

Locally Duplicated Ohnologs Evolve Faster Than Nonlocally Duplicated Ohnologs in *Arabidopsis* and Rice

Yupeng Wang*

Computational Biology Service Unit, Cornell University

*Corresponding author: E-mail: wyp1125@gmail.com.

Accepted: January 22, 2013

Abstract

Whole-genome duplications (WGDs) have recurred in the evolution of angiosperms, resulting in many duplicated chromosomal segments. Local gene duplications are also widespread in angiosperms. WGD-derived duplicates, that is, ohnologs, and local duplicates often show contrasting patterns of gene retention and evolution. However, many genes in angiosperms underwent multiple gene duplication events, possibly by different modes, indicating that different modes of gene duplication are not mutually exclusive. In two representative angiosperm genomes, *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*), we found that 9.6% and 11.3% of unique ohnologs, corresponding to 15.5% and 17.1% of ohnolog pairs, were also involved in local duplications, respectively. Locally duplicated ohnologs are widely distributed in different duplicated chromosomal segments and functionally biased. Coding sequence divergence between duplicated genes is denoted by nonsynonymous (K_a) and synonymous (K_s) substitution rates. Locally duplicated ohnolog pairs tend to have higher K_a , K_a/K_s , and gene expression divergence than nonlocally duplicated ohnolog pairs. Locally duplicated ohnologs also tend to have higher interspecies sequence divergence. These observations indicate that locally duplicated ohnologs evolve faster than nonlocally duplicated ohnologs. This study highlights the necessity to take local duplications into account when analyzing the evolutionary dynamics of ohnologs.

Key words: local gene duplication, whole-genome duplication, ohnolog, divergence, colinearity.

Introduction

With tens of plant genomes sequenced, it is now clear that whole-genome duplications (WGDs) are widespread in angiosperm (flowering plant) genomes (Bowers et al. 2003; Paterson et al. 2010). WGDs resulted in many duplicated chromosomal segments. However, computational identification of duplicated chromosomal segments is a challenging task. Besides phylogenomics methods (Bowers et al. 2003), detection of synteny (homologous genes remaining in corresponding chromosomal regions) is often adopted (Wei et al. 2007; Tang et al. 2008a, 2008b, 2010). One general method for synteny detection is the clustering of neighboring matching gene pairs (Vandepoele et al. 2002; Hampson et al. 2003; Luc et al. 2003; Ling et al. 2009; Vergara and Chen 2009). However, gene losses, tandem duplicates, and chromosomal rearrangements often complicate synteny detection. Collinearity, a more specific form of synteny that requires conserved gene order, can facilitate more accurate detection of synteny. Thus, another general method for synteny detection uses dynamic programming to chain pair-wise collinear genes, often based on a certain scoring scheme that rewards

adjacent collinear genes and penalizes the distance difference between them (Haas et al. 2004; Soderlund et al. 2006; Wang et al. 2006, 2012; Tang et al. 2008b; Rodelsperger and Dieterich 2010).

In angiosperm genomes, the duplicated genes located at syntenic/collinear positions between duplicated chromosomal segments were thought to arise from WGDs (Bowers et al. 2003; Paterson et al. 2004) and often referred to as ohnologs (Wolfe 2000). Single-gene duplications including local (tandem or proximal), transposed, and dispersed duplications are also widespread in angiosperms (Freeling 2009; Wang Y, et al. 2011; Wang Y, Wang X, et al. 2012). Indeed, the different modes of gene duplication represent different evolutionary origins of duplicated genes, depending on which, duplicated genes often show different evolutionary patterns and consequences (Wang Y, Wang X, et al. 2012). The retention of ohnologs and local duplicates is functionally biased and often reciprocal in many gene families: WGDs tend to retain transcription factors, whereas local duplicates tend to retain the genes functioning in stress response (Freeling 2009; Rodgers-Melnick et al. 2012). Single-gene

duplications frequently result in more divergent gene expression between duplicated genes than WGDs (Casneuf et al. 2006; Ganko et al. 2007; Wang Y, et al. 2011). Moreover, local duplicates were suggested to be important for lineage-specific adaptive evolution to rapidly changing environments (Hanada et al. 2008).

However, the evolutionary origins of many duplicated genes are not unique. In many synteny detection tools, tandem duplicates are collapsed before synteny detection (Lyons et al. 2008; Tang et al. 2008b; Proost et al. 2012; Wang et al. 2012), which indicates that in a genome, duplicated genes may be simultaneously located within syntenic positions and tandem arrays, further suggesting that duplicated genes may have undergone both WGDs and local duplications. The MCScanX software has enabled the recovery of “collinear tandem arrays” after synteny detection (Wang et al. 2012). However, a genome-wide evolutionary analysis of the duplicated genes with multiple origins has yet to be reported.

In two representative angiosperm genomes, *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*), we identified the ohnologs that were also locally duplicated, which were named locally duplicated ohnologs (the remaining ohnologs were named nonlocally duplicated ohnologs). We examined whether locally duplicated and nonlocally duplicated ohnologs have different levels of coding sequence and expression divergence. The evolutionary impact of local duplications of ohnologs was investigated.

Materials and Methods

Genome Annotations

Genome annotations for *A. thaliana* (Swarbreck et al. 2008), *B. rapa* (Wang X, et al. 2011), *O. sativa* (Ouyang et al. 2007), and *Bra. distachyon* (International Brachypodium Initiative 2010) were obtained from Phytozome 8.0 (<http://www.phytozome.net>, last accessed February 6, 2013). Transposable element-related genes in rice were removed. If a gene has multiple transcripts, only the first transcript in the genome annotation was used in related analyses.

Homology Search

To search for potential duplicated genes (i.e., the homology input of MCScanX) in *Arabidopsis* and rice, the protein-coding genes were compared against themselves within their genomes using BLASTP (Altschul et al. 1990). For a protein sequence, the best five nonself-hits that met an *E*-value threshold of 10^{-10} were kept. For each *Arabidopsis* or rice gene, to search for its closest ortholog in an outgroup species, the best BLASTP hit with *E* value $\leq 10^{-10}$ was retained.

Coding Sequence Divergence

Coding sequence divergence was measured by nonsynonymous (*Ka*) and synonymous (*Ks*) substitution rates.

To generate a coding sequence alignment for a pair of homologous genes, an alignment of their protein sequences was first generated using Clustalw (Thompson et al. 1994) with default parameters. Then, the protein alignment was transformed into a coding sequence alignment using the “Bio::Align::Utilities” module of the BioPerl package (<http://www.bioperl.org/>, last accessed February 6, 2013). *Ka* and *Ks* were computed using Nei–Gojobori method (Nei and Gojobori 1986), available in the “Bio::Align::DNASStatistics” module of the BioPerl package. Note that extremely high levels of sequence divergence between homologous genes may cause the “Bio::Align::DNASStatistics” module to generate invalid *Ka* or *Ks*, which were excluded from related analysis. *Ks* < 3.0 were involved in related statistical analyses.

GO Term Enrichment Analysis

GO term enrichment analysis was performed using Fisher’s exact test. The *P* value was corrected with the total number of terms (i.e., Bonferroni correction) to account for multiple comparisons. GO terms with corrected *P* value < 0.05 were returned.

Gene Expression Data

Processed *Arabidopsis* and rice microarray data measured by the Affymetrix *Arabidopsis* ATH1 Genome Array (GEO platform GPL198) and GeneChip Rice Genome Array (GEO platform GPL2020), respectively, were obtained from a previous study (Wang Y, et al. 2011). The microarray data contained 4,566 and 508 samples in *Arabidopsis* and rice, respectively. To avoid the expression values generated by cross-hybridization, probe sets that were mapped to multiple genes were excluded from analysis. (Note that the probe sets that were mapped to multiple transcripts of the same gene were kept.) For a gene, there may be multiple probe sets available on the array, and the first probe set according to alphabetic sorting was used to represent its expression profile. In *Arabidopsis*, 448 (60.4%) locally duplicated ohnolog pairs and 3,015 (74.7%) nonlocally duplicated ohnolog pairs have expression profiles. In rice, 419 (70.4%) locally duplicated ohnolog pairs and 2,373 (82.1%) nonlocally duplicated ohnolog pairs have expression profiles. Similarity between the expression profiles of two genes was initially measured by Pearson’s correlation coefficient (denoted by *r*). Their corresponding expression divergence was measured by $1 - r$ (Liao and Zhang 2008; Liao et al. 2010).

Results

Identification of Locally Duplicated Ohnologs

BLASTP (Altschul et al. 1990) was used to generate the whole set of duplicates in *Arabidopsis* and rice (see Materials and Methods). MCScanX (Wang et al. 2012) was then implemented (parameters set as default except maximum

gaps = 40) to detect gene synteny/collinearity and duplicated chromosomal segments. A total of 4,775 and 3,484 ohnolog pairs, corresponding to 7,839 and 5,968 unique ohnologs, were identified in *Arabidopsis* and rice, respectively.

To detect locally duplicated ohnologs, local gene duplications, defined as two duplicated genes within 20 annotated genes of each other on the same chromosomes without any paralog (BLASTP hit) between them, were identified. If an ohnolog was involved in a local duplication, it was classified as a locally duplicated ohnolog. If at least one ohnolog in an ohnolog pair was locally duplicated, this ohnolog pair was deemed a locally duplicated ohnolog pair. A total of 754 (9.6%) and 675 (11.3%) ohnologs, corresponding to 741 (15.5%) and 595 (17.1%) ohnolog pairs, were identified as locally duplicated ohnologs in *Arabidopsis* and rice, respectively (listed in [supplementary table S1, Supplementary Material](#) online).

For locally duplicated ohnologs, it would be interesting to distinguish between ohnologs that were locally duplicated after WGDs and locally duplicated genes that were retained following WGDs. Synonymous (K_s) substitution rates are often used as a proxy of time since divergence (Li 1997). However, using K_s as a proxy of time has great uncertainty. For example, although ohnologs belonging to the same WGD event were created simultaneously, their K_s indeed have a wide range (Zhang et al. 2002; Blanc and Wolfe 2004b). In addition, K_s increase approximately linearly with time only for relatively low levels of sequence divergence (Li 1997). Thus, for two gene duplications, the order of their K_s levels is not a reliable indicator of the sequence of the duplications. However, it may be reasonable to compare the relative ages between large groups of genes based on K_s because the variances of K_s can be estimated. For locally duplicated ohnologs, local duplications have significantly smaller K_s than ohnolog pairs in both *Arabidopsis* and rice (P value $\leq 1.875 \times 10^{-6}$, two-sample t -test), suggesting that local duplications are relatively younger than WGDs, which is consistent with previous studies (Blanc and Wolfe 2004b; Yu et al. 2005). Note that this observation may not suggest that the rate of local duplications increases after WGDs but rather confirm the fact that the turnover rate of duplicates following local duplications is much higher than that following WGDs, making a very small portion of local duplicates survive over long evolutionary times (Lynch and Conery 2000; Blanc and Wolfe 2004b).

Between the syntenic positions defined by a locally duplicated ohnolog pair, there may be multiple duplicate pairs. For each locally duplicated ohnolog pair, the best duplicate pair, defined as the one with the smallest K_s among all possible duplicate pairs between its syntenic positions, was also recorded. The terms describing different relationships among the duplicated genes in duplicated chromosomal segments are explained in figure 1.

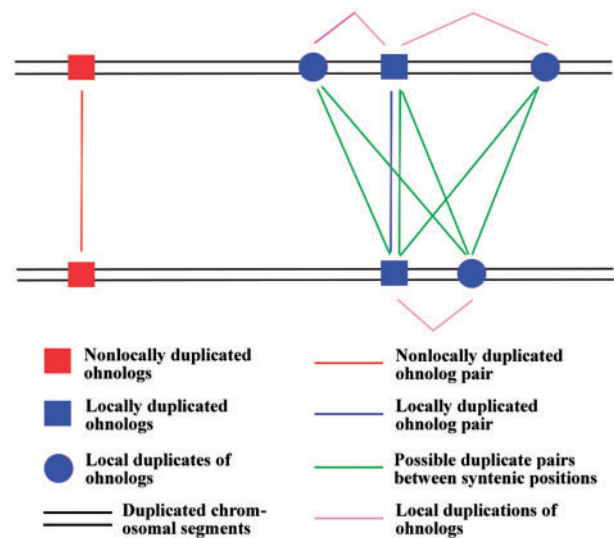


Fig. 1.—Different relationships among the duplicated genes in duplicated chromosomal segments and their terms in this study.

Chromosomal Distributions of Locally Duplicated Ohnologs

Genome circle plots were made to visualize the chromosomal distributions of locally duplicated ohnologs. In both *Arabidopsis* and rice, locally duplicated ohnologs are widely distributed in different duplicated chromosomal segments that cover all chromosomes (fig. 2), suggesting that local duplications of ohnologs are not specific to certain chromosomal regions.

Locally Duplicated Ohnologs Are Functionally Biased

The retention of duplicated genes following WGDs in terms of gene functions has been extensively studied. There is a consensus that WGDs tend to retain genes involved in transcription regulation and signal transduction (Blanc and Wolfe 2004a; Maere et al. 2005). Some studies also reported that metabolic genes are likely to be over-retained following WGDs (Rensing et al. 2007; Gout et al. 2009). To evaluate the functional characteristics of locally duplicated ohnologs, GO term enrichment analysis was performed using all ohnologs as the reference database. In *Arabidopsis* and rice, there were 22 and 26 enriched GO terms in locally duplicated ohnologs, respectively ([supplementary table S2, Supplementary Material](#) online). The common enriched GO terms between *Arabidopsis* and rice included heme binding (GO:0020037), electron carrier activity (GO:0009055), iron ion binding (GO:0005506), oxidation–reduction process (GO:0055114), monooxygenase activity (GO:0004497), and recognition of pollen (GO:0048544). This analysis suggests that locally duplicated ohnologs are functionally biased.

When all local duplicates were used as the reference database, there were no enriched GO terms in *Arabidopsis* locally

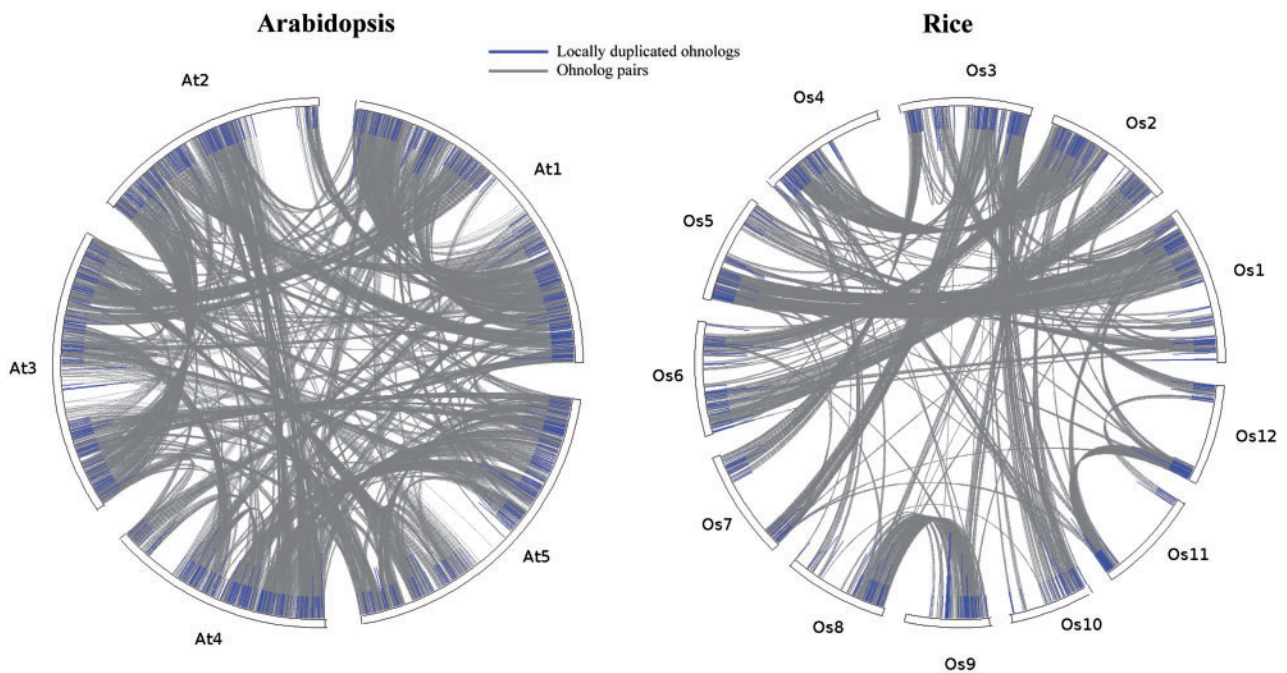


Fig. 2.—Genomic distributions of locally duplicated ohnologs in *Arabidopsis* and rice. At, *Arabidopsis thaliana*; Os, *Oryza sativa*.

duplicated ohnologs but 21 enriched GO terms in rice locally duplicated ohnologs (supplementary table S3, Supplementary Material online). Thus, in *Arabidopsis*, the functional biases of locally duplicated ohnologs reflect the general trend of local duplicates, while in rice, locally duplicated ohnologs show a high level of functional specificity.

Locally Duplicated Ohnologs Evolve Faster Than Nonlocally Duplicated Ohnologs

To study the evolutionary impact of local duplications of ohnologs, we first compared the divergence of ohnologs between locally duplicated and nonlocally duplicated ohnolog pairs. To ensure reliable comparison, the divergence of duplicates in the best duplicate pairs between the syntenic positions defined by locally duplicated ohnolog pairs and the local duplications of ohnologs was also assessed. In both *Arabidopsis* and rice, locally duplicated ohnolog pairs have significantly higher nonsynonymous (K_a) substitution rates than nonlocally duplicated ohnolog pairs (fig. 3A), indicating that locally duplicated ohnolog pairs tend to have more divergent protein sequences. Locally duplicated ohnolog pairs also show higher K_a/K_s (fig. 3B), suggesting that purifying selection acting on locally duplicated ohnologs is relatively relaxed. Comparison of gene expression divergence shows that locally duplicated ohnolog pairs have more divergent expression profiles than nonlocally duplicated ohnolog pairs (fig. 3C). The divergence of duplicates is comparable between locally duplicated ohnolog pairs and the best duplicate pairs between the syntenic positions defined by locally duplicated ohnolog pairs (fig. 3A–C),

indicating that the higher divergence between locally duplicated ohnologs was not caused by incorrect selection of ohnolog pairs between the syntenic positions with local duplications. The divergence of duplicates in local duplications of ohnologs is higher than that in nonlocally duplicated ohnolog pairs (fig. 3A–C), implying that local duplicates, though are in general younger, tend to evolve faster than ohnologs.

Next, we compared interspecies divergence between locally duplicated and nonlocally duplicated ohnologs. To this end, the K_a and K_s values between each *Arabidopsis* or rice ohnolog and its closest ortholog in *Brassica rapa* or *Brachypodium distachyon*, respectively, were computed. In both *Arabidopsis* and rice, locally duplicated ohnologs have significantly higher K_a and K_a/K_s than nonlocally duplicated ohnologs (fig. 4), suggesting that locally duplicated ohnologs also show higher levels of interspecies sequence divergence. In partial summary, locally duplicated ohnologs tend to have higher levels of sequence and expression divergence than nonlocally duplicated ohnologs, indicating that locally duplicated ohnologs evolve faster than nonlocally duplicated ohnologs.

Discussion

Previous studies for comparing different modes of gene duplication often assigned a unique origin to each duplicate gene (Guan et al. 2007; Hakes et al. 2007; Wang Y, et al. 2011). Although WGDs and local duplications were found to have different functional biases and evolutionary patterns (Freeling 2009; Rodgers-Melnick et al. 2012), this study

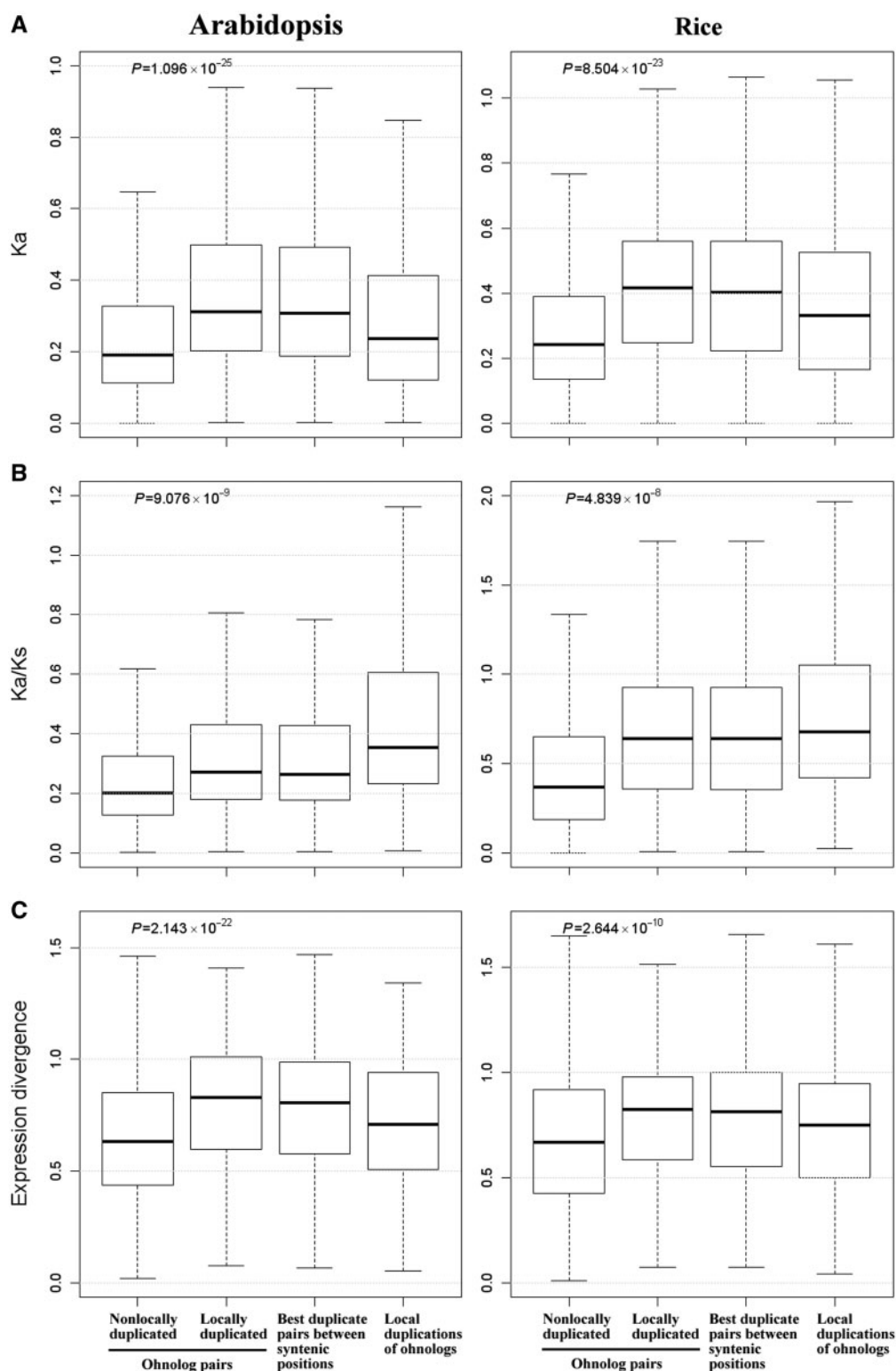


FIG. 3.—Comparisons of divergence of duplicates between locally duplicated and nonlocally duplicated ohnolog pairs in *Arabidopsis* and rice. Statistical comparisons between locally duplicated and nonlocally duplicated ohnolog pairs were performed using two-sample *t*-test, whose *P* values are indicated. The divergence of duplicates in the best duplicate pairs between the syntenic positions defined by locally duplicated ohnolog pairs and the local duplications of ohnologs is also shown. (A) Comparison of *Ka*. (B) Comparison of *Ka/Ks*. (C) Comparison of expression divergence.

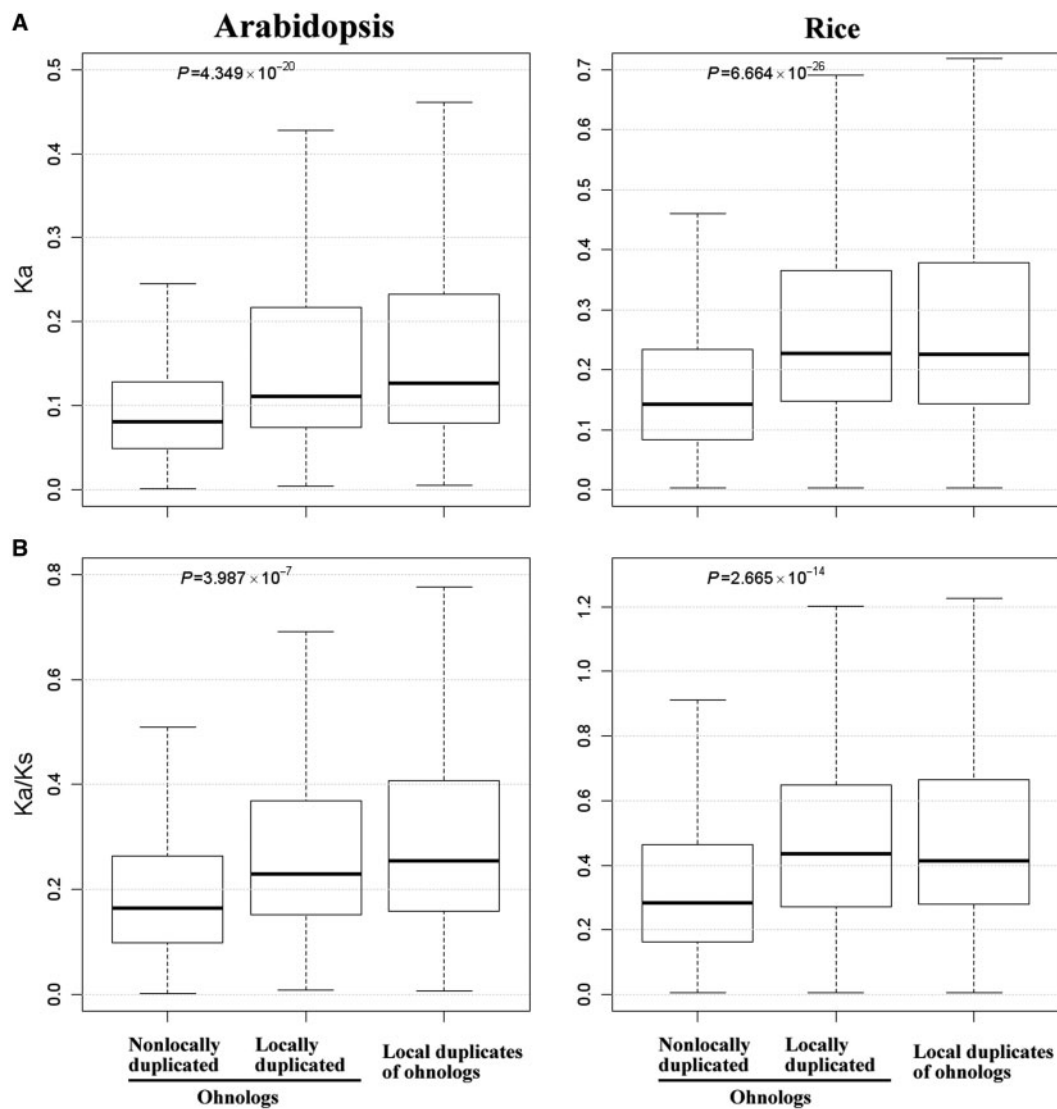


Fig. 4.—Comparisons of interspecies divergence between locally duplicated and nonlocally duplicated ohnologs. *Arabidopsis* and rice ohnologs were compared with their closest orthologs in *Brassica rapa* and *Brachypodium distachyon*, respectively. Statistical comparisons between locally duplicated and nonlocally duplicated ohnologs were performed using two-sample *t*-test, whose *P* values are indicated. (A) Comparison of *Ka*. (B) Comparison of *Ka/Ks*.

showed that approximately 10% of ohnologs have also experienced local duplications and the ohnologs that underwent local duplications have different evolutionary patterns than those that did not. In *Arabidopsis* and rice, local duplicates account for approximately 18.5% of the genes in the genome. The fact that approximately 10% of ohnologs also experienced local duplications might not indicate that ohnologs are less likely to be locally duplicated. If all duplicates in the local duplications of ohnologs are counted in ohnologs similar to a previous study (Bowers et al. 2003), the percentage of locally duplicated ohnologs increases to approximately 18%, suggesting that ohnologs and other types of genes may have comparable probabilities of being locally duplicated. Thus, to analyze the evolutionary dynamics

of ohnologs, it is necessary to take local duplications into account.

Higher levels of *Ka/Ks* in locally duplicated ohnologs than nonlocally duplicated ohnologs can be explained by the gene balance hypothesis, which postulates that any successful genome has evolved an optimal balance of gene products that bind with one another to form protein complexes or are involved in multiple steps of biological pathways (Birchler and Veitia 2007; Freeling 2009). WGDs duplicate all nuclear genes at once, and dosage balance among genes is not altered. To maintain proper gene networks following WGDs, the dosage balance constraints impose strong purifying selection on dosage-sensitive ohnologs. Thus, among different types of duplicated genes, ohnologs tend to have smaller

Ka/Ks values. In contrast, local duplications create only one more duplicated gene each time, and dosage imbalance is likely to occur. Thus, the local duplicate has a higher chance of being removed through accumulation of degenerative mutations, which implies a higher level of Ka/Ks. However, during accumulation of degenerative mutations, the local duplicate may occasionally evolve modified/new functions, rendering it finally preserved. If an ohnolog is locally duplicated, its dosage balance is broken and the evolutionary fate of the ohnolog becomes similar to local duplicates, resulting in an elevated level of Ka/Ks. Elevated Ka/Ks levels for local duplicates may provide them with greater potential of evolving lineage-specific functions (Hanada et al. 2008). Thus, there is a possibility that a part of ohnologs contribute to lineage-specific evolution by undergoing additional local duplications.

Supplementary Material

Supplementary tables S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>).

Acknowledgments

The author thanks Dr Andrew Paterson (PGML, University of Georgia) for direction during his Ph.D. studies, Drs Qi Sun and Hsiao-Pei Yang (CBSU, Cornell University) for helpful discussion, and Dr Katie Hyma (CBSU, Cornell University) for critical reading of the manuscript. This work was supported by a National Science Foundation grant (IOS #1127017) to Qi Sun.

Literature Cited

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Birchler JA, Veitia RA. 2007. The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19:395–402.
- Blanc G, Wolfe KH. 2004a. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16:1679–1691.
- Blanc G, Wolfe KH. 2004b. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667–1678.
- Bowers JE, Chapman BA, Rong JK, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438.
- Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y. 2006. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biol.* 7:R13.
- Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol.* 60:433–453.
- Ganko EW, Meyers BC, Vision TJ. 2007. Divergence in expression between duplicated genes in *Arabidopsis*. *Mol Biol Evol.* 24:2298–2309.
- Gout JF, Duret L, Kahn D. 2009. Differential retention of metabolic genes following whole-genome duplication. *Mol Biol Evol.* 26:1067–1072.
- Guan Y, Dunham MJ, Troyanskaya OG. 2007. Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* 175:933–943.
- Haas BJ, Delcher AL, Wortman JR, Salzberg SL. 2004. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20:3643–3646.
- Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL. 2007. All duplicates are not equal: the difference between small-scale and genome duplication. *Genome Biol.* 8:R209.
- Hampson S, McLysaght A, Gaut B, Baldi P. 2003. LineUp: statistical detection of chromosomal homology with application to plant comparative genomics. *Genome Res.* 13:999–1010.
- Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH. 2008. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148:993–1003.
- International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463:763–768.
- Li WH. 1997. *Molecular evolution*. Sunderland (MA): Sinauer Associates.
- Liao BY, Weng MP, Zhang J. 2010. Contrasting genetic paths to morphological and physiological evolution. *Proc Natl Acad Sci U S A.* 107:7353–7358.
- Liao BY, Zhang J. 2008. Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc Natl Acad Sci U S A.* 105:6987–6992.
- Ling X, He X, Xin D. 2009. Detecting gene clusters under evolutionary constraint in a large number of genomes. *Bioinformatics* 25:571–577.
- Luc N, Rislis JL, Bergeron A, Raffinot M. 2003. Gene teams: a new formalization of gene clusters for comparative genomics. *Comput Biol Chem.* 27:59–67.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Lyons E, et al. 2008. Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rodents. *Plant Physiol.* 148:1772–1781.
- Maere S, et al. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A.* 102:5454–5459.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 3:418–426.
- Ouyang S, et al. 2007. The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* 35:D883–D887.
- Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci U S A.* 101:9903–9908.
- Paterson AH, Freeling M, Tang H, Wang X. 2010. Insights from the comparison of plant genome sequences. *Annu Rev Plant Biol.* 61:349–372.
- Proost S, et al. 2012. i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* 40:e11.
- Rensing SA, et al. 2007. An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol Biol.* 7:130.
- Rodgersperger C, Dieterich C. 2010. CYNTENATOR: progressive gene order alignment of 17 vertebrate genomes. *PLoS One* 5:e8861.
- Rodgers-Melnick E, et al. 2012. Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. *Genome Res.* 22:95–105.
- Soderlund C, Nelson W, Shoemaker A, Paterson A. 2006. SyMAP: a system for discovering and viewing syntenic regions of FPC maps. *Genome Res.* 16:1159–1168.
- Swarbreck D, et al. 2008. The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* 36:D1009–D1014.

- Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci U S A*. 107:472–477.
- Tang H, et al. 2008a. Synteny and collinearity in plant genomes. *Science* 320:486–488.
- Tang H, et al. 2008b. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res*. 18: 1944–1954.
- Thompson JD, Higgins DG, Gibson TJ. 1994. Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.
- Vandepoele K, Saeys Y, Simillion C, Raes J, Van De Peer Y. 2002. The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between *Arabidopsis* and rice. *Genome Res*. 12:1792–1801.
- Vergara IA, Chen N. 2009. Using OrthoCluster for the detection of synteny blocks among multiple genomes. *Curr Protoc Bioinformatics*. 27: 6.10.11–16.10.18.
- Wang X, et al. 2006. Statistical inference of chromosomal homology based on gene colinearity and applications to *Arabidopsis* and rice. *BMC Bioinformatics* 7:447.
- Wang X, et al. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*. 43:1035–1039.
- Wang Y, et al. 2011. Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. *PLoS One* 6:e28150.
- Wang Y, et al. 2012. MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 40:e49.
- Wang Y, Wang X, Paterson AH. 2012. Genome and gene duplications and gene expression divergence: a view from plants. *Ann N Y Acad Sci*. 1256:1–14.
- Wei F, et al. 2007. Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS Genet*. 3:e123.
- Wolfe K. 2000. Robustness—it's not where you think it is. *Nat Genet*. 25: 3–4.
- Yu J, et al. 2005. The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol*. 3:e38.
- Zhang L, Vision TJ, Gaut BS. 2002. Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. *Mol Biol Evol*. 19:1464–1473.

Associate editor: Michael Purugganan