academicJournals

Vol. 6(1), pp. 41-49, January 2014
DOI: 10.5897/IJBC2012.090
ISSN 2141-243X © 2014 Academic Journals
http://www.academicjournals.org/IJBC

International Journal of Biodiversity and Conservation

Review

Gene duplication: A major force in evolution and bio-diversity

Chandan Roy* and Indra Deo

Department of Genetics and Plant Breeding, G. B. Pant University of Agriculture and Technology, Pantnagar U S Nagar- 263145, Uttarakhand, India.

Accepted 26 August, 2013

Bridges reported one of the earliest observations of gene duplication from the doubling of a chromosomal band in a mutant of the fruit fly, *Drosophila melanogaster*, which exhibited extreme reduction in eye size. Based on whole-genome analysis of *Arabidopsis thaliana*, there is compelling evidence that angiosperms underwent two whole-genome duplication events early during their evolutionary history. Recent studies have shown that these events were crucial for the creation of many important developmental and regulatory genes found in extant angiosperm genomes. Recent studies provide strong indications that even yeast (*Saccharomyces cerevisiae*), with its compact genome, is in fact an ancient tetraploid. Gene duplication is providing new genetic material for mutation, drift and selection to act upon, the result of which is specialized or new gene functions. Without gene duplication, the plasticity of a genome or species in adapting to changing environments would be severely limited. The era of whole genome sequencing of model organisms suggests a number of duplication events take place while evolving modern species.

Key words: Evolution, genome duplication, diversity.

INTRODUCTION

An event in which one gene gives rise to two genes is generally known as duplication; in this, the two genes cannot be operationally distinguished from each other. Duplicated genes may remain in the same genome (known as paralogs) from where they arose and their presence may be in different genome (known as orthologs) after duplication. Gene duplication is believed to play an important role in evolutionary process by providing a chance to evolve new genes. Duplicated gene generates new opportunity for natural selection. At first Darwin published his idea about the “Origin of Species” but still it is a major issue. In 1900, when the rediscovery of Mendel’s law was proposed it gave us a better understanding of how genetic variations exist for traits. However, Bridges in 1936 first identified bar eye locus in drosophila and its effect on eye shape when duplicated. Besides, recombination duplication followed by diversification is one of the great paves for creation of variation. In 1970, Ohno, in his book, “Evolution by Gene Duplication”, had given a clear-cut idea about the origin of duplicated genes and the possible fate of gene duplication. He concluded that gene duplication is the only means by which a new gene can arise and argues that in the past whole genomes have been duplicated. Duplication may take place either in single genes, a segment of chromosome, whole chromosome or even the whole genome of a species. Transition from invertebrates to vertebrate could occur only if whole genomes were duplicated (Ohno, 1970). It has been found that all the present day angiosperm has undergone large scale gene duplication or whole genome

*Corresponding author. E-mail: chandan.roy43@gmail.com.

duplication (Bodt et al., 2005). Ohno's representation of duplication as evolutionary force opened up a new window to find out evolutionary consequences through research. The idea that gene duplication has a fundamental role in the origin of diversity suggested numerous proposals for knowing how a new gene copy can emerge from its predecessor and evolve a novel function. The use of molecular markers technologies and the sequence information of the model organisms opened up new windows for carrying out research on duplication analysis, and determining the evolutionary pathways of organisms became interesting work.

MOLECULAR MECHANISMS OF GENE DUPLICATION

Duplicated gene may be produced by unequal crossing over, retro-transposition, duplicated DNA transposition and polyploidization.

Unequal crossing over

This produces tandem repeated sequences, that is, continuous repeats of DNA sequence. Depending on the position of crossing over, the duplicated regions can contain part of a gene, an entire gene, or several genes (Zhang, 2003). Unequal crossing over may lead to the evolvement of paraloguous gene through concerted evolution (Hurst and Smith, 1998; Li, 1997). Crossing over in a bivalent carrying duplication in one of the two chromosomes may lead to different consequences. If the duplicated segment pairs with its homologous segment in the other chromosome ignoring other homologous segments, then the unequal crossing over produces duplication of other segments Figure 1. If the duplicated segment is present in reverse orientation Figure 2 of the original segment or if duplication is present on the other arm, then the pairing followed by crossing over forms dicentric and acentric fragment. If there are duplicated segments on another non homologous chromosome Figure 3, crossing over with this duplicated region will produce two interchanged chromosomes (Gupta, 2007).

Retroposition

This is a process where messenger RNA (mRNA) of a gene is reverse transcribed to complementary DNA (cDNA) and then inserted into the genome. There are several molecular features of retroposition: lack of introns and regulatory sequences of gene, presence of poly- A sequence and presence of flanking short direct repeats. The major difference of this mechanism from unequal crossing over is the presence of introns. Introns are the short DNA sequence present in between the coding sequence of gene that splices out after transcription. If introns are present in the original genes, they will also be present in the duplicated genes through unequal crossing over, but absent in retrogenes. A duplicated gene generated by retroposition is usually unlinked to the original gene, because the insertion of cDNA into the genome is more or less random. Recent studies have found that retrogenes that land near other coding regions or even in the introns of expressed coding sequences are much more likely to be expressed than those that land far from coding sequences (Vinckenbosch et al., 2006). mi-RNAs are reported to be found in the intron, exons and intergeneric regions of human genome. Duplication of mi- RNAs is one of the mechanisms for their evolution into human genome. mi-RNAs are arranged mostly in 5000-nt clusters and their copies are scattered randomly throughout the genome at an average distance of 4.3×10^6 bp. Comparison of miRNAs copies with the transposable elements (TEs) revealed that most miRNAs homologues (96%) propagate by DNA transposons and retroelements (Titov and Vorozheykin, 2011).

Duplicative transposition

Duplicative transposition of DNA sequences can be accomplished by one of two main pathways: nonallelic homologous recombination (NAHR) or non- homologous end joining (NHEJ). The difference between two pathways is based on whether homologous sequences are used as a template during double-strand break repair, and this difference can also be used to infer the mechanism by which individual genes are duplicated. Recombination between these nonallelic homologous sequences can result in the duplication of the intervening sequences, which can then lead in turn to more duplications because of pairing between the new paralogs (Bailey et al., 2003). But other studies in human being have also found multiple cases with no repetitive DNA or long stretches of homologous sequence at duplication breakpoints, suggesting the action of NHEJ (Linardopoulou et al., 2005). Due to the relatively low proportion of duplicated sequences arranged in tandem in the human genome, it has been proposed that duplicative transposition is the major mode of duplication in humans (Samonte and Eichler, 2002). The number of retrogenes maintained in both mammals (Pan and Zhang, 2007) and Drosophila is lower than the number maintained by DNA-based intermediates (that is, unequal crossing-over and duplicative transposition), despite the fact that the mutation rate forming new retrocopies is higher (Pan and Zhang, 2007). The lack of functional regulatory DNA is likely to be the reason that very few of these paralogs are maintained for long; only 120 functional retrotransposed gene copies have been maintained in the human genome over the past 63 million years (Vinckenbosch et al., 2006).

Polyploidization

The fourth major mechanism of duplicate gene formation is

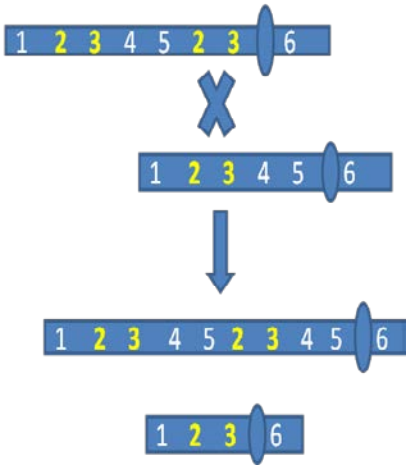

Figure 1. Duplication present on the same arm of chromosome.

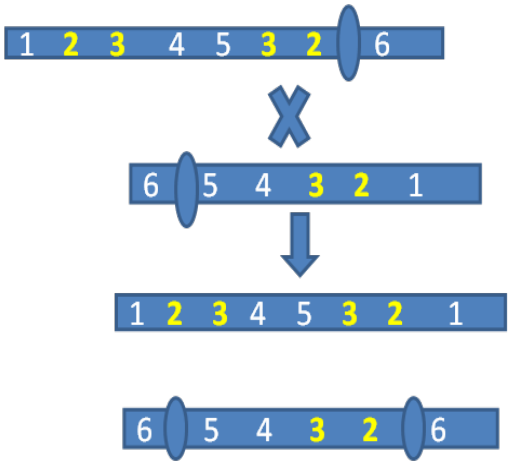

Figure 2. Duplication in reverse orientation.

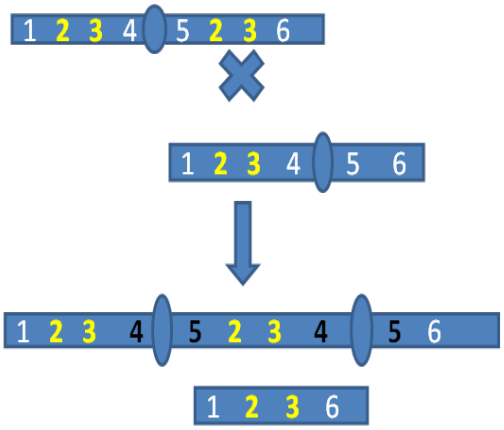

Figure 3. Duplication on another arm.

is polyploidization. Polyploidy is an evolutionary process whereby two or more genomes are brought together into the same nucleus, usually by hybridization followed by chromosome doubling. Ohno, in his book, pointed out that two rounds of genome duplication had taken place for the evolution of vertebrates. Recent studies provide strong indications about the importance of gene duplication in the origin of organisms. Even yeast (*Saccharomyces cerevisiae*), with its compact genome, is in fact an ancient tetraploid. A whole-genome duplication followed by massive gene loss and specialization has taken place during its evolutionary process (Kellis et al., 2004). In plants, polyploidy was proposed to have occurred in the lineage of at least 70% of angiosperms (Masterson, 1994) and in 95% of pteridophytes (Grant, 1981). Arabidopsis (Arabidopsis Genome Initiative, 2000) and rice (Goff et al., 2002) considered as classical diploids are apparently ancient polyploids (paleopolyploids). Many higher plant species considered as diploids because of their genetic and cytogenetic behavior are ancient polyploids that have undergone a process of extensive diploidization. Thus, polyploidy is one of the major processes that has driven and shaped the evolution of higher organisms.

DOES GENE DUPLICATION PROVIDE THE ENGINE FOR EVOLUTION?

How life evolved from a few primordial genes to the more than twenty thousands of genes in higher organisms was a major issue in Darwinism. The current primary hypothesis is that it occurred via gene duplication (Hurles, 2004). Shanks (2004) concluded that 'duplication is the way of acquiring new genes by organisms'. They appear as the result of duplication. Ohno concluded that "gene duplication is the only means by which a new gene can arise". Not only genes but whole genomes have been duplicated in the past, causing 'great leaps in evolution such as the transition from invertebrates to vertebrates, which could occur only if whole genomes were duplicated' (Ohno, 1970). Kellis et al. (2004) agree that 'whole-genome duplication followed by massive gene loss and specialization has long been postulated as a powerful mechanism of evolutionary innovation'. Genome duplication has been proved to be major events for angiosperm evolution (Bodt *et al*. 2005). The two major branches of the angiosperms (eudicots and monocots), estimated to have diverged 125–140 to 170–235 mya (Davies et al., 2004), show much more rapid structural evolution. This difference appears to be largely due to the tendency of angiosperms for chromosomal duplication and subsequent gene loss (Coghlan et al., 2005). Recent analyses of genome sequences suggest that genome duplication in angiosperms may be not merely episodic but truly cyclic, which causes various fitness advantages that erode over time, favoring new polyploidizations (Chapman et al., 2006).

THE FATES OF DUPLICATE GENES

Whole-genome duplications result in new gene copies of every gene in a genome and, obviously, all the flanking regulatory sequences. All the genes after duplication may not undergo fixation as most of the genes get lost from the genome. The birth and death of genes is a common theme in gene family and genome evolution with those genes involved in the physiologies that vary greatly among species (e.g. immunity, reproduction and sensory systems) probably having high rates of gene birth and death. After fixation, the fate of the gene (s) is determined by the function of that gene(s) in the genome (Zhang, 2003).

Pseudogenation

Generally, carrying out two identical genes in a particular genome is not advantageous, as duplicated genes produce functional redundancy (Zhang, 2003). Pseudogenization, the process by which a functional gene becomes a pseudogene Figure 4, usually occurs in the first few million years after duplication if the duplicated gene is not under any selection (Walsh, 1995; Lynch and Conery, 2000; Lynch et al., 2001; Harisson et al., 2002). There are two major forces through which duplicate genes undergo pseudogenation. These are mutation and deletion where changes in pseudogenes occur through promoter mutation, splicing junction lost, nonsense mutation or missense mutation (Harisson et al., 2002**)**. Mutation distracting the structure and function of one of the two genes is not removed by selection (Lynch and Conery, 2000; Lynch et al., 2001). Gradually, the mutation containing gene becomes a pseudogene, which is either not expressed or become non-functional. After a long time, pseudogenes will either get deleted from the genome or become more diverged from the parental genes that they are no longer identifiable with the original genes. In humans and mice, the size of the olfactory receptor gene family (~1000) is similar but the percentage of pseudogenes is >60% in humans and only 20% in mice (Rouquier et al., 2000; Mombaerts, 2001; Zhang and Firestein, 2002). This may be due to the lesser use of olfaction since the origin of hominoids, which can be compensated for by other sensory mechanisms such as better vision (Rouquier et al. 2000).

Occasionally, it has been observed that pseudegenes may also serve some functions. One functional gene (VH1) that encodes the heavy chain variable region of immunoglobulin in chicken. Immunoglobulin diversity is generated by gene conversion (Hurst and Smith, 1998) of the VH1 gene (Ota and Nei, 1995).

Conservation of gene function

There are several known proteins present in cell where large quantity is required by the cell for proper functioning. The first mechanism for maintaining a duplicate copy of gene proposed by Ohno (1970) was to simply increase the number of protein coding genes, where both loci maintain their original functions. Ohno (1970) proposed two possible models why these duplicates would maintain the original functions. The first model states that a second gene could provide functional redundancy if the original locus was disabled by mutation. The second possibility for why exact copies of duplicated genes are maintained is that there is an advantage of producing more of a gene to accomplish the increased levels of protein production in the cells. The most commonly cited examples are the highly duplicated ribosomal RNAs needed for development and histone proteins. Now the question arises: “how can two paralogous genes maintain the same function after duplication?’’ One of the possible mechanisms is concerted evolution (Li, 1997) and another is purifying selection (Nei et al., 2000). Concerted evolution: a mode of gene family evolution through which members of a family remain similar in sequence and function because of frequent gene conversion and/or unequal crossing over (Hurst and Smith, 1998; Li, 1997). Whereas strong purifying selection plays against mutations that modify gene function which can prevent duplicated genes from diverging. The difference between gene conversion and purifying selection can be described through synonymous or silent mutation; where, a synonymous nucleotide difference does not change the function of genes as the change in nucleotide in DNA sequence. Synonymous differences are more or less immune to selection and cannot be reduced by purifying selection whereas gene conversion homogenizes DNA sequences regardless of whether the differences are synonymous or non synonymous (Nei et al., 2000; Piontkivska et al., 2002; Hurst and Smith, 1998).

Sub-functionalization

In general, the duplicate gene is deleterious for the genome or species (some exceptions like histone protein coding genes). Two genes with identical functions are not maintained generally in the genome unless duplicated gene product is advantageous (Nowak et al., 1997; Lynch and Conery, 2000). After duplication, both the daughter genes are maintained in the genome for a period of time when they differ in some aspects of their functions. This can occur by subfunctionalization Figure 4, in which each daughter gene adopts part of the functions of their parental gene (Hughes, 1994; Force et al., 1999; Lynch and Force, 2000). For example, a pair of transcription factor genes in zebrafish is engrailed-1 and engrailed-1b generated by a chromosomal segmental duplication. Engrailed-1 is expressed in the pectoral appendage bud, whereas engrailed-1b is expressed in the neurons of hindbrain/spinal cord. Despite the sole engrailed-1 gene of mouse, orthologous to both genes (engrailed-1 and engrailed-1b) of the zebrafish is expressed in both pectoral appendage bud and hindbrain/spinal cord (Force

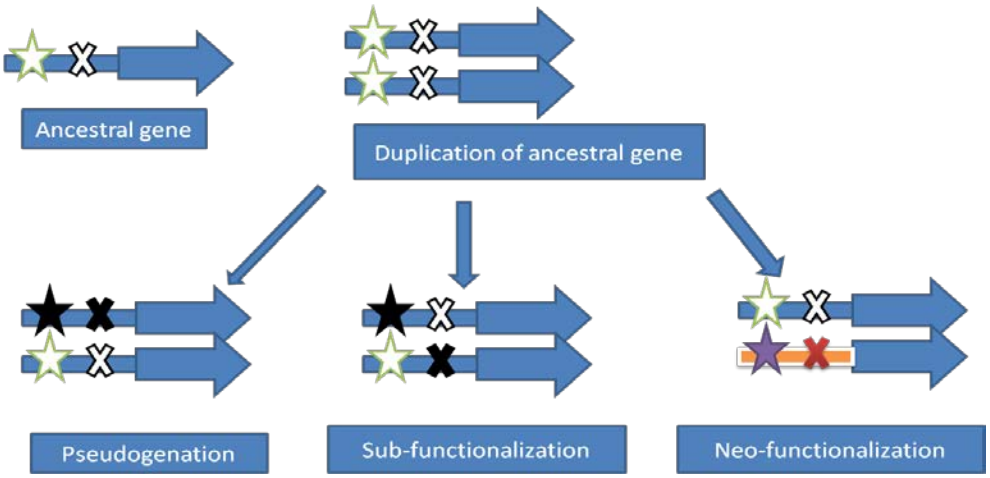

Figure 4. Diagrammatical representation of consequences of duplicated genes and the role of duplication evolving different gene function.

et al., 1999). Subfunctionalization of homeologous genes has great importance for speciation. If duplicated genes are subfunctionalized or reciprocally lost in geographically isolated populations and when such individuals from each population are united, it can lead to hybrids that lack both copies of a duplicated gene pair, resulting in hybrid inviability, reproductive isolation and speciation (Werth and Windham, 1991; Lynch and Force, 2000; Taylor et al., 2001).

Neo-functionalization

Origin of novel gene function is one of the most important outcomes of gene duplication. Gene duplication allows the evolution of genes with new functions Figure 4. Duplication followed by selection plays a major role in evolution as the selection maintains the initial amplification and beneficial mutant alleles where the less improved genes get lost from the genome (Nasvall, 2012). The evolution of a novel fruit shape in tomato (*Solanum lycopersicum*) *SUN* and its progenitor (*IQD12*) evolved by the chance of duplication. A plant specific protein (67 amino acid motif called *IQ67*) produced the gene (*SUN*) into a new regulatory context belonging to a gene family that is involved in calmodulin signaling. *SUN* is expressed at much higher levels during the early stages of fruit development, and up-regulation correlates with an elongated fruit shape instead of round type fruit produced by gene *IQD12* (Xiao et al., 2008). In contrast to that, the natural allopolyploid *Arabidopsis suecica* is readily resynthesized in the laboratory from its model progenitors, *A. thaliana* and *Arabidopsis arenosa*. An interesting feature of this allopolyploid was found; it grows to a larger stature and produces more biomass than either of its parents. Most of the genes up-regulated in allotetraploid were *CCA1* (*circadian clock associated 1*), which showed that *CCA1* and *LHY* were epigenetically suppressed in the allopolyploid and that this suppression strongly correlates with increased starch synthesis and chlorophyll content, ultimately leading to greater plant biomass (Ni et al., 2009).

DUPLICATION IN RELATION TO DIVERSITY AND SPECIATION

Duplication may take place in a part of gene, the whole gene, part of genome or the whole genome. Whole genome duplication leads to doubling of large quantity of genes at once and this provides large potential source of novelty. Selection pressure (neither completely randomly nor deterministically) would play out in different ways in different populations in different climatic situation Figure 5, potentially leading to increased rates of speciation (Christian et al., 2007). Genes that are duplicated by polyploidy could be expressed at equal levels, or there could be unequal expression or silencing of one copy. Most gene pairs formed by a WGD have only a brief lifespan before one copy becomes deleted, leaving the others to survive as a single-copy locus (Wang et al., 2004). Studies of newly created synthetic polyploids revealed that silencing of some duplicated genes often resulted in the onset of allopolyploidy (Wang et al., 2004, Kashkush et al., 2002; He et al., 2003), indicating that gene silencing is a common response to polyploidy. It is expected that the probability of retention is initially equal for both duplicates following WGD, but recent findings have suggested that one duplicate may be more susceptible to loss than others. It has been found that in *Arabidopsis thaliana*, one paralogon (duplicated genomic region) tends to contain significantly more genes than the others (Thomas et al., 2006). Silencing of genes can take place immediately at the first generation following polyploidy, although some genes are not silenced until later generations (Wang et al., 2004). Silencing and expression

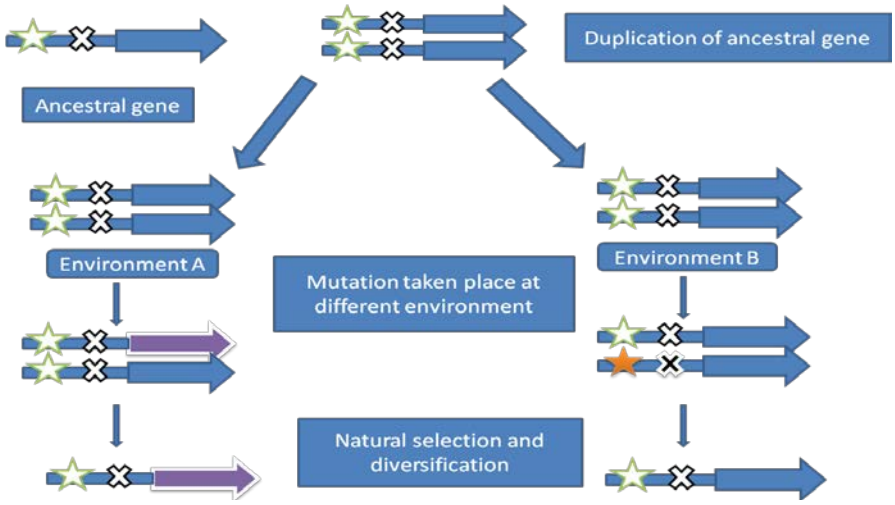

Figure 5. The process of gene diversification after duplication through natural selection.

of genes are many complex phenomenons as varying levels of gene expression are observed at organ level. Some duplicated genes are silenced immediately upon allopolyploidy in some organs of the plant but remain expressed in other organs (Adams et al., 2003). There is strong evidence for one round of genome doubling after the eudicot divergence and a second polyploidization event sometimes following the divergence of Arabidopsis and Brassica from their common ancestor with the Malvaceae, represented by cotton (Adams and Wendel, 2005).

It has been demonstrated recently that most eudicot plants are descendents of an ancient hexaploid ancestor (Jaillon et al., 2007), subsequently followed by lineage-specific tetraploidizations in some taxa: Populus (Tuskan et al., 2006; Jaillon et al., 2007), Arabidopsis (Bowers et al., 2003; Blanc et al., 2003 and Simillion et al. 2002), legumes (Cannon et al., 2006), but not recorded in Vitis (Jaillon et al. 2007). WGD has been proposed to be a lineage splitting force because of the subsequent occurrence of gene losses independently in different populations. In particular, reciprocal gene loss (RGL) occurs when two paralogs created by WGD are retained until speciation, after which each species loses a different copy (Scannell et al., 2006; Semon and Wolfe, 2007). After duplication, one of the two redundant copies of a gene should be free to accumulate mutation and become lost from the genome or gain some function without any consequence (Ohno, 1970). One analysis was performed just after artificial allopolyploidization in cotton where one paralog was silenced or down regulated in 5% of the gene pairs and that silencing was often organ-specific (Adams et al. 2004).

GENOME DUPLICATION AND THE ORIGIN OF ANGIOSPERM

It has been suggested that large-scale gene duplication or whole-genome duplication events can be associated with important evolutionary transitions and a major leaps in development of modern species. Angiosperms appear rather suddenly in the fossil record during the Jurassic (208–145 million years ago), with no obvious ancestors for a period of 80–90 million years before their appearance (Doyle and Donoghue, 1993). This ancestral lineage is coined ‘angiophytes’. It is presumed that angiophytes went through a period of little diversification during the late Triassic (220 Mya) and Jurassic, probaby because of the diversity-enhancing features, such as flowers (Wing and Boucher, 1998).

The recent transitional–combinational theory of the angiosperm origin suggests an evolution from Jurassic seed ferns through three fundamental transitions: (i) evolution of the carpel; (ii) emergence of double fertilization; and (iii) origin of the flower. The extant (or modern) angiosperms did not appear until the Early Cretaceous (145–125 Mya), when the final combination of these three angiosperm features occurred, as supported by evidence from micro- and macrofossils (Stuessy, 2004). The fossil record provides supporting evidence for this rapid diversification in floral form during the earliest phases of recorded flowering plant history. This diversification of angiosperms occurred during a period (the Aptian, 125-112 Mya) when their pollen and mega fossils were rare components of terrestrial floras and species diversity was low (Crane et al., 1995). Angiosperm fossils show a dramatic increase in diversity between the Albian (112–99.6 Mya) and the Cenomanian (99.6-93.5 Mya) at a global scale (Crane et al., 2004).

In 1996, when the sequencing of the flowering plant *A. thaliana* (Brassicaceae) genome began with its small genome, it was not expected to be an ancient polyploid. However, five years after the release of its genome sequence, there is compelling evidence that the Arabidopsis genome, or rather that of its ancestors, has been duplicated three times (events referred as 1R, 2R and 3R) during

the past 250 million years along with small scale continuous duplication (Sterck et al., 2007). Ancient polyploidy events might have directly influenced the increase in the number of plant species and plant complexity observed since the Early Cretaceous. Blanc and Wolfe (2004) studied the relationship between gene function and duplicate loss after the most recent polyploidy event (3R). Recently, Maere et al. (2005) developed an evolutionary model based on the KS distribution of the *Arabidopsis paranome* where they took into account the three major genome-wide duplication events (1R, 2R and 3R) and a continuous mode of small scale gene duplications (referred to as 0R). These studies all concluded that genes involved in transcriptional regulation and signal transduction have been preferentially retained following genome duplications. Similarly, developmental genes have been observed to be retained following genome duplications, particularly following the two oldest events, that is, 1R and 2R (Maere et al., 2005). Overall, the three polyploidy events in the ancestors of Arabidopsis might have been responsible for >90% of the transcription factors, signal transducers and developmental genes created during the past 250 million years.

S-adenosyl-l-methionine (SAM) dependent O-methyl transferases (OMTs) proteins are involved in the methylation of various secondary metabolites. Phylogeny across land plant lineages showed that OMT genes were distributed in two main classes, also suggesting that they have evolved by a gene duplication that had happened in the ancestor of land plants (Barakat et al., 2011). Soybean undergoes two separate polyploidy events resulting in 75% of genes present in multiple copies. Multiple events have taken place over the duplicated genes where sub functionalization, neo functionization, non functionalization or even the epigenetic or positional regulations play a role for gene regulation (Roulin et al., 2012).

DUPLICATION ANALYSIS IN MODEL ORGANISMS

Since 1990, the sequencing project has been launched in different organisms at different period of time; it revealed to us how to analyze the evolutionary pattern of different species by different chromosome rearrangements. The similarity and colinearity analysis of different species or within species among different chromosome has clearly shown the process of genome duplication over time and their role in species diversification. A cluster of resistance genes namely Tak703-1, Lrr703, Tak703, and Lrk703 have been identified in the D genome of wheat, where the structural cluster unit is conserved in nine grass genomes. Duplication has played major role in the *Tak*/*Lrk* evolution in oats, maize, barley, wheat, sorghum, and *Brachypodium*, while tandem duplication drove the expansion of this locus in *japonica* rice (Wang et al. 2013). Duplication analysis of some of the model organisms based on the genome sequencing data or comparing them with other species is described briefly as follows.

Duplication in arabidopsis genome (The Arabidopsis Genome Initiative, 2000)

The Arabidopsis Genome Initiative in 2000 published sequence analysis on model flowering plant, Arabidopsis. They used large-insert bacterial artificial chromosome (BAC), phage (P1) and transformation-competent artificial chromosome (TAC) libraries as the primary substrates for sequencing. The Arabidopsis genome sequence provides a complete view of chromosomal organization and clues to its evolutionary history. It revealed 1,528 tandem arrays containing 4,140 individual genes covering 17% of all genes of Arabidopsis. All the five chromosomes of Arabidopsis were aligned with each other in both orientations using MUMmer; and all segments were identified at least 1,000 bp in length; and 50% identity which revealed 24 large duplicated segments of 100 kb or larger, comprising 65.6Mb or 58% of the genome. But using TBLASTX to identify collinear clusters of genes in large duplicated chromosomal segments showed duplicated regions encompassing 67.9Mb, 60% of the genome. This study revealed a tetraploid ancestor was the progenitor of present day Arabidopsis as the majority of the Arabidopsis genome is represented in duplicated segments (Gaut and Doebley, 1997). A comparative sequence analysis of Arabidopsis and tomato estimated that duplication occurred in 112 Myr ago to form a tetraploid. The degrees of conservation of the duplicated segments might be due to divergence from an ancestral autotetraploid form, or might reflect differences present in an allotetraploid ancestor (Ku et al., 2000).

Duplication in *Saccharomyces cerevisiae*

Wolfe and Shields (1997) interpreted the presence and distribution of such regions in the *S. cerevisiae* genome as supporting a model of WGD. Kellis et al. (2004) showed that *S. cerevisiae* arose from complete duplication of eight ancestral chromosomes, and subsequently returned to functionally normal ploidy by massive loss of nearly 90% of duplicated genes in small deletions. They identified 145 paired regions in *S. cerevisiae*, tilling 88% of the genome and containing 457 duplicated gene pairs. The experiment was conducted by using *Kluyveromyces waltii*, closer to *S. cerevisiae* to identify orthologous regions. In contrast to the 1:1 mapping seen for close relatives19, most local regions in *K. waltii* are mapped to two regions in *S. cerevisiae*, with each containing matches to only a subset of the *K. waltii* genes. This clearly proved the evidence that ancient whole genome duplication would occur in the previous lineages of yeast.

Gene and chromosome duplication in rice (Report of IRGSP)

The International Rice Genome Sequencing Project was organized to achieve >99.99% accurate sequence using

Table 1. Duplication analysis in rice genome.

Chromosome number	Gene	Paralog
1	4,467	956 (21.4%)
2	3,011	616 (20.5%)
3	3197	493 (15.4%)
4	2,679	689 (25.7%)
5	2,426	472 (19.55)
6	2,342	484 (20.7%)
7	2,507	568 (22.7%)
8	2,286	489 (21.4%)
9	1,618	323 (20.4%)
10	1,724	433 (25.1%)
11	1,834	557 (30.4%)
12	1,870	497 (26.6%)
TOTAL	29,961	6577 (22.0%)

Table 2. Rice Arabidopsis synteny.

Chromosome number	Significant threshold (99.99%)
1	41
2	34
3	31
4	11
5	20
Total	137

a mapped based cloned sequencing strategies. More than 104,000 EST from a variety of rice tissue has been developed in EST database. Goff et al. (2002) describe a random shotgun sequencing of *Oryza sativa* L. ssp. *Japonica* (cv. Nipponbare) to discover rice genes, molecular markers for breeding and to mapped sequences for association of candidate genes. Using BLAST for comparing all H genes and M genes it was found that 77% were homologous to at least one other predicted genes. Chromosomal duplication was identified by comparing (BLASTN) more than 2000 mapped rice cDNA markers to the anchored portion of Syd. and it was observed that locally duplicated genes ranged from 15.4 to 30.4%, depending on the chromosome Table 1. The largest duplication is on the chromosomes 11 and 12. The amino acid substitution rate (d_A) was used to estimate the whole genome duplication that occurred in rice around 40 – 50 million years ago. Synteny analysis between rice and Arabidopsis chromosome showed 137 high confidence sytnenic groups at 99.9% threshold level (Table 2).

CONCLUSION

The most important contribution of gene duplication towards evolution is providing new genetic material for different mechanisms of evolutions, that is, mutation, drift and selection to act upon, the result of which is specialized or new gene functions. Duplication increases buffering activity of genome or species in adapting to changing environments where no more than two variants (alleles) exist at any locus within a (diploid) individual. Although, duplicated genes and genomes can provide the raw material for evolutionary diversification and the functional divergence of duplicated genes might offer a selective advantage to polyploids over a long time period, a beneficial effect of these duplications is assumed shortly after the duplication event. Lynch has suggested that differential gene duplication and pseudogenization in geographically isolated populations cause reproductive isolation and speciation, although this intriguing hypothesis awaits empirical evidence.

ACKNOWLEDGEMENT

We acknowledge our HOD, Department of Genetics Plant Breeding, GBPUAT, Pantnagar for his kind support and guide to prepare this article.

REFERENCES

Adams KL, Cronn R, Percifield R, Wendel JF (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. USA. 100:4649-4654.

Adams KL, Wendel JF (2005). Allele-specific, bi-directional silencing of an alcohol dehydrogenase gene in different organs of interspecific diploid cotton hybrids. Genetics 171:2139–2142.

Adams KL, Percifield R, Wendel JF (2004). Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics 168: 2217-2226.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815.

Bailey JA, Liu G, Eichler EE (2003). An *Alu* transposition model for the origin and expansion of human segmental duplications. Am. J. Hum. Genet. 73:823–834.

Barakat A, Choi A, Yassin NB, Park JS, Sun Z, Carlson JE (2011). Comparative genomics and evolutionary analyses of the O-methyltransferase gene family in Populus. Gene 479 (1-2):37-46.

Blanc G, Hokamp K, Wolfe KH (2003). A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res. 13:137-144.

Blanc G, Wolfe KH (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell. 16: 1679–1691.

Bowers JE, Chapman BA, Rong J, Paterson AH (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433-438.

Bridges CB (1936). The Bar 'gene' a duplication. Science 83:210–211.

Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T et al. (2006). Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc. Natl Acad. Sci. 103:14959-14964.

Chapman BA, Bowers JE, Feltus FA, Paterson AH (2006). Buffering crucial functions by paleologous duplicated genes may impart cyclicality to angiosperm genome duplication. Proc. Natl. Acad. Sci. USA 103:2730–35.

Christian roth; Shruti Rastogi;, Lars Arvestad;, Katharina Dittmar;, Sara Light;, Diana Ekman;, And David A. Liberles (2007). Evolution After Gene Duplication: Models, Mechanisms, Sequences, Systems, and Organisms. J. Exp. Zool. B Mol. Dev. Evol. 308(1):58-73.

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L (2005). Chromosome evolution in eukaryotes: a multi-kingdom perspective.

Trends Genet. 21:673–82.
Crane PR, Friis EM, Pedersen KR (1995). The origin and early diversification of angiosperms. Nature 374: 27–33.
Crane PR, Herendeen P, Friis EM (2004). Fossils and plant phylogeny. Am. J. Bot. 91:1683–1699.
Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V (2004). Darwin's abominable mystery: Insights from a super tree of the angiosperms. Proc. Natl. Acad. Sci. USA 101:1904–09.
Doyle JA, Donoghue MJ (1993). Phylogenies and angiosperm diversification. Paleobiology 19: 141–167.
Force A, Cresko WA, Pickett FB, Proulx SR, Amemiya C, Lynch M (1999). The origin of subfunctions and modular gene regulation. Genetics 170:433–46.
Gaut BS, Doebley JF (1997). DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl Acad. Sci. 94: 6809-6814.
Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H et al. (2002) A draft sequence of the rice genome (*Oryza sativa L. ssp. japonica*). Science 296: 92–100.
Grant V (1981). Plant Speciation. Columbia University Press, New York.
Gupta PK (2007). Duplication and Deficiencies. In: Cytogenetics (7th ed.). Rastogi Publication, Meerut. pp.19-43.
Harrison PM, Hegyi H, Balasubramanian S, Luscombe NM, Bertone P, Echols N, Johnson T, Gerstein M (2002). Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res. 12: 272–280.
He P, Friebe B, Gill B, Zhou J-M (2003). Allopolyploidy alters gene expression in the highly stable hexaploid wheat. Plant Mol Biol. 52:401-414.
Hughes AL (1994). The evolution of functionally novel proteins after gene duplication. Proc. Biol. Sci. 256:119–124.
Hurles M (2004). Gene duplication: the genomic trade in spare parts. PloS Biol. 2(7):900–904.
Hurst LD, Smith NGC (1998). The evolution of concerted evolution. Proc. R. Soc. Lond. Ser. B 265: 121–127.
Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C et al.(2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467.
Kashkush K, Feldman M, Levy AA (2002). Gene loss, silencing, and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651-1659.
Kellis M, Birren BW, Lander ES (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cereviseae*. Nature 428:617-624.
Ku HM, Vision T, Liu J, Tanksley SD (2000). Comparing sequenced segments of the tomato and Arabidopsis genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. Proc. Natl Acad. Sci. 97: 9121-9126.
Li WH (1997). Molecular Evolution, Sinauer.
Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ (2005). Human sub-telomeres are hot spots of inter chromosomal recombination and segmental duplication. Nature 437:94–100.
Lynch M, Force A (2000). The probability of duplicate gene preservation by subfunctionalization. Genetics 154: 459–473.
Lynch M, Conery JS (2000). The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155.
Lynch M, O'Hely M, Walsh B, Force A (2001). The probability of preservation of a newly arisen gene duplicate. Genetics 159: 1789–1804
Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005). Modeling gene and genome duplications in eukaryotes. Proc. Natl. Acad. Sci. USA 102:5454-5459.
Masterson J (1994). Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264: 421–423.
Mombaerts P (2001). The human repertoire of odorant receptor genes and pseudogenes. Annu. Rev. Genomics Hum. Genet. 2: 493–510.
Nei M (2009) Gene duplication and nucleotide substitution in evolution. Nature 221: 40–42.
Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature 457: 327–331.
Nowak MA, Boerlijst MC, Smith JM (1997). Evolution of genetic redundancy. Nature 388:167–171.
Ohno S (1970). Evolution by Gene Duplication. Springer-Verlag.
Ota T, Nei M (1995). Evolution of immunoglobulin VH pseudogenes in chickens. Mol. Biol. Evol. 12: 94–102.
Pan D, Zhang L (2007). Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: a novel strategy to estimate gene duplication rates. Genome Biol. 8:R158.
Piontkivska H, Alejandro PR, Masatoshi N (2002). Purifying selection and birth-and-death evolution in the histone H4 gene family. Mol. Biol. Evol. 19: 689–697
Roulin A, Paul LA, Marc L, Jessica S, Andrew F, Greg M, Gary S, Rebecca WD, Scott AJ (2012). The fate of duplicated genes in a polyploid plant genome. Plant J. 73(1): 143-53.
Rouquier S, Blancher A, Giorgi D (2000). The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. Proc. Natl. Acad. Sci. 97: 2870–2874.
Samonte RV, Eichler EE (2002). Segmental duplications and the evolutionof the primate genome. Nat Rev Genet. 3:65–72.
Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature 440:341-345.
Semon M, Wolfe KH (2007). Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. Trends Genet. 23:108-112.
Shanks N (2004). *God, the Devil, and Darwin*, Oxford University Press, NY. p.74.
Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y (2002). The hidden duplication past of Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA 99:13627-13632.
Sterck L, Rombauts S, Vandepoele K, Rouze P, Peer YV (2007). How many genes are there in plants (. . . and why are they there)? Curr. Opin. Plant Biol. 10:199–203.
Stuessy TF (2004). A transitional–combinatorial theory for the origin of angiosperms. Taxon. 53:3–16.
Taylor JS, Peer VD, Meyer A (2001). Genome duplication, divergent resolution and speciation. Trends Genet. 17:299–301.
Thomas BC, Pedersen B, Freeling M (2006). Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homologue leaving clusters enriched in the sensitive genes. Genome Res. 16:934-946.
Titov II, Vorozheykin PS (2011). Analysis of miRNA duplication in the human genome and the role of transposon evolution in this process. Russian J. Genet. 1(4): 308-314.
Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A et al. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313:1596-1604.
Vinckenbosch N, Dupanloup I, Kaessamann H (2006). Evolutionary fate of retroposed gene copies in the human genome. PNAS 103 (9): 3220-25.
Walsh JB (1995). How often do duplicated genes evolve new functions? Genetics 139: 421–428.
Wang J, Tian L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ (2004). Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics 167:1961-1973.
Wang Z, Banik M, Cloutier S (2013). Divergent evolutionary mechanisms of co-located *Tak*/*Lrk* and *Glu-D3* loci revealed by comparative analysis of grass genomes. Genome 56(4):195-204.
Werth CR, Windham MD (1991). A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. Am. Nat. 137:515–526.
Wing SL, Boucher LD (1998). Ecological aspects of the cretaceous flowering plant radiation. Annu. Rev. Earth Planet Sci. 26: 379–421.
Wolfe KH, Shields DC (1997). Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387:708–713.
Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science 319:1527–1530.
Zhang J (2003). Evolution by gene duplication: an update. Trends Ecol. Evol. 18:192-98.
Zhang X, Firestein S (2002). The olfactory receptor gene superfamily of the mouse. Nat. Neurosci. 5: 124–133.