

Pamela S. Soltis
Douglas E. Soltis *Editors*

Polyploidy and Genome Evolution

 Springer

Polyploidy and Genome Evolution

Pamela S. Soltis · Douglas E. Soltis
Editors

Polyploidy and Genome Evolution

 Springer

Editors

Pamela S. Soltis
Florida Museum of Natural History
University of Florida
Gainesville, FL
USA

Douglas E. Soltis
Florida Museum of Natural History and
Department of Biology
University of Florida
Gainesville, FL
USA

ISBN 978-3-642-31441-4 ISBN 978-3-642-31442-1 (eBook)
DOI 10.1007/978-3-642-31442-1
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012945474

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Polyploidy (whole-genome duplication; WGD) is common in plants and has long been considered as both an important speciation mechanism and a crucial component of plant genome structure. Analyses of chromosome numbers and hypothesized breaks between diploid and polyploid base numbers have suggested anywhere from 30 to 80 % of all angiosperms are polyploid. While recent polyploids may be easily detected through comparison of chromosome numbers, various processes of diploidization or fractionation may substantially alter chromosome numbers and structure, ultimately masking the evolutionary history of duplication events. In contrast, other footprints of ancient WGD may remain in the genome, even when chromosome numbers no longer carry the signature of past WGDs. Genome sequences and other sources of genomic data tell us that, in fact, all angiosperms, as well as all seed plants, have undergone one or more rounds of polyploidy. Furthermore, ancient WGD characterizes all vertebrates, with subsequent, more recent polyploidization in fishes and amphibians. Ancient WGD is also evident in the genomes of yeast and other fungi. While more common in plants than other major lineages of life, polyploidy is now recognized as a fundamental process in all crown eukaryotes. Polyploidy plays a major role in shaping genome structure and organization and in establishing patterns and mechanisms of gene regulation. In fact, it is now impossible to construct models of genome evolution that do not account for genomic content and genetic interactions contributed by WGD.

It has been over 30 years since the publication of a comprehensive treatment of polyploidy [*Polyploidy: Biological Relevance*, W. H. Lewis (ed.), 1980]. The intervening years have witnessed a technological revolution with a transition from the early days of recombinant DNA to nearly routine genome sequencing of non-model organisms and from limited biological computing to high-performance computing networks for the biological sciences. These transformations in methodology and computation permit fresh perspectives on polyploidy and the ability to ask old questions with new tools.

Over the past decade, it has been a dream of ours to publish a book that synthesizes the rapid progress in understanding the role of polyploidy in genome

evolution, and this book is now a reality. In the current volume, we have compiled the expertise of scientists studying polyploid genome evolution from multiple perspectives in phylogenetically diverse organisms. Topics range from the conceptual and theoretical underpinnings of polyploidy (chapters by McGrath and Lynch, Birchler) to processes at work in polyploid genomes (Zielinski and Mittelsten Scheid, Finigan et al., Evans et al.), to patterns of ancient polyploidy and its detection (Burleigh, Paterson et al.), to a series of case studies that both document attributes of genome evolution in focal species and address general properties of polyploid genomes, from ancient polyploids [maize (Schnable and Freeling), legumes (Doyle), vertebrates (Cañestro), fishes (Braasch and Postlethwait), yeast (Hudson and Conant)] to classic model polyploids [cotton (Wendel et al.), tobacco (Kovarik et al.), wheat (Feldman et al.)] to very recent ones [*Spartina* (Ainouche et al.), *Senecio* (Hegarty et al.), and *Tragopogon* (Soltis et al.)]. The emerging paradigm from these studies is that polyploidy—through alterations in genome structure and gene regulation, some of which occur shortly after polyploid formation—generates genetic and phenotypic novelty that manifests itself at the chromosomal, physiological, and organismal levels, with long-term ecological and evolutionary consequences.

We thank our many colleagues, students, and postdocs for lively and challenging discussions on polyploidy and its many evolutionary consequences. We further acknowledge the support of the U.S. National Science Foundation (Grants 9624643, 0346437, 0614421, 0919254, and 0922003) and thank the National Evolutionary Synthesis Center for its hospitality during the preparation of this book.

Gainesville, April 2012

Pamela S. Soltis
Douglas E. Soltis

Contents

1	Evolutionary Significance of Whole-Genome Duplication	1
	C. L. McGrath and M. Lynch	
2	Genetic Consequences of Polyploidy in Plants	21
	James A. Birchler	
3	Meiosis in Polyploid Plants	33
	Marie-Luise Zielinski and Ortrun Mittelsten Scheid	
4	Origins of Novel Phenotypic Variation in Polyploids	57
	Patrick Finigan, Milos Tanurdzic and Robert A. Martienssen	
5	Identifying the Phylogenetic Context of Whole-Genome Duplications in Plants	77
	J. Gordon Burleigh	
6	Ancient and Recent Polyploidy in Monocots	93
	Andrew H. Paterson, Xiyin Wang, Jingping Li and Haibao Tang	
7	Genomic Plasticity in Polyploid Wheat	109
	Moshe Feldman, Avraham Levy, Boulos Chalhoub and Khalil Kashkush	
8	Maize (<i>Zea Mays</i>) as a Model for Studying the Impact of Gene and Regulatory Sequence Loss Following Whole-Genome Duplication	137
	James C. Schnable and Michael Freeling	
9	Polyploidy in Legumes	147
	Jeff J. Doyle	

10	Jeans, Genes, and Genomes: Cotton as a Model for Studying Polyploidy	181
	Jonathan F. Wendel, Lex E. Flagel and Keith L. Adams	
11	Evolutionary Implications of Genome and Karyotype Restructuring in <i>Nicotiana tabacum</i> L.	209
	Ales Kovarik, Simon Renny-Byfield, Marie-Angèle Grandbastien and Andrew Leitch	
12	Polyploid Evolution in <i>Spartina</i>: Dealing with Highly Redundant Hybrid Genomes.	225
	M. Ainouche, H. Chelaifa, J. Ferreira, S. Bellot, A. Ainouche and A. Salmon	
13	Allopolyploid Speciation in Action: The Origins and Evolution of <i>Senecio cambrensis</i>	245
	Matthew J. Hegarty, Richard J. Abbott and Simon J. Hiscock	
14	The Early Stages of Polyploidy: Rapid and Repeated Evolution in <i>Tragopogon</i>	271
	Douglas E. Soltis, Richard J. A. Buggs, W. Brad Barbazuk, Srikar Chamala, Michael Chester, Joseph P. Gallagher, Patrick S. Schnable and Pamela S. Soltis	
15	Yeast as a Window into Changes in Genome Complexity Due to Polyploidization.	293
	Corey M. Hudson and Gavin C. Conant	
16	Two Rounds of Whole-Genome Duplication: Evidence and Impact on the Evolution of Vertebrate Innovations.	309
	Cristian Cañestro	
17	Polyploidy in Fish and the Teleost Genome Duplication.	341
	Ingo Braasch and John H. Postlethwait	
18	Polyploidization and Sex Chromosome Evolution in Amphibians	385
	Ben J. Evans, R. Alexander Pyron and John J. Wiens	
19	Erratum to—Polyploid Evolution in <i>Spartina</i>: Dealing with Highly Redundant Hybrid Genomes	E1
	M. Ainouche, H. Chelaifa, J. Ferreira, S. Bellot, A. Ainouche and A. Salmon	
	Index	411

Chapter 1

Evolutionary Significance of Whole-Genome Duplication

C. L. McGrath and M. Lynch

Abstract Whole-genome duplication (WGD) appears to be a widespread phenomenon, occurring in diverse taxa including many of the model organisms used in molecular, cellular, and developmental biology. It is therefore essential to understand the potential evolutionary consequences for individual duplicated genes, as well as for the lineage as a whole. For example, duplicate genes may undergo pseudogenization or may be maintained due to neofunctionalization, subfunctionalization, or selection for increased dosage or dosage balance. Duplicates created via WGD are maintained at higher rates than single-gene duplicates, perhaps due to dosage-balance constraints. Duplicate-gene maintenance may lead to heterodimerization of an existing homodimer or to the divergence of an entire duplicated network or pathway. Allopolyploids and autopolyploids are likely to undergo different evolutionary pressures due to increased divergence between allopolyploid paralogs and an increased prevalence of multivalent formation at meiosis in autopolyploids. Perhaps most importantly, duplicate-gene loss following a WGD may significantly increase the rate of reproductive isolation between geographically isolated subpopulations and may therefore temporarily increase the speciation rate within polyploid lineages.

1.1 Introduction

One of the major findings of the new field of evolutionary genomics is that duplication events involving individual genes or multigene segments arise at rates comparable to the rate of mutation at single-nucleotide sites (Lynch and Conery

C. L. McGrath · M. Lynch (✉)

Department of Biology, Indiana University, Bloomington, IN, USA

e-mail: milync@indiana.edu

2000, 2003a, b), or possibly at even higher rates (Lipinski et al. 2011). Such observations lend credibility to Ohno's (1970) early speculation that gene duplication is a major resource for the origin of evolutionary novelties. Moreover, it is now clear that whole-genome duplication (WGD) events have occurred in a wide diversity of phylogenetic lineages, including most of the model systems relied upon in molecular, cellular, and developmental biology. For example, budding yeast is a descendant of an ancient genome duplication (Wolfe and Shields 1997; see Chap. 15, this volume), as is the frog *Xenopus laevis* (Morin et al. 2006; see Chap. 18, this volume) and the zebrafish (Postlethwait et al. 2000; see Chap. 17, this volume). Many ray-finned fish lineages have experienced additional rounds of WGD (Meyer and Van de Peer 2005; see Chap. 17, this volume), and Ohno's (1970) suggestion that two WGD events preceded the radiation of the vertebrate lineage has become increasingly credible (Panopoulou and Poustka 2005; Hughes and Liberles 2008; Putnam et al. 2008; see Chap. 16, this volume). Finally, three WGD events are recorded within the genome of *Arabidopsis thaliana* (Simillion et al. 2002), and nearly all other land-plant genomes appear to harbor a legacy of at least one polyploidization event (Doyle et al. 2008), with a proposed WGD in the ancestor of all seed plants and another in the ancestor of all angiosperms (Jiao et al. 2011). Thus, it is clear that understanding the mechanisms of origin and preservation of duplicate genes promises to reveal not only the ways in which genes acquire new functions and organisms respond to natural selection, but also the roots of organismal diversity across the tree of life.

Because genome duplication adds thousands of duplicate genes to the genome, understanding the evolutionary forces that act on individual duplicate genes is critical to our understanding of polyploidization. Processes such as neofunctionalization and subfunctionalization have the potential to influence all gene duplicates, whether created through polyploidization or smaller scale duplication events. It has become increasingly clear, however, that duplicates that arise via polyploidization are subject to unique evolutionary forces, such as increased retention due to dosage-balance constraints. Further, there may be processes that are exclusive to gene duplicates that arise via specific types of polyploidization, such as changes in duplicate-gene expression due to the genomic merger that occurs with allopolyploidization. The relative contributions of these evolutionary forces that give rise to the maintenance and evolution of duplicate genes that arise via WGD, or to the evolution of the genome or species as a whole, are currently unknown. However, discriminating between these forces and their effects is likely to be the subject of much research over the next several years.

1.2 Fates of Duplicate Genes

The fate of the vast majority of duplicate genes arising by segmental duplication is nonfunctionalization of one member of the pair (Lynch and Conery 2000, 2003a, b), and this is expected to occur within a few million years in the absence of any

intrinsic advantage of a duplicate copy (Watterson 1983; Lynch et al. 2001). Despite this, most genomes that have been studied contain a large number of duplicate genes, some of which are clearly quite ancient (Lynch and Conery 2000). Based on this observation, several mechanisms have been proposed for the permanent preservation of duplicate genes (Hughes 1994; Force et al. 1999; Lynch et al. 2001; Taylor and Raes 2004; Lynch 2007; Innan and Kondrashov 2010): (1) neofunctionalization, whereby one copy acquires a novel, beneficial function at the expense of an essential ancestral function; (2) subfunctionalization, whereby complementary mutations lead to a partitioning of independently mutable subfunctions in the ancestral gene; (3) selection for increased gene product; and (4) the masking of nonfunctional alleles.

When a duplicate is maintained by selection for increased gene product, it experiences purifying selection (and may also undergo repeated gene conversion) in order to maintain its ancestral function; this process is likely responsible for the multiple copies of ribosomal RNA genes present in many genomes (e.g., Pinhal et al. 2011). Neofunctionalization, on the other hand, is thought to involve positive selection for the mutation(s) responsible for the new function, generally arising at the expense of an essential original function, thereby preserving both copies. There are many examples of neofunctionalization giving rise to novel gene functions in a variety of organisms, including *Arabidopsis* (Erdmann et al. 2010), fish (Ngai et al. 1993), vertebrates (Layeghifard et al. 2009), and yeast (Byrne and Wolfe 2007; Tirosh and Barkai 2007). Because one duplicate is undergoing positive selection for a new function while the other is under purifying selection to maintain the ancestral function, asymmetric evolutionary rates between duplicates are often thought to be a hallmark of neofunctionalization (Johnson and Thomas 2007; Han et al. 2009), though purely stochastic mechanisms can also give rise to apparent rate asymmetry (Lynch and Katju 2004).

Subfunctionalization may involve positive selection acting on both duplicates if the partitioning of the ancestral functions leads to relaxation of pleiotropic constraints, enabling each ancestral function to be fine-tuned and improved through mutation independently in each copy (Piatigorsky and Wistow 1991; Hughes 1994; Des Marais and Rausher 2008). Alternatively, subfunctionalization may be a completely neutral process if each duplicate copy simply acquires a degenerative mutation that renders it unable to perform one of the ancestral functions (Force et al. 1999). At this point, both copies are needed in order to provide the organism with all of the functionality of the original, single-copy gene, and so both will be maintained in the genome by selection. Although identifying definitive cases of subfunctionalization requires determining that the ancestral gene carried multiple functions that have been partitioned in the daughter duplicates, there are nonetheless several compelling examples (e.g., Force et al. 1999; Altschmied et al. 2002; Yu et al. 2003; Adams and Liu 2007; MacNeil et al. 2008; Semon and Wolfe 2008; Buggs et al. 2010; Deng et al. 2010; Hickman and Rusche 2010; Colon et al. 2011; Froyd and Rusche 2011).

In addition to these cases of qualitative subfunctionalization, where duplicates eventually come to be expressed in different tissues or at different times or carry out

different functions relative to each other, quantitative subfunctionalization, in which reduction-of-expression (Force et al. 1999) or activity-reducing mutations (Stoltzfus 1999; Scannell and Wolfe 2008) affect both duplicates, is also possible. In quantitative subfunctionalization, both duplicates acquire partial loss-of-function mutations that affect the same function, again rendering both copies essential for the proper dosage or activity of the gene products. In this case, both copies are preserved while retaining the ancestral gene function. Although few studies demonstrating quantitative subfunctionalization exist, Qian et al. (2010) estimated that this process has been responsible for the maintenance of a large proportion of duplicates in yeast and mammals, whereas Woolfe and Elgar (2007) postulated that sequence evolution in *cis*-regulatory elements may have caused quantitative subfunctionalization among *Fugu* duplicates.

A related consequence of gene duplication is that it can allow for the differentiation of multimeric subunits, such as the evolution of heterodimers from homodimers. Consider a gene whose protein product forms a homodimer. After duplication of this gene, protein subunits produced by the two duplicates (denoted *A* and *B*) may randomly associate to make mixtures of dimers in the ratio 1 *AA*: 2 *AB*: 1 *BB*. If the duplicate genes are identical initially, the *AA*, *AB*, and *BB* dimers will be identical as well. However, subsequent differentiation of the duplicate genes causes the three types of dimer to become distinct. This differentiation could be neutral, or it could be selective. If, for example, there were pleiotropic constraints on the form or function of the pre-duplication homodimer, duplication could allow for escape from these constraints in the *AB* heterodimer, as each subunit (*A* and *B*) can now evolve independently. This can be viewed as a special type of subfunctionalization of duplicates. Winter et al. (2002) showed that a class-B floral protein heterodimer had evolved from an ancestral homodimer via this mechanism during the gymnosperm/angiosperm transition. In gymnosperms, GGM2-like genes form homodimers, while the duplicated homologs in eudicots, DEF-like genes and GLO-like genes, form heterodimers. Monocots also have duplicated DEF-like genes and GLO-like genes, but, interestingly, it appears the GLO-like proteins of monocots can both homodimerize and heterodimerize with DEF-like proteins, perhaps representing the transition between the homo- and heterodimerized states (Winter et al. 2002; Kanno et al. 2003; Soltis et al. 2006).

A similar process appears to have occurred several times in the evolution of the DUF606 family of transmembrane proteins in bacteria (Lolkema et al. 2008). In bacteria with a single DUF606 gene, the DUF606 proteins are able to insert into the membrane in both orientations, and functional homodimers are formed by two subunits in opposite (antiparallel) orientations. Other species of bacteria, however, have duplicated DUF606 genes located tandemly in an operon. In all of these latter cases, the two protein subunits each have a fixed but opposite orientation in the membrane, and they heterodimerize to form the necessary antiparallel two-domain complex. A phylogenetic analysis of the DUF606 gene family reveals that this process of duplication followed by heterodimerization likely occurred five different times in the history of this gene family lineage (Lolkema et al. 2008). Other proposed examples of this mechanism include SMC proteins (Surcel et al. 2008),

adenyl cyclases (Sinha et al. 2005), and mitochondrial peptidases (Brown et al. 2007), all gene families that contain duplicates that form heterodimers in eukaryotes (or eukaryotic mitochondria) with single-copy homologs that form homodimers in prokaryotes.

1.3 Fates of Duplicate Genes Arising via WGD

In addition to the general preservational processes just mentioned, paralogs resulting from WGD events are subject to unique mechanisms of duplicate-gene maintenance and evolution (Force et al. 1999; Lynch and Conery 2000; Yang et al. 2003; Davis and Petrov 2005; Veitia et al. 2008). Well-studied polyploid species commonly exhibit 25–75 % retention of paralogous gene pairs from the most recent WGD event (reviewed in Lynch 2007; Otto 2007), budding yeast being an exception with only ~8 % duplicate-gene preservation (Wolfe and Shields 1997). These are surprisingly high preservational levels, when, as discussed above, the fate of the vast majority of duplicate genes arising by segmental duplication is nonfunctionalization of one duplicate (Lynch and Conery 2000, 2003a, b). Although it is possible that many polyploid species have not yet reached equilibrium and are still in an ongoing phase of duplicate-gene loss, it has become increasingly clear that there are likely to be additional forces acting to preserve duplicate genes arising via WGD.

A simple explanation for the large number of preserved duplicates within polyploids is that, unlike single-gene duplicates, WGD duplicates exhibit complete conservation of surrounding regulatory sequences, chromosomal environments, etc. Although this likely contributes somewhat to the pattern of higher duplicate retention in polyploids, it does not explain the observation that different types of genes seem to be preserved following WGD compared to smaller scale duplications. This fact can be better explained by selection for dosage balance among proteins. Due to stoichiometric relationships with other interacting genes (e.g., multi-subunit complexes and numerous pathways involved in metabolism and transcriptional regulation), the functions of a subset of protein-coding loci can be highly influenced by dosage imbalances (Veitia 2002; Papp et al. 2003; Birchler et al. 2005; Veitia et al. 2008). In such cases, duplication of just a single member of a gene interaction may be detrimental and actively selected against. In contrast, following a WGD event, most stoichiometric relationships are initially intact, and therefore subsequent losses of interacting paralogs will be inhibited by selection for proper dosage relationships. Thus, for dosage-dependent genes, the dosage-balance hypothesis predicts an under-representation among duplicates created by single-gene duplications, but an over-representation among those created by WGD (Yang et al. 2003; Davis and Petrov 2005; Veitia et al. 2008). For example, Davis and Petrov (2005) showed that the pool of preserved duplicates from the WGD event in *S. cerevisiae* is enriched for ribosomal genes (which form a large complex) and regulatory genes encoding transferases, kinases, and transcription factors, while

those involved in ion transport are under-represented. Likewise, the *Paramecium tetraurelia* genome exhibits elevated retention of duplicate genes involved in known complexes (Aury et al. 2006) and in metabolic pathways (Gout et al. 2009). As in yeast, ribosomal genes, transferases, and kinases are over-represented among surviving paralogs, while ion-transport genes are underrepresented. In *Paramecium*, there also appears to be an additional effect whereby highly expressed genes are over-retained in duplicate following the most recent polyploidization event (Gout et al. 2010). That certain types of genes are maintained preferentially following a WGD has achieved fairly convincing empirical support from other studies as well (Papp et al. 2003; Yang et al. 2003; Barker et al. 2008; Liang et al. 2008; Qian and Zhang 2008; Edger and Pires 2009), including studies in *Arabidopsis* (Blanc and Wolfe 2004; Maere et al. 2005; Thomas et al. 2006), vertebrates (Makino and McLysaght 2010), and across divergent species (Paterson et al. 2006). Selection to maintain dosage balance following WGD has also been hypothesized to be the driving force behind the original selective advantage of the WGD in the *Saccharomyces cerevisiae* lineage (Conant and Wolfe 2007). In this scenario, the maintenance of glycolytic genes and the loss of non-glycolytic genes following WGD might have increased the relative dosage of glycolytic genes, thereby increasing flux through the glycolysis pathway and providing polyploid yeast with a growth advantage over non-polyploids due to increased glucose fermentation ability.

Duplicate genes that arise via WGD are further unique in that entire (or partial) duplicated pathways or networks of interacting proteins can diverge in concert. For example, Evlampiev and Isambert (2007) modeled the evolution of protein–protein interaction networks following WGD and concluded that such networks grow under exponential, rather than time-linear, dynamics following WGD. Interestingly, they also found that these exponential dynamics relied on asymmetric divergence between duplicates.

Another intriguing possibility is that following WGD, a whole ancestral network may become neofunctionalized or subfunctionalized following polyploidization, with one set of paralogs carrying out one task or reaction and a parallel set of paralogs carrying out a related, but largely independent, task. Obviously, such innovations require the establishment of multiple mutations and the avoidance of pathway crosstalk. Although the essential population genetic theory remains to be worked out, several examples of such paralog coevolution appear to have followed the WGD in yeast: parallel paralogous networks have been identified where the expression of each gene is highly correlated with the other genes within its network but poorly correlated with its paralog (Blanc and Wolfe 2004; Conant and Wolfe 2006). In this way, polyploidy provides a unique mechanism for the evolution of gene networks with new (or subdivided) functions.

A final consideration in duplicate-gene evolution is whether the forces that act to preserve duplicates change over evolutionary time. For example, it seems possible that following WGD, a large proportion of genes could be initially maintained due to dosage-balance constraints. Subsequently, however, over longer periods of evolutionary time, some duplicates might accumulate mutations that could lead to neofunctionalization or subfunctionalization. Because these genes

are dosage sensitive (hence their initial preservation due to selection for dosage balance), it is likely that such neo- or subfunctionalizing mutations would need to be preceded or rapidly followed by mutations affecting the dosage of one or both copies. For a more detailed example, imagine proteins *A* and *B* that must interact in a 1:1 ratio for proper functioning. Both genes become duplicated during a WGD, giving rise to duplicates *A1* and *A2* and *B1* and *B2*. Initially, all four genes are preserved by selection for dosage balance, as loss of any one gene interrupts the 1:1 interaction ratio. Over evolutionary time, however, slightly deleterious mutations in the *A1* promoter that decrease its expression level become fixed due to drift. To compensate, mutations in the *A2* promoter that increase its expression level are fixed, which helps to restore the 1:1 *A/B* ratio. At this point, *A1* is contributing fewer products to the overall *A* protein pool. A subsequent mutation that changes the function of *A1*, allowing it to take on a new role completely, is now more easily accommodated, as *A2* is better able to compensate and take on the full load of the ancestral *A* activity. Note that, instead of *A1* and *A2* dosage evolving in concert, as above, *A1* and *B1* dosage could also evolve in concert to maintain the proper 1:1 *A/B* ratio, allowing both *A1* and *B1* to take on new functions.

While still just a verbal theory, this scenario has two advantages in terms of allowing for neofunctionalization (or subfunctionalization) of WGD duplicates. First, there is a longer time frame in which neo- or subfunctionalizing mutations can arise, as duplicates are maintained for longer time-scales without becoming nonfunctionalized. This is important because neofunctionalization requires the accumulation of beneficial mutation(s), which are thought to be rare. Second, this process would allow for neo- or subfunctionalization of dosage-sensitive duplicates, both of which might otherwise be constrained to maintain their ancestral function indefinitely following WGD.

1.4 Autopolyploidy Versus Allopolyploidy

Whether polyploidization occurs by autopolyploidy or allopolyploidy can have a significant impact on the expression and evolution of duplicate genes. Autopolyploids arise when there is an increase in ploidy within a single species (often within a single individual), while allopolyploids are created by hybridization between two different species, each of which contributes a full complement of chromosomes to the hybrid, thus doubling the genome (reviewed in Coyne and Orr 2004). Many plant and frog polyploids are the result of allopolyploidization (Adams 2007; Evans 2008), while the yeast WGD appears to have been an autopolyploidization event (Scannell et al. 2007), although in practice it is difficult to ascertain the ancestral state once paralog divergence has become high.

It has long been assumed that autopolyploids would initially form multivalents at meiosis, with all four homologous chromosomes pairing randomly, while allopolyploids would be more likely to form bivalents, with homologous

chromosomes from each diploid ancestor pairing independently. This would mean that duplicate copies in autopolyploids would not represent true paralogs as the term is usually understood, but would instead represent a doubling of the number of homologs (i.e., four homologs instead of two). The presence of multivalents is significant biologically, as multivalent pairing can lead to intergenomic recombination via segregation, crossing-over, and double reduction. Certain duplicates from one diploid parent could be lost completely via this process, leaving only duplicates from the other diploid parent. This would not represent gene silencing as it is typically understood then, but would rather be a byproduct of multivalent formation and segregation. Evidence from plants indicates that multivalent pairing is indeed more prevalent among autopolyploids, though the difference between the two forms of polyploidy is perhaps less than originally expected: a survey of plant polyploids indicated that the mean percent occurrence of multivalents is 28.8 % in autopolyploids and 8.0 % in allopolyploids (Ramsey and Schemske 2002). Although multivalent formation occurs at a lower rate in allopolyploids, it may be more biologically significant than multivalent formation in autopolyploids, as intergenomic recombination is likely to have a greater effect when genomes are more divergent. Over time, divergence between duplicated chromosomes would lead to increased bivalent formation.

Because allopolyploids are the result of a genomic merger between two species, duplicate genes in allopolyploids are already differentiated to some extent immediately after polyploidization, while duplicates in autopolyploids are likely to be more similar in sequence and may even be identical. Allopolyploids often exhibit immediate changes in gene expression due to the genetic differentiation present between homeologs. This can lead to changes in methylation (Salmon et al. 2005; Gaeta et al. 2007), changes in heterochromatin formation and transposable element suppression (Josefsson et al. 2006), biased expression of homeologs (Adams et al. 2003; Bottley et al. 2006; Tate et al. 2006; Udall et al. 2006; Rapp et al. 2009), and non-additive expression effects between homeologs (Hegarty et al. 2006; Wang et al. 2006; Rapp et al. 2009). These initial expression differences between homeologs can, in turn, impact the long-term evolution of duplicates, as selection pressures may be expected to act differently on genes that are differentially expressed. For example, Anderson and Evans (2009) showed that in octoploid and dodecaploid *Xenopus* species, paralogs of *RAG1 β* were more likely to become pseudogenized than paralogs of *RAG1 α* (the homeolog of *RAG1 β* from an earlier allopolyploidy event), and they inferred that this was due to differences in ancestral expression between *RAG1 α* and *RAG1 β* .

Many of these effects seen in allopolyploids are believed to be due to the hybridization between two divergent genomes, rather than genome doubling *per se*. Flagel et al. (2008) estimated that of the genes with biased expression between homeologs in the allopolyploid *Gossypium hirsutum*, 24 % exhibit a bias due to the genomic merger (i.e., the bias existed immediately when the allopolyploidization occurred, at time zero), while the bias in the remaining 76 % is due to long-term evolutionary forces such as neofunctionalization and subfunctionalization. The relationship between the magnitude of these alterations in gene expression and the

genetic divergence between the two parental genomes is not well understood, however, as demonstrated by *Brassica* allopolyploids (Pires and Gaeta 2011). While the parental species that gave rise to the allopolyploid *Brassica napus* are more similar to each other than those that gave rise to *B. juncea*, resynthesized *B. napus* polyploids exhibit more genomic rearrangements, changes in gene expression, and epigenetic alterations than do resynthesized *B. juncea* polyploids.

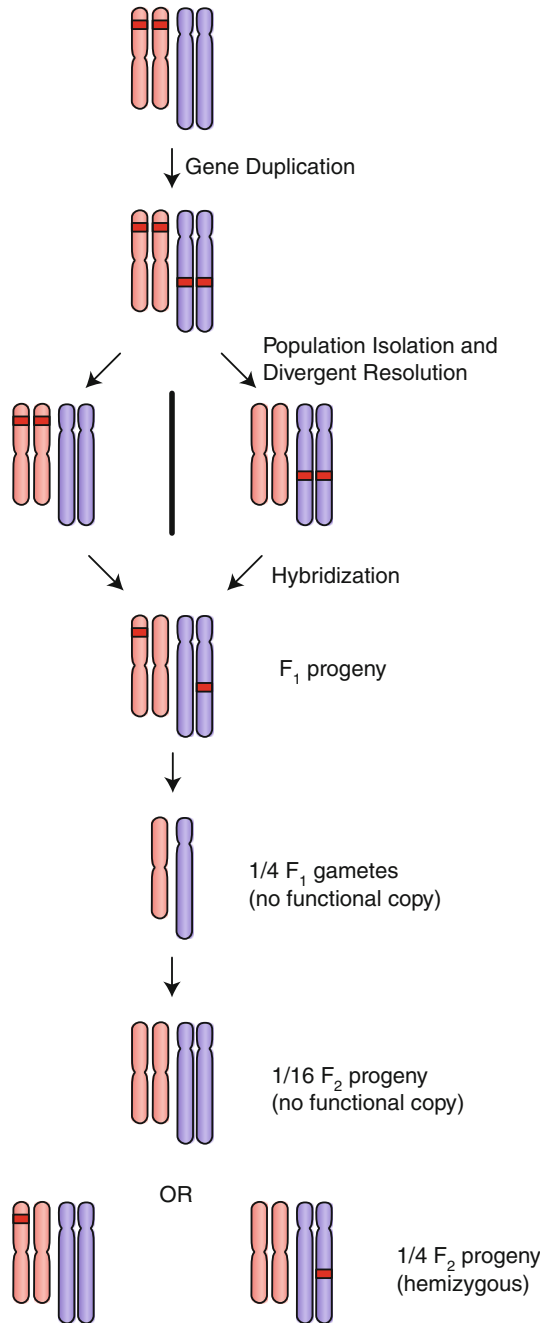
1.5 Polyploidization and Speciation

Perhaps the most pivotal role that polyploidization plays in evolution is in the creation of new species. The polyploidization event itself can lead to instantaneous reproductive isolation and speciation, as the cross between a new tetraploid ($4n$) and its diploid progenitor ($2n$) yields triploid ($3n$) offspring, which are often sterile due to problems with chromosome pairing/segregation during meiosis and the production of aneuploid gametes (reviewed in Coyne and Orr 2004). It is for this reason that models predict that species capable of self-fertilization are more likely to give rise to a successful polyploid lineage (Rodriguez 1996; Baack 2005; Rausch and Morgan 2005).

Perhaps more importantly, however, once a polyploid lineage is established, subsequent silencing of duplicate genes can lead to further reproductive isolation among subpopulations of the polyploids themselves and, therefore, give rise to additional daughter species (Oka 1988; Werth and Windham 1991; Lynch and Conery 2000; Lynch and Force 2000). In this model, we assume a pair of fully functional and redundant duplicate genes, *A* and *B*, in an ancestral population, such that each member of the initial population has the genotype *AABB* (Fig. 1.1). If two subpopulations become geographically isolated and one duplicate becomes nonfunctionalized in each subpopulation, there is a 50 % probability that a different duplicate copy will be lost in each of the two groups. This reciprocal gene loss (or divergent resolution) would result in the genotypes *aaBB* and *AAbb* for the two subpopulations, where *a* and *b* denote null alleles. Hybridization between the two groups would then lead to offspring with the genotype *AaBb*. Gametes produced by these F_1 individuals would have a 1/4 probability of carrying an *ab* genotype and would therefore be inviable if a functional copy of the *A/B* gene were essential for gamete survival or function. Even if this were not the case, 1/16 of the F_2 individuals would have the *aabb* genotype and, if a functional copy were essential for zygote viability or sterility, would be inviable or sterile, whereas another 1/4 would have three null alleles and might experience reduced viability or sterility. Up to 50–65 % of the genes encoding transcription factors, membrane receptors, and members of macromolecular protein complexes are estimated to be haploinsufficient (Jimenez-Sanchez et al. 2001; Veitia 2002), suggesting that only one functional allele of such genes is indeed likely to be deleterious.

An appealing aspect of the divergent resolution model is that it is a natural consequence of degenerative mutations, requiring no adaptive evolution at the

Fig. 1.1 Divergent resolution of duplicate genes can lead to hybrid sterility/inviability and reproductive isolation. *Red bar* represents a gene that becomes duplicated. See text for details



molecular level for speciation to occur. Moreover, in addition to genes that are reciprocally silenced, duplicate pairs that undergo neofunctionalization or subfunctionalization may also contribute to hybrid sterility/inviability in a similar fashion—for example if a different duplicate becomes neofunctionalized and loses its ancestral function in each subpopulation, or if the two duplicates become subfunctionalized in complementary ways in the two subpopulations (Lynch and Force 2000).

The process of reciprocal gene loss has been shown to be responsible for male sterility between hybrids of *Drosophila melanogaster* and *D. simulans* (Masly et al. 2006). A gene essential for male fertility, *JYAlpha*, is located on the fourth chromosome in *D. melanogaster* and on the third chromosome in *D. simulans*. This translocation presumably occurred via duplication of the *JYAlpha* gene and subsequent silencing of one copy. The difference in chromosomal location of the gene in the two species causes a proportion of hybrids to completely lack *JYAlpha*, leading to their sterility.

Two similar cases were recently identified in rice. The first involves reproductive isolation between two subspecies of *Oryza sativa*. The ancestral *O. sativa* genome appears to have had a pair of duplicates termed *DOPPELGANGER1* (*DPL1*) and *DOPPELGANGER2* (*DPL2*) (Mizuta et al. 2010). The subspecies *japonica* and *indica* have experienced independent losses of one copy each: *DPL1* has become a pseudogene in *indica*, while *DPL2* has been nonfunctionalized in *japonica*. Hybrid pollen lacking a functional copy of either *DPL1* or *DPL2* is nonfunctional and does not germinate, contributing to the partial reproductive isolation present between the subspecies. This validates an earlier hypothesis by Oka (1988) that F₁ sterility between *japonica* and *indica* was caused by “duplicate gametophytic sterility genes”, *japonica* being homozygous for one nonfunctional copy and *indica* being homozygous for another nonfunctional copy. In the second rice example, reciprocal loss of one of the duplicated nuclear genes encoding mitochondrial ribosomal protein L27 in *O. sativa* and *O. glumaepatula* again causes a proportion of the pollen produced by F₁ hybrids to be sterile (Yamagata et al. 2010).

The final example of reproductive isolation through reciprocal gene loss comes from *A. thaliana*, where the histidinol-phosphate amino-transferase gene appears in different chromosomal locations (as in the *Drosophila* example, presumably via duplication and subsequent silencing of one copy) in the Columbia and Cape Verde Island accessions (Bikard et al. 2009). F₂ offspring homozygous for both null alleles completely lack the gene’s product, HPA, which results in arrested embryo development and seed abortion. In addition, in at least one intermediate heterozygote, a quantitative phenotype termed “weak root” was observed, suggesting that the presence of three null alleles is somewhat deleterious in this cross. As these four examples constitute ~1/3 of the dozen or so successful searches for the genes underlying the speciation process (most in *Drosophila* species; Presgraves 2010), there now seems little question that the passive non-functionalization of duplicate genes is a major mechanism of speciation.

These examples demonstrate that the divergent resolution of even one duplicated gene can lead to detectable reproductive isolation. However, genetic incompatibility between two populations can be magnified substantially when reciprocal gene loss occurs at hundreds or thousands of duplicated loci simultaneously, as is the case in polyploid lineages. The probability that an F_2 offspring obtained by outcrossing will be double null for at least one of n pairs of divergently resolved loci is $1 - (15/16)^n$, which takes on values of 0.063, 0.276, 0.476, and 0.998 for $n = 1, 5, 10,$ and 100 , respectively. Moreover, in species that undergo autogamy or selfing, such as *Paramecium*, this probability can be as high as $1 - (3/4)^n$, giving probabilities of 0.250, 0.763, 0.944, and ≈ 1 for $n = 1, 5, 10,$ and 100 . Speciation events will continue to occur as long as duplicates are still being resolved between subpopulations, leading to nested rounds of speciation, and, because a large number of duplicates are thought to be silenced quickly following WGD (Scannell et al. 2006), a cluster of speciation events might occur within a brief period of time. The net result is the expected generation of a species radiation following a WGD event.

It has been suggested that this nested speciation process might be responsible for the radiations of the polyploid yeast species (Scannell et al. 2006), teleost fishes (Semon and Wolfe 2007; see Chap. 15, this volume), angiosperms (Soltis et al. 2009; though see Mayrose et al. 2011), and the *Paramecium aurelia* species complex (Aury et al. 2006). This mechanism may also be responsible for reproductive isolation between mutagenized lines of an experimentally derived allotetraploid created by hybridizing two species of *Saccharomyces* (Maclean and Greig 2010).

1.6 Unsolved Problems

The maintenance of duplicate genes via selection for increased gene product, neofunctionalization, and subfunctionalization has been hypothesized for nearly 40 years (Ohno 1970). Recent genetic and genomic data have now identified compelling examples of these processes and have further contributed to our understanding of the prevalence of whole-genome duplications and the dosage-balance theory of duplicate maintenance. However, a number of unresolved questions related to WGDs and duplicate maintenance merit further scrutiny.

The first avenue for future study involves a more comprehensive understanding of the relative importance of the forces behind duplicate-gene maintenance, including maintenance for increased dosage, dosage-balance constraints, neofunctionalization, and subfunctionalization. All of these mechanisms have been demonstrated to be responsible for duplicate maintenance in certain cases, but it remains unclear which, if any, is responsible for maintaining the majority of duplicate genes or how such contributions vary among phylogenetic lineages. Most likely, there will be no single driving force for duplicate maintenance but the relative strength of these forces will differ among taxonomic groups or among functional classes of genes. For example, subfunctionalization of duplicates may be more

likely within species that have evolved a modular (and therefore independently mutable) regulatory structure. Such modular systems are predicted to arise more easily within species with smaller population sizes (Force et al. 2005), demonstrating how species-level features of an organism may influence the evolutionary forces acting upon duplicate genes. More comprehensive studies of large numbers of duplicates from a variety of organisms are required to address what other features might influence the relative strengths of mechanisms of duplicate maintenance. Such studies must not only detail what genes remain duplicated vs. single-copy, but must also detail whether existing duplicates have the same function as each other (to assess rates of neofunctionalization) or share functions with the pre-duplicated ancestor (to assess rates of subfunctionalization) (Fig. 1.2).

There are few data on the rate of duplicate-gene loss over time following a WGD, though data from polyploid yeast species suggest that the rate changes over time (Scannell et al. 2006). Data from additional taxa would aid in determining whether this is a general pattern among all WGDs (Fig. 1.3). A related unresolved question is whether the evolutionary forces controlling duplicate maintenance change over time following a WGD, e.g., whether a dosage-sensitive gene may initially be preserved due to selection for dosage balance, but then evolve a new function concurrent with its release from such dosage constraints. An analysis of this question could be made by comparing the fates of duplicate genes in multiple lineages descended from a single WGD event. Such an analysis might identify duplicates that had been maintained due to dosage constraints in the majority of daughter lineages but that had become neofunctionalized in one lineage, perhaps suggesting a secondary mechanism of retention.

The unsolved question that promises to be the hardest to answer is why certain lineages or taxonomic groups appear to contain more WGD events than others. This is not the same as asking why certain groups contain more polyploid species, as this may be a simple reflection of the fact that WGD may promote subsequent reproductive isolation and speciation. The pattern remains, however, that some phylogenetic groups seem to contain more independent WGD events than others in their evolutionary pasts. For example, a recent analysis estimated that among ferns, 31 % of speciation events involve polyploidization, while the value for angiosperms is only 15 % (Wood et al. 2009). Similarly, in the history of the *Xenopus* lineage, there are many more instances of WGD events than compared to, say, mammals. Several factors could contribute to such patterns, such as the ability to hybridize and form allopolyploids, or the ability to self-fertilize (at least transiently) or undergo asexual reproduction, which helps a polyploid lineage become abundant in a surrounding world of diploids. It is not even understood whether mechanistic reasons (at meiosis, say) or differences in developmental programs would facilitate or hinder creation of a viable polyploid in certain lineages, or whether discrepancies in ecological persistence of polyploid species alone are able to explain the patterns that we see. Perhaps the best way to approach such a question is to study closely related lineages where one exhibits several WGD events and the other does not, though teasing apart the mechanistic and ecological differences is certain to remain a challenge for decades to come.

Fig. 1.2 Distinguishing between the evolutionary forces that maintain duplicate genes. Panel A shows the expression level of a gene (*purple*) across different conditions or tissues before duplication. The *bottom* six panels (B–G) show patterns that might be seen for the two copies (*red* and *blue*) once the gene has been duplicated and what evolutionary processes these patterns would indicate. Note that panel D might indicate maintenance for increased dosage in the case of a single-gene duplicate or maintenance for dosage balance in the case of a duplicate arising via WGD

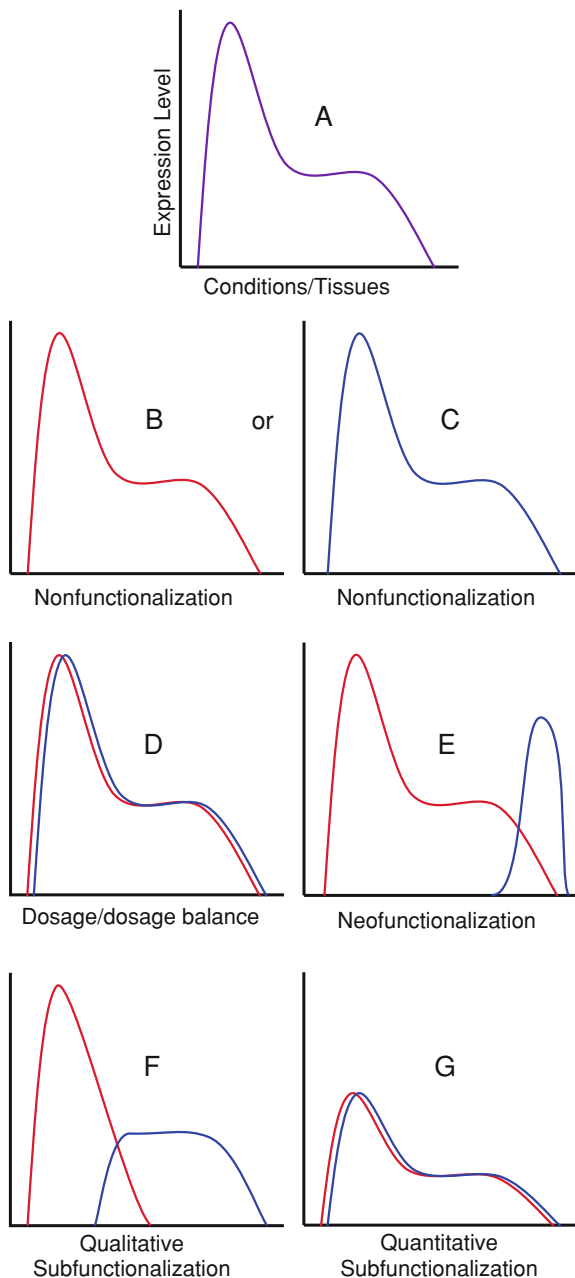
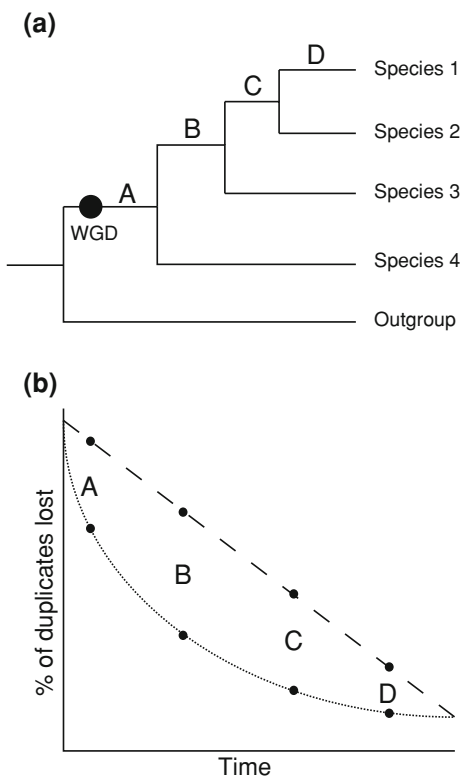


Fig. 1.3 Determining the rate of duplicate-gene loss over time. **a** An example of a tree for four species that share a whole-genome duplication (WGD). Duplicate-gene presence/absence information for each of the four species could be used to infer the number of duplicate genes lost on each of the branches labeled *A*, *B*, *C*, and *D*. **b** The data on gene retention/loss gathered in (a) could be used to plot the percentage of duplicate genes lost per unit time (or divergence). Depending on where points *A*, *B*, *C*, or *D* land on the graph, the data may indicate that the rate of gene loss remains constant (*top dashed line*) or changes (*bottom dotted line*) over time



1.7 Conclusions

Whole-genome duplications are widespread across the tree of life and appear in the evolutionary history of a large number of model organisms. Processes such as neo- and subfunctionalization affect retention of individual gene duplicates, and dosage-balance constraints promote the retention of large sets of genes following polyploidization. Allopolyploidization, through hybridization and subsequent changes or biases in homeolog expression, has the ability to instantaneously create a population of individuals that are ecologically and epigenetically unique from either parent lineage, providing a new lineage upon which natural selection can act. Both allo- and autopolyploidization provide a unique opportunity for the differentiation of new gene networks and pathways through concerted evolution of duplicated, interacting proteins. Most importantly, WGD can lead to reproductive isolation through divergent resolution of duplicated genes, thus creating new species and species groups. Further understanding of the relative importance and the temporal properties of the forces acting on polyploid species and the duplicate genes within their genomes promises to enhance our knowledge of the origins of species as well as genetic, protein network, and organismal complexity.

References

- Adams KL (2007) Evolution of duplicate gene expression in polyploid and hybrid plants. *J Hered* 98(2):136–141
- Adams KL, Cronn R et al (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Nat Acad Sci U S A* 100:4649–4654
- Adams KL, Liu Z (2007) Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr Biol* 17(19):1669–1674
- Altschmied J, Delfgaauw J et al (2002) Subfunctionalization of duplicate *mitf* genes associated with differential degeneration of alternative exons in fish. *Genetics* 161(1):259–267
- Anderson DW, Evans BJ (2009) Regulatory evolution of a duplicated heterodimer across species and tissues of allopolyploid clawed frogs (*Xenopus*). *J Mol Evol* 68:236–247
- Aury J-M, Jaillon O et al (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444:171–178
- Baack EJ (2005) To succeed globally, disperse locally: effects of local pollen and seed dispersal on tetraploid establishment. *Heredity* 94:538–546
- Barker MS, Kane NC et al (2008) Multiple paleopolyploidizations during the evolution of the compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol Biol Evol* 25(11):2445–2455
- Bikard D, Patel D et al (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* 323:623–626
- Birchler JA, Riddle NC et al (2005) Dosage balance in gene regulation: biological implications. *Trends Genet* 21(4):219–226
- Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16:1679–1691
- Bottley A, Xia GM et al (2006) Homoeologous gene silencing in hexaploid wheat. *Plant J* 47:897–906
- Brown MT, Goldstone HMH et al (2007) A functionally divergent hydrogenosomal peptidase with protomitochondrial ancestry. *Mol Microbiol* 64(5):1154–1163
- Buggs RJA, Elliott NM et al (2010) Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytol* 186:175–183
- Byrne KP, Wolfe KH (2007) Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. *Genetics* 175(3):1341–1350
- Colon M, Hernandez F et al (2011) *Saccharomyces cerevisiae* Bat1 and Bat2 aminotransferases have functionally diverged from the ancestral-like *Kluyveromyces lactis* orthologous enzyme. *PLoS ONE* 6(1):e16099
- Conant GC, Wolfe KH (2006) Functional partitioning of yeast co-expression networks after genome duplication. *PLoS Biol* 4(4):0545–0554
- Conant GC, Wolfe KH (2007) Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol Syst Biol* 3:129
- Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Inc, Sunderland, MA
- Davis JC, Petrov DA (2005) Do disparate mechanisms of duplication add similar genes to the genome? *Trends Genet* 21(10):548–551
- Deng C, Cheng CHC et al (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc Nat Acad Sci U S A* 107(50):21593–21598
- Des Marais DL, Rausher MD (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454(7205):762–765
- Doyle JJ, Flagel LE et al (2008) Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet* 42:443–461
- Edger PP, Pires JC (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res* 17:699–717

- Erdmann R, Gramzow L et al (2010) GORDITA (AGL63) is a young paralog of the Arabidopsis thaliana B (sister) MADS box gene ABS (TT16) that has undergone neofunctionalization. *Plant J* 63(6):914–924
- Evans BJ (2008) Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Front Biosci* 13:4687–4706
- Evlampiev K, Isambert H (2007) Modeling protein network evolution under genome duplication and domain shuffling. *BMC Syst Biol* 1:49
- Flagel LE, Udall JA et al (2008) Duplicate gene expression in allopolyploid gossypium reveals two temporally distinct phases of expression evolution. *BMC Biol* 6:16
- Force A, Cresko WA et al (2005) The origin of subfunctions and modular gene regulation. *Genetics* 170(1):433–446
- Force A, Lynch M et al (1999) The preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545
- Froyd CA, Rusche LN (2011) The duplicated deacetylases sir2 and hst1 subfunctionalized by acquiring complementary inactivating mutations. *Mol Cell Biol* 31(16):3351–3365
- Gaeta RT, Pires JC et al (2007) Genomic changes in resynthesized brassica napus and their effect on gene expression and phenotype. *Plant Cell* 19(11):3403–3417
- Gout J-F, Duret L, et al. (2009) Differential retention of metabolic genes following whole-genome duplication. *Mol Biol Evol* 26(5):1067–1072
- Gout J-F, Kahn D, et al. (2010) The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genetics* 6(5):e1000944
- Han MV, Demuth JP et al (2009) Adaptive evolution of young gene duplicates in mammals. *Genome Res* 19(5):859–867
- Hegarty MJ, Barker GL et al (2006) Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. *Curr Biol* 16:1652–1659
- Hickman MA, Rusche LN (2010) Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. *Proc Nat Acad Sci U S A* 107(45):19384–19389
- Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci* 256:119–124
- Hughes T, Liberles DA (2008) Whole-genome duplications in the ancestral vertebrate are detectable in the distribution of gene family sizes of tetrapod species. *J Mol Evol* 67(4):343–357
- Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* 11:97–108
- Jiao YN, Wickett NJ et al (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–100
- Jimenez-Sanchez G, Childs B et al (2001) Human disease genes. *Nature* 409:853–855
- Johnson DA, Thomas MA (2007) The monosaccharide transporter gene family in Arabidopsis and rice: a history of duplications, adaptive evolution, and functional divergence. *Mol Biol Evol* 24(11):2412–2423
- Josefsson C, Dilkes B et al (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Curr Biol* 16:1322–1328
- Kanno A, Saeki H et al (2003) Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). *Plant Mol Biol* 52:831–841
- Layeghifard M, Pirhaji L et al (2009) Adaptive evolution in the Per gene family of vertebrates: neofunctionalization by positive Darwinian selection after two major gene duplications. *Biol Rhythm Res* 40(6):433–444
- Liang H, Plazonic KR et al (2008) Protein under-wrapping causes dosage sensitivity and decreases gene duplicability. *PLoS Genet* 4(1):0072–0077
- Lipinski KJ, Farslow JC et al (2011) High spontaneous rate of gene duplication in *Caenorhabditis elegans*. *Curr Biol* 21:306–310
- Lolkema JS, Dobrowolski A et al (2008) Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family. *J Mol Biol* 378:596–606
- Lynch M (2007) The origins of genome architecture. Sinauer Associates, Sunderland, MA

- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–1155
- Lynch, M, Conery JS (2003a). The evolutionary demography of duplicate genes. In: Meyer A, Van de Peer Y (eds) *Genome evolution*. Kluwer Academic Publishers, Dordrecht, 35–44
- Lynch M, Conery JS (2003) The origins of genome complexity. *Science* 302:1401–1404
- Lynch M, Force AG (2000) The origin of interspecific genomic incompatibility via gene duplication. *Am Naturaliste* 156(6):590–605
- Lynch M, Katju V (2004) The altered evolutionary trajectories of gene duplicates. *Trends Genet* 20:544–549
- Lynch M, O’Hely M et al (2001) The probability of preservation of a newly arisen gene duplicate. *Genetics* 159:1789–1804
- Maclean CJ, Greig D (2010) Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast. *Evolution* 65(4):932–945
- MacNeil AJ, McEachern LA et al (2008) Gene duplication in early vertebrates results in tissue-specific subfunctionalized adaptor proteins: CASP and GRASP. *J Mol Evol* 67(2):168–178
- Maere S, De Bodt S et al (2005) Modeling gene and genome duplications in eukaryotes. *Proc Nat Acad Sci U S A* 102(15):5454–5459
- Makino T, McLysaght A (2010) Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Nat Acad Sci U S A* 107:9270–9274
- Masley JP, Jones CD et al (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* 313:1448–1450
- Mayrose I, Zhan SH et al (2011) Recently formed polyploid plants diversify at lower rates. *Science* 333(6047):1257
- Meyer A, Van de Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* 27:937–945
- Mizuta Y, Harushima Y et al (2010) Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc Nat Acad Sci U S A* 107(47):20417–20422
- Morin RD, Chang E et al (2006) Sequencing and analysis of 10,967 full-length cDNA clones from *Xenopus laevis* and *Xenopus tropicalis* reveals post-tetraploidization transcriptome remodeling. *Genome Res* 16:796–803
- Ngai J, Dowling MM et al (1993) The family of genes encoding odorant receptors in the channel catfish. *Cell* 72:657–666
- Ohno S (1970) *Evolution by gene duplication*. Springer, Berlin
- Oka HI (1988). Functions and genetic bases of reproductive barriers. Origin of cultivated rice. Japan Scientific Societies Press/Elsevier, HI Oka, Tokyo, pp 181–209
- Otto SP (2007) The evolutionary consequences of polyploidy. *Cell* 131:452–462
- Panopoulou G, Poustka AJ (2005) Timing and mechanism of ancient vertebrate genome duplications—the adventure of a hypothesis. *Trends Genet* 21:559–567
- Papp B, Pal C et al (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424:194–197
- Paterson AH, Chapman BA et al (2006) Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends Genet* 22(11):597–602
- Piatigorsky J, Wistow G (1991) The recruitment of crystallins—new functions precede gene duplication. *Science* 252(5009):1078–1079
- Pinhal D, Yoshimura TS et al (2011) The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. *BMC Evol Biol* 11:151
- Pires JC, Gaeta RT (2011) Structural and functional evolution of resynthesized polyploids. In: Schmidt R, Bandcroft I (eds) *Genetics and genomics of the brassicaceae*, Springer, New York, 9:195–214
- Postlethwait JH, Woods IG et al (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* 10:1890–1902

- Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nat Rev Genet* 11:175–180
- Putnam NH, Butts T et al (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071
- Qian W, Zhang J (2008) Gene dosage and gene duplicability. *Genetics* 179:2319–2324
- Qian WF, Liao BY et al (2010) Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet* 26(10):425–430
- Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. *Annu Rev Ecol Syst* 33:589–639
- Rapp RA, Udall JA et al (2009) Genomic expression dominance in allopolyploids. *BMC Biol* 7:18
- Rausch JH, Morgan MT (2005) The effect of self-fertilization, inbreeding depression, and population size on autopolyploid establishment. *Evolution* 59(9):1867–1875
- Rodriguez DJ (1996) A model for the establishment of polyploidy in plants. *Am Naturalist* 147(1):33–46
- Salmon A, Ainouche ML et al (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Mol Ecol* 14:1163–1175
- Scannell DR, Byrne KP et al (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440:341–345
- Scannell DR, Frank AC et al (2007) Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc Nat Acad Sci U S A* 104(20):8397–8402
- Scannell DR, Wolfe KH (2008) A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. *Genome Res* 18(1):137–147
- Semon M, Wolfe KH (2007) Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. *Trends Genet* 23(3):108–112
- Semon M, Wolfe KH (2008) Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. *Proc Nat Acad Sci U S A* 105(24):8333–8338
- Simillion C, Vandepoele K et al (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc Nat Acad Sci U S A* 99:13627–13632
- Sinha SC, Wetterer M et al (2005) Origin of asymmetry in adenylyl cyclases: structures of *Mycobacterium tuberculosis* Rv1900c. *EMBO J* 24:663–673
- Soltis DE, Albert VA et al (2009) Polyploidy and angiosperm diversification. *Am J Bot* 96(1):336–348
- Soltis PS, Soltis DE et al (2006) Expression of floral regulators in basal angiosperms and the origin and evolution of ABC-function. *Adv Bot Res* 44:483–506
- Stoltzfus A (1999) On the possibility of constructive neutral evolution. *J Mol Evol* 49:169–181
- Surcel A, Zhou X et al (2008) Long-term maintenance of stable copy number in the eukaryotic SMC family: origin of a vertebrate meiotic SMC1 and fate of recent segmental duplicates. *J Syst Evol* 46(3):405–423
- Tate JA, Ni Z et al (2006) Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics* 173:1599–1611
- Taylor JS, Raes J (2004) Duplication and divergence: The evolution of new genes and old ideas. *Annu Rev Genet* 38:615–643
- Thomas BC, Pedersen B et al (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome Res* 16:934–946
- Tirosh I, Barkai N (2007) Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. *Genome Biol* 8(4):R50
- Udall JA, Swanson JM et al (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* 173:1823–1827
- Veitia RA (2002) Exploring the etiology of haploinsufficiency. *BioEssays* 24(2):175–184

- Veitia RA, Bottani S et al (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic, and proteomic effects. *Trends Genet* 24(8):390–397
- Wang J, Tian L et al (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172:507–517
- Watterson GA (1983) On the time for gene silencing at duplicate loci. *Genetics* 105(3):745–766
- Werth CR, Windham MD (1991) A model for divergent, allopatric speciation of polyploid Pteridophytes resulting from silencing of duplicate-gene expression. *Am Naturalist* 137(4):515–526
- Winter KU, Weiser C et al (2002) Evolution of class B floral homeotic proteins: obligate heterodimerization originated from homodimerization. *Mol Biol Evol* 19(5):587–596
- Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387:708–713
- Wood TE, Takebayashi N et al (2009) The frequency of polyploid speciation in vascular plants. *Proc Nat Acad Sci US A* 106(33):13875–13879
- Woolfe A, Elgar G (2007) Comparative genomics using *Fugu* reveals insights into regulatory subfunctionalization. *Genome Biol* 8(4):R53
- Yamagata Y, Yamamoto E et al (2010) Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. *Proc Nat Acad Sci U S A* 107(4):1494–1499
- Yang J, Lusk R et al (2003) Organismal complexity, protein complexity, and gene duplicability. *Proc Nat Acad Sci U S A* 100(26):15661–15665
- Yu W-P, Brenner S et al (2003) Duplication, degeneration and subfunctionalization of the nested *synapsin-Timp* genes in *Fugu*. *Trends Genet* 19:180–183

Chapter 2

Genetic Consequences of Polyploidy in Plants

James A. Birchler

Abstract Most eukaryotes have a history of whole-genome multiplication events followed by a progressive return to a more diploid state. The initial state of polyploidization, in which more than two copies of the genome are present, is considered here and the various types of genetic consequences that occur depending on the nature of the polyploid formed. The degree of association of chromosomes in meiosis is determined by the relative homology and will affect the segregation of the chromosome which determines the genetic properties. If all the chromosomes are quite similar and form associations of like type, this situation is referred to as autopolyploidy. If the different sets of multiple chromosomes are sufficiently dissimilar to each other, then the homologs will pair in meiosis with themselves and segregate independently of the different but related chromosome pair. This situation is referred to as allopolyploidy. Gene expression in ploidal series typically follows a per cell level correlated more or less with the number of sets of chromosomes present. Variation of individual chromosomes, or aneuploidy, produces a greater number of modulations of gene expression in parallel to classical studies noting that aneuploids have greater impact on the phenotype than changes in the copy number of the whole genome. The genetic properties of odd-number ploidies, such as triploids, are also described as well as higher ploidal levels such as hexaploidy and octoploidy.

J. A. Birchler (✉)

Division of Biological Sciences, University of Missouri, 311 Tucker Hall,
65211 Missouri, Colombia
e-mail: birchlerj@missouri.edu

2.1 Introduction

Most eukaryotes have a history of polyploidization followed by fractionation back to a near diploid level (Wolfe and Shields 1997; Simillion et al. 2002; Bowers et al. 2003; Blanc and Wolfe 2004; Chapman et al. 2006; Maere et al. 2005; Blomme et al. 2006; Freeling and Thomas 2006; Barker et al. 2008). Thus, at the least, polyploidy in essence is a matter of degree, and it has played an important role in the composition of the gene repertoire of many species. Typically, it is defined as the presence of more copies of the whole genome than the normal two that constitute a diploid (Stebbins 1947). However, from the standpoint of gene content, the determination of whether a species is a polyploid is somewhat arbitrary and dependent on the time before the present when the copy number of the genome was increased. Nevertheless, for evolutionarily “recent” events, certain principles can apply which will be summarized in this chapter.

In the polyploidy literature, the basic chromosome number is designated by x and consists of the complete set of chromosomes, or a genome. The number of chromosomes in the gametophyte generation and hence the gametes is referred to as the gametic chromosome number or n . In diploids, $x = n$, but at higher levels of polyploidy, this is not the case.

Polyploidy is typically divided into at least two categories that are determined by the type of chromosome pairing in meiosis I and the distribution of chromosomes during this process. Indeed, the type of chromosome pairing that occurs in meiosis affects the genetic properties of the species so such classifications have value. If the increase in genome copy number results from the combination of chromosome sets from divergent species, the different types of chromosomes will usually not pair with each other in prophase of meiosis I. In the case of tetraploids, if both divergent genomes are doubled by whatever means, those sets of chromosomes that are similar or identical will preferentially pair with each other to the exclusion of the other genome. This type of pairing is referred to as “disomic” in analogy with the situation in a diploid. A species with this type of scenario is referred to as an allopolyploid because the contributing genomes are different from each other.

2.2 Allopolyploids

Genetic ratios in an allotetraploid depend on the constitution of each genome (Clausen and Goodspeed 1925; Clausen 1941; Clausen and Cameron 1944). The different sets of related chromosomes are referred to as homoeologues. If both the homoeologues possess the homoeologous gene copies that are expressed similarly, then both would need to be mutant in order to express a recessive phenotype. Under these circumstances, duplicate gene ratios would typically be observed. In other words, recessive phenotypes would be found in 1/16 ($1/4 \times 1/4$) of the F_2