A peer-reviewed version of this preprint was published in PeerJ on 7 April 2016.

Fuchs EJ, Meneses Martínez A, Calvo A, Muñoz M, Arrieta-Espinoza G. 2016. Genetic diversity in *Oryza glumaepatula* wild rice populations in Costa Rica and possible gene flow from *O. sativa*. PeerJ 4:e1875 https://doi.org/10.7717/peerj.1875

Genetic structure of *Oryza glumaepatula* wild rice populations and evidence of introgression from *O. sativa* in Costa Rica

Eric Fuchs, Allan Meneses Martínez, Amanda Calvo A, Melania Muñoz, Griselda Arrieta-Espinoza

Wild crop relatives are an important source of genetic diversity for crop improvement. However, gene flow from cultivated species into wild species may prove detrimental. Introgression may lead to changes in wild species by incorporating alleles from domesticated species, which may increase the likelihood of extinction. The objective of the present study is to analyze how genetic diversity is distributed within and among populations of the wild rice species *Oryza glumaepatula* in Costa Rica. We also evaluated if there is evidence of introgression between wild rice and commercial varieties of *O. sativa* since it is cultivated commonly in close proximity to wild rice populations. Individuals from all known *O. glumaepatula* populations in Costa Rica were collected. With the aid of 455 AFLP markers, we characterized the genetic diversity and structure among seven populations in northern Costa Rica. Given the dominant nature of our markers, Bayesian estimates of genetic structure were used. We also compared genetic diversity estimates between *O. glumaepatula* individuals and *O. sativa* commercial rice. Our results show that *O. glumaepatula* populations in Costa Rica have moderately high levels of genetic diversity, comparable to those found in South American populations. This is likely a result of large population size. Despite the restricted distributions of this wild species, in Costa Rica most populations are composed of several thousand individuals, thus reducing the effects of drift on genetic diversity. Our results also found low but significant structure (\theta=0.03±0.001) among populations that are separated by ~10 Km within a single river. The position of the population along the river did not influence genetic diversity estimates or differences among populations. This river does not have a strong current and meadows or seeds may easily move upstream, thus homogenizing genetic diversity across populations regardless of river position. Ample gene flow through pollen, seeds or detached culms within the same river reduces genetic structure. A Bayesian structure analysis showed that individuals from two populations share a significant proportion of their genomes with *O. sativa* genome. These results suggest that the low levels of genetic structure found in these populations are likely the result of introgression from cultivated *O. sativa* populations. These results expose an important biohazard as recurrent hybridization may reduce genetic diversity of this wild rice species. Introgression may transfer

commercial traits into the only populations of *O. glumaepatula* in Costa Rica, which in turn could alter genetic diversity and increase the likelihood of local extinction. These results have important implications for *in situ* conservation strategies of the only wild populations of *O. glumaepatula* in Costa Rica.

Genetic structure of *Oryza glumaepatula* wild rice populations and evidence of introgression from *O. sativa* in Costa Rica

Eric J. Fuchs. Escuela de Biología, Universidad de Costa Rica. 2060 San José, Costa Rica.
Allan Meneses Martínez. Centro de Investigación en Biología Celular y Molecular, Universidad de Costa Rica, 2060 San José, Costa Rica.
Amanda Calvo. Centro de Investigación en Biología Celular y Molecular, Universidad de Costa Rica, 2060 San José, Costa Rica.
Melania Muñoz. Centro de Investigación en Biología Celular y Molecular, Universidad de Costa Rica, 2060 San José, Costa Rica.
Griselda Arrieta- Espinoza. Centro de Biología Celular y Molecular, Universidad de Costa Rica, 2060 San José, Costa Rica.

Corresponding author: Griselda Arrieta-Espinoza. Centro de Biología Celular y Molecular, Universidad de Costa Rica, 2060 San José, Costa Rica.
griselda.arrieta@ucr.ac.cr

Abstract

Wild crop relatives are an important source of genetic diversity for crop improvement. However, gene flow from cultivated species into wild species may prove detrimental. Introgression may lead to changes in wild species by incorporating alleles from domesticated species, which may increase the likelihood of extinction. The objective of the present study is to analyze how genetic diversity is distributed within and among populations of the wild rice species *Oryza glumaepatula* in Costa Rica. We also evaluated if there is evidence of introgression between wild rice and commercial varieties of *O. sativa* since it is cultivated commonly in close proximity to wild rice populations. Individuals from all known *O. glumaepatula* populations in Costa Rica were collected. With the aid of 444 AFLP markers, we characterized the genetic diversity and structure among seven populations in northern Costa Rica. Given the dominant nature of our markers, Bayesian estimates of genetic structure were used. We also compared genetic

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

diversity estimates between *O. glumaepatula* individuals and *O. sativa* commercial rice. Our results show that *O. glumaepatula* populations in Costa Rica have moderately high levels of genetic diversity, comparable to those found in South American populations. This is likely a result of large population size. Despite the restricted distributions of this wild species, in Costa Rica most populations are composed of several thousand individuals, thus reducing the effects of drift on genetic diversity. Our results also found low but significant structure (θ=0.03±0.001) among populations that are separated by ~10 Km within a single river. The position of the population along the river did not influence genetic diversity estimates or differences among populations. This river does not have a strong current and meadows or seeds may easily move upstream, thus homogenizing genetic diversity across populations regardless of river position. Ample gene flow through pollen, seeds or detached culms within the same river reduces genetic structure. A Bayesian structure analysis showed that individuals from two populations share a significant proportion of their genomes with *O. sativa* genome. These results suggest that the low levels of genetic structure found in these populations are likely the result of introgression from cultivated *O. sativa* populations. These results expose an important biohazard as recurrent hybridization may reduce genetic diversity of this wild rice species. Introgression may transfer commercial traits into the only populations of *O. glumaepatula* in Costa Rica, which in turn could alter genetic diversity and increase the likelihood of local extinction. These results have important implications for *in situ* conservation strategies of the only wild populations of *O. glumaepatula* in Costa Rica.

Keywords:

Crop Wild Relatives, Gene flow, hybridization, biosafety, AFLP, Guanacaste, Medio Queso

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

Introduction

Crop wild or weedy relatives (CWR) are considered an important source of genetic diversity for modern agriculture (Hajjar & Hodgkin, 2007). The genetic variability contained in these wild populations is highly important for breeding programs and genetic crop improvement (Brondani et al., 2005). For instance, introgression of specific traits from wild to cultivated varieties may result in novel gene combinations which may increase the productivity of the crop, however introgression in the opposite direction may be detrimental for wild species (Song et al., 2003). Conservation of wild populations should be a priority for preserving these genetic resources for future breeding programs.

There are 21 identified wild rice species, six of them are diploid with the AA genome which makes hybridization possible with commercial *O. sativa* (Khush, 1997; Vaughan, Morishima & Kadowaki, 2003). These wild species have proven to be important sources of novel and commercially traits. Interspecific crosses between wild relatives *O. rufipogon*, *O. longistaminanta* and cultivated *O. sativa* resulted in the introgression of traits that improve tolerance to acid soils, drought and increase yield (Brar, 2005; Hajjar & Hodgkin, 2007) which have allowed the cultivation of rice in former unsuitable areas.

The name *O. glumaepatula* was originally assigned to a cultivated rice species from Suriname, and then considered *Oryza rufipogon*, due to the lack of consistent morphological differences between both species (Tateoka, 1962). However, sterility barriers between Asian wild species and *O. glumaepatula* confirmed its species status (Juliano, Naredo & Jackson, 1998) and its difference from *O. rufipogon* (Ren et al., 2003). *Oryza glumaepatula* is the only diploid ($A^{gp}A^{gp}$) native rice species in America. *O. glumaepatula* hybridizes with *O. sativa* and genes related to tillering and panicle size have been successfully bred into commercial rice (Brondani et al., 2002) which supports its potential as a source of novel genetic variability that may be used to improve commercial varieties in future breeding programs.

Genetic diversity in *O. glumaepatula* has been primarily studied in South American (Brazil) populations (Karasawa et al., 2007b). In Central America populations of this species are naturally and antropogenically fragmented since this species depends on specific habitats (rivers, wetlands and inundated areas), that have extensively been destroyed for changes in land use (Vaughan et al., 2005). Some of these marginal populations may be of special interest as they are small and likely with limited dispersal. Populations with these conditions are generally characterized by lower levels of genetic

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

diversity within population and significant genetic structure among them (Lesica & Allendorf, 1995), which increase their risk of extinction due to genetic or ecological factors (Newman & Pilson, 1997). These small, peripheral populations may often contain important or novel and exclusive genetic traits not present anywhere else.

Oryza glumaepatula is predominantly selfing (Ge et al., 1999; Karasawa et al., 2007a), however variation in selfing rates across populations suggest a mixed mating system (Goodwillie, Kalisz & Eckert, 2005). Autogamous mating is likely to reduce gene flow among populations resulting in a highly structured genetic diversity. Consistently, previous studies in *O. glumaepatula* found that a larger proportion of genetic diversity was structured among populations, a likely effect of limited gene flow (Akimoto, Shimamoto & Morishima, 1998; Buso, Rangel & Ferreira, 1998; Karasawa et al., 2007b; Veasey et al., 2008, 2011). However, these studies also show significant spatial and temporal variability in within population estimates of genetic diversity, and high gene flow among populations (Akimoto, Shimamoto & Morishima, 1998) which suggests that seeds could play an important role in gene flow patterns. As a hydrochorous species, seed dispersal through waterways or river systems is likely to contribute significantly to homogenize diversity of *O. glumaepatula* either among populations connected by river tributaries. Seed dispersal via animals has also been documented, however is likely to represent a negligible proportion of seed dispersal. As a mixed mating species, gene flow through pollen dispersal is also likely to contribute to the observed levels of genetic diversity.

Gene flow between related interfertile species may result in introgression of genes (Chu & Oka, 1970). Introgressive hybridization is thought to have played an important role in the evolution of plants (Grant, 1981; Arnold, 1997) by introducing novel genetic variation which allowed species to exploit new adaptive landscapes (Ellstrand, Prentice & Hancock, 1999). However, introgression often increases genetic differences between wild and introgressed populations (Jiang et al., 2012), and may pose a threat if hybridization reduces local adaptation (outbreeding depression), increases the risk of extinction through reduction of autochthonous genetic diversity (Ellstrand, Prentice & Hancock, 1999). Introgression from cultivated rice into wild or weedy rice species has been documented in controlled and natural settings (Song et al., 2003; Chen et al., 2004), and though introgression may lead to a short-term increase in genetic diversity within populations, original or autochthonous diversity is often lost in the process (Lu, 2013). Estimation of

introgression in *O. glumaepatula* is highly relevant for biosafety regulations. Gene flow from cultivated species into wild relatives may also pose a biosafety threat if transgenes from GM crops are able to move via gene flow into wild populations (Snow, 2002; Ellstrand, 2003).

The aim of this study is to describe the distribution of genetic diversity within and among Costa Rican populations of the wild rice species *Oryza glumaepatula*, and determine if populations within a single river are genetically structured and if river direction plays a role in the magnitude of genetic diversity. Finally, we want to determine if there is evidence of introgression between wild rice and cultivated varieties of *O. sativa*, as commercial rice species are planted in the vicinity of wild rice populations. These results have important implications for *in situ* conservation strategies of the only wild populations of *O. glumaepatula* in Costa Rica.

Methods

Study Species

Oryza glumaepatula L. is a wild rice species distributed throughout Central, South America, and the Caribbean (Vaughan, Morishima & Kadowaki, 2003). Detailed morphological descriptions of *Oryza* wild species in Costa Rica are found in (Zamora et al., 2003). *O. glumaepatula* is found in flooded areas, marshes, rivers or other wetlands, and prefers clay or loam soils. It is a perennial, tufted and scrambling grass with a brittle culm near the base of plants. Culms may detach and float downstream creating new populations. Flowering occurs between October and November, immediately followed by a brief (2-3 weeks) fruiting episode. This species is likely wind pollinated (anemophily) however, previous reports (Ge et al. 1999 reviewed in Karasawa 2007a) suggest that autogamous pollination is common. Propagule dispersal occurs via water (hydrochory), with seeds and vegetative propagules being carried by streams and rivers. The amount of vegetative reproduction is still not yet quantified.

Study Sites

We analyzed all known population of *O. glumaepatula* in Costa Rica, which were located in two geographic sites. The first site is in northwestern Costa Rica (10º57’03.3” N / 85º36’59.2” W) in a seasonal wetland placed in a farm along the road to Murcielago

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

sector of the Guanacaste Conservation Area in the Guanacaste Province (Figure 1). During the rainy season this area is flooded allowing the growth of *O. glumaepatula* forming a small population with a few individuals (n < 100) in an area of less than 1 Ha. The second site is located in the Medio Queso wetland (MQ) in northeastern Costa Rica (11º01'34.1" N / 84º40'42.8" W)(Figure 1). This site is a palustrine wetland of 5000 ha, irrigated by the Medio Queso River (Jimenez, 2004). This wetland is about 10-20 km in length with abundant gramineous vegetation throughout, and some tree species (e.g., *Pachira aquatic*). Along riverbanks farms with livestock production, maize, beans and rice plantations are common. *O. glumaepatula* is commonly found in MQ with large patches of individuals on both banks of the river and in multiple sites along the MQ wetland. Populations along the river are unconnected patches of > 100 individuals along the 10km river. Separation between populations may increase during the rainy season when river volume increases (Jimenez, 2004).

Sampling

Guanacaste population has less than a 100 *O. glumaepatula* individuals located in a few separated patches along a marsh. Therefore, we were only able to collect 30 individuals. In MQ, we sampled populations along the Medio Queso River. A population was defined as a continuous meadow with a few hundred individuals and clearly separated from other populations by at least 100 meters. Within populations we sampled at least 10 individuals that were separated by 10-20 m from each other, in order to avoid collecting genets. A boat was used to collect individuals and care was taken to collect the entire individual (including the rhizome). Sites were geo-referenced using a GPS. We transplanted individuals to a greenhouse at Universidad de Costa Rica and plants were kept alive for further analyses. A total of 120 and 30 individuals were collected at MQ and Guanacaste respectively. Due to mortality in the greenhouse only 79 individuals were analyzed (Table1). Additionally, six individuals of cultivated rice varieties (*O. sativa*) were also analyzed. Seeds were donated by Centro para Investigaciones en Granos y Semillas (CIGRAS) from University of Costa Rica. Comisión Institucional de Biodiversidad issued permit VI-3150-2010 to Griselda Arrieta-Espinoza authorizing field collections.

DNA extraction and AFLP genotyping

Flag leafs from plants were used for DNA extraction. Total DNA was extracted from 50 to 150 mg of dry leaf material or from 100 to 250 mg of fresh tissue. The flag leaf was collected, dried and grinded using the FastPrep tissue grinder following the manufacturer instructions. Genomic DNA was extracted using the FastDNA® (MP Biomedicals, CA, USA) kit protocol. DNA quantity and quality were determined by gel electrophoresis and via Nanodrop quantification.

Individual genotypes were determined using Amplified Fragment Length Polymorphism (AFLP) following (Vos et al., 1995) with modifications indicated by the AFLP kit provided by Applied Biosystems. Based on the AFLP Plant Mapping Kit for small genomes (Applied Biosystems, Inc., Foster City CA) and the AFLP™ Plant Mapping Protocol 50 ng of genomic DNA were digested with EcoRI and MseI (New England Biolabs, Inc., Ipswich, MA) at 37 ºC overnight. Then double-stranded adaptors were ligated to the ends of DNA fragments, generating template DNA for subsequent PCR amplifications. After incubation, each sample was diluted 20-fold with TE buffer. The ligated adaptors served as primer binding sites for low-level selection in pre-selective amplification of the restriction fragments. Pre-selective amplification mixture was prepared by adding 4 µl of 20-fold diluted DNA from the restriction-ligation reaction, 1µl AFLP pre-selective primers (Applied Biosystems), and 15 µl AFLP core mix. The pre-amplified DNA was diluted again 20-fold with low TE buffer and polymerase chain reactions (PCR) for selective amplification were carried out in 10 µl volume using different fluorescence-labeled EcoRI-primers and MseI primer combinations.

Selective amplification standardization: Thirty-two of the 64 available combinations were randomly used to determinate which primers generated the most polymorphic peaks in samples of *O. glumaepatula* and aDNA control from AB kit, water was used as negative control. The selective amplification of the samples was carried out using these primer combinations twice in independent experiments and in different thermal cycler (Hybaid and GeneAmp 9700) to assure reproducibility. Based on this data, eight combinations E-TC/M-CTC, E-TA/M-CAA, E-AT/M-CAC, E-TG/M-CTA, E-AC/M-CTG, E-AG/M-CAT, E-AA/M-CTC and E-TT/CAC were chosen for this study.

AFLPs band analysis: Amplified fragments were sized on an ABI 3100 sequencer and band scoring was conducted using GeneMarker v 9.1 software. Only AFLP peaks between 150 and 500 bp were considered. The same individual was scored at least three

times, to determine false positives. We only kept bands that had a consistent intensity (intensity higher than 500) that were reproducible and were separated by at least 2 bp from each other. Bands were scored manually as 1 (present) or 0 if absent for each fragment within individuals. Two different analysts scored the electropherograms and the consensus between both was recorded as the multilocus genotype for each individual. We combined data from all three marker combinations (primer and selective nucleotide combination) for subsequent diversity analyses.

Statistical analyses

For comparison purposes band frequencies and the Shannon diversity index were estimated using GenAlExv 6.5 (Peakall & Smouse, 2006). Bayesian estimates of intrapopulation heterozygosity were obtained using AFLPSurv (Vekemans, 2002) with non-uniform priors and Fis=0.7 as estimated by Karasawa et al. (2007a). A Spearman correlation analysis was used to determine if genetic diversity was associated with the position of the population on the river, given a value of 0 to the population in the uppermost part of the river and subsequent numbers to populations downriver.

Genetic structure and inbreeding coefficients were also estimated using the program Hickory v1.1 (Holsinger, Lewis & Dey, 2002). Lack of HWE may bias traditional diversity estimates in dominant markers (Lynch & Milligan, 1994). To this end we selected Hickory as it uses a Bayesian approach to estimate these parameters without assuming HWE within populations and takes advantage of all the information provided by dominant markers (Holsinger & Wallace, 2004). In all cases we used the default priors, burnin values and number of MCMC chains. Runs with twice as many simulations yielded comparable results (data not shown). We compared the full model to a model were the inbreeding coefficient was fixed to $f = 0$ and a model where $f = 0$ and genetic diversity is not structured among populations $\theta^{II} = 0$. A model where Hickory chooses a random *f*-value from the posterior distribution (i.e., *f*-free model), while estimating other parameters was also tested. This last model (f-free model) may circumvent the bias introduce in *f* estimates from small samples and dominant markers (Holsinger, Lewis & Dey, 2002)(Hickory Manual). The best model was chosen based on Deviance Information Criterion (DIC). The model with the smallest DIC value was chosen as the most appropriate model. Models that differed by less than five DIC units were considered comparable. We ran Hickory on two different data sets. The first data set included *O.*

glumaepatula populations in MQ and in Guanacaste. A second data set included all *O. glumaepatula* populations and a sample of 10 *O. sativa* individuals. This second data set allowed us to determine genetic structure among cultivated and wild species.

Genetic structure was also estimated using Bayesian clustering algorithms implemented in the programs Structure (Pritchard, Stephens & Donnelly, 2000; Falush, Stephens & Pritchard, 2007) and Geneland (Guillot, Mortier & Estoup, 2005). Structure and Geneland both assign individuals to clusters that minimize HWE disequilibrium and linkage disequilibrium, regardless of the population were they were collected. Geneland uses a similar structuring algorithm as Structure, however Geneland also uses geographic information as *a priori* information to assist in cluster definition. This type of clustering may be more realistic as individual position may aid in cluster identification. In both Structure and Geneland we estimated cluster number using only MQ populations. A second analysis included all *O. glumaepatula* populations from MQ and the geographically separated Guanacaste. A final analysis included all *O. glumaepatula* populations and a sample of six *O. sativa* individuals to determine if AFLP markers were able to distinguish between commercial and wild species.

In Structure we selected the admixture model with correlated allele frequencies to estimate the most likely number of clusters. We discarded the first 20 000 iterations as a burnin, while preserving 50 000 iterations for cluster estimation. Previously we had determined that 25000 iterations (10 000 burnin) were sufficient for stable estimates of Q and α. Multiple runs of Structure were conducted to test for the most likely number of possible clusters (K). We changed the number of clusters from K=1 to K=10 and twenty replications were conducted for each K value. The most likely number of clusters was inferred using Structure harvester (Evanno, Regnaut & Goudet, 2005). We also performed a visual inspection of Q value bar graphs, to assess the likelihood of a single cluster. Once the number of clusters was found, we used CLUMPP to determine the most likely configuration of admixture for graphs (Jakobsson & Rosenberg, 2007). CLUMPP's default settings and 10 000 chains were selected. Admixture bar graphs were created using R (R Development Core Team, 2012).

For Geneland we used correlated allele frequencies and performed 100 000 iterations, recorded data every 100 iterations (thining=100) and discarded the initial 200 iterations (burnin). Increasing iteration numbers or burnin values tenfold, did not affect the outcome of the analysis. The number of clusters was inferred from the modal number of

clusters from the posterior distribution which is regarded as the maximum a posteriori estimate of K (Guillot, Mortier & Estoup, 2005). Twenty different runs were performed of each model, to determine if the MCMC was converging to the same value in all runs. We also conducted an analysis without spatial information that included *O. sativa* individuals. This last analysis is comparable to analyses performed in Structure, which do not use spatial information for cluster identification.

Results

We analyzed a total of 79 individuals of *O. glumaepatula* from seven populations, six populations in MQ and one population in Guanacaste. We also resolved AFLP bands on six cultivated *O. sativa* individuals. A total of 444 AFLP bands were scored in all individuals. Genetic diversity parameters for *O. glumaepatula* populations are shown in Table 1. All pooled populations had 84.2% polymorphic loci. Heterozygosity estimates were similar across populations and were not associated with river position ($r < 0.05$, $p > 0.05$). Populations located on opposite sides of the Medio Queso River had comparable levels of genetic diversity. Guanacaste was the smallest and most isolated population, but our results did not suggest that isolation or population size affected genetic diversity estimates. Both Shannon Indexes and Bayesian estimates of heterozygosity were similar for Guanacaste and all other populations (Table 1). If heterozygosity is estimated assuming HWE all estimates are similar and lower than He=0.09 (data not shown). Bayesian estimates of heterozygosity from AFLPSurv were higher and are likely to be better estimates because they incorporate the species breeding system (Vekemans, 2002; Vekemans et al., 2002).

In Hickory a model that included inbreeding ($f \neq 0$) had a better fit to the data (DIC=6758.65) than a model without inbreeding (DIC=6809.64), suggesting a significant departure from HWE. Average inbreeding was f=0.61 ±0.42 (CI: 0.06-0.99) across *O. glumaepatula* populations (including Guanacaste). This expected deviation from HWE suggests a significant heterozygote deficit in all populations. Similar results were obtained for MQ populations and when MQ and Guanacaste were pooled in a single data set. All of our analyses suggest that populations experience significant inbreeding, consistent with the mating strategy of this species (tm=0.116) (Karasawa et al., 2007a). Genetic diversity estimates of commercial *O. sativa* individuals consistently showed low

levels of genetic diversity with I=0.135, Hj=0.08 and P%=34.34; as expected for a commercial variety.

For *O. glumaepatula* populations in MQ and Guanacaste we found that a model that included genetic structure ($\theta \neq 0$) described the data better (DIC=6758.65) than a model without genetic structure (DIC=7143.2). Populations in MQ and in northern Costa Rica exhibited a low but significant genetic structure among them ($\theta = 0.03 \pm 0.001$). If Guanacaste population is excluded from the model, genetic structure estimates did not differ, suggesting that the average differences in allele frequencies among populations in MQ are similar to differences in allele frequencies between MQ and Guanacaste populations. When commercial *O. sativa* individuals are included in the analysis, as expected, there is a slight increase in genetic structure ($\theta = 0.04 \pm 0.001$), suggesting differences in allele frequencies among wild and cultivated species.

Bayesian structuring algorithms found little evidence of structure among populations in *O. glumaepatula* populations in MQ. Both Structure and Geneland grouped populations into a single cluster (K=1) as the most likely result. Including the geographic location of populations to aid in cluster identification (i.e., Geneland) did not change the outcome. When Guanacaste was included into the analysis, both Structure and Geneland concurrently suggest that the most likely number of clusters is K=2 (Figure 2). A detailed analysis of admixture values shows that two populations: El Muro and Guanacaste populations (Figure 1) are grouped separately from other populations, and show admixture levels that are genetically distinct from other MQ populations (Figure 2). A few individuals in the remaining populations are clustered in the same group as El Muro and Guanacaste suggesting that individuals in other populations may be the result of migration or gene flow originating from individuals in Guanacaste or El Muro. These results are congruent with results from Hickory, which found a small but significant structure among populations independent of the presence of Guanacaste in the analysis. This structure is likely caused by the differences in allele frequencies detected between Muro and Guanacaste and the rest of the MQ populations.

When commercial *O. sativa* individuals were included in the sample, STRUCTURE and Geneland both resulted on K=2 as the most likely number of clusters. However in this case, individuals from the commercial species are assigned to one cluster, while the other cluster is composed of *O. glumaepatula* populations. These results suggest that one cluster represents the *O. sativa* genome while the second cluster likely represents the wild

rice genome. Structure results also indicate that *O. glumaepatula* individuals from El Muro and Guanacaste shared a significant proportion of their genomes with *O. sativa* (Figure 2). Individuals in these two *O. glumaepatula* populations are admixed with commercial rice varieties. Introgression from cultivated rice into its wild ancestor appears to cause genetic structure among sympatric *O. glumaepatula* populations.

Discussion

We present the first assessment of genetic structure in populations of *O. glumaepatula* in Costa Rica and in Mesoamerica. Our results suggest intermediate levels of genetic diversity, significant evidence of inbreeding, as well as a lack of structure among populations. This differs from expectations and previous findings in South American populations, as we only found small differences among populations separated by more than100 km. Results also suggest that introgression from cultivated rice is the predominant force driving genetic differences among Costa Rican populations of *O. glumaepatula*. These results are of great importance for this CWR as it may be in danger of losing its genetic diversity through introgressive hybridization.

Previous studies on *O. glumaepatula* have shown great variability in estimates of within population diversity (reviewed in (Karasawa et al., 2007b). These differences are likely the result of differences in population size, connectivity and the nature of markers used to estimate genetic diversity. Generally, studies using hyper-variable markers such as SSR find higher diversity levels (Brondani et al., 2005; Karasawa et al., 2007b, 2012). Genetic diversity in Costa Rican populations is comparable to studies in Brazil using co-dominant and dominant markers (Buso, Rangel & Ferreira, 1998). This suggest that *O. glumaepatula* populations in Costa Rica have moderate levels of genetic diversity, which are lower than expected for a monocot species with predominant autogamous mating, but similar to those expected for a mixed annual or short-lived species (Hamrick & Godt, 1996).

Intrapopulation genetic levels in *O. glumaepatula* are likely influenced by the mating system of this species, which in turn will have an important effect on gene flow patterns among populations. As in previous reports we also find a significant deficit of heterozygotes in all populations (Karasawa et al. 2007b), attributed mostly to its mating system. As a predominantly selfing species, with a significant proportion of clonal growth, we expect *O. glumaepatula* populations to display high rates of inbreeding. In

equilibrium our Fis estimates in MQ and Guanacaste suggest a mixed-mating system skewed towards selfing ($\hat{t}_m = \frac{1-F_{IS}}{1+F_{IS}} = 0.24$). A mixed-mating breeding strategy has also been proposed by other studies (Brondani et al., 2005; Vaz et al., 2009). Progeny arrays and maximum likelihood estimates of the mating system reached similar results (*tm* = 0.01-0.223. Karasawa et al. 2007a). In a mixed-mating breeding system (Goodwillie, Kalisz & Eckert, 2005) outcrossing is a likely occurrence. Occasional outcrossing in predominately autogamous populations may have a significant effect on genetic diversity by introducing novel diversity into populations increasing genetic diversity estimates. Therefore it is likely that the moderate levels of genetic diversity that we measured in MQ and Guanacaste are probably positively influenced by intermediate outcrossing rates, which facilitate gene flow across populations.

Genetic diversity estimates may also respond to differences in population size or gene flow patterns. Large interconnected populations are more likely to have higher levels of genetic diversity than smaller isolated populations (Frankham, 1996). *O. glumaepatula* populations in this study are the only remaining populations of this wild rice in Costa Rica. These populations are limited to very specific areas (inundated wetlands with a characteristic dry season) and are separated by large distances between each other and from other populations in Central America or the Caribbean (Vaughan et al., 2005). This distribution increases the effect of genetic drift reducing within population genetic diversity. However, MQ is a very large metapopulation with more than 10K individuals, all of which are interconnected by the Medio Queso River. This population may be large enough to lessen the effects of genetic drift, thus preserving most of the original genetic diversity. Similar results were found for other larger and similarly interconnected populations in Brazil (Buso, Rangel & Ferreira, 1998; Vaz et al., 2009).

In *O. glumaepatula,* propagules and seed dispersal may occur via hydrochory and zoochory (Gastezzi, Martínez & Villareal, 2012). If the river carries seeds or detached culms, we would expect gene flow to occur predominantly downstream leading to greater admixture in downstream populations, resulting in higher levels genetic diversity. For hydrochorous species, upstream populations are expected to suffer an erosion of genetic diversity, caused by random extinction of populations (metapopulation dynamics) and drift. However, we were unable to detect any effect of river direction on genetic diversity estimates. One way of explaining this result is that the current of the MQ River is not strong enough to avoid upstream migration of seeds and vegetative propagules. Upstream

gene flow is very likely to occur in MQ as this river does not have a strong current (Jimenez, 2004) and during the rainy season floating meadows or seeds may easily move upstream, thus homogenizing genetic diversity estimates across populations regardless of river position. Other studies have suggested that waterfowl are also likely to dispersed seeds among rivers (Pollux et al., 2007) and may disperse into upstream populations. Personal observation of bird fauna in MQ found gramineous seeds in *Leptotila plumbeiceps* (Gastezzi, Martínez & Villareal, 2012). Given the mixed-mating system of *O. glumaepatula* in MQ, outcrossing between populations separated by less than 15 Km may also contribute to upstream gene flow. Therefore, populations interconnected via vegetative dispersal, seed-dispersal and pollen flow should have comparable levels of genetic diversity regardless of their position along the river. This should also lead to low levels of genetic structure as observed in our sample (See below).

Karasawa et al. (2007b) reviewed several studies analyzing genetic structure estimates for different Brazilian populations of *O. glumaepatula*. These studies consistently find significant differences in allele frequencies among populations as expected for an autogamous, short-lived species (Hamrick & Godt, 1996)). In contrast we found a significant but dismissible structure ($\theta = 0.03$) among populations separated by less than 15 Km within the same river. The reason for these low levels of genetic structure is likely the result of high rates of gene flow among these populations. The Medio Queso River connects all of our MQ populations. These populations are perennial and since river fluctuations are small, these populations are able to live year around. Migration between populations is likely at the beginning of the rainy season (May) when seeds and culms are likely carried along the river. Gene flow via pollen and seed dispersal are also possible, all of which should homogenize allele frequencies among populations. Vaz et al. (2009) had similar conclusions when analyzing the largest population in Brazil. They found no evidence of structure among samples collected along the same river separated by ~ 10 km. Previous studies that found structure (Akimoto, Shimamoto & Morishima, 1998; Buso, Rangel & Ferreira, 1998; Brondani et al., 2005; Karasawa et al., 2007b) consistently analyzed populations that were separated by large distances (>100 km) and in most cases in different unconnected river systems. Even when populations where in the same river and separated by a few Km, they were in different tributaries and therefore, gene flow via seeds or vegetative propagules was unlikely. This suggests that in MQ all

452 populations likely form a single meta-population with ample levels of gene flow among
453 individuals and sites.
454 An interesting and somewhat contradictory result is that Guanacaste, a population
455 separated by more than 100 km and a mountain range from MQ, shows almost no genetic
456 structure with MQ (Figure 2) as would be expected from previous studies that find
457 structure in populations separated by comparable distances (reviewed in Karasawa
458 2007a). It is then likely that the population of *O. glumaepatula* in Guanacaste likely
459 originated from MQ by means of founder effect. A detailed analysis of the environmental
460 requirements of *O. glumaepatula* shows that the only site where this species is likely to
461 occur is MQ (Gastezzi, Martínez & Villareal, 2012). Guanacaste is a very small
462 population composed by less than 100 individuals, in contrast to the MQ, which holds
463 more than 10K plants. However, allelic frequencies do not deviate significantly from
464 those in MQ suggesting that this population has only been recently founded and genetic
465 drift has not had enough time to reduce genetic diversity. Guanacaste province is the area
466 with the largest rice plantations in Costa Rica and the largest processing plants (Lomas &
467 Herrera, 1985). Therefore, cultivated rice from different parts of the country is
468 transported to facilities in Guanacaste. The Guanacaste population may have been
469 established from *O. glumaepatula* seeds originating in MQ that were transported by
470 dispersers or humans into Guanacaste in the recent past (Mack & Lonsdale, 2001; Nathan
471 et al., 2008).
472 The major factor driving genetic structure in Costa Rican populations of *O.*
473 *glumaepatula*, was the potential of some individuals to hybridize with cultivated rice (*O.*
474 *sativa*). Our analyses showed that two *O. glumeapatula* populations in our sample shared
475 a part of their genome with cultivated *O. sativa* (Figure 3) resulting in a significant
476 increase in genetic structure. Wild rice species *O. latifolia* and *O. grandiglumis* (CCDD
477 genome) as well as *O. glumaepatula* (AA) are all present in Costa Rica, however only *O.*
478 *glumaepatula* is diploid and able to hybridize with diploid commercial rice. Introgression
479 between *O. sativa* and *O. glumaepatula* has been previously documented and gene flow
480 across species was positively influenced when commercial cropping occurs in proximity
481 of wild rice (Brondani et al., 2005). Cultivated rice and *O. glumaepatula* are sympatric at
482 El Muro in MQ and also in Guanacaste. In MQ farms increasingly invade the pristine
483 wetland and in Guanacaste *O. glumaepatula* grows within a rice farm. However, we
484 expected limited cross-pollination given that the flowering phenology of cultivated rice

and *O. glumaepatula* have little overlap; wild rice flowers once a year between October and November about two months after *O. sativa*. Our results show nonetheless, that hybridization between both species is likely to have occurred at these sites. Hybrids between *O. glumaepatula* and *O. sativa* are male sterile, however stigmas are still receptive and can produce viable seeds (Yamagata et al., 2010). Therefore, hybrids are able to backcross to either progenitor, which in turn may result in the introgression of parts of the *O. sativa* genome into *O. glumaepatula*. These hybrids and backcrosses are apparently morphologically similar to wild rice species. During our collecting trips we may have inadvertently collected hybrids and 3rdor 4thgeneration backcrosses as well as wild rice individuals without introgression, all of which are morphologically similar.

Introgression of genes from cultivated species into their wild relatives has been studied for a large number of crops and it has been established to occur when crops and wild species come into contact (Ellstrand, Prentice & Hancock, 1999). Genes that increase fitness are more likely to introgress into weedy or wild species because they are not selected against and are thus more likely to persist in wild populations. However, if gene flow rates are high (m > s), even deleterious alleles may persist in the recipient population (Ellstrand, Prentice & Hancock, 1999). Crop-to-wild gene flow in rice has been documented (Brondani et al., 2002; Chen et al., 2004; Song et al., 2006), even across reproductive barriers such as F1 hybrid sterility (Chu & Oka, 1970). Song et al. (2006) determined that hybrids between *O. sativa* and *O. rufipogon* had lower reproductive fitness but higher vegetative growth, which resulted in similar composite fitness, allowing genes from cultivated species to persist in wild rice species, as documented in the present study. Introgression may lead to significant differentiation between populations of wild species with important implications for *in situ* conservation of natural resources (Xia et al., 2011; Jiang et al., 2012). Introgressive hybridization may increase the likelihood of extinction of wild species, by reducing the genetic integrity of the species through outbreeding depression or because the wild species becomes integrated with the cultivated species by repeated introgression (Ellstrand, Prentice & Hancock, 1999).

Given the lack of structure found in populations of the wild rice species *O. glumaepatula*, we conclude that all sites may be grouped into a single meta-population. We found moderate to high levels of intrapopulation genetic diversity, likely a result of large population size. Conserving this population should preserve its genetic diversity as

a valuable resource for future breeding programs. Our results highlight the importance for both *in situ* and *ex situ* conservation strategies and breeding programs. We also observe that the main cause of genetic differences among populations is likely introgression from cultivate *O. sativa* populations. Introgression may represent an important biohazard as it may have a significant impact on autochthonous genetic diversity. It may also allow for commercial traits to be transferred into the only populations of *O. glumaepatula* in Costa Rica, which in turn could increase the likelihood of local extinction. Recently, introgression has become a major focus of interest in regards to biosafety issues, particularly relative to gene flow from genetically modified organisms (GMO's) into non-GM cultivars and their wild/weedy relatives (Lu & Snow, 2005; Sanchez-Olguin et al., 2009). Biosafety authorities and regulators may use our results to establish zones of exclusion for the eventual release of GM rice in these areas. The information presented in the present study, should be used to delineate conservation strategies and implement better planting practices to allow for the long-term viability of *O. glumaepatula* populations in Costa Rica. Although wetlands are under international protection by Ramsar Convention (http://www.ramsar.org/wetland/costa-rica) this site is at present endangered by the construction of a road along the northern border of Costa Rica, which could put in peril the largest *O. glumaepatula* population in Costa Rica and a likely stepping-stone between Costa Rica and Nicaraguan wild rice populations.

Acknowledgements

The authors would like to express their gratitude to Dr. Ana M. Espinoza and Dr. Elizabeth Johnson for their valuable contribution to initiate this research; and to Dr. Federico Albertazzi Castro for the general coordination of LAC-Biosafety Project in Costa Rica. In addition, special thanks to Melania Muñoz, Amanda Calvo, Cindy Aguilar & Genuar Muñoz for their technical support and to G. Barrantes and J Ross-Ibarra for comments on the manuscript.

References

Akimoto M, Shimamoto Y, Morishima H. 1998. Population genetic structure of wild rice *Oryza glumaepatula* distributed in the Amazon flood area influenced by its life-history traits. *Molecular Ecology* 7:1371–1381.

Arnold ML. 1997. *Natural Hybridization and Evolution*. Oxford University Press, USA.

Brar DS. 2005. Broadening the genepool and exploiting heterosis in cultivated rice. In: Toriyama K, Heong KL, Hardy B eds. *Rice is life: scientific perspectives for the 21st Century*. Proceedings of the World Rice Research Conference. Tokyo and Tsukuba, Japan, 157–160.

Brondani C, Rangel P, Brondani R, Ferreira M. 2002. QTL mapping and introgression of yield-related traits from Oryza glumaepatula to cultivated rice (Oryza sativa) using microsatellite markers. *Theoretical and Applied Genetics* 104:1192–1203.

Brondani RPV, Zucchi MI, Brondani C, Rangel PHN, Borba TCDO, Rangel PN, Magalhães MR, Vencovsky R. 2005. Genetic structure of wild rice *Oryza glumaepatula* populations in three Brazilian biomes using microsatellite markers. *Genetica* 125:115–123.

Buso GSC, Rangel PH, Ferreira ME. 1998. Analysis of genetic variability of South American wild rice populations (Oryza glumaepatula) with isozymes and RAPD markers. *Molecular Ecology* 7:107–117.

Chen LJ, Lee DS, Song ZP, Suh HS, Lu BR. 2004. Gene Flow from Cultivated Rice (Oryza sativa) to its Weedy and Wild Relatives. *Annals of Botany* 93:67.

Chu Y-E, Oka H-I. 1970. Introgression Across Isolating Barriers in Wild and Cultivated Oryza Species. *Evolution* 24:344–355.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

574 Ellstrand NC. 2003. Current knowledge of gene flow in plants: implications for transgene flow.
575 *Philosophical Transactions of the Royal Society of London B: Biological Sciences*
576 358:1163–1170.
577 Ellstrand NC, Prentice HC, Hancock JF. 1999. Gene Flow and Introgression from Domesticated
578 Plants into Their Wild Relatives. *Annual Review of Ecology and Systematics* 30:539–563.
579 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the
580 software structure: a simulation study. *Molecular Ecology* 14:2611–2620.
581 Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus
582 genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7:574–578.
583 Frankham R. 1996. Relationship of Genetic Variation to Population Size in Wildlife.
584 *Conservation Biology* 10:1500–1508.
585 Gastezzi P, Martínez D, Villareal J. 2012. *Monitoreo biológico de aves acuáticas en la zona*
586 *norte de Costa Rica*. San José, Costa Rica: Universidad de Costa Rica, Banco Mundiar y
587 GEF.
588 Ge S, Oliveira GCX, Schaal B a., Gao L-Z, Hong D. 1999. RAPD variation within and between
589 natural populations of the wild rice *Oryza rufipogon* from China and Brazil. *Heredity*
590 82:638–644.
591 Goodwillie C, Kalisz S, Eckert CG. 2005. The evolutionary enigma of mixed mating systems in
592 plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of*
593 *Ecology, Evolution, and Systematics* 36:47–79.
594 Grant V. 1981. *Plant Speciation*. Columbia University Press.
595 Guillot G, Mortier F, Estoup A. 2005. Geneland: a computer package for landscape genetics.
596 *Molecular Ecology Notes* 5:712–715.

Hajjar R, Hodgkin T. 2007. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156:1–13.

Hamrick JL, Godt MJW. 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences* 351:1291—1298.

Holsinger KE, Lewis PO, Dey DK. 2002. A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology* 11:1157–1164.

Holsinger KE, Wallace LE. 2004. Bayesian approaches for the analysis of population genetic structure: an example from Platanthera leucophaea (Orchidaceae). *Molecular Ecology* 13:887–894.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.

Jiang Z, Xia H, Basso B, Lu B-R. 2012. Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. *Theoretical and Applied Genetics* 124:309–322.

Jimenez A. 2004. Cooperación transfronteriza ambiental para la conservación de humedales en cuencas compartidas: Un análisis desde la frontera Costa Rica – Nicaragua. Magister Scientia Thesis. Universidad de Costa Rica.

Juliano AB, Naredo MEB, Jackson MT. 1998. Taxonomic status of Oryza glumaepatula Steud. I. Comparative morphological studies of New World diploids and Asian AA genome species. *Genetic Resources and Crop Evolution* 45:197–203.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

Karasawa MMG, Vencovsky R, Silva CM, Zucchi MI, Oliveira GCX, Veasey EA. 2007a. Mating System of Brazilian Oryza glumaepatula Populations Studied with Microsatellite Markers. *Annals of Botany* 99:245–253.

Karasawa MMG, Vencovsky R, Silva CM, Zucchi MI, Oliveira GCX, Veasey EA. 2007b. Genetic structure of Brazilian wild rice (Oryza glumaepatula Steud., Poaceae) populations analyzed using microsatellite markers. *Genetics and Molecular Biology* 30:400–410.

Karasawa MMG, Vencovsky R, Silva CM, Cardim DC, Bressan E de A, Oliveira GCX, Veasey EA. 2012. Comparison of microsatellites and isozymes in genetic diversity studies of Oryza glumaepatula (Poaceae) populations. *Revista de Biología Tropical* 60:1463–1478.

Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. In: Sasaki T, Moore G eds. *Oryza: From Molecule to Plant*. Springer Netherlands, 25–34.

Lesica P, Allendorf FW. 1995. When Are Peripheral-Populations Valuable For Conservation. *Conservation Biology* 9:753—760.

Lomas J, Herrera H. 1985. Weather and rice yield relationships in tropical Costa Rica. *Agricultural and Forest Meteorology* 35:133–151.

Lu B-R. 2013. Introgression of transgenic crop alleles: Its evolutionary impacts on conserving genetic diversity of crop wild relatives. *Journal of Systematics and Evolution* 51:245–262.

Lu B-R, Snow AA. 2005. Gene Flow from Genetically Modified Rice and Its Environmental Consequences. *BioScience* 55:669–678.

Lynch M, Milligan BG. 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3:91–99.

Mack RN, Lonsdale WM. 2001. Humans as Global Plant Dispersers: Getting More Than We Bargained For Current introductions of species for aesthetic purposes present the largest single challenge for predicting which plant immigrants will become future pests. *BioScience* 51:95–102.

Nathan R, Schurr FM, Spiegel O, Steinitz O, Trakhtenbrot A, Tsoar A. 2008. Mechanisms of long-distance seed dispersal. *Trends in Ecology & Evolution* 23:638–647.

Newman D, Pilson D. 1997. Increased probability of extinction due to decreased genetic effective population size: experimental populations of \textitClarkia pulchella. *Evolution* 51:354—362.

Peakall R, Smouse PE. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288–295.

Pollux BJA, Jong MDE, Steegh A, Verbruggen E, Van Groenendael JM, Ouborg NJ. 2007. Reproductive strategy, clonal structure and genetic diversity in populations of the aquatic macrophyte Sparganium emersum in river systems. *Molecular Ecology* 16:313–325.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155:945–959.

R Development Core Team. 2012. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ren F, Lu B-R, Li S, Huang J, Zhu Y. 2003. A comparative study of genetic relationships among the AA-genome Oryza species using RAPD and SSR markers. *Theoretical and Applied Genetics* 108:113–120.

Sanchez-Olguin ER, Arrieta-Espinoza G, Lobo JA, Espinoza-Esquivel AM. 2009. Assessment of gene flow from a herbicide-resistant indica rice (Oryza sativa L.) to the Costa Rican

weedy rice (Oryza sativa) in Tropical America: factors affecting hybridization rates and characterization of F1 hybrids. *Transgenic Research* 18:633–647.

Snow AA. 2002. Transgenic crops—why gene flow matters. *Nature Biotechnology* 20:542–542.

Song ZP, Lu B-R, Zhu YG, Chen JK. 2003. Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. *New Phytologist* 157:657–665.

Song Z, Zhu W, Rong J, Xu X, Chen J, Lu B-R. 2006. Evidences of introgression from cultivated rice to Oryza rufipogon (Poaceae) populations based on SSR fingerprinting: implications for wild rice differentiation and conservation. *Evolutionary Ecology* 20:501–522.

Tateoka T. 1962. Taxonomic studies of Oryza. IO latifolia complex. II. Several species complexes. *Bot. Mag., Tokyo* 75:418–27.

Vaughan DA, Kadowaki K, Kaga A, Tomooka N. 2005. On the Phylogeny and Biogeography of the Genus *Oryza*. *Breeding Science* 55:113–122.

Vaughan DA, Morishima H, Kadowaki K. 2003. Diversity in the Oryza genus. *Current Opinion in Plant Biology* 6:139–146.

Vaz AR de C, Borba TC de O, Brondani C, Rangel PHN, Camargo GS de O, Telles MP de C, Filho JAFD, Brondani RPV. 2009. Genetic analysis of a local population of Oryza glumaepatula using SSR markers: implications for management and conservation programs. *Genetica* 137:221–231.

Veasey EA, Cardin D, Silva RM, Bressan E de A, Vencovsky R. 2008. Assessing the genetic structure of Oryza glumaepatula populations with isozyme markers. *Brazilian Archives of Biology and Technology* 51:873–882.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

687 Veasey EA, Bressan E de A, Zucchi MI, Vencovsky R, Cardim DC, Silva RM da. 2011. Genetic
688 diversity of American wild rice species. *Scientia Agricola* 68:440–446.
689 Vekemans X. 2002. *AFLP-Surv*. Distributed by the author.: Labortoire de Genetique et Ecologie
690 Vegetale, Universite Libre de Bruxelles, Belgium.
691 Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I. 2002. Data from amplified fragment
692 length polymorphism (AFLP) markers show indication of size homoplasy and of a
693 relationship between degree of homoplasy and fragment size. *Molecular Ecology*:139–
694 151.
695 Vos P, Hogers R, Bleeker M, Reijans M, Lee T van de, Hornes M, Friters A, Pot J, Paleman J,
696 Kuiper M, Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic*
697 *Acids Research* 23:4407–4414.
698 Xia H-B, Wang W, Xia H, Zhao W, Lu B-R. 2011. Conspecific Crop-Weed Introgression
699 Influences Evolution of Weedy Rice (Oryza sativa f. spontanea) across a Geographical
700 Range. *PLoS ONE* 6:e16189.
701 Yamagata Y, Yamamoto E, Aya K, Win KT, Doi K, Sobrizal, Ito T, Kanamori H, Wu J,
702 Matsumoto T, Matsuoka M, Ashikari M, Yoshimura A. 2010. Mitochondrial gene in the
703 nuclear genome induces reproductive barrier in rice. *Proceedings of the National*
704 *Academy of Sciences* 107:1494–1499.
705 Zamora A, Barboza C, Lobo J, Espinoza AM. 2003. Diversity of native rice (Oryza Poaceae:)
706 species of Costa Rica. *Genetic Resources and Crop Evolution* 50:855–870.
707
708
709
710
711
712

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

713

714

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

Table 1. Genetic diversity estimates for six *O. glumaepatula* populations in Costa Rica. N: number of individuals, I:Shannon's index estimated withGenAlEx. H_jBayesian estimate of expected heterozygosity and %P of polymorphic loci p>0.05 from AFLPsurv

Population	**N**	**I**	**H_j**	**P%**
Delirio	22	0.166	0.137	58.79
Pta.Cortez	6	0.145	0.219	32.93
Refugio	11	0.152	0.167	41.82
Robles	8	0.156	0.194	41.01
Luna	4	0.142	0.266	30.30
El Muro	17	0.171	0.148	57.37
Guanacaste	11	0.145	0.162	41.21
Average			0.184	

Commercial individuals (n=6) had the lowest Shannon index: 0.135 and P% 34.34

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

Figure Legend

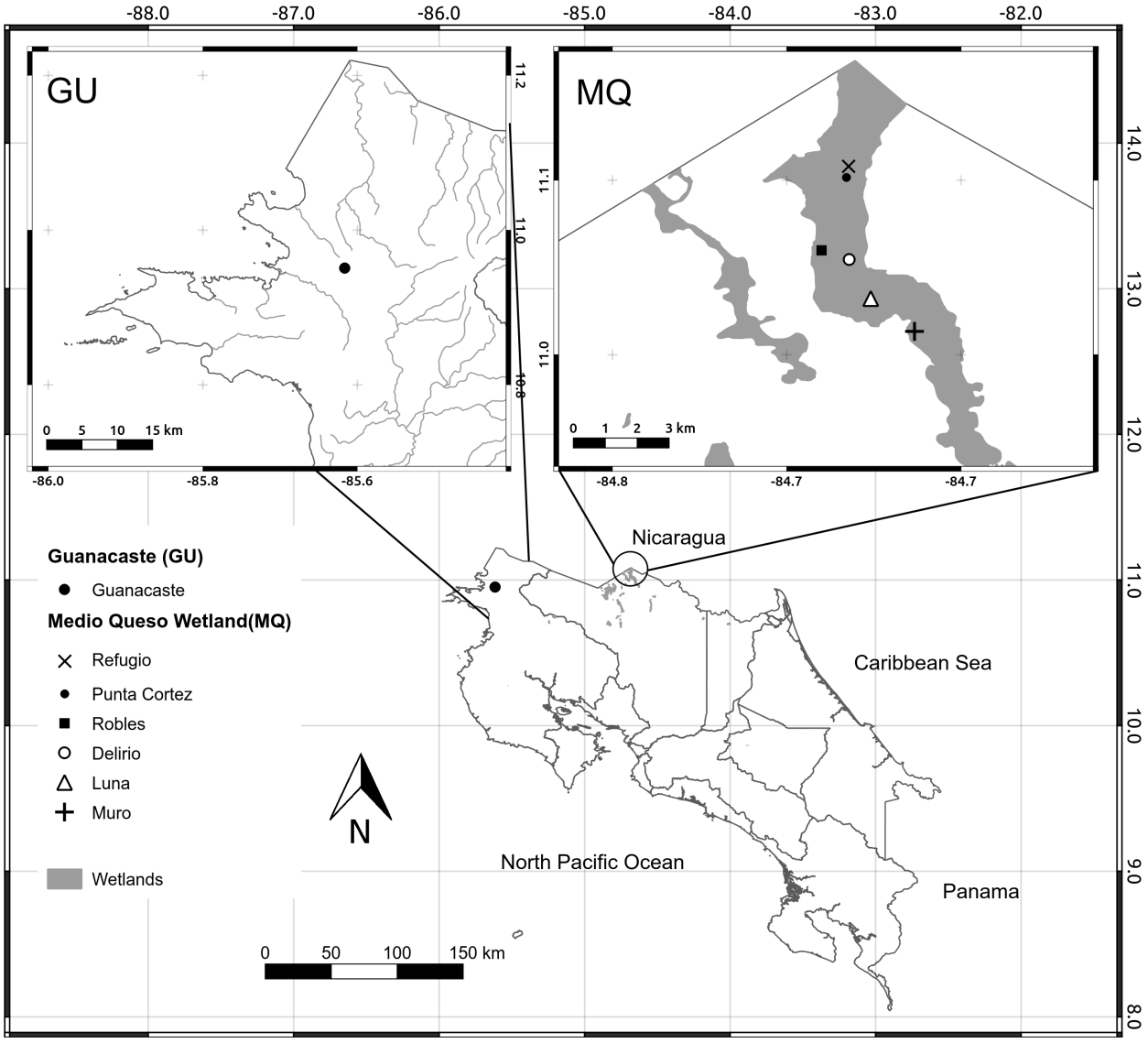
Figure 1. Location of *O.glumaepatula* populations in northern Costa Rica along the Medio Queso River.

Figure 2. Assignment of *O. glumaepatula* samples in MQ and Guanacaste to two clusters (K = 2) using the Bayesian clustering algorithm in Structure.

Figure 3. Assignment of wild *O. glumaepatula* and commercial *O. sativa* individuals into two clusters (K=2) using the Bayessian algorithm implemented in Structure.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

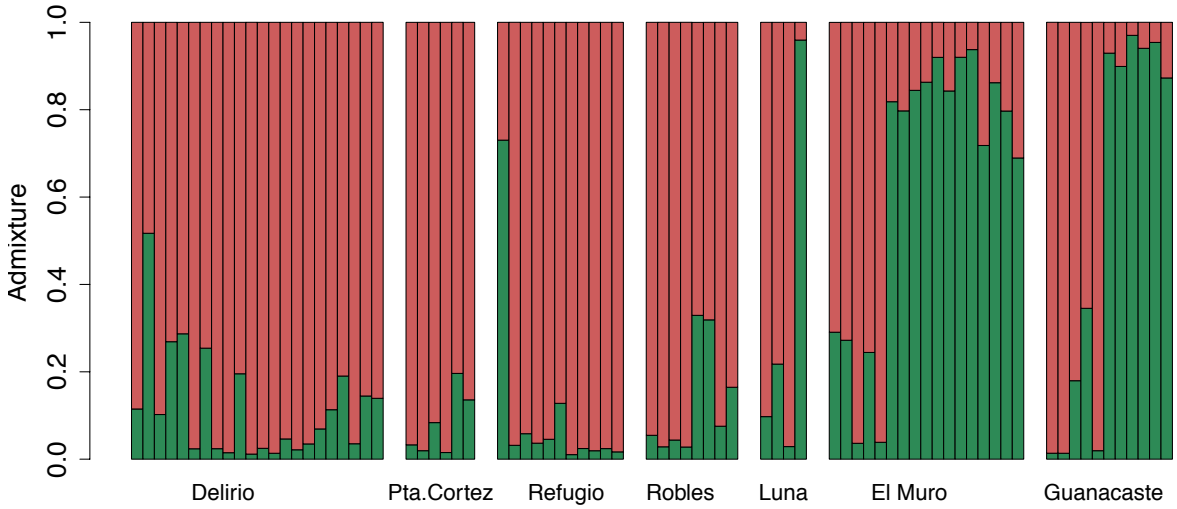
736 Figure 1
737



738

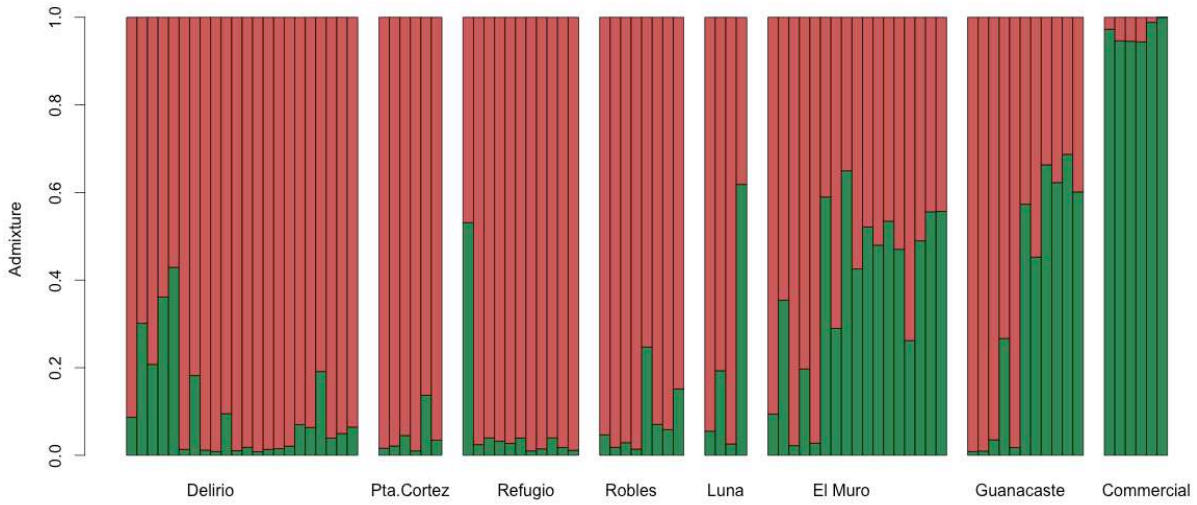
PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

Figure 2.



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015

746 Figure 3.
747
748



749
750
751
752

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1095v1 | CC-BY 4.0 Open Access | rec: 19 May 2015, publ: 19 May 2015