Purifying selection and gene conversion in polyploid wheat evolution

Akhunov ED[1,5], Akhunova AR[1,5], Anderson OD[2], Anderson JA[3], Blake N[6], Clegg MT[4], Coleman-Derr D[1], Conley EJ[3], Crossman CC[1], Deal KR[1], Dubcovsky J[1], Gill BS[5], Gu YQ[2], Hadam J[5], Heo HY[6], Huo N[1], Lazo GR[2], Luo MC[1], Ma YQ[1], Matthews DE[7], McGuire PE[8], Morrell P[4], Qualset CO[8], Renfro J[1], Reynolds S[3], Tabanao D[3], Talbert LE[6], Tian C[1], Toleno D[4], Warburton M[9], You FM[1], Zhang W[1], Dvorak J[1]

[1]Plant Sciences, UC Davis, Davis CA; [2]GGD, USDA/ARS WRRC, Albany CA;[3]Agronomy and Plant Genetics, University of Minnesota, St. Paul MN; [4]Ecology and Evolutionary Biology, UC Irvine, Irvine, CA;[5]Plant Pathology, KSU, Manhattan KS; [6]Plant Sciences and Plant Pathology, Montana State University, Bozeman MT; [7]USDA-ARS, Cornell University, Ithaca NY;[8]UC GRCP, Davis CA; [9]CIMMYT, Mexico City, Mexico

ABSTRACT

Functional redundancy in polyploid organisms creates possibilities for evolution of new gene functions by reducing purifying selective pressure acting on genes. During wheat SNP discovery, more than 2,000 gene fragments were sequenced in polyploid wheat and its diploid ancestors. It was shown that orthologous genes in polyploid wheat genomes diverged from each other faster than in the wheat diploid ancestors. This observation could be explained either by relaxation of purifying selection, acceleration of the nucleotide substitution rate after polyploidization or by inter-genomic gene conversion. The latter process would generate shared polymorphism between wheat genomes. The number of shared mutations between orthologous genes in the wheat genomes was negligible suggesting that inter-genomic gene conversions were rare during polyploid wheat evolution. The ratio of nonsynonymous (*dN*) to synonymous (*dS*) mutation rates in coding sequences was used as a measure of relaxation of purifying selection in polyploid wheat. The *dN/dS* ratio was significantly increased in polyploid wheat compared to that in the wheat diploid ancestors indicating relaxation of purifying selection, which is reflected by an increase in the amino acid substitution rate during polyploid wheat evolution.

INTRODUCTION

Substitutions of single nucleotides (single nucleotide polymorphism, SNP) are well suited for the investigation of evolutionary dynamics. While studies of DNA diversity are straightforward in diploids, they are complicated in recently evolved polyploids by gene duplication. For this reason, only a few whole genome diversity studies have been reported in recently evolved polyploids.

Genetic redundancy in polyploids is expected to shelter deleterious mutations from purifying (negative) selection. This effect is extreme in allopolyploids, which behave meiotically as diploids with crossovers taking place only between homologous chromosomes. The absence of crossovers between homoeologous chromosomes precludes segregation of recessive mutations unless a recessive mutation with a similar effect is also present at the other orthologous locus or loci. The expected consequence of relaxation of purifying selection is increased mutation load and higher rate of divergence in coding sequences in allopolyploids compared to that in related diploids.

The strength of purifying selection acting on codons has been investigated in tetraploid frog, *Xenopus laevis*, and several plant paleopolyploids[1,2]. The strength of selection was assessed by calculating the *dN/dS* ratio. Contrary to what would be expected, the *dN/dS* rate ratio in *X. laevis* did not differ from that in diploid vertebrate lineages, leading to the conclusion that purifying selection was not relaxed during polyploid evolution. A similar conclusion was reached for ancient polyploid plants, such as maize, rice, sorghum, barley, *Medicago*, and soybean[2]. No estimate of purifying selection exists for a recently evolved polyploid.

The presence of diverged orthologous genes in allopolyploids creates an opportunity for recombination between them, either reciprocally via crossovers or nonreciprocally via gene conversion. The latter process is of particular interest because of its potential for generating allelic variation within genomes of allopolyploids without altering the gross structure of their chromosomes. Like relaxation of purifying selection, orthologous gene conversions are expected to accelerate accumulation of nucleotide diversity and rate of divergence in polyploid lineages relative to those in diploids. However, unlike the effects of relaxation of purifying selection, orthologous gene conversions would lead to shortening of genetic distances between orthologous nucleotide sequences at the polyploid level compared to those in the same genomes at the diploid level. The existence of orthologous gene conversions has been investigated at 16 loci in allotetraploid cotton, *Gossypium hirsutum*, but none were detected[3].

Here we used several wheat populations and two of the diploid ancestors of wheat to examine diversity dynamics during the early stages of polyploid evolution. We

developed genome-specific primer pairs for the amplification of genomic DNA with polymerase chain reaction (PCR) from close to 2,000 polyploid wheat genes (wheat.pw.usda.gov/SNP/new/index.shtml) and either all of them or a subset was used to amplify target DNA from a representative panel of wild emmer, durum, *Triticum aestivum,* and the diploid *T. urartu*, and *Aegilops tauschii* and sequence the amplicons. The *dN/dS* ratios were computed and used to assess the strength of purifying selection during diploid and polyploid evolution. The existence of gene conversions between orthologus genes in wheat was assessed.

RESULTS

Purifying selection: The ratio of nonsynonymous (*dN*) to synonymous (*dS*) mutation rates in coding sequences was used to assess the relaxation of purifying selection due to gene duplication in polyploid wheat. Exonic sequences in the A, B, and D genomes were concatenated for the estimation of *dN* and *dS* rates. The total length of exonic sequences was 66369 bp, 65841 bp, and 69699 bp, in the A, B, and D genomes, respectively. The *dN/dS* rates of the two internal branches that correspond to the divergence of a polyploid wheat genome from its diploid ancestors were computed. The *dN/dS* rates for the lineages of *T. urartu*, *Ae. speltoides*, and *Ae. tauschii,* the putative ancestors of the A, B, and D genomes of polyploid wheat, were 0.12, 0.15, and 0.12, respectively. The *dN/dS* rates in the exonic sequences ranged from 0.31 to 0.37 in the polyploid wheat A genome and from 0.16 to 1.08 in the polyploid wheat B genome. In the D genome, the ratio was 0.42.

Orthologous gene conversion: The numbers of fixed (mutations differentiating two genomes from each other), shared (polymorphisms present in both genomes), and private (rare mutations found only in one of the genomes) mutations in the A, B, and D genomes of polyploid wheat were determined in each of the three pair-wise comparisons between wheat genomes. The numbers of fixed, shared, and private mutations were also determined for pair-wise comparisons of *T. urartu*, *Ae. speltoides*, and *Ae. tauschii*. We hypothesized that if orthologous gene conversions played an important role in the origin of wheat gene diversity, they should be detectable by reduced divergence per nucleotide site between genomes of polyploid wheat compared to that between the genomes of diploid ancestors. This was not found. In fact, the opposite trend was observed. The divergence between genomes in polyploid wheat was 25% greater in each pair-wise comparison than the divergence between the genomes of diploid ancestors. The obvious cause of this acceleration of orthologous gene divergence in polyploid wheat is relaxation of purifying selection confirmed by estimating the *dN/dS* ratio. Purifying selection affects diversity in coding sequences but not intronic sequences, whereas gene conversions are expected to affect diversity in both. Inter-orthologous gene conversions would generate shared polymorphism between wheat genomes. The numbers of such polymorphisms between wheat genomes were negligible, which also suggests that inter-orthologous gene conversions are rare. For the A-B genome comparison the ratio of shared polymorphisms to fixed mutations was 4.4×10^{-4}, for the A-D genome comparison it was 1.6×10^{-4}; and for the B-D genome comparison the ratio was 1.7×10^{-4}.

DISCUSSION

Increase in the rate of divergence between orthologous genes in polyploid wheat genomes suggests that polyploid evolution is accompanied by an accelerated rate of mutation accumulation. The *dN/dS* rate ratios for diploid lineages ranged from 0.12 to 0.15. The internodes of the phylogenetic trees of polyploid wheat showed *dN/dS* rates ranging from 0.16 to 1.08. In the A and D genomes trees they ranged from 0.31 to 0.42. In the B genome tree, they varied from 0.16 to 1.08. The *dN/dS* rate estimates would be free of bias if one would be able to use only mutations that originated in polyploid wheat, which could not be done. The more that mutations originating at the diploid level are included into the calculations the greater is the underestimation of the magnitude of the actual *dN*/dS ratio. Variation in *dN/dS* rates in the phylogenetic tree of the B genome is consistent with this prediction. The ratio is 0.16 prior to the split of wild emmer and *T. aestivum* but 0.85 and 1.08 in the more terminal internodes, which include private mutations for hexaploid wheat. This indicates that the overall *dN/dS* ratios reported for polyploid wheat are underestimated and that purifying selection has been even weaker in polyploid wheat than indicated by these estimates. No evidence was obtained to show that the relaxation was greater at the hexaploid than tetraploid level.

Relaxation of purifying selection during evolution of polyploid wheat is evident from frequencies of gene deletions[4]. Single-copy loci have been deleted during evolution of polyploid wheat 10 times faster than during evolution of wheat diploid ancestors[5]. Evidence for relaxation of purifying selection acting on replacement positions in codons found here parallels the evidence obtained earlier for single-copy locus deletions. In polyploid *X. laevis* and a number of polyploid plants, purifying selection was not relaxed[2]. All these polyploids were ancient; the youngest originated 10 MYA. To reconcile these conflicting data, we suggest that purifying selection is relaxed in early stages of polyploid evolution, exemplified by wheat, because most of the duplicated genes are equivalent and can substitute for each other. Ancient polyploids become "diploidized" by deletions of duplicated genes, and neofunctionalization and subfunctionalization of duplicated genes[6]. Diverging duplicated genes are subjected to increasing strength of purifying selection. In fact, most of the ancient polyploid species investigated were viewed as diploid by classical genetics not many years ago[2].

It has been pointed out that empirical findings on the fate of duplicated genes are inconsistent with the expectations based on the classical model of duplicated gene evolution formulated by Ohno[7], namely, relaxation of purifying selection making it possible for one gene copy

to mutate and frequent evolution of nulls. The evidence contradicting these predictions was mostly obtained by studying ancient polyploids or ancient gene duplications. However, recently evolved polyploid wheat shows both, relaxation of purifying selection (present data) and an order of magnitude elevated frequency of deletions[4,5]. While this does not mean that Ohno's model of new gene evolution by gene duplication is correct, it does highlight the need to focus studies on evolution of duplicated genes in recently evolved polyploids and recent paralogous sets.

A consequence of relaxation of purifying selection is acceleration of nucleotide substitution rate in coding sequences at the initial stages of polyploid evolution. Cronn et al. (1999)[3] examined the relative rate of divergence at 16 loci in tetraploid *G. hirsutum* and did not observe acceleration of the rate of evolution compared to that in diploid lineages[3]. Seven of the 16 loci studied were *Pst*I clones with no homology to coding sequences in databases. Additionally, distinction was not made between coding and noncoding sequences in the rate estimation. It is therefore possible that if coding sequences were compared, acceleration of the rate may have been observed in tetraploid cotton too. Because 0.2 to 0.5 MY have passed since the origin of wild emmer[5], accelerated nucleotide substitution rate in coding sequences may have broadened the level of diversity in wild emmer and may account for its surprisingly high diversity and limited evidence of polyploidy bottleneck.

A potential cause of the observed increase in the rate of divergence between orthologous genes in polyploid wheat are orthologous gene conversions. However, search for orthologous conversions in 962 genes was unsuccessful to detect conversion tracts in wheat. Likewise, no orthologous gene conversions were detected in 16 loci in tetraploid cotton[3]. A legitimate question is why conversions between orthologues have been so rare during evolution of wheat and cotton. A most likely explanation for their paucity is the activity of genes suppressing homoeologous recombination, like *Ph1*[8]. A similar genetic activity was inferred in polyploid cotton[9]. This hypothesis is further supported by the recent discovery of a clear example of gene conversion between the Psy-A1 and the Psy-B1 genes in the tetraploid wheat Cappelli carrying the ph1c mutation[10].

According to the current view, the decision resulting in a crossover (CO), which may or may not be accompanied by gene conversion, or gene conversion without crossover (noncrossover recombination, NCR) is the result of two different pathways after the initial double strand DNA break and 5' to 3' recession of the ends[10]. The decision whether CO or NCR will take place is made prior to onset of stable strand exchange. In yeast, the path leading to NCR does not require the Holliday intermediate structure[11] and requires uninterrupted homology of 25 to 60 bp whereas that leading to CO requires the Holliday intermediate structure and uninterrupted homology of 150 to 250 bp. Since *Ph1* precludes CO between orthologues, it also precludes gene conversions accompanying them. This leaves only NCR as the potential source of gene conversions in polyploid wheat and cotton. Whether or not this path takes place between orthologues and why it does not generate diversity requires more research. An expected consequence of conversions via NCR would be an increase in homology between wheat homoeologous chromosomes and an increase in heterogenetic chromosome pairing with time, which has not been observed. Hence, both relative rate of divergence in nucleotide sequences and cytogenetic behavior of wheat chromosomes are consistent in showing no evidence for convergence of gene sequences in homoeologous chromosomes.

In the absence of evidence for a significant contribution of orthologous gene conversions to allelic diversity in polyploid wheat, only relaxation of purifying selection with the concomitant accumulation of new diversity may have contributed to the rate of divergence between orthologous genes and relatively high diversity of wild emmer. The novel diversity was undoubtedly combined with the pre-existing diversity that was contributed by gene flow from wheat ancestors to wheat counteracting the effects of a polyploidy bottleneck[12].

REFERENCES

1. Cannatella DC and RO de Sa (1993) Syst. Biol. 42:476-507
2. Schlueter et al. (2004) Genome 47: 868-876
3. Cronn RC et al. (1999) *PNAS* 96: 14406-14411
4. Dvorak J et al. (2004) Genetics 168: 1665-1675
5. Dvorak J and ED Akhunov (2005) Genetics 171: 323-332
6. Moore RC and MD Purugganan (2005) Curr. Opin in Plant Biol. 8:122–128
7. Ohno S (1970) Evolution by gene duplication Springer Verlag, New York.
8. Sears ER and Okamoto M (1958) Proc. X Int. Congress Genet. 2: 258-259
9. Kimber G (1961) Nature 191: 98-100
10. Zhang W & J Dubcovsky(2008). Theor Appl Genet 116:635-645.
11. Allers T and Lichten M (2001) Cell 106: 47-57
12. Dubcovsky J and J. Dvorak. 2007. Science316: 1862-1866