Perspective & Hypotheses

Vol. 7, No. 1 (2024)
ISSN: 2532-5876

DOI: 10.13133/2532-5876/18152

Polyploidization, Gene Duplication, and the Origin of Variability. Where is the Evidence?

(RIP) Milton H. Gallardo[†a] *& Elkin Y. Suárez-Villota*[b*]

[a] *Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Casilla 567, Valdivia, Chile*

[b] *Instituto de Ciencias Naturales, Universidad de las Américas, Av. Jorge Alessandri 1160, 5 Piso, Campus El Boldal, Concepción, Chile*

***Corresponding author:** Elkin Y. Suárez-Villota, Email: esuarezv@gmail.com

Abstract

The emergence of molecular complexity and its impact on evolutionary organismal innovations, achieved through the duplication of ancestral states and resulting in the creation of new structures and functions, precedes the acclaimed work of Ohno *et al.* (1968) and the advent of the genome research era. The insights of Margaret O. Dayhoff (1966) let her to conclude that ferrodoxins and other proteins were derived by doubling of short peptides and that proteins' first folded domains arose by duplication, fusion, and diversification of shorter, ancestral peptides. Here, we survey some important milestones related to gene duplication and polyploidy, starting with the contribution of Dayhoff, which reveals a complex landscape where gene duplication emerges as a fundamental evolutionary force (sections 1, 2). We subsequently address polyploidy as one of the factors contributing to gene duplication, collectively driving the evolution of molecular complexity (section 3). In this context, we explore the patterns of gene/genome duplications in the expansion of Hox gene clusters, which serve as signatures of ancient polyploidization, highlighting their role in significant evolutionary transitions (section 4). We then examine the empirical findings of synthetic polyploids and the genetic variation in heat shock proteins in wheat (sections 5, 6). These investigations offer critical insights, suggesting that lineage crossings involving chromosome set duplication (allopolyploidy) are consequential phenomena in hybrid speciation, significantly contributing to the emergence of a substantial portion of macroevolutionary diversity. Allopolyploidy and ancient gene transfers among the three domains of life generate such a variation that mutation rates based on common descent lose preponderance and the notion of tree of life gets suffocated in an entangled genomic bush.

Keywords: macroevolution, genome duplication, allopolyploidy, hybridization, Hox genes

Citation: Gallardo, MH, & Suárez-Villota, EY 2024, "Polyploidization, Gene Duplication, and the Origin of Variability. Where is the Evidence?", *Organisms: Journal of Biological Sciences*, vol. 7, no. 1, pp. 13-22. DOI: https://doi.org/10.13133/2532-5876/18152

1. Margaret O. Dayhoff and the Evolution of Protein Complexity by Gene Duplication

The seminal idea that complex macromolecules and organismal novelties are derived by duplication of ancestral states predates by far the advent of genome research (Eck and Dayhoff 1966). Over 50 years ago, the pioneering work of Margaret O. Dayhoff (1925-1983) led her to conclude that functional proteins evolved through gene duplication, resulting in the the doubling and self-assembly of short peptides. Around the same time, Dayhoff *et al.* (1965) had started the Atlas of Protein Sequence and Structure, a book series, where a model of evolutionary change based on gene mutations and natural selection was

advanced (Dayhoff *et al.* 1978). Surprisingly, the idea that protein complexity is achieved by duplication of simpler units was overlooked by Ohno *et al.* (1968) in his acclaimed work on evolution from fish to mammals by gene duplication.

Dayhoff postulated that sequence homology within domains of tertiary and quaternary structures of ferrodoxins and other ancient proteins resulted from gene duplications (Eck and Dayhoff 1966; Hunt *et al.* 1974; Schwartz and Dayhoff 1978). Following duplication, the fusion of monomers leads to the observed sequence homology within protein domains, which gradually diminishes as each monomer evolves independently (Romero *et al.* 2016). Natural selection is involved in this evolutionary process, restricting sequence variation on the primary structure of proteins (Hunt and Dayhoff 1970).

Ferredoxins, studied by Dayhoff, are crucial enzymes in photosynthesis and exhibit significant internal sequence homology attributed to duplication, fusion, and peptide diversification processes. (Eck and Dayhoff 1966; Schwartz and Dayhoff 1978). Notably, clostridial-type ferredoxins provide compelling evidence for a widespread gene duplication event shared among anaerobic, heterotrophic bacteria near the root of the evolutionary tree. Comparative analyses using ferredoxins, 5S RNA, and c-type cytochromes suggest a secondary duplication event preceding the radiation of eukaryotes and involved in cellular respiration (Schwartz and Dayhoff 1978). Recent investigations into the origins of oxygenic photosynthesis confirm that the Type I photosynthetic reaction centre comprises a heterodimeric core consisting of two homologous subunits (PsaA and PsaB), arising from gene duplication (Cardona 2017). This stands in contrast to the reaction centre of anoxygenic phototrophs, which features a homodimeric core (Liebl *et al.* 1993). A compelling hypothesis regarding the evolution of a heterodimeric Type I reaction centre suggests that the gene duplication enabling the divergence of PsaA and PsaB was a response to incorporate photoprotective mechanisms against the formation of reactive oxygen species, occurring after the origin of water oxidation to oxygen (Cardona 2018). This is consistent with the hypothesis that the event originating the duplicated heterodimeric condition occurred in the early Archean, before the Great Oxidation Event, approximately two billion years ago (Cardona 2018; Oliver *et al.* 2021). In this scenario, marked by a significant increase in biological complexity, aerobic respiration preceded oxygen-releasing photosynthesis (Sousa *et al.* 2013; Cardona 2018; Soo *et al.* 2019).

Dayhoff's concepts regarding gene duplication, fusion, and diversification, along with natural selection, should serve as the foundational framework for elucidating the extensive molecular complexity found in proteins. However, the fundamental process of duplication, which underlies the emergence of complex proteins, has often been overlooked over the years and much of the research has focused on studying the adaptive scenarios of protein evolution (*e.g.* Bloom and Arnold 2009; Jayaraman *et al.* 2022). Indeed, there remains a considerable gap in our understanding of how and when short peptides and domains duplicate, originate, and combine (Buljan and Bateman 2009; Schaper and Anisimova 2015). Gene duplication is frequently regarded as a phenomenon for elucidating macromolecular complexity, functioning both as a source and a consequence of evolutionary processes. Nonetheless, comprehending the mechanisms that lead to duplications is paramount, as they represent the genuine sources (unequal recombination, replication errors, transposition, and polyploidization).

2. Protein Evolution Through Gene Duplication

Proteins are enormously diverse agents of life. Their evolution, based on sequence similarity has provided clues about gene functions. They display substantial sequence similarity and a three-dimensional structure derived by autonomous folding of units or domains (Söding and Lupas 2003). Domains are evolutionary units whose coding sequences may undergo duplication, recombination, and divergence. These processes can occur randomly and be selected not only at the domain level. Typically, small proteins contain one domain of 100 to 250 residues and large proteins contain a combination of them. Moreover, domain families contain small proteins or parts of larger ones, descended from a common ancestor (Chothia *et al.* 2003). Thus, protein domains evolved by gene duplication, forming novel and more complex proteins (Chothia *et al.* 2003; Levy *et al.* 2008). Most models of gene duplication consider genetic redundancy and predict that after doubling, the function and structure of proteins diversify (Conant and Wolfe 2008; Kuzmin *et al.* 2020; Birchler and Yang 2022). In fact, prolonged high rates of evolution would have been determined by functional properties, acquired during, or soon after a gene duplication event (Pich and Kondrashov 2014). Interestingly, while duplications contributed to the emergence of novel traits and species diversification, phylogenetic

analyses in plants have inferred 'lag-times' between duplication events and radiations. This is supported by phylogenetic asymmetries with species-rich crown groups and species-poor sister clades occurring before duplication events. Thus, the diversification of crown groups may involve not only duplication events and novel traits but also evolutionary factors such as migration events, changing environments, and differential extinction rates (Scharanz *et al.* 2012).

At the time, since homologous structures of proteins did not exist, or could not be identified, first models were constructed from scratch. This procedure, called *ab initio* modeling, was the first approach to address the riddle of protein structure (Lee *et al.* 2009). Later on, gene duplication, mutation and recombination, became more important to address the subject. In fact, the genome sequencing of the bacterium *Haemophilus influenzae* indicated that at least one-third of this protein arose after gene duplication (Brenner *et al.* 1995). This doubling process has even been reported in viruses (Shackelton and Holmes 2004; Simon-Loriere and Holmes 2013; Willemsen *et al.* 2016).

Likewise, the detection of two paralogues of the tRNA endonuclease gene of *Methanocaldococcus jannaschii* in the genome of the crenarchaeote *Sulfolobus solfataricus* led to the identification of an unrecognized oligomeric form. Both genes code for different subunits required cleaving the pre-tRNA substrate. There are three forms of tRNA endonucleases in the Archaea, namely, a homodimer and a heterotetramer (in Euryarchaea), and a third, heterotetramer endonuclease (in Crenarchaea and Nanoarchaea). It is postulated that the last one likely resulted by gene duplication (or horizontal gene transfer), and subsequent subfunctionalization (Tocchini-Valentini *et al.* 2005).

Ribosomal protein genes constitute a large class of conserved duplicated genes in mammal with a substantial number of duplicates are transcriptionally active. Selection against dominant-negative mutations would be responsible for its unexpected retention and conservation (Dharia *et al.* 2014). Ribosomal assembly proceeds by fusing two interacting subunits, to the current atomic-resolution structures of the prokaryotic 70S and the eukaryotic 80S ribosomes (Melnikov *et al.* 2012). Large and small subunits has been captured in different functional states (Yusupova and Yusupov 2017). Although inferences about ribosomal origin are speculative (Smith *et al.* 2008; Fox 2010), a model by accretion evolution has been hypothesized using 3D comparative methods. In this model, the ribosome evolved by recursively adding expansion segments, iteratively growing, subsuming, and freezing the rRNA (Petrov *et al.* 2015).

Ancestral protein reconstruction allows the characterization of ancient macromolecules by computational analyses of modern-day protein sequences. Nevertheless, the reconstruction of protein families is limited, as exemplified by the antigen receptors of jawed vertebrates, which evolved from an extinct homodimeric ancestor through gene duplication (Rouet *et al.* 2017). Similar studies supported the idea that most domain gains in animal proteins were directly mediated by gene fusion, preceded by duplication and recombination (Marsh and Teichmann 2010).

In connection with proteins assembling, Levy *et al.* (2008) demonstrated the reversal of the process through the dilution of the denaturant and/or manipulation of the ionic strength. They observed the recovery of the original homodimer in 50% of the studied complexes. They refer to this result as "a molecular analogy to Haeckel´s evolutionary paradigm of embryonic development, where an intermediate in the assembly of a complex represents a form that appeared in its own evolutionary history". It must be said that Haeckel´s biogenetic law contrasts two timescales: ontogenetics and phylogenetics. Ontogeny recapitulates in some way the phylogeny or life history of biological units. Self-assembling, in turn refers to reversible chemical accretion of symmetric multisubunit complexes occurring in the same timescale.

The presence of diverse multigene eukaryotic families underscores the significance of gene duplication followed by the diversification of functional genomes. However, the origins of these duplication events and the myriad of new functions that emerge alongside them remain less understood. There is a scarcity of cases that clearly distinguish patterns from processes, particularly when examining the evolutionary progression of protein functional diversification. Several publications continue to echo Dayhoff's framework of gene duplication, fusion, and diversification. However, the process that may be involved in the gene expansion pattern may be polyploidy, which could entail the hybridization of lineages.

3. Gene Duplication in Polyploidization and Genome Size

Polyploidization occurs when a complete set of chromosomes is added to an existing genome from the same species (autopolyploidy) or

through hybridization (allopolyploidy). These phenomena often result from errors in meiotic and mitotic segregation, leading to chromosomal endoreduplication during gamete production (Consortium 2014; Fox and Duronio 2013). Such processes induce revolutionary and evolutionary changes in the function and structure of the genome (Feldman and Levy 2012; Van de Peer *et al.* 2017). The doubling of DNA content, a consequence of polyploidization, accompanies extensive gene duplications (Van de Peer *et al.* 2017). Given that gene duplication significantly influence protein structure, as discussed in the preceding sections, hybridization processes (allopolyploidy) or autopolyploidy could be implicated in the origin of increased protein domains.

Since polyploidy has occurred repeatedly throughout evolutionary history, genome size increases accordingly, although not in an ideal geometrical progression. Derived from genetic redundancy, molecular and cytological adjustments lead to gene losses and gains, altered regulatory genetic and epigenetic pathways (Markov and Kaznacheev 2016; Van de Peer *et al.* 2017). Genome evolution has followed varied evolutionary pathways as indicated by gradual and quantum shifts accounting for the differences in DNA content (Gallardo *et al.* 2003; Gregory and Hebert 1999). One of the most significant shifts in genome size is exemplified by the transition from haploidy to diploidy (and later, to polyploidy). Indeed, the staggering genome size variation across eukaryotes, mounts to over 64,000-fold whereas in land plants it ranges around 2,400-fold (Gregory 2024; Pellicer *et al.* 2018). Genome size has been widely recognized as instrumental to understand genome evolution (Levasseur and Pontarotti 2011; Wolfe 2006), molecular novelties (Conant and Wolfe 2008; Deng *et al.* 2010) and organismal complexity (Ferguson *et al.* 2014; McLysaght *et al.* 2002; Panopoulou and Poustka 2005). Thus, genome size variation is a complex topic with ongoing debates surrounding its evolutionary origins and impacts. Questions persist regarding whether genome size is a neutral trait or subject to selective pressures, and to what extent these pressures shape evolutionary outcomes. Duplications and deletions of genic regions can have immediate phenotypic effects, while changes in non-coding DNA may have longer-term consequences. Advances in sequencing technologies offer new insights into these mechanisms, but integrating them with traditional evolutionary experiments could provide a comprehensive understanding of genome size evolution and resolve existing debates (Blommaert 2020).

4. Genome Duplication and Hox Genes

Hox genes, a subfamily of homeobox-containing transcription factors, specify cell fate along the anterior-posterior axis of bilaterian animals (Mallo and Alonso 2013). Whole-genome duplication (WGD) is the most widely accepted explanation for the numerical increase in *Hox* gene clusters coincident with the origin of vertebrates and gnathostomes (Amores *et al.* 1998; Holland *et al.* 1994; Pascual-Anaya *et al.* 2013). In fact, the structure and gene content of the amphioxus genome corroborated the existence of two genome-wide duplications and subsequent reorganizations in the vertebrate lineage (Putnam *et al.* 2008). The first round of genome duplication would have predated the Cambrian explosion while the second would have occurred in the early Devonian (2R hypothesis). A fish-specific round of WGD is proposed to have occurred by the late Devonian (Meyer and Schartl 1999). Phylogenetic analyses suggest that tandem duplication of a *protoHox* gene produced a four-gene cluster, which was duplicated producing a four-gene *Hox* cluster, and a four-gene Para*Hox* cluster on a different chromosome (Brook *et al.* 1998). It is argued that these genome duplications were causally associated with quantum jumps in morphological complexity, body design, and adaptive radiations (reviewed in Taylor and Raes 2004). Apparently, vertebrates have undergone significant modifications since the last common ancestor of the chordates. Although some anterior genes are dated back to the ancient divergence between protostomes and deuterostomes, others have been lost from the vertebrate lineage more recently (Butts *et al.* 2010; Furlong and Holland 2002; Zhong and Holland 2011). Indeed, the family of posterior *Hox* genes is claimed to probably originated through independent tandem duplication events at the origin of each of the ambulacrarian, cephalochordate and vertebrate/urochordate lineages (Pascual-Anaya *et al.* 2013).

Phylogenetic analysis of ambulacrarian posterior genes (*Hox* 9 to 13), indicates a lack of correlation and multiple polytomies between this cluster and the posterior genes from cephalochordates and vertebrates (Ferrier *et al.* 2000). This lack of correlation pattern (one-to-one orthology assignments) is referred to as deuterostome posterior flexibility (Ferrier *et al.* 2000; Amemiya *et al.* 2008). Its far from understood causality is claimed to have resulted from dilution of selective constraints (Ferrier *et al.* 2000).

Thus, the genomic organization of bilaterian animals, reflected by a shared set of *Hox* genes is

rather confusing. In some distantly-related species, *Hox* genes are collinearly clustered, but not in others. This suggests that the urbilaterian ancestor had a *Hox* gene set with clustered genomic organization that was subsequently either maintained or lost (Duboule 2007). When discussing the reasons and mechanisms behind *Hox* gene clustering and collinear developmental organization, the phenomenon of allopolyploidy becomes crucial. Indeed, *Hox* gene clusters arranged on different chromosomes serve as a cytogenetic evidence supporting the underlying causal process of polyploidy.

5. Genome Duplication and Synthetic Polyploids

Wild and cultivated allopolyploids are well adapted and stable. Synthetic (man-made) allopolyploids are cytogenetically unstable at the beginning, exhibiting in some cases homeotic transformation (Murai *et al.* 2002; Murai 2013), but eventually leading to the establishment of biological novelties (Chester *et al.* 2012; Comai 2000). Chromosomal rearrangements, changes in chromatin constitution, fluctuations, and distribution in repeats of repetitive DNA accompany the newly synthesized allopolyploids (Liu *et al.* 1998a; Liu *et al.* 1998b). Retrotransposition is activated following polyploidization in several syntethic plants (Parisod *et al.* 2010). Moreover, regulatory abnormalities derive from ploidy changes and/or incompatible interactions between parental genomes (Jones and Pasakinskiene 2005). In this way, it has been suggested that intergenomic incompatibilities play the major role in the generation of a fertile organism (Comai 2000). Epigenetic expression patterns are altered as well as chromatin remodeling, affecting promoter´s response in the new cellular environment (Wendel *et al.* 2016). On the other hand, the impacts of polyploidy on the genomic processes of natural *Arabidopsis* populations are subtle yet far-reaching. These effects encompass reduced purifying selection efficiency, variations in linked selection, and extensive gene flow from diploids. Polyploidy initially conceals harmful mutations, accelerates nucleotide substitution rates, and facilitates interploidy introgression (Monnahan *et al.* 2019).

The importance of hybridizing the paternal species of naturally-occurring polyploids (2n, 4n, 6n) is that their genome dynamics and the phylogenetic pattern already evolved in nature (H_0), could be compared with its homologous synthetic allopolyploid combination, produced and maintained in laboratory conditions (H_1). All parameters of genetic or ecological interest can be accurately studied and any empirical result, comparatively validated. Thus, sequence elimination after polyploidization, genomic differentiation, and diploid-like meiotic behavior of the synthetic counterpart turn into predictive empirical questions that support conceptually transformative hypotheses. In this context, no inferences are made or needed since those predictions reflect the underlying process directly.

Micro and macroevolutionary changes in newly synthesized amphiploids of *Triticum* and *Aegilops* can become fixed in few generations and could give rise to evolutionary novelties (Liu *et al.* 1998a; Liu *et al.* 1998b; Mason and Wendel 2020). Nevertheless, the most synthetic allo- and autopolyploids are meiotically unstable, as evidenced by high frequencies of chromosome rearrangements in young allotetraploid species such as *Tragopogon miscellus* (Mason and Wendel 2020). Extensive karyotype variation has been observed in these species, including clear products of homoeologous recombination between the subgenomes (Chester *et al.* 2012). Additionally, studies comparing individuals and populations of synthetic lines with natural populations of the recently formed allotetraploids *Tragopogon mirus* and *T. miscellus* have detected extensive chromosomal polymorphisms (Lim *et al.* 2008). These included monosomic and trisomic individuals for particular chromosomes, intergenomic translocations, and variable sizes and expression patterns of individual rDNA loci. Chromosomal translocations, gene loss, and meiotic irregularities (i.e., quadrivalents) were detected in both synthetic lines and sibling plants (Lim *et al.* 2008). These patterns point to an explanatory meaning for cytogenetic variation and indicate that chromosomal adjustments, chromatin remodeling and elimination occur rapidly following polyploidization (Wendel *et al.* 2016). The lineage giving rise to the *Arabidopsis* genus has experienced three rounds of genome duplication in the last 250 Ma (De Bodt *et al.* 2005). Its synthetic allotetraploids also exhibit rapid epigenetic changes including gene silencing via heterochromatization and have preferentially retained development genes an others involved in signal transduction pathways (Bomblies and Madlung 2014; Del Pozo and Ramirez-Parra 2015; Shi *et al.* 2015). Thus, it appears that the species´ genetic redundancy is responsible for its rapid diversification (De Bodt *et al.* 2005; Couvreur *et al.* 2010; Schranz *et al.* 2012). In contrast to the 'genome shock' observed in synthetic polyploids, characterized by genome reorganization, altered expression, and transposition, recent research

has revealed that the genome of natural polyploid *Arabidopsis suecica* remains colinear with ancestral genomes. There is no dominance of a subgenome in expression, and transposon dynamics appear stable (Burns *et al.* 2021). This suggests that domesticated polyploids may not always accurately represent natural polyploidization processes.

The formation of allopolyploid wheat has been also accompanied by rapid nonrandom changes in low-copy noncoding, and coding DNA sequences (Liu *et al.* 1998a; Liu *et al.* 1998b; Levy and Feldman 2022). Indeed, newly synthesized amphiploids of different ploidy levels showed disappearance of parental hybridization fragments, and appearance of novel fragments. Pattern variations among individual plants of the same amphiploid level and between several synthetic and natural amphiploids occurred at random (Feldman and Levy 2012). Moreover, intergenomic recombination triggered DNA methylation and modified expression levels that led to meiotic diploidization, gene-dosage compensation and increasing variation among amphidiploid plants (Liu *et al.* 1998a; Liu *et al.* 1998b; Li *et al.* 2021). These evolutionary changes observed during the lifespan of allopolyploids increase intra-specific genetic diversity. Consequently, this enhancement leads to greater fitness and competitiveness (Feldman and Levy 2009). The scientific value of synthetic polyploids allows us to realize that the above-mentioned duplicated genomic patterns and adjustments are derived from interspecific hybridizations, precisely dated and available in the greenhouse. Thus, synthetic polyploids provide empirical tests of enormous predictive capabilities to address the otherwise overlooked transcendental evolutionary role of interlineage hybridization.

6. Small Heat Shock Proteins in Wheat

Bread wheat originated from hybridization involving genera *Triticum* and *Aegilops* to give rise to the allotetraploid emmer wheat (*Triticum turgidum*; AABB). The second hybridization event between emmer wheat and *Aegilops tauschii* (DD), occurred around 0.4 Ma and gave rise to allohexaploid wheat (*Triticum aestivum*; AABBDD). Thus, the present-day genome of wheat is a product of multiple, cyclic rounds of genome duplications (Marcussen *et al.* 2014).

Comparative analysis of small heat shock proteins (sHSPs) in bread wheat has pointed out massive intrachromosomal expansions and expression pattern diversity with polyploidization (Wang *et al.* 2017). The number of sHSPs in tetraploid wheat and in its diploid progenitors was similar, although gene copy number much higher and enriched in specific chromosome fragments of hexaploids. In fact, 25 to 31 sHSP genes were identified in diploid and tetraploid relatives whereas 117 were identified in the bread wheat; many more than the 56 to 70 copies of its tetraploid progenitors. Further genomic comparisons revealed remarkable sHSPs expansion in subgenomes A and B, but not in subgenome D, consistent with its stable gene content after tetraploidization (Wang *et al.* 2017). These findings underscore the significance of hexaploidization, alongside segmental and tandem duplications, in explaining the rise in sHSP numbers. This relationship between polyploidy and intrachromosomal segmental and tandem duplications, which contribute to sHSPs gene expansions, is also evident in *Arabidopsis* (Waters *et al.* 2008), rice (Sarkar *et al.* 2009), and soybean (Lopes-Caitar *et al.* 2013).

A detailed partitioning of chromosome 3B of bread wheat indicated that its 2,216 genes greatly surpass the gene number of homologues in rice and sorghum (Choulet *et al.* 2014). Additionally, 46% of these duplicated genes are tandemly repeated, while 56% are dispersed duplicates, resulting in an intriguingly even split. Additionally, more than twice as many duplicate genes are retained after intrachromosomal duplication relative to other grass species. The finding that 94% of the conserved genes in those grass relatives are also present in chromosome 3B indicated limited gene loss after polyploidization. Indeed, the reduction of the basic chromosome number from 12 to 7 in Triticeae proceeded by the telomeric insertion of one chromosome into a centromeric break of another; a process unaffecting gene content (Luo *et al.* 2009). The 23% of syntenic dispersed duplicates (those located at their ancestral locus) have originated from recent intrachromosomal and interchromosomal duplications at a much higher comparative rate. Interestingly, interchromosomal duplicates were evenly distributed along chromosome 3B whereas the increase of tandem duplications is only telomeric, suggesting the existence of two superimposed mechanisms of gene duplication (Choulet *et al.* 2014). These findings underscore the intricate interplay between polyploidization, gene duplication, and genomic evolution in shaping the genetic landscape of bread wheat.

7. Evolutionary Significance of Polyploidy

While polyploidization represents one of the most dramatic mutations known to occur, it is also a widespread and common phenomenon among eukaryotes, serving as a source for evolutionary innovation and species diversification (Otto and Whitton 2000; Otto 2007; Van de Peer *et al.* 2021). Indeed, the majority of flowering plants and vertebrates have descended from polyploid ancestors. Up to seven rounds of ancestral polyploidy have been suggested in major angiosperm phyla (Jaillon *et al.* 2007), and three have been proposed in the lineage that gave rise to chordates (Holland *et al.* 1994; Pascual-Anaya *et al.* 2013). Ancient polyploidies are widely recognized as events with important roles in the origin of evolutionary novelties in plants and animals, such as the origin of seeds and flowers (Clark and Donoghue 2017; Jiao *et al.* 2011), as well as the emergence of limbs and jaws (Holland 1998; Pascual-Anaya *et al.* 2013). Nevertheless, polyploidy is thought to have had a lesser role in animal evolution (Otto and Whitton 2000). The distinction between the numbers of validated polyploids in plants and animals is indeed substantial. While attributing solely to specific factors may oversimplify this issue, the increased developmental plasticity in plants, the absence of the Weisman barrier, and differences in meiotic processes that prevent rapid solutions to high crossover rates in animals post-WGD could all play crucial roles (Mable *et al.* 2004 and literature therein). Nevertheless, this difference gets blurred under genomic scrutiny. In fact, the conventional assertion that polyploidy is less feasible in animals has been reverted, since insects the most speciose class of invertebrates, has experienced massive polyploidization and extensive genome duplication (Li *et al.* 2018).

In this exposition, we explore genetic and genomic data concerning duplication, investigating the origin and evolutionary patterns of *Hox* genes. We also explore studies of synthetic and natural plant polyploids to glean insights into their evolutionary trajectories. Our survey suggests that polyploidy plays a significant role in generating genetic variability, driving protein evolution, and facilitating the emergence of macroevolutionary diversity. Moreover, contributes to the activation of transposons and the formation of tandem duplicates in diverse organisms. These intricate molecular processes, stemming from gene duplication and polyploidy, challenge conventional evolutionary paradigms and enrich our understanding of macroevolutionary diversity.

Acknowledgments

This perspective is in memory of Dr. Milton H. Gallardo (1947-2019). Several parts of the text remained intact according to the original notes found in Dr. Gallardo's house after his death. EY. Suárez-Villota would particularly like to acknowledge Mónica Barbet-Claus and Paola Solis for allowing him to take some of these notes.

The author also wishes to thank the Editor in Chief, Dr. Andràs Paldi, for encouraging him to complete and revise the article, as well as the anonymous reviewer for their detailed review and suggestions for improving the manuscript.

References

Amemiya, CT, Prohaska, SJ, Hill-Force, A, Cook, A, Wasserscheid, J, *et al.* 2008, "The amphioxus *Hox* cluster: characterization, comparative genomics, and evolution", *J Exp Zool B Mol Dev Evol*, 310, 5, pp. 465-77.

Amores, A, Force, A, Yan, YL, Joly, L, Amemiya, C, *et al.* 1998, "Zebrafish *Hox* clusters and vertebrate genome evolution", *Science*, 282, 5394, pp. 1711-4.

Birchler, JA, & Yang, H 2022, "The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation", *Plant Cell*, 34, 7, pp. 2466-74.

Blommaert, J 2020, "Genome size evolution: towards new model systems for old questions", *Proc Biol Sci*, 287, 1933, p. 20201441.

Bloom, JD, & Arnold, FH 2009, "In the light of directed evolution: pathways of adaptive protein evolution", *Proc Natl Acad Sci U S A*, 16, 106, pp. 9995-10000.

Bomblies, K, & Madlung, A 2014, "Polyploidy in the *Arabidopsis* genus", *Chromosome Res*, 22, 2, pp. 117-34

Brenner, SE, Hubbard, T, Murzin, A, & Chothia, C 1995, "Gene duplications in *H. influenzae*", *Nature*, 378, 6553, p. 140.

Brooke, N, Garcia-Fernàndez, J, & Holland, P 1998, "The Para*Hox* gene cluster is an evolutionary sister of the *Hox* gene cluster", *Nature*, 392, pp. 920-22.

Buljan, M, & Bateman, A 2009, "The evolution of protein domain families", *Biochem Soc Trans*, 37, Pt 4, pp. 751-5.

Butts, T, Holland, PW, & Ferrier, DE 2010, "Ancient homeobox gene loss and the evolution of chordate brain and pharynx development: deductions from amphioxus gene expression", *Proc Biol Sci*, 277, 1699, pp. 3381-9.

Burns, R, Mandáková, T, Gunis, J, Soto-Jiménez, LM, Liu, C, Lysak, MA, Novikova, PY, & Nordborg, M 2021, "Gradual evolution of allopolyploidy in *Arabidopsis suecica*'. *Nat Ecol Evol*, 5, pp. 1367-81.

Cardona, T 2017, "Photosystem II is a chimera of reaction centers", *Journal of Molecular Evolution*, 84, 2-3, pp. 149-51.

Cardona, T 2018, "Early archean origin of heterodimeric photosystem I", *Heliyon*, 4, 3, p. e00548.

Chester, M, Gallagher, JP, Symonds, VV, Cruz da Silva, AV, Mavrodiev, EV, Leitch, AR, Soltis, PS, & Soltis, DE 2012, "Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae)", *Proc Natl Acad Sci U S A*, 109, 4, pp. 1176-81.

Chothia, C, Gough, J, Vogel, C, & Teichmann, SA 2003, "Evolution of the protein repertoire", *Science*, 300, 5626, pp. 1701-3.

Choulet, F, Alberti, A, Theil, S, Glover, N, Barbe, V, *et al.* 2014, "Structural and functional partitioning of bread wheat chromosome 3B", *Science*, 345, 6194, p. 1249721.

Clark, JW, & Donoghue, PCJ 2017, "Constraining the timing of whole genome duplication in plant evolutionary history", *Proc Biol Sci*, 284, 1858, p. 20170912.

Comai, L 2000, "Genetic and epigenetic interactions in allopolyploid plants", *Plant Mol Biol*, 43, 2-3, pp. 387-99.

Conant, GC, & Wolfe, KH 2008, "Turning a hobby into a job: how duplicated genes find new functions", *Nat Rev Genet*, 9, p. 938.

Consortium, TIWGS 2014, "A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome", *Science*, 345, 6194, p. 1251788.

Couvreur, TL, Franzke, A, Al-Shehbaz, IA, Bakker, FT, Koch, MA, & Mummenhoff, K 2010, "Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae)", *Mol Biol Evol*, 27, 1, pp. 55-71.

Dharia, AP, Obla, A, Gajdosik, MD, Simon, A, & Nelson, CE 2014, "Tempo and mode of gene duplication in mammalian ribosomal protein evolution", *PLoS One*, 9, 11, p. e111721.

Dayhoff, MO, Eck, RV, Chang, MA, & Sochard, MR 1965, *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Silver Spring (Maryland).

Dayhoff, MO, Schwartz, RM, & Orcutt, BC 1978, "A model of evolutionary change in proteins", in MO Dayhoff, (ed), *Atlas of Protein Sequence and Structure*, pp. 345-52, National Biomedical Research Foundation, Silver Spring (Maryland).

De Bodt, S, Maere, S, & Van de Peer, Y 2005, "Genome duplication and the origin of angiosperms", *Trends Ecol Evol*, 20, 11, pp. 591-7.

Del Pozo, JC, & Ramirez-Parra, E 2015, "Whole genome duplications in plants: an overview from *Arabidopsis*", *J Exp Bot*, 66, 22, pp. 6991-7003.

Deng, C, Cheng, CHC, Ye, H, He, X, & Chen, L 2010, "Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict", *Proc Natl Acad Sci U S A*, 107, 50, pp. 21593-21598.

Duboule, D 2007, "The rise and fall of *Hox* gene clusters", *Development*, 134, 14, pp. 2549-60.

Eck, RV, & Dayhoff, MO 1966, "Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences", *Science*, 152, 3720, p. 363.

Feldman, M, & Levy, AA 2009, "Genome evolution in allopolyploid wheat – A revolutionary reprogramming followed by gradual changes", *Journal of Genetics and Genomics*, 36, 9, pp. 511-18.

Feldman, M, & Levy, AA 2012, "Genome evolution due to allopolyploidization in wheat", *Genetics*, 192, 3, pp. 763-74.

Ferguson, L, Marletaz, F, Carter, JM, Taylor, WR, Gibbs, M, Breuker, CJ, & Holland, PW 2014, "Ancient expansion of the *Hox* cluster in lepidoptera generated four homeobox genes implicated in extra-embryonic tissue formation", *PLoS Genetics*, 10, 10, p. e1004698.

Ferrier, DEK, Minguillón, C, Holland, PWH, & Garcia-Fernàndez, J 2000, "The amphioxus *Hox* cluster: deuterostome posterior flexibility and *Hox14*", *Evolution & Development*, 2, 5, pp. 284-93.

Fox, DT, & Duronio, RJ 2013, "Endoreplication and polyploidy: insights into development and disease", *Development*, 140, 1, pp. 3-12.

Fox, GE 2010, "Origin and evolution of the ribosome", *Cold Spring Harb Perspect Biol*, 2, 9, p. a003483.

Furlong, RF, & Holland, PWH 2002, "Were vertebrates octoploid?", *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357, 1420, pp. 531-44.

Gallardo, MH, Bickham, JW, Kausel, G, Kohler, N, & Honeycutt, RL 2003, "Gradual and quantum genome size shifts in the hystricognath rodents", *J Evol Biol*, 16, 1, pp. 163-9.

Gregory, TR 2024, "Animal genome size database", [online]. Available from: http://www.genomesize.com [Accessed 23 February 2024].

Gregory, TR, & Hebert, PD 1999, "The modulation of DNA content: proximate causes and ultimate consequences", *Genome Res*, 9, 4, pp. 317-24.

Holland, PWH 1998, "Major transitions in animal evolution: a developmental genetic perspective", *Integrative and Comparative Biology*, 38, 6, pp. 829-42.

Holland, PWH, Garcia-Fernàndez, J, Williams, NA, & Sidow, A 1994, "Gene duplications and the origins of vertebrate development", *Development*, 1994, Supplement, p. 125.

Hunt, LT, Barker, WC, & Dayhoff, MO 1974, "Epidermal growth factor: internal duplication and probable relationship to pancreatic secretory trypsin inhibitor", *Biochem Biophys Res Commun*, 60, 3, pp. 1020-28.

Hunt, LT, & Dayhoff, MO 1970, "The occurrence in proteins of the tripeptides Asn-X-Ser and Asn-X-Thr and of bound carbohydrate", *Biochem Biophys Res Commun*, 39, 4, pp. 757-765.

Jaillon, O, Aury, J-M, Noel, B, Policriti, A, Clepet, C, *et al.* 2007, "The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla", *Nature*, 449, p. 463.

Jayaraman, V, Toledo-Patiño, S, Noda-García, L, & Laurino, P 2022, "Mechanisms of protein evolution", *Protein Sci*, 31, 7, p. e4362.

Jiao, Y, Wickett, NJ, Ayyampalayam, S, Chanderbali, AS, Landherr, L, *et al.* 2011, "Ancestral polyploidy in seed plants and angiosperms", *Nature*, 473, 7345, pp. 97-100.

Jones, N, & Pasakinskiene, I 2005, "Genome conflict in the gramineae", *The New phytologist*, 165, 2, pp. 391-409.

Kuzmin, E, Vandersluis, B, Nguyen, AN, Wang W, Koch, EN, *et al.* 2020, "Exploring whole-genome duplicate gene retention with complex genetic interaction analysis", *Science*, 368, 6498, p. eaaz5667.

Lee, J, Wu, S, & Zhang, Y 2009, "*Ab initio* protein structure prediction", in DJ Rigden, (ed), *From Protein Structure to Function with Bioinformatics*, pp. 3-25, Springer Netherlands, Dordrecht.

Levasseur, A, & Pontarotti, P 2011, "The role of duplications in the evolution of genomes highlights the need for evolutionary-based approaches in comparative genomics", *Biol Direct*, 6, p. 11.

Levy, AA, & Feldman, M 2022, "Evolution and origin of bread wheat", *Plant Cell*, 34, 7, pp. 2549-67.

Levy, ED, Boeri Erba, E, Robinson, CV, & Teichmann, SA 2008, "Assembly reflects evolution of protein complexes", *Nature*, 453, 7199, pp. 1262-5.

Li, Z, McKibben, MTW, Finch, GS, Blischak, PD, Sutherland, BL, & Barker, MS 2021, "Patterns and processes of diploidization in land plants", *Annu Rev Plant Biol*, 72, pp. 387-410.

Li, Z, Tiley, GP, Galuska, SR, Reardon, CR, Kidder, TI, Rundell, RJ, & Barker, MS 2018, "Multiple large-scale gene and genome duplications during the evolution of hexapods", *Proc Natl Acad Sci U S A*, 115, 18, pp. 4713-18.

Liebl, U, Mockensturm-Wilson, M, Trost, JT, Brune, DC, Blankenship, RE, & Vermaas, W 1993, "Single core polypeptide in the reaction center of the photosynthetic bacterium *Heliobacillus mobilis*: structural implications and relations to other photosystems", *Proc Natl Acad Sci U S A*, 90, 15, pp. 7124-8.

Lim, KY, Soltis, DE, Soltis, PS, Tate, J, Matyasek, R, Srubarova, H, Kovarik, A, Pires, JC, Xiong, Z, & Leitch, AR 2008, "Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae)", *PLoS One*, 3, 10, p. e3353.

Liu, B, Vega, JM, & Feldman, M 1998a, "Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences", *Genome*, 41, 4, pp. 535-42.

Liu, B, Vega, JM, Segal, G, Abbo, S, Rodova, M, & Feldman, M 1998b, "Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy noncoding DNA sequences", *Genome*, 41, 2, pp. 272-77.

Lopes-Caitar, VS, De Carvalho, MC, Darben, LM, *et al.* 2013, "Genome-wide analysis of the *Hsp20* gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses", *Bmc Genomics*, 14, p. 577.

Luo, MC, Deal, KR, Akhunov, ED, Akhunova, AR, Anderson, OD, *et al.* 2009, "Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae", *Proc Natl Acad Sci U S A*, 106, 37, pp. 15780-5.

Mable, BK 2004, "Why polyploidy is rarer in animals than in plants: myths and mechanisms", *Biol J Linn Soc Lond*, 82, 4, pp. 453-66.

Mallo, M, & Alonso, C.R 2013, "The regulation of *Hox* gene expression during animal development", *Development*, 140, 19, p. 3951.

Marcussen, T, Sandve, SR, Heier, L, Spannagl, M, Pfeifer, M, Jakobsen, KS, Wulff, BBH, Steuernagel, B, Mayer, KFX, & Olsen, OA 2014, "Ancient hybridizations among the ancestral genomes of bread wheat", *Science*, 345, 6194, p. 1250092.

Markov, AV, & Kaznacheev, IS 2016, "Evolutionary consequences of polyploidy in prokaryotes and the origin of mitosis and meiosis", *Biol Direct*, 11, p. 28.

Marsh, JA, & Teichmann, SA 2010, "How do proteins gain new domains?", *Genome biology*, 11, 7, pp. 126-26.

Mason, AS, & Wendel, JF 2020, "Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution", *Front Genet*, 11, p. 1014.

Melnikov, S, Ben-Shem, A, Garreau de Loubresse, N, Jenner, L, Yusupova, G, & Yusupov, M 2012, "One core, two shells: bacterial and eukaryotic ribosomes", *Nat Struct Mol Biol*, 19, p. 560.

McLysaght, A, Hokamp, K, & Wolfe, KH 2002, "Extensive genomic duplication during early chordate evolution", *Nature Genetics*, 31, p. 200.

Meyer, A, & Schartl, M 1999, "Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions", *Curr Opin Cell Biol*, 11, 6, pp. 699-704.

Monnahan, P, Kolář, F, Baduel, P, Sailer, C, Koch, J, *et al.* 2019, "Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*", *Nat Ecol Evol*, 3, pp. 457-68.

Murai, K 2013, "Homeotic genes and the ABCDE model for floral organ formation in wheat", *Plants (Basel)*, 2, 3, pp. 379-95.

Murai, K, Takumi, S, Koga, H, Ogihara, Y 2002, "Pistillody, homeotic transformation of stamens into pistil-like structures, caused by nuclear-cytoplasm interaction in wheat", *Plant J*, 29, 2, pp. 169-81.

Ohno, S, Wolf, U, & Atkin, NB 1968, "Evolution from fish to mammals by gene duplication", *Hereditas*, 59, 1, pp. 169-87.

Oliver, T, Sánchez-Baracaldo, P, Larkumx, AW, Rutherford, AW, & Cardona, T 2021, "Time-resolved comparative molecular evolution of oxygenic photosynthesis", *Biochim Biophys Acta Bioenerg*, 1862, 6, p. 148400.

Otto, SP 2007, "The evolutionary consequences of polyploidy", *Cell*, 131, 3, pp. 452-62.

Otto, SP, & Whitton, J 2000, "Polyploid incidence and evolution", *Annu Rev Genet*, 34, pp. 401-37.

Panopoulou, G, & Poustka, AJ 2005, "Timing and mechanism of ancient vertebrate genome duplications – the adventure of a hypothesis", *Trends Genet*, 21, 10, pp. 559-67.

Parisod, C, Alix, K, Just, J, Petit, M, Sarilar, V, Mhiri, C, Ainouche, M, Chalhoub, B, & Grandbastien, MA 2010, "Impact of transposable elements on the organization

and function of allopolyploid genomes", *New Phytol*, 186, 1, pp. 37-45.

Pascual-Anaya, J, D'Aniello, S, Kuratani, S, & Garcia-Fernandez, J 2013, "Evolution of *Hox* gene clusters in deuterostomes", *BMC Dev Biol*, 13, p. 26.

Pellicer, J, Hidalgo, O, Dodsworth, S, & Leitch, IJ 2018, "Genome size diversity and its impact on the evolution of land plants", *Genes (Basel)*, 9, 2, p. 88.

Petrov, AS, Gulen, B, Norris, AM, Kovacs, NA, Bernier, CR, *et al.* 2015, "History of the ribosome and the origin of translation", *Proc Natl Acad Sci U S A*, 112, 50, pp. 15396-401.

Pich, IRO, & Kondrashov, FA 2014, "Long-term asymmetrical acceleration of protein evolution after gene duplication", *Genome Biol Evol*, 6, 8, pp. 1949-55.

Putnam, NH, Butts, T, Ferrier, DE, Furlong, RF, Hellsten, U, *et al.* 2008, "The amphioxus genome and the evolution of the chordate karyotype", *Nature*, 453, 7198, pp. 1064-71.

Romero, ML, Rabin, A, & Tawfik, DS 2016, "Functional proteins from short peptides: Dayhoff's hypothesis turns 50", *Angewandte Chemie. International Ed. In English*, 55, 52, pp. 15966-71.

Rouet, R, Langley, DB, Schofield, P, Christie, M, Roome, B, Porebski, BT, Buckle, AM, Clifton, BE, Jackson, CJ, Stock, D, & Christ, D 2017, "Structural reconstruction of protein ancestry", *Proc Natl Acad Sci U S A*, 114, 15, p. 3897.

Sarkar, NK, Kim, YK, & Grover, A 2009, "Rice *sHsp* genes: genomic organization and expression profiling under stress and development", *Bmc Genomics*, 10, p. 393.

Schaper, E, & Anisimova, M 2015, "The evolution and function of protein tandem repeats in plants", *New Phytol*, 206, 1, pp. 397-410.

Schranz, ME, Mohammadin, S, & Edger, PP 2012, "Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model", *Curr Opin Plant Biol*, 15, 2, pp. 147-53.

Schwartz, RM, & Dayhoff, MO 1978, "Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts", *Science*, 199, 4327, p. 395.

Shackelton, LA, & Holmes, EC 2004, "The evolution of large DNA viruses: combining genomic information of viruses and their hosts", *Trends Microbiol*, 12, 10, pp. 458-65.

Shi, X, Zhang, C, Ko, DK, & Chen, ZJ 2015, "Genome-wide dosage-dependent and -independent regulation contributes to gene expression and evolutionary novelty in plant polyploids", *Mol Biol Evol*, 32, 9, pp. 2351-2366.

Simon-Loriere, E, & Holmes, EC 2013, "Gene duplication is infrequent in the recent evolutionary history of RNA viruses", *Mol Biol Evol*, 30, 6, pp. 1263-69.

Smith, TF, Lee, JC, Gutell, RR, & Hartman, H 2008, "The origin and evolution of the ribosome", *Biol Direct*, 3, p. 16.

Soo, RM, Hemp, J, & Hugenholtz, P 2019, "Evolution of photosynthesis and aerobic respiration in the cyanobacteria", *Free Radic Biol Med*, 140, pp. 200-5.

Söding, J, & Lupas, AN 2003, "More than the sum of their parts: on the evolution of proteins from peptides", *BioEssays*, 25, 9, pp. 837-46.

Sousa, FL, Shavit-Grievink, L, Allen, JF, & Martin, WF 2013, "Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis", *Genome Biol Evol*, 5, 1, pp. 200-16.

Taylor, JS, & Raes, J 2004, "Duplication and divergence: the evolution of new genes and old ideas", *Annu Rev Genet*, 38, 615-43.

Tocchini-Valentini, GD, Fruscoloni, P, & Tocchini-Valentini, GP 2005, "Structure, function, and evolution of the tRNA endonucleases of Archaea: an example of subfunctionalization", *Proc Natl Acad Sci U S A*, 102, 25, pp. 8933-8.

Van de Peer, Y, Ashman, TL, Soltis, PS, & Soltis, DE 2021, "Polyploidy: an evolutionary and ecological force in stressful times", *Plant Cell*, 33, 1, pp. 11-26.

Van de Peer, Y, Mizrachi, E, & Marchal, K 2017, "The evolutionary significance of polyploidy", *Nat Rev Genet*, 18, 7, pp. 411-24.

Wang, X, Wang, R, Ma, C, Shi, X, Liu, Z, Wang, Z, Sun, Q, Cao, J, & Xu, S 2017, "Massive expansion and differential evolution of small heat shock proteins with wheat (*Triticum aestivum* L.) polyploidization", *Sci Rep*, 7, 1, p. 2581.

Waters, ER, Nguyen, SL, Eskandar, R, Behan, J, & Sanders-Reed, Z 2008, "The recent evolution of a pseudogene: diversity and divergence of a mitochondria-localized small heat shock protein in *Arabidopsis thaliana*", *Genome*, 51, 3, pp. 177-86.

Wendel, JF, Jackson, SA, Meyers, BC, & Wing, RA 2016, "Evolution of plant genome architecture", *Genome Biol*, 17, 37.

Willemsen, A, Zwart, MP, Higueras, P, Sardanyés, J, & Elena, SF 2016, "Predicting the stability of homologous gene duplications in a plant RNA virus", *Genome Biol Evol*, 8, 9, pp. 3065-82.

Wolfe, KH 2006, "Comparative genomics and genome evolution in yeasts", *Philos Trans R Soc Lond B Biol Sci*, 361, 1467, pp. 403-12.

Yusupova, G, & Yusupov, M 2017, "Crystal structure of eukaryotic ribosome and its complexes with inhibitors", *Philos Trans R Soc Lond B Biol Sci*, 372, 1716, p. 20160184.

Zhong, YF, & Holland, PW 2011, "The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in mouse and human lineages", *BMC Evol Biol*, 11, 169.