doi:10.1111/j.1558-5646.2012.01674.x

GENOME SIZE IS NOT CORRELATED WITH EFFECTIVE POPULATION SIZE IN THE *ORYZA* SPECIES

Bin Ai,[1,2] Zhao-Shan Wang,[1] and Song Ge[1,2,3]

[1]*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China*

[2]*Graduate University, Chinese Academy of Sciences, Beijing 100039, China*

[3]*E-mail: gesong@ibcas.ac.cn*

Received January 15, 2012
Accepted March 22, 2012

Genome sizes vary widely across thetree of life and the evolutionary mechanism underlined remains largely unknown. Lynch and Conery (2003) proposed that evolution of genome complexity was driven mainly by nonadaptive stochastic forces and presented the observation that genome size was negatively correlated with effective population size (N_e) as a strong support for their hypothesis. Here, we analyzed the relation between N_e and genome size for 10 diploid *Oryza* species that showed about fourfold genome size variation. Using sequences of more than 20 nuclear genes, we estimated N_e for each species after correction for the effects of demography and heterogeneity of mutation rates among loci and species. Pairwise comparisons and correlation analyses did not detect a negative relationship between N_e and genome size despite about 6.5-fold interspecies N_e variation. By calculating phylogenetically independent contrasts (PICs) for N_e, we repeated correlation analysis and did not find any correlation between N_e and genome size. These observations suggest that the genome size variation in the *Oryza* species cannot be explained simply by the effect of effective population size.

KEY WORDS: **Effective population size, genome size, *Oryza*, polymorphism.**

With a vast spectrum spanning several orders of magnitude across the tree of life, genome size remains an aspect of concerns in population genetics and comparative genomics because it serves as the most fundamental property of a genome (Petrov 2001; Lynch 2007). According to Plant DNA C-value database (Bennett and Leitch 2005a), genome sizes range nearly 2000-fold across angiosperms and about 100-fold in Poaceae. It is widely accepted that varying amounts of noncoding DNA contribute to the large-magnitude variation of genome sizes (Lynch 2007; Flowers and Purugganan 2008) and the mechanisms for both increase and decrease in DNA content have been well addressed (Lynch 2007; Hawkins et al. 2008; Whitney et al. 2010). To date, different hypotheses have been proposed to account for the tremendous diversity of the genome sizes across major lineages of organisms and invoked extensive controversy (Petrov 2002; Lynch and Conery 2003; Charlesworth and Barton 2004; Bennett and Leitch 2005b; Yi 2006; Charlesworth 2008; Hawkins et al. 2008; Whitney et al. 2010, 2011; Boussau et al. 2011; Lynch 2011).

Lynch and Conery (2003) proposed a nonadaptive theory arguing that evolution of genome complexity was driven mainly by nonadaptive stochastic forces rather than by adaptive evolution. They predicted that more harmful noncoding DNA in lineages with small N_e would accumulate due to the preponderance of random genetic drift over natural selection and thus the genome sizes in these lineages became larger. They presented the observation that genome size was negatively correlated with effective population size (N_e) across prokaryotes and eukaryotes as a strong support for their hypothesis. In a comparison between freshwater

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

and marine fish species, Yi and Streelman (2005) detected an obvious negative relationship between genome size and N_e independent of phylogeny, body size, and generation time, consistent with Lynch and Conery's (2003) hypothesis. By examining the ratio of nonsynonymous to synonymous substitutions (D_n/D_s) among recently diverged taxa that differ in genome size, Boussau et al. (2011) found that genomic duplications (causing larger genome size) occurred concomitantly with smaller N_e (higher D_n/D_s) in the mitochondrial genomes of tetrapods, again supporting the contribution of nonadaptive processes to the mitochondrial genome size.

Nevertheless, objections to the idea that genome size is determined by population size have been frequently raised based on theoretical and empirical investigations (e.g., Charlesworth and Barton 2004; Gregory and Witt 2008; Whitney and Garland 2010; Whitney et al. 2010). First, the analyzing strategies in Lynch and Conery (2003) were problematic in that (1) using heterozygosity as a proxy for N_e failed to consider substantially varying mutation rates in different lineages (Charlesworth and Barton 2004; Daubin and Moran 2004); (2) bacterial species were difficult to distinguish from each other so that polymorphism levels might be falsely elevated (Daubin and Moran 2004); (3) estimate of N_e based on polymorphism using molecular markers was sensitive to recent demographical histories (Daubin and Moran 2004; Yi 2006; Gregory and Witt 2008); (4) the negative correlation between genome size and N_e disappeared when using phylogenetically independent contrasts (PICs) (Whitney and Garland 2010). Second, empirical investigations on different lineages of organisms were not in accordance with Lynch and Conery's (2003) hypothesis, including bacteria (Daubin and Moran 2004; Kuo et al. 2009), mammals (Vinogradov 2004), *Arabidopsis thaliana* versus *A. lyrata* (Wright et al. 2002; Charlesworth and Barton 2004), maize versus fly (Charlesworth 2008), and seed plants (Whitney et al. 2010). For example, Daubin and Moran (2004) indicated that small N_e in symbiotic bacteria might result in reduced genomes through gene loss and thus the relation between N_e and genome size in bacteria was the opposite of that proposed by Lynch and Conery (2003). In a recent study based on a dataset including 205 seed plant species, Whitney et al. (2010) found no relationship between genome size and N_e using PIC analyses.

Despite these debates, relatively few studies have been conducted on related species that share a common evolutionary history and differ in much fewer properties (Charlesworth and Barton 2004; Charlesworth 2008; Flowers and Purugganan 2008). The rice genus (*Oryza* L.) has increasingly become an important model for a variety of studies on biological questions thanks to the completion of rice genome sequencing of two rice cultivars and the development of the *Oryza* Map Alignment Project (OMAP) that aims to build a genome-level experimental system for *Oryza* studies (Shimamoto and Kyozuka 2002; Wing et al. 2005). The genus *Oryza* consists of two cultivated and approximately 22 wild species distributed across the world (Vaughan 2004). These species can be classified into 10 genome groups (A-, B-, C-, E-, F-, G-, BC-, CD-, HJ-, HK-genomes) (Ge et al. 1999; Wing et al. 2005) and serve as a proper study system to test Lynch's hypothesis with the following advantages: (1) About fourfold genome size variation was found for the diploid species (Ammiraju et al. 2006; Miyabayashi et al. 2007); (2) Phylogeny of *Oryza* is fully resolved, especially for the diploid species (Ge et al. 1999; Zou et al. 2008); (3) Shared evolutionary history helps eliminate the effects of other factors like ancient demography, when comparing different species for genome size and effective population size.

In this study, we obtained sequences of more than 20 nuclear gene fragments for 10 *Oryza* species representing all diploid genome types in the genus, which avoids the confounding effect of polyploidy. We corrected for the effects of demography and heterogeneity of mutation rates among loci and species before estimating N_e. By pairwise comparisons and correlation analyses, we found no significant correlation between genome size and N_e in the genus *Oryza*, providing the first fine-scale comparison among closely related species.

Materials and Methods

SPECIES SAMPLING AND LOCI STUDIED

We sampled 10 diploid *Oryza* species, representing all six diploid genome types (A-, B-, C-, E-, F-, and G-genomes) in the genus (Table 1), of which four (B-, E-, F-, and G-) genomes each have a single species (Vaughan 2004; Zou et al. 2008). For each species, we sampled at least eight accessions that covered the distribution range of the species except for *O. rhizomatis* from which four accessions were used (Table 1) because this species is endemic to Sri Lanka. Information on the sampled accessions is listed in Table S1.

In our previous studies on population genetics of the *Oryza* species, we have obtained sequences of 10 loci for five species (*O. rufipogon*, *O. nivara*, *O. officinalis*, *O. rhizomatis*, and *O. eichingeri*) (Zhang and Ge 2007; Zhu et al. 2007), and 14 loci for *O. barthii* (Li et al. 2011). For these species, we further sequenced additional 12 or 13 loci in this study. For the remaining four species (*O. punctata*, *O. australiensis*, *O. brachyantha*, and *O. granulata*), we sequenced 20 loci (Table 1). Therefore, we obtained sequences from 20 to 26 unlinked nuclear loci for each of the 10 species (Table 1). Detailed information on the sampled loci is provided in Tables S2 and S3.

DNA EXTRACTION, AMPLIFICATION, AND SEQUENCING

Total DNA was extracted from fresh or silica gel-dried leaves, using the CTAB (hexadecyltrimethylammonium bromide) method

Table 1. Summary of genome size, nucleotide diversity, and effective population size for 10 diplod *Oryza* species.

Taxon	Genome type	C value (pg)[1]	N[2]	Number of loci	π_s[3]	N_e ($\times 10^6$)[3]
Oryza nivara	A	0.93	11	22	0.0054 (0.0053)	0.28 (0.28)
Oryza rufipogon	A	0.88	15	22	0.0054 (0.0054)	0.23 (0.23)
Oryza barthii	A	0.94	13	26	0.0021 (0.0019)	0.13 (0.12)
Oryza punctata	B	0.86	10	20	0.0011 (0.0010)	0.08 (0.07)
Oryza officinalis	C	1.28	12	23	0.0050 (0.0050)	0.35 (0.35)
Oryza rhizomatis	C	1.92	4	23	0.0063 (0.0066)	0.47 (0.49)
Oryza eichingeri	C	1.39	8	23	0.0073 (0.0069)	0.52 (0.51)
Oryza australiensis	E	1.92	10	20	0.0044 (0.0046)	0.32 (0.33)
Oryza brachyantha	F	0.61	10	20	0.0033 (0.0035)	0.24 (0.23)
Oryza granulata	G	2.38	14	20	0.0032 (0.0030)	0.19 (0.19)

[1]The genome size estimates (C values, picograms, pg) of the 10 species were obtained from Miyabayashi et al. (2007) except for that of *O. nivara* that was from Ammiraju et al. (2006).

[2]The number of accessions sampled for each species in this study. The number of accessions sampled for *Oryza nivara* and *O. rufipogon* (Zhu et al. 2007), and *O. barthii* (Li et al. 2011) was 12, 18, and 20, respectively.

[3]Average π_s and N_e across loci. The figures in parentheses were calculated based on the filtered datasets in which all sequences from nonneutral loci were removed.

as described in Ge et al. (1999). PCR amplification and purification of the products were performed generally following those in previous studies (Zhang and Ge 2007; Zhu et al. 2007). Sequencing was done on an ABI3730XL automatic sequencer (Applied Biosystems, Foster City, CA). Purified products were sequenced either directly or after cloning into *pGEM* T-easy vectors (Promega, Madison, WI) if direct sequencing failed or dual peaks were found. At least three cloned DNA fragments were sequenced for each individual. The number of clones per individual was added by three until the haplotype was shared among at least two clones, so as to exclude artificial singleton (Zhang and Ge 2007). All sequences have been deposited in GenBank, with the accession numbers JQ414289–JQ415911.

SEQUENCE ANALYSIS

Sequences were assembled with the ContigExpress program (Informax Inc., North Bethesda, MD) and aligned with ClustalX 1.83 (Thompson et al. 1997) before additional manual refinements. As well established in population genetics (Charlesworth 2009), the expected level of nucleotide diversity in a sample of a population (π) is $4N_e\mu$ under the standard model, where N_e is the effective population size and μ is the mutation rate per nucleotide. Because of the heterogeneity of evolutionary rates across genes, we first used the method of Zhang and Ge (2007) to estimate μ at silent sites for each locus by $\mu = \mu_{adh1} \times K_{sil} / K_{sadh1}$, where K_{sil} and K_{sadh1} are silent distances between the target species and its corresponding outgroup at that locus and at *Adh1* locus, respectively (Table S3). μ_{adh1} is estimated to be 7.0×10^{-9} substitutions per synonymous site per year, a fossil-calibrated synonymous rate of *Adh1* divergence in grasses (Gaut et al. 1996). Then, we calculated average pairwise difference per basepair between sequences at silent sites (π_s) for each locus using DnaSP version 5.10.00 (Librado and Rozas 2009). N_e can be estimated from silent nucleotide site diversity by $N_e = \pi_s/(4\mu)$ (Charlesworth 2009).

Because nonneutral data may bias the N_e estimate, we performed several neutrality tests to confirm whether the loci were neutral. We calculated Tajima's D (1989) and $D*$ and $F*$ of Fu and Li (1993) for each locus to test for the neutral equilibrium model of evolution across species. The associated one-tailed P-values were obtained by computing 1000 coalescent simulations, with recombination taken into account. The minimum number of recombination events (R_m) was estimated with the four-gamete test (Hudson and Kaplan 1985). These tests were performed using the program DnaSP. To discriminate between selection forces and population demography, the multilocus HKA test (Hudson et al. 1987) was performed with the HKA package (http://genfaculty.rutgers. edu/hey/software#HKA). Individual runs of the HKA test were performed for the contrast between each of 10 species and its corresponding outgroup. Detailed information about the loci and the outgroups in HKA runs is provided in Table S4.

TEST THE RELATIONSHIP BETWEEN N_E AND GENOME SIZE

Pairwise comparison of π_s and N_e among the 10 species was conducted and the significance was evaluated by paired t-test. We performed correlation analysis between N_e and genome size. Correlation coefficient and its significance were calculated for geometric mean N_e estimates across loci, and all values were $\log_{10}$ transformed prior to analysis.

Because shared phylogenetic history may violate the assumption of statistical independence, we further used PICs (Felsenstein

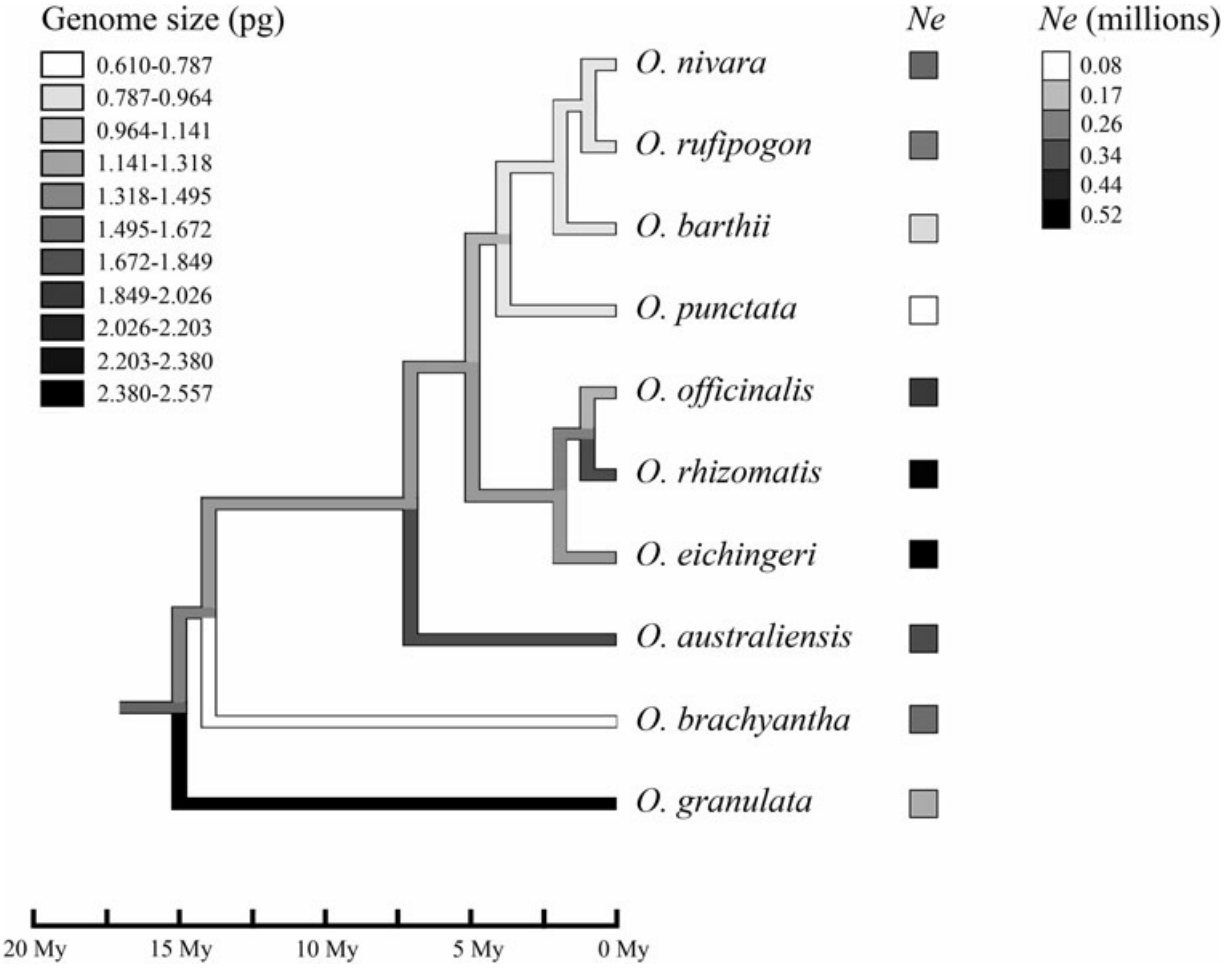

Figure 1. Phylogeny for 10 *Oryza* species with a reconstruction of ancestral genome sizes. Shades of the branches on the tree represent the genome sizes, and shades of squares after the species names indicate the mean N_e estimates for the species. Time scale is provided below the tree.

1985) to repeat the correlation analysis. The *Oryza* phylogeny of the 10 species (Fig. 1) was basically obtained from Zou et al. (2008) with slight modification in which *O. sativa* was replaced by *O. nivara* because these two species are most closely related (Zhu and Ge 2005) and *O. sativa* is a cultivated species. The *Oryza* phylogeny with branch lengths was imported to Mesquite version 2.7.4 (Maddison and Maddison 2010). We did ancestral reconstruction for genome size in *Oryza* with the parsimony ancestral state method of Mesquite (Maddison and Maddison 2010). To test whether the relationship between the traits in *Oryza* exhibited phylogenetic signals, we first used BayesTraits (Pagel and Meade 2009) to calculate λ that varies from 0 (entirely phylogenetical independence) to 1 (entirely phylogenetical dependence). Then, we obtained PICs for N_e and genome size using the PDAP:PDTREE module in Mesquite (Midford et al. 2002). We obtained standardized contrasts for further correlation analyses by dividing the raw contrasts by the standard deviations (Garland et al. 1992).

Results

A total of 2463 sequences from 29 loci were obtained from 10 diploid *Oryza* species. The length of aligned sequences for each locus ranged from 418 to 1467 bp, with a total of 23,915 bp (including 6429 bp of coding region) in length (Table S2). Of the 29 loci, 26 contained both coding and noncoding sites and the remaining three contained either intron (*Adh1*, *TFIIAγ-1*) or 5′-flanking (*CatA*) regions. The schematic diagrams of the 29 genes are shown in Figure S1. The number of loci used for each of 10 species varied from 22 to 26 (Table 1). Standard statistics of sequence variation for each locus for each species are summarized in Table S3. At the species level, average estimates of silent nucleotide diversity (π_s) across loci varied substantially among 10 species, ranging from 0.0011 (*O. punctata*) to 0.0073 (*O. eichingeri*) (Table 1; Fig. S2). Given the heterogeneity of evolutionary rates across genes, we performed the paired *t*-test between all pairs of 10 species and found substantial variation of diversity levels

among species. As shown in Figure 3a and Table S5, 23 species contrasts showed significant *t*-test values ($P < 0.05$) and the remaining 22 contrasts showed no significance. Of the significant contrasts, 15 and eight have positive or negative values, respectively (Fig. S3a). These observations indicated that the species with larger genome size might not have smaller π_s values.

Summary of the tests of Tajima's *D* (1989) and D^* and F^* of Fu and Li (1993) for each locus for each species are shown in Table S3. No significance was observed for a majority of values, except for *Adh1* in *O. nivara* and *O. rufipogon*, *Cbp1* in *O. rufipogon*, and *NP70* in *O. granulata*, in which all three tests were significant (Table S3). A significant departure from neutrality at a specific locus may not necessarily indicate the signature of selection because these statistics are sensitive to population demography. We further conducted a multilocus HKA test that is robust to population structure and demography. Signature of departure from the neutral model was detected in four contrasts (*O. rhizomatis*/*O. punctata*, *O. eichingeri*/*O. punctata*, *O. brachyantha*/*O. granulata*, and *O. granulata*/*O. brachyantha*) (Table S4). The HKA test was repeated with exclusion of the loci with the largest contribution to the overall statistics until the statistic dropped below the critical value (Table S4). Significant departure can be explained by a larger variance in the polymorphism/divergence ratio than that expected under a neutral equilibrium model, which might be attributed to selection on some loci for some species.

N_e estimates for each locus for each species are listed in Table S3. Consistent with the π_s values, the average N_e estimates across loci varied substantially among the 10 species, with the maximum difference between species being about 6.5-fold (Table 1). When pairwise comparison of N_e was performed with paired *t*-test, similar pattern to the π_s values was detected as expected. Twenty-three of 45 pairs of species comparisons were significant ($P < 0.05$), with 14 being positive and nine negative (Fig. S3b, Table S5). To assess the relationship between N_e and genome size, we first performed correlation analysis between N_e and genome size using the log-transformed trait values. Using geometric mean N_e across loci, we did not detect significant relationship between N_e and genome size (Fig. 2a). Then, we repeated correlation analysis to account for the phylogenetic nonindependence. We obtained $\lambda = 0.95$, indicating strong phylogenetic signal for the relationship between the two traits. We thus calculated PICs for the N_e and C after standardization (Table S6). The correlation patterns using PICs (Fig. 2b) were similar to those using the phylogenetically uncorrected data (Fig. 2a).

As indicated above, a few of loci did not evolve neutrally in some species, including *Adh1* in *O. nivara* and *O. rufipogon*, *Adh1-C* in *O. eichingeri*, *Cbp1* in *O. rufipogon*, *Lhs1* in *O. brachyantha* and *O. granulata*, *NP70* in *O. granulata*, and *Waxy* in *O. rhizomatis*, *O. brachyantha*, and *O. granulata* (Tables S3 and S4). To account for the potential impact of the nonneutral loci on evaluation of the correlation between genome size and effective population size, we filtered the dataset by removing the sequences from nonneutral loci and obtained the π_s/N_e values using the filtered dataset (Table 1). Although the π_s/N_e estimates were slightly different, paired *t*-test generated similar results, in which about half of total 45 species comparisons was not significant for both π_s and N_e (Figs. S3c and S3d, Table S5). Correlation analysis with the filtered data did not find significant relationship between N_e and genome size either (Fig. S5).

Lynch (2011) indicated that $N_e\mu$ (as estimated by π_s), rather than N_e was the relevant predictor variable to correlate with genome size. Thus, we performed correlation analyses between genome size and π_s based on the filtered datasets. The results indicated that no significant correlation was detected between genome size and π_s, both with ($r = 0.277$, $P = 0.470$) and without ($r = 0.317$, $P = 0.372$) phylogenetical corrections. Together, all analyses above provide no evidence that genome size is correlated with effective population size in the genus *Oryza*.

Discussion

The negative correlation between genome size and effective population size hypothesized by Lynch and Conery (2003) was based on nearly neutral theory and population genetic principles and seemingly applied in many cases across the web of life (Yi 2006; Lynch 2007; Boussau et al. 2011). The explanation power of N_e for several evolutionary questions is commonly accepted and selection efficiency related to N_e may explain for variation of some other biological attributes (e.g., codon bias, evolutionary rate, proportion of adaptive substitution, etc.) among different lineages of organisms (Lynch 2007; Charlesworth 2009). However, as pointed out by many authors (Charlesworth and Barton 2004; Daubin and Moran 2004; Yi 2006; Gregory and Witt 2008), the potential biases for obtaining reliable estimates of N_e such as mutation rate heterogeneity, taxonomy difficulties, sensitivity of molecular markers to demography should be considered cautiously. Phylogenetical nonindependence should be also taken into consideration in multiple species analysis (Whitney et al. 2010; Whitney and Garland 2010). In the present study, we tested Lynch and Conery's (2003) hypothesis using 10 diploid *Oryza* species in which fourfold genome size variation is present across species. We did not detect a negative relationship between N_e and genome size as predicted by Lynch and Conery (2003), despite the fact that alternate analyses were performed to account for the potential biases mentioned above.

It should be noted that genome size evolution is a complex process influenced by numerous evolutionary forces (Charlesworth and Barton 2004; Lynch 2007; Hawkins et al. 2008; Whitney et al. 2010). For instance, using multiple

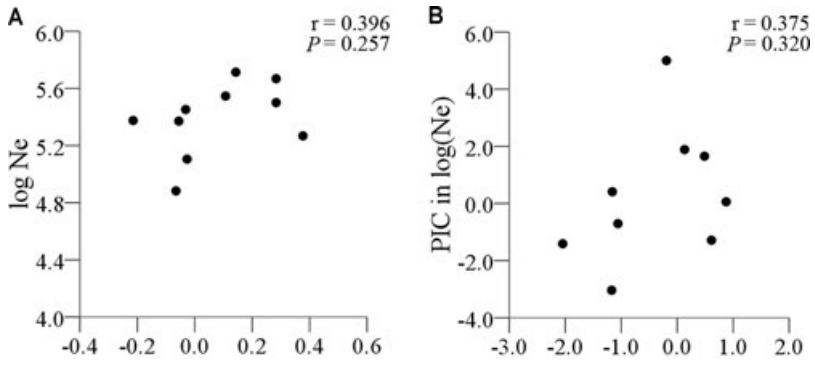

Figure 2. Plot of the correlation between mean N_e and genome size (C) without (a) and with (b) phylogenetical corrections based on all sequences.

regression analysis with PICs for the relation among N_e, genome size, and outcrossing rate in seed plants, Whitney et al. (2010) found no relationship between N_e and genome size but a weak relationship between outcrossing and genome size. In our case, 10 *Oryza* species diverged within the last 15 million years (Fig. 1 and Tang et al. 2010) with a majority of features in common. However, different mating systems have been recorded for the *Oryza* species (Vaughan 2004) despite few extensive investigations. Based on previous studies, we chose seven *Oryza* species with available records and divided them into two groups according to their mating systems: (1) predominantly inbreeding species, including *O. nivara* (Barbier 1989), *O. barthii* (Vaughan 2004), *O. eichingeri* (Jayasuriya and Vaughan 2003), and *O. granulata* (Qian et al. 2001); and (2) outcrossing species, including *O. rufipogon* (Barbier 1989), *O. officinalis* (Gao et al. 2001; Jayasuriya and Vaughan 2003), and *O. rhizomatis* (Jayasuriya and Vaughan 2003). Then, we compared the genome sizes between the two species groups with contrasting mating systems and found no significant relationship ($P = 0.917$), suggesting that the genome size variation in *Oryza* cannot be simply explained by mating system either.

A number of studies have detected recent bursts of several LTR-retrotransposon families in some *Oryza* species, which was considered as the main cause of the fourfold genome size variation in the diploid *Oryza* species (e.g., Piegu et al. 2006; Zuccolo et al. 2007). Fine-scale comparative analyses based on orthologous *Oryza* BAC sequences also supported the TE activity (Ammiraju et al. 2008; Lu et al. 2009; Sanyal et al. 2010). These reports indicated the important contribution of the noncoding elements like TE, which is one of the prerequisites for Lynch and Conery's (2003) hypothesis. However, the lineage-specific genome expansion could not be exclusively explained by the effect of N_e, as demonstrated by this study. As reviewed in Whitney et al. (2010), the principle theories for genome expansion could be classified as adaptive (Gregory and Hebert 1999; Bennett and Leitch 2005b), neutral (Petrov 2002; Oliver et al. 2007), and maladaptive ("junk DNA" theories) (Doolittle and Sapienza 1980; Lynch and Conery 2003). Although adaptive effects of larger genome size and numerous functions of noncoding DNA were proposed (Gregory and Hebert 1999; Bennett and Leitch 2005b), no direct evidence has been found in *Oryza*.

Petrov (2002) argued that the process of genome size evolution might fit a mutational equilibrium model, in which all insertions and deletions were neutral and organisms got their optimum genome sizes until DNA loss through small deletions was equal to DNA gain through long insertions. Although indel bias might be efficient for long-term evolution (Petrov 2001), it might not explain the vast change in genome sizes within such a short time scale in diversification of *Oryza*. Oliver et al. (2007) showed that the rate of genome size change was proportional to genome size, with a faster rate occurring in the larger genome. It was predicted that smaller genomes were more difficult to become large whereas larger genomes were easier to become small, and thus a skewed distribution toward smaller values would be found. Because this hypothesis might be proper for generalization over long time scales, it is expected that the extant C-values and reconstructed values at ancestral nodes in *Oryza* did not show such a skewed pattern (Fig. 1).

It is worth noting that a number of limitations might arise in our detection of correlation between genome size and effective population size. First, we used the method described in Zhang and Ge (2007) to correct for the heterogeneity of mutation rates across genes before estimating N_e. However, this correction was based on the divergence data between the target species and its corresponding outgroup under the molecular clock assumption that would be violated to some extents. Second, no significant correlation in our estimates may result partly from small sample size (10 data points). This cannot be avoided in our case that sampled all major lineages in *Oryza*, but should be taken into consideration in further investigations using closely related species. Finally, it is likely that N_e reflects the coalescence history while changes in genome size (e.g., due to TE proliferation, Piegu et al. 2006; Zuccolo

et al. 2007) might be rapid and recent. Therefore, evolutionary time scale might have different impacts on the estimates of effective population size and genome size, and thus should be considered with caution. To sum up, genome size evolution is a complex process influenced by several evolutionary forces, and therefore the genome size variation in the diploid *Oryza* species could not be simply explained by one of the above theories. Investigation of the fitness significance of important LTR-retrotransposon families with large contribution to genome size should be considered in future study in *Oryza*. Particularly, multivariate analysis based on reliable estimates of N_e, in conjunction with phylogenetic correction, is required at different taxonomic levels to distinguish the relative contribution of correlated variables to genome size variation.

ACKNOWLEDGMENTS

We thank X. h. Zou, F. M. Zhang, L. L. Zang, L. Huang, X. M. Zheng, and other members of Ge's group for technical assistances, and A. Cutter, the associate editor, and two anonymous reviewers for valuable comments and suggestions on the manuscript. We are grateful to D. A. Vaughan for providing some leaf samples, and to the International Rice Research Institute (Los Banos, Philippines) for providing seed samples. This work was supported by the National Natural Science Foundation of China (30990240) and the National Basic Research Program of China (2007CB815704).

LITERATURE CITED

Ammiraju, J. S. S., M. Luo, J. L. Goicoechea, W. Wang, D. Kudrna, C. Mueller, J. Talag, H. Kim, N. B. Sisneros, B. Blackmon, et al. 2006. The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. Genome Res. 16:140–147.

Ammiraju, J. S. S., F. Lu, A. Sanyal, Y. Yu, X. Song, N. Jiang, A. C. Pontaroli, T. Rambo, J. Currie, K. Collura, et al. 2008. Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical dataset. Plant Cell 20:3191–3209.

Barbier, P. 1989. Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. II. influence of the mating system and life-history traits on the genetic structure of populations. Jap. J. Genet. 64:273–285.

Bennett, M. D., and I. J. Leitch. 2005a. Plant DNA C-values database (release 4.0, Dec. 2005). Royal Botanic Gardens, Kew. Available at: http://data.kew.org/cvalues/.

———. 2005b. Genome size evolution in plants. Pp. 89–162 *in* T. R. Gregory, ed. The evolution of the genome. Elsevier, Amsterdam.

Boussau, B., J. M. Brown, and M. K. Fujita. 2011. Nonadaptive evolution of mitochondrial genome size. Evolution 65:2706–2711.

Charlesworth, B. 2008. The origin of genomes—not by natural selection? Curr. Biol. 18:R140–R141.

———. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10:195–205.

Charlesworth, B., and N. Barton. 2004. Genome size: does bigger mean worse? Curr. Biol. 14:R233–R235.

Daubin, V., and N. A. Moran. 2004. Comment on "The Origins of Genome Complexity". Science 306:978a.

Doolittle, W. F., and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Flowers, J. M., and M. D. Purugganan. 2008. The evolution of plant genomes—scaling up from a population perspective. Curr. Opin. Genet. Dev. 18:565–570.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.

Gao, L. Z., S. Ge, and D. Y. Hong. 2001. High levels of genetic differentiation of *Oryza officinalis* Wall. et Watt. from China. J. Heredity 92:511–516.

Garland, T., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 41:18–32.

Gaut, B. S., B. R. Morton, B. C. McCaig, and M. T. Clegg. 1996. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc. Natl. Acad. Sci. USA 93:10274–10279.

Ge, S., T. Sang, B. R. Lu, and D. Y. Hong. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. USA 96:14400–14405.

Gregory, T. R., and P. D. N. Hebert. 1999. The modulation of DNA content: proximate causes and ultimate consequences. Genome Res. 9: 317–324.

Gregory, T. R., and J. D. S. Witt. 2008. Population size and genome size in fishes: a closer look. Genome 51:309–313.

Hawkins, J. S., C. E. Grover, and J. F. Wendel. 2008. Repeated big bangs and the expanding universe: directionality in plant genome size evolution. Plant Sci. 174:557–562.

Hudson, R. R., and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA-sequences. Genetics 111:147–164.

Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159.

Jayasuriya, A. H. M., and D. A. Vaughan. 2003. Conservation and use of crop wild relatives. Proceedings of the joint Department of Agriculture, Sri Lanka and National Institute of Agrobiological Sciences, Japan workshop. Plant Genetic Resource Center, Department of Agriculture, Peradeniya, Sri Lanka.

Kuo, C. H., N. A. Moran, and H. Ochman. 2009. The consequences of genetic drift for bacterial genome complexity. Genome Res. 19:1450–1454.

Li, Z. M., X. M. Zheng, and S. Ge. 2011. Genetic diversity and domestication history of African rice (*Oryza glaberrima*) as inferred from multiple gene sequences. Theor. Appl. Genet. 123:21–31.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Lu, F., J. S. S. Ammiraju, A. Sanyal, S. L. Zhang, R. T. Song, J. F. Chen, G. S. Li, Y. Sui, X. Song, Z. K. Cheng, et al. 2009. Comparative sequence analysis of *MONOCULM1*-orthologous regions in 14 *Oryza* genomes. Proc. Natl. Acad. Sci. USA 106:2071–2076.

Lynch, M. 2007. The origins of genome architecture. Sinauer Associates, Sunderland, MA.

———. 2011. Statistical inference on the mechanisms of genome evolution. PLoS Genet. 7: e1001389.

Lynch, M., and J. S. Conery. 2003. The origins of genome complexity. Science 302:1401–1404.

Maddison, W. P., and D. R. Maddison. 2010. Mesquite: a modular system for evolutionary analysis, v. 2.74. Available at: http://mesquiteproject.org.

Midford, P. E., T. Garland Jr, and W. Maddison. 2002. PDAP:PDTREE package for Mesquite, v. 1.00. Available at: http://mesquiteproject.org/pdap_mesquite/.

Miyabayashi, T., K. I. Nonomura, H. Morishima, and N. Kurata. 2007. Genome size of twenty wild species of *Oryza* determined by flow cytometric and chromosome analyses. Breed. Sci. 57: 73–78.

Oliver, M. J., D. Petrov, D. Ackerly, P. Falkowski, and O. M. Schofield. 2007. The mode and tempo of genome size evolution in eukaryotes. Genome Res. 17:594–601.

Pagel, M., and A. Meade. 2009. BayesTraits. University of Reading, UK. Available at: http://www.evolution.rdg.ac.uk/BayesTraits.html.

Petrov, D. A. 2001. Evolution of genome size: new approaches to an old problem. Trends Genet. 17:23–28.

———. 2002. Mutational equilibrium model of genome size evolution. Theor. Popul. Biol. 61:531–544.

Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, K. Collura, D. S. Brar, S. Jackson, R. A. Wing, et al. 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res. 16:1262–1269.

Qian, W., S. Ge, and D. Y. Hong. 2001. Genetic variation within and among populations of a wild rice *Oryza granulata* from China detected by RAPD and ISSR markers. Theor. Appl. Genet. 102:440–449.

Sanyal, A., J. S. S. Ammiraju, F. Lu, Y. Yu, T. Rambo, J. Currie, K. Kollura, H. Kim, J. F. Chen, J. X. Ma, et al. 2010. Orthologous comparisons of the *Hd1* region across genera reveal *Hd1* Gene lability within diploid *Oryza* species and disruptions to microsynteny in *Sorghum*. Mol. Biol. Evol. 27:2487–2506.

Shimamoto, K., and J. Kyozuka. 2002. Rice as a model for comparative genomics of plants. Annu. Rev. Plant Biol. 53:399–419.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Tang, L., X. H. Zou, G. Achoundong, C. Potgieter, G. Second, D. Y. Zhang, and S. Ge. 2010. Phylogeny and biogeography of the rice tribe (*Oryzeae*): evidence from combined analysis of 20 chloroplast fragments. Mol. Phylogen. Evol. 54:266–277.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.

Vaughan, D. A. 2004. The wild relatives of rice: a genetic resources handbook. International Rice Research Institute, Manila, Philippines.

Vinogradov, A. E. 2004. Evolution of genome size: multilevel selection, mutation bias or dynamical chaos? Curr. Opin. Genet. Dev. 14:620–626.

Whitney, K. D., and T. Garland Jr. 2010. Did genetic drift drive increases in genome complexity? PLoS Genet. 6:e1001080.

Whitney, K. D., E. J. Baack, J. L. Hamrick, M. J. Godt, B. C. Barringer, M. D. Bennett, C. G. Eckert, C. Goodwillie, S. Kalisz, I. J. Leitch, et al. 2010. A role for nonadaptive processes in plant genome size evolution? Evolution 64:2097–2109.

Whitney, K. D., B. Boussau, E. J. Baack, and T. Garland Jr. 2011. Drift and genome complexity revisited. PLoS Genet. 7: e1002092.

Wing, R. A., J. S. Ammiraju, M. Luo, H. Kim, Y. Yu, D. Kudrna, J. L. Goicoechea, W. Wang, W. Nelson, K. Rao, et al. 2005. The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. Plant Mol. Biol. 59:53–62.

Wright, S. I., B. Lauga, and D. Charlesworth. 2002. Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. Mol. Biol. Evol. 19:1407–1420.

Yi, S. V. 2006. Non-adaptive evolution of genome complexity. Bioessays 28:979–982.

Yi, S. V., and J. T. Streelman. 2005. Genome size is negatively correlated with effective population size in ray-finned fish. Trends Genet. 21:643–646.

Zhang, L. B., and S. Ge. 2007. Multilocus analysis of nucleotide variation and speciation in *Oryza officinalis* and its close relatives. Mol. Biol. Evol. 24:769–783.

Zhu, Q. H., and S. Ge. 2005. Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytol. 167:249–265.

Zhu, Q. H., X. M. Zheng, J. C. Luo, B. S. Gaut, and S. Ge. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol. Biol. Evol. 24:875–888.

Zou, X. H., F. M. Zhang, J. G. Zhang, L. L. Zang, L. Tang, J. Wang, T. Sang, and S. Ge. 2008. Analysis of 142 genes resolves the rapid diversification of the rice genus. Genome Biol. 9:R49.

Zuccolo, A., A. Sebastian, J. Talag, Y. Yu, H. Kim, K. Collura, D. Kudrna, and R. A. Wing. 2007. Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. BMC Evol. Biol. 7:152.

Associate Editor: A. Cutter

Supporting Information

The following supporting information is available for this article:

Figure S1. Schematic diagrams of the nuclear loci and locations of the regions sequenced.
Figure S2. Boxplots of π_s (filled box) and N_e (open box, in millions) values for 10 diploid *Oryza* species.
Figure S3. Summary of the significance of the paired *t*-test for comparison of π_s (a and c) and N_e (b and d) among 10 *Oryza* species, based on all sequences (a and b) and the filtered datasets in which all sequences from nonneutral loci were removed (c and d).
Figure S4. Tree with node numbers (used in Table S6), and extant and reconstructed genome size values (in parentheses).
Figure S5. Plot of the correlation between mean N_e and genome size (C) without (a) and with (b) phylogenetical corrections, based on the filtered datasets in which all sequences from nonneutral loci were removed.
Table S1. List of plant materials sampled in this study.
Table S2. Summary of the loci surveyed and the primer sequences used in this study.
Table S3. Summary of silent nucleotide polymorphism, neutrality tests, estimates of mutation rate, and effective population size for each locus and each species.
Table S4. Summary of the results of multilocus HKA test.
Table S5. Summary of the paired *t*-test for comparison of π_s (a and c) and N_e (b and d) values based on all sequences (a and b) and the filtered datasets in which all sequences from nonneutral loci were removed (c and d). The numbers of loci (upper triangle) and the significance (lower triangle) are indicated.
Table S6. Columns in files of independent contrasts (FIC) and standardized contrasts.

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