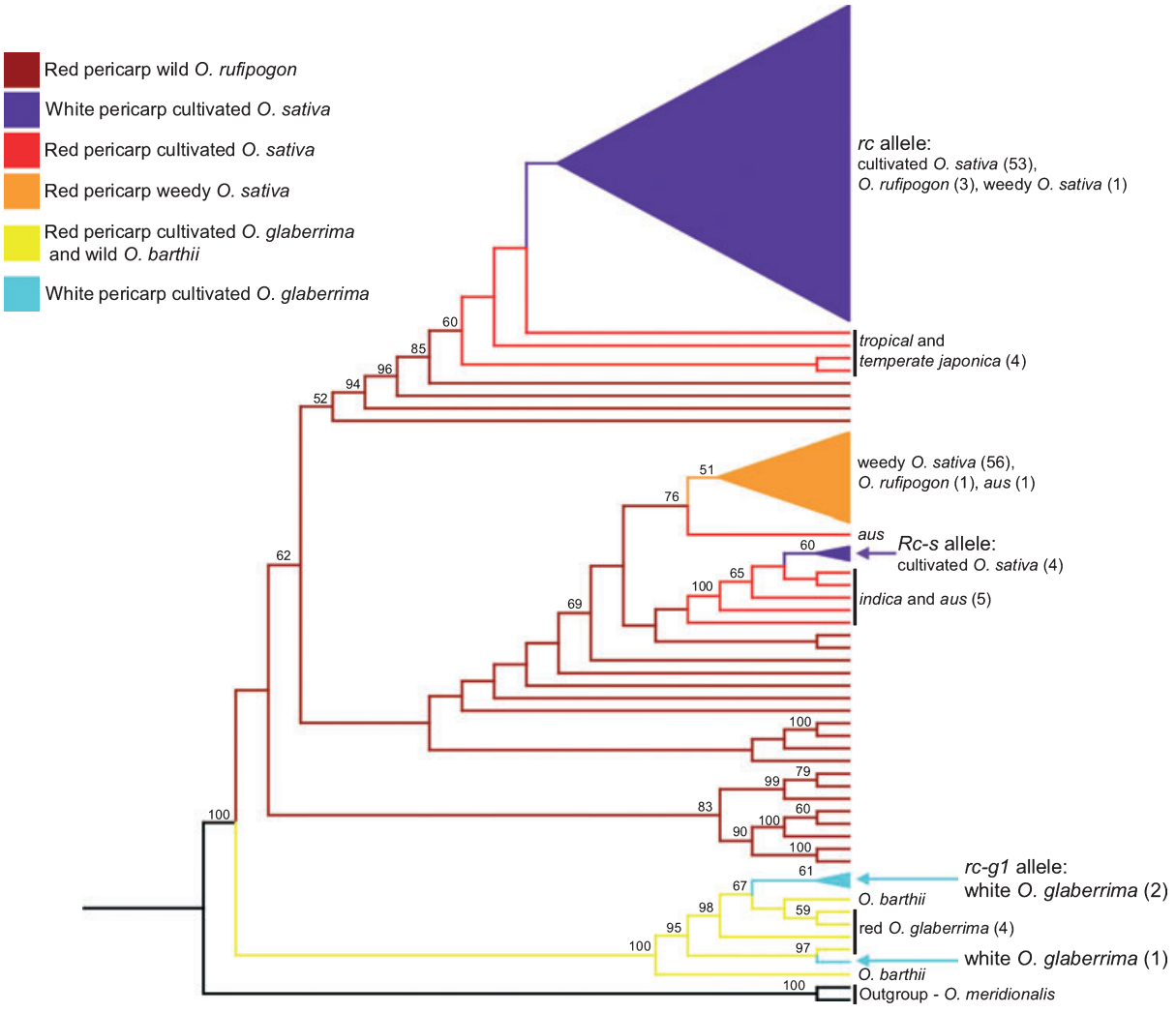

Fig. 2 Maximum likelihood (ML) tree of *Rc* haplotypes and all African rice samples. All sequences except *Oryza glaberrima* and *Oryza barthii* were collapsed to the smallest number of unique haplotypes to generate the ML tree. Haplotypes carrying similar alleles are grouped as triangles, with the size of the triangle roughly proportional to the number of unique haplotypes in the group. Colours indicate the identities of accessions carrying a haplotype, as indicated in the key. Numbers following labels indicate the number of individuals sharing a single haplotype or appearing in a group of similar haplotypes. Haplotypes represented by only one individual are labelled but not numbered, except for haplotypes unique to *Oryza rufipogon*, which are neither labelled nor numbered. Numbers at nodes represent percent bootstrap support; bootstrap values over 50% are shown. For simplicity, only *Oryza meridionalis* is shown as the outgroup.

for *O. rufipogon* and *O. sativa*. The two white pericarp *O. glaberrima* with *rc-g1* alleles group closely with each other, and the third white pericarp *O. glaberrima* groups closely with a red pericarp *O. glaberrima*.

Discussion

Origin and spread of white pericarps in African rice

Oryza sativa (Asian rice) and *O. glaberrima* (African rice) represent two phylogenetically close but evolutionarily independent domestication events. Comparison of these species thus provides an opportunity to examine whether human selection for domestication traits has targeted the same underlying genes in these two lineages. Our findings for the pericarp colour gene, *Rc*, indicate that remarkably similar genetic mechanisms can be at play. We find that two of the three white pericarp *O. glaberrima* varieties in our sample harbour a unique point mutation that is predicted to result in a premature stop codon in exon 7 of the *Rc* gene (*rc-g1*; Fig. 1); this newly described mutation occurs in the same gene and even the same exon as the mutations characterizing the *rc* and *Rc-s* white pericarp alleles in *O. sativa* (Fig. 1) (Sweeney *et al.*,

2006, 2007). Thus, at least in some cases, human selection may be acting on nearly identical molecular genetic changes in separate domesticated species.

One white pericarp *O. glaberrima* sample showed no differences in the coding region when compared to the red pericarp African rice samples, nor did it show any differences from the sequence of the most closely related red pericarp *O. glaberrima* in the sequenced upstream region or first intron, both of which are potential regulatory regions. In this case, it is possible that mutations at other loci, or potentially undetected changes in the *cis*-regulatory region of the *Rc* gene, are responsible for the accession's white pericarps. Interestingly, we find no evidence that the white pericarp phenotype in *O. glaberrima* reflects hybridization and introgression from *O. sativa* (Figs 1 and 2), despite evidence of hybridization and gene flow in previous surveys of African domesticates (Chu & Oka, 1970; Semon *et al.*, 2005). At the time that samples for this study were acquired, the International Rice Research Institute germplasm bank listed 1345 samples of *O. glaberrima* for which pericarp colour had been recorded, 48 of which were described as having white pericarps (roughly 3.6% of the samples). A thorough survey of these samples could potentially yield more evidence of *glaberrima*-specific mutations (either at *Rc* or other genes), or potentially provide evidence of *rc* or *Rc-s* introgression from *O. sativa*.

Our finding that all known mutations in the *Rc* gene are restricted to exon 7 suggests some constraint on how this gene can be rendered nonfunctional. That is, premature stop codons occurring near the beginning of the gene have apparently not persisted in any cultivars, perhaps because there is some selective disadvantage to that type of mutation (assuming that mutations are not limiting in this system). It is possible that the *Rc* gene performs another function outside of proanthocyanidin biosynthesis that is interrupted if the gene product is truncated at an earlier point. This study also shows that white pericarps might be achieved by mutations in *cis*-regulatory regions of *Rc* or potentially mutations in other genes, a pattern that was not observed in an extensive survey of *O. sativa* cultivars (Sweeney *et al.*, 2007). Future studies of gene expression and an exploration of the other genes in the proanthocyanidin synthesis pathway are necessary to fully understand the basis of the white pericarp phenotype in this accession.

One major question that remains unanswered by this study is whether the white pericarps seen in some accessions of *O. glaberrima* were the result of intentional selection or simply neutral mutations that have persisted in some cultivars. White pericarps were selected for in both Asian rice and wheat, indicating that a lighter grain colour may be generally preferred by humans, perhaps because of the novelty of a new colour or because the lighter coloured pericarp was softer and thus was easier to cook or remove during milling (Sweeney *et al.*, 2007). It has also been suggested that white pericarps are simply a by-product of selection for lower dormancy in domesticated grains; several studies have indicated that darker pericarp (or testa) colour is associated with stronger dormancy (Gfeller & Svejda, 1960, Debeaujon *et al.*, 2000, Groos *et al.*, 2002), and QTL mapping in rice has shown that a dormancy locus is located in the genomic region where the *Rc* locus is found (Gu *et al.*, 2004, 2005, 2006). Of course, these explanations for selection on grain colour are, by necessity, somewhat speculative and may not apply to African rice. However, overall, the limited number of African accessions with white pericarps suggests that the trait was either unnoticed or not valued under domestication. Unfortunately, the very fact that the *rc-g1* alleles resulting in white pericarps are so rare in African rice precludes the use of population genetic tests of selection to evaluate patterns of selection.

Alternatively, it is possible that the white pericarps would have been favoured and spread across *O. glaberrima* domesticates had the process of domestication not been 'interrupted' by the introduction of *O. sativa* to western Africa in the 16th century (Linares, 2002). *Oryza glaberrima* was likely domesticated 2000–4000 years ago, whereas *O. sativa* is thought to have been domesticated between 10 000 and 12 000 years ago. The two domesticated species appear to have undergone selection for some of the same features early in the domestication process, such as higher yield and lower rates of shattering (also seen in other domesticated grains) (Morishima *et al.*, 1963; Katayama & Sumi, 1995). Therefore, it is possible that grain colour would have followed the same pattern had *O. glaberrima* continued as the major African rice crop for a longer period of time instead of being largely replaced by *O. sativa* (Linares, 2002). Although it is impossible to know with certainty when the white pericarp *O. glaberrima* samples first appeared, the patterns are consistent with a scenario wherein the trait arose only shortly before the introduction of white pericarp Asian rice, preventing the spread of the *rc-g1* allele. A similar scenario has been suggested to explain the restricted distribution of the *Rc-s* allele in *aus* varieties of *O. sativa* (Sweeney *et al.*, 2007). Thus, the distribution of both the African alleles and the *Rc-s* allele may reflect of the fact that a widely available white pericarp variety with the *rc* allele made careful breeding of the local white pericarp varieties redundant.

The genetic basis of repeated phenotypic evolution in crops

General rules for the genetic basis of repeated phenotypic evolution at the interspecific level seem to be correlated with whether the trait is involved in the initial domestication changes or in the subsequent improvement or diversification of traits among varieties of a crop (Gross & Olsen, 2010). QTLs controlling classical domestication traits such as reduced shattering and increased apical dominance appear to map to different genomic locations in separate species (Li & Gill, 2006; Sood *et al.*, 2009), and

the domestication genes that have been cloned to date have only a minor, if any, effect on the same traits in other species (Doust *et al.*, 2004). In contrast, phenotypes that contribute to diversification within a domesticated crop have occasionally been shown to be controlled by the same gene in several different species (e.g. grain amylose content) (Fukunaga *et al.*, 2002; Olsen & Purugganan, 2002; Kawase *et al.*, 2005; Fan *et al.*, 2008).

At the intraspecific level, repeated phenotypic evolution during domestication can be explored in two contexts; one is the repeated evolution of a trait in varieties resulting from a single domestication event, and the other is repeated evolution of a trait in varieties that result from multiple independent domestication events within the same species. In some cases, it has been possible to use complementation and mapping to show that a single phenotype is controlled by different loci in different varieties of a crop (Komatsuda *et al.*, 2004; Azhaguvel & Komatsuda, 2007). However, mapping and complementation are only effective if the loci or mutations controlling a phenotype can be separated by recombination events, which makes it difficult to determine whether a phenotype might be caused by different mutations at the same locus. Answering this question has only recently been made possible because of the availability of cloned and characterized domestication and improvement genes for several crops. Interestingly, sequencing surveys of worldwide samples have shown that some identical phenotypes are produced by alternative mutations at the same locus, at least in rice (Sweeney *et al.*, 2007; Yu *et al.*, 2008; Gross *et al.*, 2009; Kovach *et al.*, 2009; this study). Just as for repeated evolutionary events at the interspecific level, this mainly seems to be the case for genes contributing to improvement or diversification – in rice, for example, 5 of 11 cloned improvement or diversification genes harbour multiple alleles that can produce the same phenotype (Gross & Olsen, 2010). It is possible that repeated evolution of functionally similar diversification and improvement alleles reflects the fact that these phenotypes are more frequently caused by loss-of-function alleles (compared to alleles responsible for domestication traits), and such null alleles could be achieved through a variety of point mutations, insertions or deletions. On the other hand, this pattern might also be considered to be surprising, in that it should also be possible to 'break' pathways at multiple points (in the absence of strong constraint), rather than through repeated mutations at the same gene (Gross & Olsen, 2010).

To our knowledge, only the *Waxy* gene (which contributes to crop diversification in grain amylose content) shows patterns of repeated evolution that are comparable to those seen at the *Rc* gene (which is classified as both an improvement and domestication gene). The glutinous phenotype of domesticated foxtail millet is caused by several unique mutations at the *Waxy* gene (Fukunaga *et al.*, 2002; Kawase *et al.*, 2005), just as white pericarps can be caused by either the *rc* or *Rc-s* allele in Asian domesticated rice (Sweeney *et al.*, 2007). Looking across species, glutinous grains are achieved in several different species of domesticated cereals because of independent mutations in the *Waxy* gene (Fukunaga *et al.*, 2002; Olsen & Purugganan, 2002; Fan *et al.*, 2008), and independent mutations in the *Rc* gene are responsible for white grains in both African and Asian domesticated rice. Whether other genes will be shown to be important for controlling domestication-related traits, both within species and between species, remains to be determined. However, the relative importance of both *Rc* and *Waxy* appears to be the exception rather than the rule at this point, given that so many domestication genes are uniquely important for each species (Gross & Olsen, 2010).

Acknowledgments

We thank members of the Olsen lab for constructive criticism on an earlier version of this manuscript. B.L.G. was supported by a National Institutes of Health Ruth L. Kirschstein Postdoctoral Fellowship (1F32GM082165). F.T.S. was supported by the Biology Summer Scholars Program funded by an Undergraduate Biological Sciences Education Program grant from the Howard Hughes Medical Institute to Washington University in St. Louis. K.M.O. was supported by a National Science Foundation PGRP award (DBI-0638820).

References

Azhaguvel, P. & Komatsuda, T. 2007. A phylogenetic analysis based on nucleotide sequence of a marker linked to the brittle rachis locus indicates a diphyletic origin of barley. *Ann. Bot.* **100**: 1009–1015.

Burger, J.C., Chapman, M.A. & Burke, J.M. 2008. Molecular insights into the evolution of crop plants. *Am. J. Bot.* **95**: 113–122.

Burke, J.M., Burger, J.C. & Chapman, M.A. 2007. Crop evolution: from genetics to genomics. *Curr. Opin. Genet. Dev.* **17**: 525–532.

Caicedo, A., Williamson, S., Hernandez, R.D., Boyko, A., Fledel-Alon, A., York, T., Polato, N., Olsen, K., Nielsen, R., McCouch, S.R., Bustamante, C.D. & Purugganan, M.D. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* **3**: e163.

Chu, Y.-E. & Oka, H.-I. 1970. Introgression across isolating barriers in wild and cultivated *Oryza* species. *Evolution* **24**: 344–355.

Debeaujon, I., Leon-Kloosterziel, K.M. & Koornneef, M. 2000. Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Phys.* **122**: 403–414.

Doebley, J.F., Gaut, B.S. & Smith, B.D. 2006. The molecular genetics of crop domestication. *Cell* **127**: 1309–1321.

Doust, A.N., Devos, K.M., Gadberry, M.D., Gale, M.D. & Kellogg, E.A. 2004. Genetic control of branching in foxtail millet. *Proc. Natl Acad. Sci. USA* **101**: 9045–9050.

Fan, L., Quan, L., Leng, X., Guo, X., Hu, W., Ruan, S., Ma, H. & Zeng, M. 2008. Molecular evidence for post-domestication selection in the *Waxy* gene of Chinese waxy maize. *Mol. Breed.* **22**: 329–338.

Felsenstein, J. 2005. *PHYLIP (Phylogeny Inference Package) version 3.6*. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle.

Fukunaga, K., Kawase, M. & Kato, K. 2002. Structural variation in the *Waxy* gene and differentiation in foxtail millet [*Setaria italica* (L.) P. Beauv.]: implications for multiple origins of the waxy phenotype. *Mol. Genet. Genomics* **268**: 214–222.

Gfeller, F. & Svejda, F. 1960. Ineritance of post-harvest seed dormancy and kernel colour in spring wheat lines. *Can. J. Plant Sci.* **40**: 1–6.

Glémin, S. & Bataillon, T. 2009. A comparative view of the evolution of grasses under domestication. *New Phytol.* **183**: 273–290.

Groos, C., Gay, G., Perretant, M.R., Gervais, L., Bernard, M., Dedryver, F. & Charmet, G. 2002. Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a whitexred grain bread-wheat cross. *Theor. Appl. Genet.* **104**: 39–47.

Gross, B.L. & Olsen, K.M. 2010. Genetic perspectives on crop domestication. *Trends Plant Sci.* **15**: 529–537.

Gross, B.L., Skare, K.J. & Olsen, K.M. 2009. Novel *Phr1* mutations and the evolution of phenol reaction variation in US weedy rice (*Oryza sativa*). *New Phytol.* **184**: 842–850.

Gross, B.L., Reagon, M., Hsu, S.-C., Caicedo, A.L., Jia, Y. & Olsen, K.M. 2010. Seeing red: the origin of grain pigmentation in US weedy rice. *Mol. Ecol.* **19**: 3380–3393.

Gu, X.-Y., Kianian, S.F. & Foley, M.E. 2004. Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). *Genetics* **166**: 1503–1516.

Gu, X.-Y., Kianian, S.F., Hareland, G.A., Hoffer, B.L. & Foley, M.E. 2005. Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). *Theor. Appl. Genet.* **110**: 1108–1118.

Gu, X.Y., Kianian, S.F. & Foley, M.E. 2006. Isolation of three dormancy QTLs as Mendelian factors in rice. *Heredity* **96**: 93–99.

Katayama, T.C. & Sumi, A. 1995. Studies on agronomic traits of African rice (*Oryza glaberrima* Steud.). *Jpn. J. Crop Sci.* **64**: 807–814.

Kawase, M., Fukunaga, K. & Kato, K. 2005. Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol. Genet. Genomics* **274**: 131–140.

Komatsuda, T., Maxim, P., Senthil, N. & Mano, Y. 2004. High-density AFLP map of nonbrittle rachis 1 (*btr1*) and 2 (*btr2*) genes in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **109**: 986–995.

Kovach, M.J., Calingacion, M.N., Fitzgerald, M.A. & McCouch, S.R. 2009. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proc. Natl Acad. Sci. USA* **106**: 14444–14449.

Li, W. & Gill, B. 2006. Multiple genetic pathways for seed shattering in the grasses. *Funct. Integr. Genomics* **6**: 300–309.

Linares, O.F. 2002. African rice (*Oryza glaberrima*): history and future potential. *Proc. Natl Acad. Sci. USA* **99**: 16360–16365.

Londo, J.P., Chiang, Y.-C., Hung, K.-H., Chiang, T.-Y. & Schaal, B.A. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestication of cultivated rice *Oryza sativa*. *Proc. Natl Acad. Sci. USA* **103**: 9578–9583.

Morishima, H., Hinata, K. & Oka, H.-I. 1963. Comparison of modes of evolution of cultivated forms from two wild rice species, *Oryza breviligulata* and *O. perennis*. *Evolution* **17**: 170–181.

Olsen, K.M. & Purugganan, M.D. 2002. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* **162**: 941–950.

Paterson, A.H., Lin, Y.-R., Li, Z., Schertz, K.F., Doebley, J.F., Pinson, S.R.M., Liu, S.-C., Stansel, J.W. & Irvine, J.E. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**: 1714–1718.

Purugganan, M.D. & Fuller, D.Q. 2009. The nature of selection during plant domestication. *Nature* **457**: 843–848.

Semon, M., Nielsen, R., Jones, M.P. & McCouch, S. 2005. The population structure of African cultivated rice *Oryza glaberrima* (Stued.): evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. *Genetics* **169**: 1639–1647.

Sood, S., Kuraparthy, V., Bai, G. & Gill, B. 2009. The major threshability genes soft glume (*sog*) and tenacious glume (*Tg*), of diploid and polyploid wheat, trace their origin to independent mutations at non-orthologous loci. *Theor. Appl. Genet.* **119**: 341–351.

Stamatakis, A. (2006a) Phylogenetic models of rate heterogeneity: a high performance computing perspective. In: *Proceedings of the IPDP2006*, Rhodos, Greece.

Stamatakis, A. 2006b. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Stamatakis, A., Hoover, P. & Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**: 758–771.

Sweeney, M. & McCouch, S. 2007. The complex history of the domestication of rice. *Ann. Bot.* **100**: 951–957.

Sweeney, M.T., Thomson, M.J., Pfeil, B.E. & McCouch, S. 2006. Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**: 283–294.

Sweeney, M.T., Thomson, M.J., Cho, Y.G., Park, Y.J., Williamson, S.H., Bustamante, C.D. & McCouch, S.R. 2007. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet.* **3**: e133.

Yu, Y., Tang, T., Qian, Q., Wang, Y., Yan, M., Zeng, D., Han, B., Wu, C.-I., Shi, S. & Li, J. 2008. Independent losses of function in a polyphenol oxidase in rice: differentiation in grain discoloration between subspecies and the role of positive selection under domestication. *Plant Cell* **20**: 2946–2959.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Accessions included in the study.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 26 May 2010; revised 15 July 2010; accepted 31 August 2010