

Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity

Michael Freeling^{1,3} and Brian C. Thomas²

¹Department of Plant and Molecular Biology, University of California–Berkeley, Berkeley, California 94720, USA; ²College of Natural Resources, University of California–Berkeley, Berkeley, California 94720, USA

Controversy surrounds the apparent rising maximums of morphological complexity during eukaryotic evolution, with organisms increasing the number and nestedness of developmental areas as evidenced by morphological elaborations reflecting area boundaries. No “predictable drive” to increase this sort of complexity has been reported. Recent genetic data and theory in the general area of gene dosage effects has engendered a robust “gene balance hypothesis,” with a theoretical base that makes specific predictions as to gene content changes following different types of gene duplication. Genomic data from both chordate and angiosperm genomes fit these predictions: Each type of duplication provides a one-way injection of a biased set of genes into the gene pool. Tetraploidies and balanced segments inject bias for those genes whose products are the subunits of the most complex biological machines or cascades, like transcription factors (TFs) and proteasome core proteins. Most duplicate genes are removed after tetraploidy. Genic balance is maintained by not removing those genes that are dose-sensitive, which tends to leave duplicate “functional modules” as the indirect products (spandrels) of purifying selection. Functional modules are the likely precursors of coadapted gene complexes, a unit of natural selection. The result is a predictable drive mechanism where “drive” is used rigorously, as in “meiotic drive.” Rising morphological gain is expected given a supply of duplicate functional modules. All flowering plants have survived at least three large-scale duplications/diploidizations over the last 300 million years (Myr). An equivalent period of tetraploidy and body plan evolution may have ended for animals 500 million years ago (Mya). We argue that “balanced gene drive” is a sufficient explanation for the trend that the maximums of morphological complexity have gone up, and not down, in both plant and animal eukaryotic lineages.

A controversial trend in morphological complexity

Much controversy surrounds the general topic of increases in maximums of morphological complexity over the last 800,000 yr of evolution. The commonly accepted view is that amphioxus, a living representative of the urochordate sister group to vertebrates, is simpler morphologically than any vertebrate; that a liverwort, thought to be similar to the earliest land plant (Wellman et al. 2003), is less morphologically complicated than a fern than a pine than a sunflower. However, this popular view is not universally accepted. D.W. McShea (1996) has defined and evaluated various sorts of potentially rising complexities for Metazoans. His overall conclusion was that only some types of complexity have risen at all, and those only for a short time: “The evidence so far supports only agnosticism, indeed it supports an emphatic agnosticism” (McShea 1996). In a more recent and broader treatment, McShea (1998) identified two sorts of increasing developmental complexities, “developmental depth” and “structural depth,” among his eight potential largest-scale evolutionary trends, and both involve the subdivision of one developmental compartment or area into subareas (making a nest with new boundaries), permitting subsequent evolution involving division of labor and more diverse elaborations on a finer scale. New developmental boundaries (developmental depth) are only

seen as phenotype if they organize molecules that elaborate new shapes (structural depth).

We use a particular definition of “increasing morphological complexity” (Text Box 1). This definition uses terms that also have exact meanings: “developmental boundary” and “gene functional module,” also defined in our Glossary (Text Box 1). “Developmental boundary” and “gene functional module” are genotypic (informational) only, and are invisible to selection. Only when genetic information specifies something tangible (phenotype), in the form of morphological elaborations, can Darwinian selection operate. Examples of such morphological elaborations that form under the control of new developmental boundaries are segment landmarks, glands, cuticle/epidermal wax ornamentation, hairs, growth foci in space or time, branch foci, developmental identity switchpoints, and the like.

A plant example of nested developmental areas, and concomitant new developmental boundaries (Text Box 1), is the evolution from single apical cells of basal plants represented by fern (Banks 1999) to the zoned, layered shoot apical meristem of higher plants (Carles and Fletcher 2003). Although incompletely understood, there is general agreement that a segmented-ribbon early Metazoan (zootype), with its few protoHOX genes, evolved into the multioverlapping vertebrate segmental arrays, facilitated by their four HOX clusters (Garcia-Fernandez 2005). The general concept: Until developmental areas replicate and subdivide, generating a finer set of boundaries or time points (increased genetic potential), there is no way to evolve elaborations and divide labor up into more specialized temporal or spatial units (pheno-

³Corresponding author.

E-mail freeling@nature.berkeley.edu; fax (510) 642-4995.

Article is online at <http://www.genome.org/cgi/doi/10.1101/gr.3681406>.

type); finer-scale elaborations comprise fossil evidence of morphological complexity. Focusing on this particular type of developmental complexity is not new. In the words of S.B. Carroll (2001), “The main innovation that enabled large, modular organisms to evolve was the evolution of regional specification systems that subdivide growing embryos into semi-autonomous units” (Davidson et al. 1995).

The evolution of new boundaries nested within pre-existing boundaries probably involves the duplication of gene sets, but the products of these genes do not always bind to one another into a molecular complex of dedicated function as do, for example, genes encoding a ribosome or proteasome, or a transcription factor complex. It was not long ago that geneticists and evolutionary biologists moved away from individual gene products to “gene functional modules” (defined in Text Box 1) as the unit of cellular function (Hartwell et al. 1999). As pointed out by Ravasz et al. (2002), a tendency to cluster coexpressed genes on chromosomes, for which there is excellent evidence (Lercher et al. 2002; Williams and Hurst 2002; Hurst et al. 2004; Williams and Bowles 2004; Schmid et al. 2005) implies some degree of modularity. Ravasz et al. (2002) went on to successfully model metabolic networks in 43 different genomes, demonstrating topological modules connected into larger units in a hierarchical manner. Multiple algorithms—conserved gene neighborhood, gene fusions, and common phylogenetic distributions—were later used to predict functional associations, “functional mod-

ules,” that extended beyond metabolism (von Mering et al. 2003). Similar methods of module discovery have been applied to 10 genomes, including five eukaryotes, leading to 37 cellular systems of “parallel functional modules” including new functions not readily predicted by protein homology (Li et al. 2005). Recent genomics work within yeast has shown that novel functional specificities have happened in the evolution of yeast by duplication of functional modules (Pereira-Leal and Teichmann 2005); this work is particularly relevant to this discussion because they attempt to differentiate between simultaneous and stepwise evolution of a duplicated module (as will be reviewed).

The “functional module” of gene products—associated into a unit by being a molecular complex, a pathway, a cascade, or a network—would be a natural unit to be recruited or co-opted (Text Box 1) to a new developmental boundary during positive selection for morphological adaptations. The heart of this review is to present and explain a mechanism that naturally increases (duplicates) gene functional modules with each tetraploidy or large-scale segmental duplication generation. For a preview of this mechanism, see Text Box 1, where a new term, “balanced gene drive,” is defined.

In their major treatise, Maynard-Smith and Szathmáry (1995) assumed rising complexity at every level from chemical networks to networks of human neurons, including a transition involved in the sort of developmental–morphological complexity we use here. R. Dawkins (1986) devoted most of the classic

Text Box 1. Glossary

New term: Increasing morphological complexity. An increase in the number and nestedness of developmental areas as evidenced by **morphological elaborations** reflecting area boundaries. Increased are (1) the number of **developmental boundaries** in space or time, thus permitting a division of labor, and (2) the diversity of **gene functional modules** expressed in boundary cells.

New term: Balanced gene drive. A natural tendency to enrich the genome for genes encoding regulatory products—transcription factors and components of signal transduction—and genes encoding subunits of life’s most complex machines, during each tetraploidy or segmental-duplication generation. The mechanism is **purifying selection** for gene balance, leaving paired genes that, if fractionated, would have decreased fitness. This is selection for the status quo. This drives the duplication of **gene functional modules (spandrels of purifying selection)**, and this duplicated regulatory potential limits **increasing morphological complexity**.

Developmental boundary. A line within a population of cells—often drawn in response to morphogen concentration—generating two functionally different sides. Boundaries are maintained and refined by cell–cell interactions. Once distinct, specialized cells develop along both sides of the line, cells that influence the pattern and fates of nearby cells and, perhaps, the specification of progressively finer subdivisions. (Adapted

from Irvine and Rauskolb 2001.) Originally, a line that prevented cell migration and adhesion within the primordial insect wing (Garcia-Bellido et al. 1973). This definition is now expanded to include lines in time (early-late, juvenile-adult). Regulatory genes are expressed at pre-boundaries and in boundary cells. A boundary is invisible unless boundary cells elicit downstream **morphological elaborations** and new phenotypes.

Purifying selection. Usually, the elimination of nonstandard alleles by strong directional selection against them (adapted from Wilkins 2002). Also, the loss of non- or low-functional sequence by neutral substitutions, deletions, or conversions. Result: to maintain status quo. Synonym: negative selection.

Spandrel. A metaphor (Gould and Lewontin 1979) used to indicate a phenotype that originated as a necessary by-product of selection, either positive (as an adaptation) or negative (along with **purifying selection**). Once extant, spandrels may evolve functions, such as those that fuel **increasing morphological complexity**.

Gene recruitment. A mutation that results in the novel usage of a pre-existing gene. Much recruitment involves changes in enhancers, leading to new patterns of transcriptional activation. Synonym: co-option. Adapted from Wilkins (2002).

Coadapted gene complex. A group of genes comprised of alleles that coevolved and that, collectively, increased fitness in a set environment. This complex is (was) a functional unit of selection. Originally applied to every

allele in a species as part of a definition of “species” (E. Mayr; see Schilthuisen 2000). The genetic synonym is **gene network of function**, using evidence based on coexpression, double mutant phenotypes, two-hybrid screens, dosage dependencies, and the like. The genomic synonym is **gene functional module**, used here as a hypothetical group of genes encoding products that work together as deduced by algorithms measuring chromosomal linkage, coevolution, domain fusions, co-loss, motif-sharing, and the like. It is essential that members of a functional module have more in common with themselves than with those of other modules (Hartwell et al. 1999); see text for recent references and examples.

Drive. Used in an evolutionary context, drive implies an adapted molecular mechanism(s) that must (predictably), by its nature, change the genotype of a population over time in one particular direction until the trend is saturated or mitigated. Synonym, used incredulously: **predictable drive** (Gould 1995).

Meiotic drive. A well-established example of **drive**. A mechanism that generates gametes that are distorted from expected Mendelian ratios. Segregation distortion may be accomplished by a near-neutral interaction of particular chromosomal sites with parts of the meiotic apparatus, thereby escaping **purifying selection** operating to maintain the status quo. Example: increased frequency of chromosomes with “knobs” is a **spandrel** of meiotic drive.

text *The Blind Watchmaker* to explain the trend of increasing developmental and morphological complexity defined in ways compatible with our own. Dawkins explained the rise in complexity as part of a conceptual process he called “cumulative selection,” devised to explain how the repetitive selection of simple, single chance events can lead to an outcome that is directional and of increased complexity. Each step was selected positively for increased fitness owing directly to phenotype specified by the selected alleles.

There are ways in which mutation and natural selection together can lead, over the long span of geological time, to a building up of complexity that has more in common with addition than with subtraction. There are two main ways in which this buildup can happen; The first of these goes under the name of “co adapted genotypes”; the second under the name of “arms races.” (Dawkins 1986)

In summary, Dawkins (1) finds data in support of a trend involving rising morphological complexity and (2) attempts to explain this trend, without any sort of “drive” (defined in Text Box 1), by positive, stepwise selection as first proposed by Darwin (1859, *Origin*, exemplified in Chapter 6, “Organs of Extreme Perfection and Complication”).

Dawkins’ explanation of rising morphological complexity rests on the assumption that the sort of genetic variation necessary to fuel rising morphological complexity somehow existed naturally. S.J. Gould saw Dawkins’ explanation for rising complexity as a “just so” story, where the origin of an evolutionary outcome is simply assumed to have something to do with its fully evolved functionality (see Gould and Lewontin 1979). Gould argues:

Much of evolution is downward in terms of morphological complexity, rather than upward. We are not marching toward some greater thing. The actual history of life is awfully damn curious in the light of our usual expectation that there’s some predictable drive toward generally increasing complexity in time. If that’s so, life certainly took its time about it: five-sixths of the history of life is the story of single-celled creatures only.

[and. . .]

. . . the small bit of the history of life that we can legitimately see as involved in progress arises for an odd structural reason and has nothing to do with an predictable drive toward it. (Gould 1995)

Putting aside the choice to use loaded metaphysical words like “purpose” and “greater thing” and the fully enigmatic “odd structural reason,” and putting aside for now data relating to downward evolutionary lineages (possible trends of decreasing morphological complexity that might counterbalance rising trends), we assume that the discovery of a mechanism that provides “predictable drive” toward increasing morphological complexity would have changed Gould’s mind.

At least one serious student of complexity finds the biological trend being discussed to be potentially trivial. McShea (1998), discussing the transitions over time posited by Maynard-Smith and Szathmáry, suggested that the overrepresentation of increases over decreases in complexity might be expected because “. . . as seems plausible, decreases were limited at some low level by a boundary, a lower limit on hierarchical depth.” This argument may reduce to “there’s no way to go but up.” Similarly, rising complexity may be trivial because total complexity variance may be increasing in the clade, but there is a lower limit (Gould 1988).

We are hoping to transcend the discussion that would natu-

rally ensue at this point by examining critically recent data from molecular biology, genetics, and genomics from the “predictable drive” perspective. We will use the word “drive” in a stringent, genetic, and causal sense (defined in Text Box 1), not in the way complexity theorists sometimes divide trends into “passive” and “driven” categories. We call this new drive “balanced gene drive” (defined in Text Box 1). This drive mechanism derives from recent research progress on gene and genome duplication in eukaryotes and phenotypes specified by altered gene dosage. Balanced gene drive fits best a “mutationist” explanation of big evolutionary trends (Ni 2005). We review this research progress in the next few pages, and follow with an argument for balanced gene drive.

Eukaryotic gene duplication and loss

The gene content of living things changes over time by (1) gene addition by horizontal transfer, (2) gene addition involving some form of duplication followed by divergence, and (3) gene loss by removal or by being copied over by gene conversion. For multicellular eukaryotes, horizontal transfer will be assumed to be negligible. Gene loss is not easy to measure without the appropriate whole-genome sequences, and is best measured when there is one or more whole-genome outgroups in a phylogenetic tree relationship. There has been an important advance in how we look at gene loss by conversion. Gao and Innan (2004) used the phylogenetic tree of various yeast whole genomes to establish when local duplications originated. They concluded that base-substitution clock estimates of tandem duplication occurrence over time, estimated to be 0.01 per gene per million years (Myr) by (Lynch and Conery 2000), were overestimated by ~100-fold because of the prevalence of gene conversion. In general, gene family loss and concomitant loss of coregulated genes characterizes eukaryotes (Koonin et al. 2004, and citations therein). The tetraploidies that decorate the eukaryotic phylogenetic tree of Figure 1 have been interpreted to have generally added genes to a lineage. Following tetraploidy, there is certainly “gene loss,” but the assumption has been that both homeologs are not lost. Except for the yeast tetraploidy, none of the tetraploidies of Figure 1 can be simultaneously evaluated for complete (both homeologs) gene loss because of the lack of an appropriate outgroup genome sequence. We now assume that complete gene loss, had it happened, was random with respect to gene functional category.

Gene content over time is also influenced by gene duplication. Most biologists think of gene duplication followed by divergence as an important source of information for the evolution of novel adaptation (Haldane 1933; Ohno 1970; Taylor and Raes 2004). The work of E.B. Lewis has been particularly important since he first demonstrated a causal link between genes duplicated in the genome and segments duplicated and diversified along the anterior–posterior axis of a metazoan (Lewis 1951). Lewis’s general scheme of duplication, repression of one of the duplicates, accumulation of mutants, followed by either loss or de-repression with novel possibilities, clearly presaged the modern scheme of duplication and gene/motif co-option (True and Carroll 2002) or gene recruitment (Text Box 1; Wilkins 2002) to novel function. Genes generated by duplication certainly fuel the evolution of most anything new (e.g., the diversification of angiosperms during the Cretaceous, De Bolt et al. 2005).

In general, genes can duplicate (1) locally, usually in tandem; (2) as part of a chromosomal segment; or (3) via a tetra-

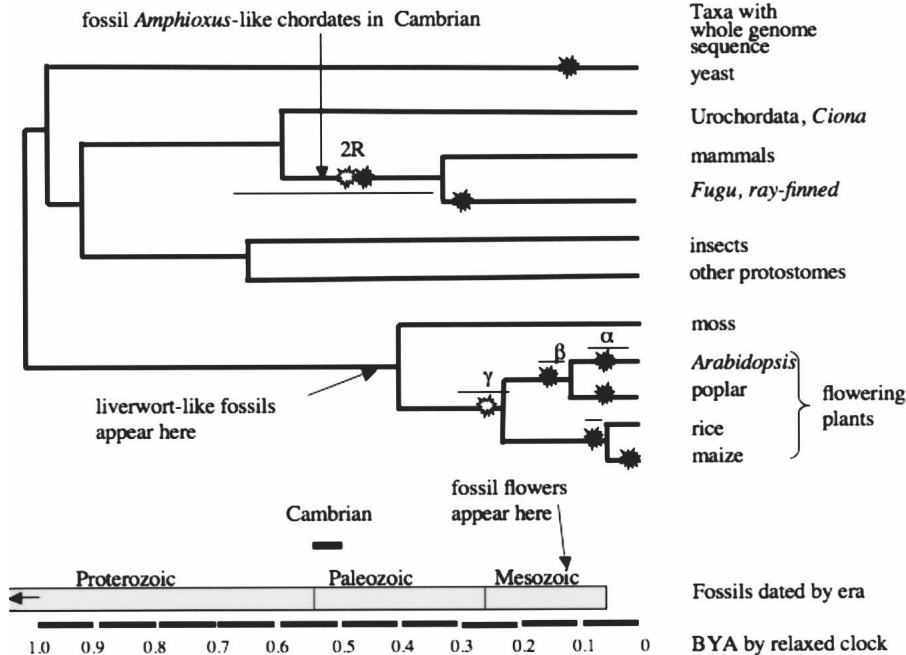


Figure 1. Paleopolyploidy in eukaryotes plotted onto a phylogenetic tree (relaxed clock by Douzery et al. 2004) of eukaryotes using species that have a majority of genome sequenced. Tetraploidy events are denoted with a black starburst, and large-scale segmental events (possible tetraploidies) with an outlined starburst. Outlined starbursts and times previous are in the “twilight zone” (Simillion et al. 2004). Each event has a range of suggested time points indicated by the length of the thin line. The geological eras (Freeman and Herron 2004) are indicated by rectangles. The position of amphioxus is inferred from *Hox* gene research (Furlong and Holland 2004). The first appearance of a liverwort-like plant reflects a recent fossil find (Wellman et al. 2003). The angiosperm fossil record and first appearance of flowers in the Cretaceous have been reviewed (Friis et al. 2005); note the huge discrepancy between fossil-based and relaxed clock ages for the appearance of flowering plants. Major tetraploidy references include for yeast (Wolfe and Shields 1997; Kellis et al. 2004); chordates (McLysaght et al. 2002; Simillion et al. 2004); ray-finned fish (see Taylor et al. 2003); *Arabidopsis* α , β , γ , identified with Greek letters on the tree (Bowers et al. 2003), using a comparative gene-tree approach; Maere et al. (2005) used a molecular clock to deduce three similar events called 1R, 2R, and 3R, and general *Arabidopsis* large-scale duplication (Blanc et al. 2000; Ku et al. 2000; Simillion et al. 2002); poplar (D. Rokhsar, Joint Genomes Institute, DOE, >60 Mya; pers. comm.); rice (Paterson et al. 2005; Tian et al. 2005; Yu et al. 2005), and maize (Gaut and Doebley 1997). (BYA) Billion years ago.

ploidy, sometimes called whole-genome duplication. In terms of gene balance, local duplication and tetraploidy have very different consequences, and the consequences of a segmental duplication depend, as will be seen, on whether or not the genes on the segment encode products that cooperate in the same macromolecular complex or network (gene functional module, Text Box 1).

The Gene Balance Hypothesis, and its rich experimental context

Although the phrase “balance hypothesis” was first used by Papp et al. (2003) in their functional genomic work on haploinsufficiency in yeast and humans, there is a rich context within which this hypothesis has meaning. Important to the gene balance hypothesis is the body of work and analysis of Birchler and coworkers on dosage effects, inverse dosage effects, and compensations in maize and *Drosophila* (Birchler et al. 2001), and Vieta’s more theoretical work on haploinsufficiency and transcriptional machinery (Veitia 2002). Together, these three citations properly credit the Gene Balance Hypothesis (Text Box 2).

The relationship between gene regulation and the Gene Balance Hypothesis was reviewed (Birchler et al. 2005). Using data in

a functional genomics database containing growth curves for single-gene knockouts of otherwise diploid yeast, compared with a diploid control, Papp et al. (2003) found a significant correlation between genes whose products participate in subunit–subunit interactions and a slower-growth phenotype. The degree of product interaction was deduced from data in the MIPS Comprehensive Yeast Genome Database: <http://mips.gsf.de/genre/proj/yeast/>. Papp and coworkers used genes that, when homozygous null, did not support growth. In addition, they showed that transcriptional regulators and proteins that are part of signal transduction in humans were significantly oversensitive to gene dose (Papp et al. 2003). These investigators noted that not all genes retained from the yeast tetraploidy event (thought to have occurred ~100 Mya) (Fig. 1), have been equally retained as pairs in modern yeast: Genes encoding ribosomal proteins were significantly over-retained. This was expected of the Gene Balance Hypothesis because tetraploidy does not change gene balance, thus connected genes should be difficult to remove from a tetraploid one at a time because of haploinsufficiency, and impossible to remove in concert. As predicted, connected genes should be over-retained following tetraploidy. The Birchler review is especially thorough in showing a consensus result among many experiments designed to better understand how whole-genome duplications did relatively little to phenotype compared to phenotypes caused by altered

gene balance, and how the exact stoichiometry of regulatory factors must explain these results. Of particular importance are explanations of a dosage phenomenon called “inverse dosage effect,” in which positive and negative regulatory components interact to achieve gene regulation (see Box 1 of Birchler et al. 2005). The result is a model that is inherently dose-sensitive. Veitia (2002) discusses the many mechanisms that might explain susceptibility and resistance to haploinsufficiency phenotypes and their relevance to concepts of dominance, and concludes that most cases of haploinsufficiency can be accounted for by nonlinear interactions between or among subunits at the time of assembly of, for example, a transcription factor complex. Veitia shows experimental and simulation evidence that typical transcription factor complexes are composed of multiple subunits, and that the activity of the complex relative to the concentration of any one of the subunits is sigmoidal; this indicates maximum concentration-dependence at the inflection point that constitutes a de facto threshold or switch. For a theoretical example, the assembly of an active ABA heterotrimer is shown to be hypersensitive (nonlinear, cooperative) to a 50% reduction of A subunit if the assembly pathway specifically goes A to AB to ABA but shows a linear dosage response if A to AA to ABA. Essentially,

Text Box 2. Statement of the Gene Balance Hypothesis

Genes whose products participate in product–product interactions, or whose product concentrations limit the expression of downstream genes that themselves participate in product–product interactions, or when downstream genes in a cascade participate in product–product interactions (or through more complicated balances of positive and negative regulatory mechanisms) (see Birchler et al. 2005), are predicted to be particularly dosage-sensitive. These genes, lumped together, are referred to as “connected genes” in order to best-fit data for all groups of genes that display dosage sensitivity. Dosage sensitivity is predicted to manifest itself as, in comparison to a diploid wild type, haploinsufficient or triploinsufficient phenotypes, phenotypes that almost always lower fitness or cause disease. The selective consequences of duplication events during evolution may be modeled on the basis of this hypothesis. Because multicellular eukaryotes have particularly complex machinery that assembles and operates synergistically to regulate transcription, signal transduction, and protein degradation, those connected genes that specify the components of these regulatory machines—and any other similarly interconnected machine—are predicted to behave over evolutionary time in accordance with this hypothesis. [A synthesis of Birchler et al. 2001; Papp et al. 2003 (Hurst lab); and Veitia 2002.]

particular assembly routes titrate the limiting subunit, so that fully active product is cooperatively reduced (Veitia 2002). How such haploinsufficient kinetics of transcriptional complex assembly can be mitigated, at least using mathematical simulations based essentially on the mass action law of physics, is further explored (Veitia 2003). Thanks to these kinetic experiments, both wet and in silico, the Gene Balance Hypothesis has a firm theoretical foundation.

Predicted changes in gene content by type of duplication

1. Autotetraploidy.⁴ Connected genes (like proteasome core or transcription factor genes in higher plants and animals, defined in Text Box 2), should be over-retained, and unconnected genes under-retained, after an autotetraploid has fractionated to a stable version of diploidy. Papp et al. (2003) predicted that those genes encoding transcription factors in multicellular eukaryotes might exceed ribosomal protein genes in complexity, and thus be over-retained after tetraploidy.
2. Local duplication. Connected genes should be underrepresented among genes in clusters of local (mostly tandem) duplicates because increasing the concentration of but one sub-

⁴Autopolyploidy, where one chromosomal set doubles, is the easiest sort of tetraploid to model because the new tetraploid is assumed to be the sum of its genomes. Allotetraploidy, where two different genomes combine, adds complications to these predictions, complications that probably reflect reality. Selection for polyploidy in the first place, given its expected lowered fitness due to mis-segregations, is easier to explain if the parents are of different genotypes, and the special characteristics of the tetraploid increase fitness. Recent studies on synthetic allopolyploids in plants show that gene silencing is common, and “subfunctionalized” silencing occurs; in general, the parental genotypes are not equally expressed (Adams and Wendel 2005). Evidence for rapid intrachromosomal genome changes following allopolyploidy, and ideas about mechanisms that may be involved, have been reviewed (Osborn et al. 2003). Genes from one of the two parents in an allotetraploid might be preferentially coadapted. As has and will be further documented, coexpressed genes tend to be positioned together in chromosomes. Such clusters might be coregulated at the chromatin level, and might tend to stay together during tetraploid fractionation. Even though chromosome-level regulation, such as incomplete or organ-specific silencing, could buffer gene dosage in allotetraploids, the gene dosage hypothesis accurately predicts changes in gene content following tetraploidy.

unit in a complex by 50% should be much like halving the dose, and also reduce fitness. Conversely, unconnected genes are predicted to preferentially occur in local arrays.

3. Segmental duplication. To the extent that genes participating in the same machine or network (functional module) are linked on a chromosomal segment, as is common in prokaryotes, segmental duplication and whole-genome duplication have similar consequences. Retained, duplicated segments should tend to carry one gene of each dose-sensitive component of each machine in which they participate.
4. There should be selection for any innovation that mitigates dosage effects.

Data regarding these predictions

Following the most recent tetraploidies in both the *Arabidopsis* and rice lineages, here called α -tetraploidies, most gene pairs are reduced to one gene. The process is called diploidization, or in reference to any sort of duplication, fractionation (Lockton and Gaut 2005).

All flowering plants (angiosperms) are paleopolyploids. The evidence for this was deduced from intragenomic BLAST comparisons of proteins organized by their gene’s map position. The dots, falling-into lines of best nonself BLAST hits, were plotted by chromosomal position (see references in Fig. 1 legend). Syntenous chromosomal stretches were visualized as lines of dots. The *Arabidopsis* genome has been reduced to a dot-plot, which reveals the most recent tetraploidy event, called α . Within these α -syntenous regions are more degraded syntenous lines providing evidence for an earlier β tetraploidy, and, nested within, an even earlier segmental duplication or tetraploidy (γ) (Bowers et al. 2003). The differentiation between separate contemporaneous duplication events has been accomplished using a gene-tree phylogenetic approach to nest older lines within presumably more recent lines (Chapman et al. 2004). Figure 1 plots these three events on the eukaryotic phylogenetic tree. Three discreet tetraploidies in the *Arabidopsis* lineage have independent support (Maere et al. 2005) using third-codon-position decay measurements. There have been several discoverers of the large-segmental or whole-genome duplication past of *Arabidopsis* (Fig. 1 legend).

Blanc and Wolfe (2004) found that percent retention from the tetraploidy in the *Arabidopsis* lineage differed by GO category. Their results are now presented in terms of percent above or below expectations. They used two independently derived pairs lists. One list was 3800 pairs among 26,000 total GenBank genes compiled by Bowers et al. (2003). Percent retention cross-referenced with GO category ranged significantly from a low of 29% below expected for genes involved in DNA repair (GO:0006281; $n = 86$) to a high of 243% above expected for genes annotated as encoding sodium:hydrogen antiporter activity (GO:0015385; $n = 20$). Most interesting were the two larger categories of genes that were significantly over-retained: transcription factor activity (TF genes) at 156% above expected (GO:0003700; $n = 552$) and two classes of protein kinases at 153% above expected (GO:0004713 and 0004674; $n = 1251$). The value n above indicates Blanc and Wolfe’s estimate of pre-tetraploid gene numbers. Seoighe and Gehring (2004) prepared a reduced edition of the Bowers gene pairs. They found that genes in GO categories “transcription regulator” and “signal transducer” were significantly over-retained. These workers took advantage of the unique phylogenetic tree method used by Bowers et al. (2003) to show that lineages of genes that included over-

retained genes from an earlier tetraploidy event tended to contribute genes retained in a later event, thus showing that biased retention may repeat itself through tetraploidy generations. “Transcription regulators” were significantly over-retained at 121% above expected, and “signal transducer” genes, which included protein kinases, were significantly over-retained at 128% above expected.

A recent study by Maere et al. (2005) confirms and extends the conclusions of previous *Arabidopsis* investigators, and correlates time of tetraploidy (million years ago [Mya]) with important evolutionary transitions. These workers also show that whole-genome and local duplications change gene content in reciprocal fashion. As predicted by the Gene Balance Hypothesis, categories of genes we have called “connected,” such as TF, signal transducer, and developmental genes, are over-retained following the middle tetraploidy (their 2R, which is most similar to β of Fig. 1) and often the other two tetraploidies as well, and are under-retained as tandem duplicates. Conversely, genes described as encoding “conserved biological functions,” as those in categories involving DNA metabolism, nuclease activity, and RNA-binding, tend to be under-retained following tetraploidies and over-retained among local duplicates.

The over-retention following tetraploidy of genes encoding upstream regulators, connected genes as predicted by the Gene Balance Hypothesis, is not confined to *Arabidopsis*. The date for grass family radiation has been estimated to be 50–70 Mya (Kellogg 2001), or perhaps before Godwana split apart (ca. >80 Mya) (Prasad et al. 2005). The grass lineage of monocot flowering plants had a whole genome duplication (Fig. 1 legend) before this radiation. TF genes are preferentially retained as α -pairs in rice; while average retention is 16%, TF genes were retained at 50%: 312% above expectations (Tian et al. 2005).

Conversely, the Gene Balance Hypothesis predicts that local duplication, because of the out-of-balance phenotype that should arise from gene hyperploidy, will preferentially include genes encoding monomers and “less-connected” genes. Those very GO categories over-retained from tetraploidy are predicted to be under-retained in the local duplicate data set. There is a correlation between the number of subunits in the quaternary gene product and whether or not the gene is found as a local duplicate—as subunit count goes up, local duplication tendency goes down; this conclusion is true for both yeast and humans (Papp et al. 2003; Yang et al. 2003), although humans had a far greater proportion (>2 \times) of genes that were locally duplicated, and only the most “connected” human genes were reliably singlets. The underrepresentation of TF genes in local clusters, as predicted by the Gene Balance Hypothesis, is also true for plants (see above; Maere et al. 2005). Of the ~28,271 protein-coding genes that are not within transposons in TIGR assembly Version 5 *Arabidopsis*, 4167 are in local clusters (Haberer et al. 2004). Therefore, the average gene has approximately a 14.4% chance to be in a local array. Of the 1827 genes annotated with GOSLIM term “transcription factor activity,” 182 are on Haberer and co-workers’ local duplicate list: 10.0% local duplication is significantly below what is expected by 31% (our calculations).

More highly expressed genes in yeast tend to have been retained following tetraploidy (Seoighe and Wolfe 1999), thus subunit–subunit interactions are not the only functional features to correlate positively with retention.

The term “connected genes” (Text Box 2) is necessarily inexact because it must denote genes that are dose-sensitive for more than one reason, including protein–protein interactions

and regulatory cascades/circuitry. One way to evaluate the prediction “retained genes tend to be connected genes” is to generate tetraploid retention data per GO category in an unbiased way, and examine all GO categories and especially the extreme categories. There are 510 GO terms (obtained from The *Arabidopsis* Information Resource Web site 6/05) that include at least 20 of the 25,219 *Arabidopsis* genes in a minimized *Arabidopsis* gene list (Supplemental material 1). Retention (calculated from a pairs list edited from Bowers et al. 2003; we use 3178 pairs) frequencies ranged from a high of 0.75 (GO: proteasome core complex, *sensu* Eukaryota; $n = 20$) through an average of 0.20 defined by GO: molecular function unknown ($n = 7832$) to a low of 0.0 (GO: de novo pyrimidine base biosynthesis; $n = 20$). There are a few individual GO categories that display retention frequencies that do not seem predictable by any sort of fuzzy “connectedness” model; for example, GO: toxin catabolism is retained at a high 0.5. More importantly, a trend from more connected down to less connected, a trend reported by all researchers, seems supported by the data in this complete list. Examples with a rich experimental history include, in order of descending retention, 0.43 (GO: ribosome biogenesis; $n = 102$), 0.40 (GO: protein serine/threonine kinase activity; $n = 502$), 0.38 (GO: transcription factor activity; $n = 1719$), 0.32 (GO: motor activity; $n = 241$), 0.28 (GO: structural constituent of cell wall), 0.26 (GO: RNA binding; $n = 519$), 0.20 (average), 0.17 (GO: tRNA processing), 0.13 (GO: cysteine-type peptidase activity; $n = 108$), 0.12 (GO: damaged DNA binding; $n = 50$), 0.09 (GO: ATP-dependent DNA helicase activity; $n = 44$), and the penultimate 0.08 (GO: DNA methylation; $n = 38$). As pointed out by Maere et al. (2005), ancient gene categories, as well as “unconnected” ones (from Gene Balance Hypothesis predictions), seem under-retained following tetraploidy. However, predictions on the connectedness of any particular GO category based on retention frequency alone should constitute a hypothesis to be tested.

Papp et al. (2003) estimated “connectedness” by protein–protein interaction. Protein–DNA interactions could also contribute to some measure of connectedness. Comparisons of orthologous genes from maize and rice found that, using alignment settings that found the average gene encoding an enzyme to have 2.4 conserved non-coding sequences (CNSs; pairwise phylogenetic footprints), genes encoding TFs had nine, and much longer CNSs as well (Inada et al. 2003). As already reviewed, TF genes are also over-retained following tetraploidy. It is possible that DNA–protein as well as protein–protein binding contribute to the connectedness of regulatory genes. Micro and small RNAs could bind as well. Much of a gene’s nonexon space in mammals is conserved over evolutionary time. Jareborg et al. (1999) found that non-coding mammalian space was filled with CNSs: 36% of promoters, 50% of 5′-UTRs, 23% of introns, and 56% of 3′-UTRs. In general, transcription factors are thought to bind over large stretches of animal gene promoters (Yuh et al. 2001; Bolouri and Davidson 2002; Levine and Tjian 2003). Comparisons among *cis*-acting sequences in well-studied genes of yeast, nematodes, fruit flies, mosquitoes, sea squirts, pufferfish, mice, and humans (in order) evidence a general increase in length and a “progressively more elaborate regulation of gene expression” (Levine and Tjian 2003).

Purifying selection after tetraploidy and the fate of retained pairs

In light of the Gene Balance Hypothesis, selection works on a new tetraploid to preserve the status quo: balanced gene expres-

sion. Purifying selection (Text Box 1) against any change upsetting gene balance tends to leave pairs of connected genes, and consequently tends to duplicate each (dose-sensitive) gene in a gene functional module. Thus, tetraploid fractionation by some sort of gene removal mechanism tends to duplicate functional modules; module duplication is a “spandrel” (Text Box 1) or indirect by-product of fractionation. It is not clear whether or not a functional module duplicates by duplicating all or just some of its constituents’ genes. The evolution of complexity requires divergence of duplicated functional modules or gene networks (duplicate precursors to coadapted gene complexes); all of these near synonyms are defined operationally in Text Box 1. Since functional duplication should relax selection on at least one of the duplicates of any one module, the rate of divergence is expected to increase with duplication events.

There are many studies that compare retained duplicates after some evolutionary time. The generalized result for all eukaryotes is that duplicates diverge rapidly, although it is usually difficult to clearly differentiate subfunctionalization⁵ from gain of function (Gu et al. 2002; Wagner 2002; Makova and Li 2003; Raes and Van de Peer 2003; Gu et al. 2004; Haberer et al. 2004; He and Zhang 2005). Co-option or gene recruitment (True and Carroll 2002; Wilkins 2002) is the stepwise, coadaptation process within which this divergence might be best understood. It may be safely concluded that evolution will use duplicate genes, and duplicate functional modules, for different purposes once they exist.

Drive and balanced gene drive (Text Box 1)

The term “drive” has been abused to the extent that its usage should probably be reserved for cases essentially like “meiotic drive.” Increasing morphological complexity is driven in this stringent sense. M.M. Rhoades (1942) found that certain naturally occurring maize chromosomes were preferentially transmitted to progeny. The cause of this biased transmission of information into the gene pool was that particular chromosomes with “knobs” acted like neocentromeres, attached to the spindle and were pulled early into one of the terminal cells of the column of megaspores, and thus tended to transmit to the single egg. The adapted machine causing drive is the unidirectional meiotic divisions and stereotypical pattern of female meiotic product cells. Orderly meiosis is a status quo process of most eukaryotic life, and is maintained by purifying selection. The biased transmission of genes that happen to be on the early-segregating chromosome is unadapted, but undoubtedly used by evolution, as has been shown in *Zea mays* (Buckler et al. 1999). The biased transmission of certain genes and the phenotypes they encode is a spandrel. Rhoades’ case and other cases of segregation distortion (Lyttle 1993; Buckler et al. 1999) are called “meiotic drive.”

⁵Subfunctionalization (Force et al. 1999) was originally put forth specifically as a mechanism to explain over-retention of duplicates following tetraploidy, and has generated much theory (Lynch and Force 2000; Prince and Pickett 2002; Lynch and Conery 2003; Force et al. 2005). It is a neutral process where different, dispensable *cis*-functional parts of a gene are compensatorily lost such that both duplicates are required to specify the original function. Subfunctionalization is a two-hit mutational mechanism that locks in pairs only after the tetraploid evolves. Alternatively, the Gene Balance Hypothesis predicts that pairs will be preserved—by resisting purifying selection, a zero-hit mechanism—because particular multi-subunit machines or cascades cannot end up in a haploinsufficient state, and, as we have shown, predicts correctly the GO-term content of retained genes. Therefore, subfunctionalization is probably not a primary mechanism for pair retention, although it certainly occurs once a pair is retained.

Natural selection for preservation of balanced gene activity as the tetraploid fractionates to a more-diploid state injects genes into the gene pool that are significantly biased toward connected genes, and this bias is compounded over tetraploidy (or perhaps segmental duplication) generations. The result is the incidental duplication of functional modules. The duplication of functional modules occurs as an unadapted by-product, a spandrel, of purifying selection for gene balance. Thus, meiotic drive and what we now call “balanced gene drive” are essentially the same, except one operates each sexual generation, while the other each duplication generation.

The duplication outcomes of balanced gene drive are predictable from the rules of physics and chemistry governing subunit-subunit interactions (Veitia 2002, 2003) and probably predictable from gene products participating in cascades and networks (Birchler et al. 2005). We see duplication of functional modules as the limiting step to rising morphological complexity, and that such complexity would not rise as a trend without balanced gene drive. In other words, once there is a supply of diverged, duplicated functional modules—the precursors to coadapted gene complexes—then Darwinian stepwise selection may be expected to deliver new morphological elaborations some of the time.

The advantage of using the pompous word “drive” or phrase “balanced gene drive” is that, if a trend is driven in the meiotic drive sense, it is not trivial.

Balanced gene drive in an adaptationist scenario

Using population genetics theory, trends in the evolution of morphological complexity might be explained, essentially, in one of three ways: (1) by natural selection in small, ever-positive steps (e.g., adaptationists Darwin and Dawkins); (2) by neutral fixations (e.g., Gould); or (3) by imagining or hoping for special mutations to limit or direct evolution with natural selection playing an important but secondary and nondirectional role (e.g., geneticists Morgan 1926, 1932 and Goldschmidt 1953). This latter theoretical category, sometimes called “mutationism” (explanation 3 above), would own balanced gene drive as a supporting mechanism, but all three sorts of selection would play a part in any evolutionary scenario. Ni (2005) has carefully reviewed these theoretical matters from a population genetics perspective.

Populations of species compete with one another for necessary but limiting resources, and sometimes morphological innovations are a part of successful adaptation, for example, the evolution of alternative plant leaf anatomies to more efficiently fix CO₂ in the tropics, and the evolution of different sorts of woody stems (trunks) to support the heights necessary to survive an “arms race” toward the sun. As explained previously, increasing morphological complexity requires the recruitment of new functional modules to cells recognizing new developmental boundaries. However, duplicating the genes that would become coadapted gene complexes, duplicate gene functional modules, is a formidable task, and one that neither Darwin nor Dawkins addresses. Balanced gene drive operates at each tetraploidy or balanced gene segment duplication to generate duplicate functional modules. These spandrels are not the result of direct selection. This review has shown that genes, once duplicated, diverge naturally. Once diverged, duplicate functional modules are components of what Dawkins (quoted above) called “co-adapted genotypes.” We see the trend of rising morphological complexity as the predictable outcome. However, duplication of functional

modules does not automatically circumvent the dose-sensitivity problem. In order to recruit a diverged, duplicate functional module to a new boundary, gene dosage sensitivity must be avoided or mitigated.

For flowering plants, tetraploidies have happened at least every 100 Myr (see Fig. 1 and legend) over at least the last 300 Myr; it seems unlikely that this basal rate of tetraploidy—with current levels of balanced gene drive—could continue indefinitely without excessive costs. Of the many ways to mitigate costs, evolving obligate sexual reproduction is one that characterizes many higher animals. (See Supplemental material 2: Plants vs. Animals.)

Gould (quoted above) did not see any “predictable drive” toward increasing morphological complexity. We offer balanced gene drive. Gould also argued that upward trends in complexity, if they existed at all, were counterbalanced by downward trends. We found little case support for decreasing complexity as a trend. There are certainly obligatory neotenus lineages, as in amphibians (Pierce and Smith 1979) and plant lineages in which the embryonic root has been lost; these could be seen as simplifications because a developmental stage or embryonic domain has been partially abandoned. However, sister lineages that do undergo metamorphosis or use complete embryos evolved in parallel. Specific cases of simplification of organs do exist (Roth et al. 1997), but most simplification trends seem more on the surface, like land animals to cetaceans, or are expected, as with parasites or symbionts. A trend toward reducing the number of defined areas (developmental compartments defined by developmental boundaries; Text Box 1) in plant meristems or organ primordia, in lineages with repeated large-scale genome duplications, would challenge the ideas presented here. We found only upward morphological trends in the green plant lineage.

Plants versus animals, body plans, and comparative evolution of morphological complexity

See Supplemental material 2. Figure 1 shows and its legend documents how animals probably had their last tetraploidy(s) just before the Cambrian explosion 495–543 Mya, and how all flowering plants are repeatedly paleopolyploids (or have repeatedly suffered large segmental duplications), with their most recent or α tetraploidy being younger than ~70 Mya. Because of vast biological differences between plants and animals and difficulties in comparing their body plans, comparing morphological complexity between plants and animals is about as much light-hearted speculation as it is logical deduction. It is possible that body plan evolution ended for animals when tetraploidy ended, but that body plan evolution for flowering plants is still happening, as is tetraploidy. Having neither cell migration nor cell rotation, some sorts of complexity are not available to plants. Morphological complexity, as defined here, fits plants well.

Limitations and prospects

We know too little about epigenetic involvements following large-scale duplications. A particularly effective adaptation to the overload of regulatory machinery expected following repeated tetraploidies would be the evolution of mechanisms to underexpress or silence, organ-specifically, one or the other duplicate chromosomal region; Footnote 4 discusses this particularly important possible consequence of balanced gene drive especially involving allotetraploidy. The reason epigenetic mechanisms are

relegated to a footnote in this synthesis is not because they are unimportant. Rather, it is clear from the gene content of retained-gene-pair GO categories (as reviewed here and Supplemental material 1) that whatever innovations evolved to buffer the genome from balanced gene drive have been, at least until now, incompletely effective. These potential obstacles have been “driven through.” Nevertheless, our understanding of balanced gene drive will not be robust until epigenetic involvement is better understood. Epigenetic mechanisms are also implicated by the general result that closely linked genes in both mammals (Lercher et al. 2002; Williams and Hurst 2002; Hurst et al. 2004) and higher plants (Williams and Bowles 2004; Schmid et al. 2005) tend to be coexpressed. Although chromosomal linkage has been used in algorithms that have successfully computed functional modules, linkage is not the only indicator (Ravasz et al. 2002; von Mering et al. 2003). To the extent that dose-sensitive genes participating in a functional module are linked together on the same chromosomal segment, a segmental duplication could duplicate the module without causing triplo-insufficiency, an unfit out-of-balance phenotype.

The lack of useful genome sequence, genomes strategically placed in the phylogenetic tree, is a primary limitation of this review and of the general field of comparative genomics. For example, only by comparing several genomes that could be arranged in an unambiguous phylogenetic tree of yeasts could Gao and Innan (2004) distinguish rates of local gene duplication from rates of loss by gene conversion. Flowering plants are particularly well suited to the continued study of balanced gene drive. However, none of the plant tetraploidies of Figure 1 have useful outgroups with sequenced genomes. Useful outgroups would branch just before the tetraploidy event, and would ideally not have undergone large-scale duplications in their own lineages. The most useful outgroup is not likely to be in a taxon that includes a commodity. For example, the best outgroup for the pre-grass family α -tetraploidy is probably one of the two known *Joinvillea* species, the sole representatives of the sister family to the grasses.

Our knowledge of the duplication of functional modules is rudimentary, and now limits our understanding of the more innovative aspects of developmental/morphological evolution. There has been a serious attempt to understand how functional modules have duplicated in yeast (Pereira-Leal and Teichmann 2005), where 6%–20% of protein complexes have a homology with at least one other such complex, but with different binding specificities or regulation. Yeast, a unicellular organism, had a tetraploidy ~100 Mya (Fig. 1 legend). These investigators show that duplication usually did not include duplication of all subunits in a protein complex, and that such duplication usually did not happen at the tetraploidy event. Rather, they support a model for partial, stepwise duplication of metabolic functional models. We predict that, when similar research is complete in multicellular lineages showing increasing morphological complexity, as in plants from mosses to modern Angiosperms, the abundance of duplicate functional modules will be much greater than in yeast, and that component duplications will often correspond to known tetraploidy events.

Even with the limitations of this synthesis, the application of balanced gene drive should end the controversy over whether or not there was a trend to increase, and not decrease, eukaryotic morphological complexity over time. Morphological complexity has increased because it was driven to do so, using conventional and experimentally validated molecular mechanisms and obeying accepted selection theory.

Acknowledgments

We thank NSF Plant Genome Project (DBI-0337083) and NSF Bioinformatics (DBI-034937) for funding (M.F.) the work that engendered the ideas of this review and synthesis. We thank the College of Natural Resources, University of California, Berkeley, for its partial subsidy of the Statistics and Bioinformatics Consulting Service.

References

- Adams, K.L. and Wendel, J.F. 2005. Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **8**: 135–141.
- Banks, J.A. 1999. Gametophyte development in ferns. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**: 163–186.
- Birchler, J.A., Bhadra, U., Bhadra, M.P., and Auger, D.L. 2001. Dosage-dependent gene regulation in multicellular eukaryotes: Implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev. Biol.* **234**: 275–288.
- Birchler, J.A., Riddle, N.C., Auger, D.L., and Veitia, R.A. 2005. Dosage balance in gene regulation: Biological implications. *Trends Genet.* **21**: 219–226.
- Blanc, G. and Wolfe, K.H. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**: 1667–1678.
- Blanc, G., Barakat, A., Guyot, R., Cooke, R., and Delseny, M. 2000. Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**: 1093–1101.
- Bolouri, H. and Davidson, E.H. 2002. Modeling DNA sequence-based cis-regulatory gene networks. *Dev. Biol.* **246**: 2–13.
- Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438.
- Buckler, E.S., Phelps-Durr, T.L., Buckler, C.S., Dawe, R.K., Doebley, J.F., and Holtsford, T.P. 1999. Meiotic drive of chromosomal knobs reshaped the maize genome. *Genetics* **153**: 415–426.
- Carles, C.C. and Fletcher, J.C. 2003. Shoot apical meristem maintenance: The art of dynamic balance. *Trends Plant Sci.* **8**: 394–401.
- Carroll, S.B. 2001. Chance and necessity: The evolution of morphological complexity and diversity. *Nature* **409**: 1102–1109.
- Chapman, B.A., Bowers, J.E., Schulze, S.R., and Paterson, A.H. 2004. A comparative phylogenetic approach for dating whole genome duplication events. *Bioinformatics* **20**: 180–185.
- Darwin, C. 1859. *On the origin of the species*, 6th ed. In *Darwin*, 1970 (ed. P. Appleman), p. 674. W.W. Norton, New York.
- Davidson, E., Peterson, K., and Cameron, R. 1995. Origin of bilaterian body plans: Evolution of developmental regulatory mechanisms. *Science* **270**: 1319–1325.
- Dawkins, R. 1986. *The Blind Watchmaker*. Longman, London.
- De Bolt, S., Maere, S., and Van de Peer, Y. 2005. Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* **20**: 591–597.
- Douzery, E.J., Snell, E.A., Bapteste, E., Delsuc, F., and Philippe, H. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci.* **101**: 15386–15391.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.
- Force, A., Cresko, W.A., Pickett, F.B., Proulx, S.R., Amemiya, C., and Lynch, M.M. 2005. The origin of subfunctions and modular gene regulation. *Genetics* **170**: 433–446.
- Freeman, S. and Herron, J.C. 2004. *Evolutionary analysis*. Prentice-Hall, Upper Saddle River, NJ.
- Friis, E.M., Pedersen, K.R., and Crane, P.R. 2005. When Earth started blooming: Insights from the fossil record. *Curr. Opin. Plant Biol.* **8**: 5–12.
- Furlong, R.F. and Holland, P.W.H. 2004. Polyploidy in vertebrate ancestry: Ohno and beyond. *Biol. J. Linn. Soc. Lond.* **82**: 425–430.
- Gao, L.Z. and Innan, H. 2004. Very low gene duplication rate in the yeast genome. *Science* **306**: 1367–1370.
- Garcia-Bellido, A., Ripoll, P., and Morata, G. 1973. Developmental compartmentalization of the wing disc of *Drosophila*. *Nat. New Biol.* **245**: 251–253.
- Garcia-Fernandez, J. 2005. Hox, ParaHox, ProtoHox: Facts and guesses. *Heredity* **94**: 145–152.
- Gaut, B.S. and Doebley, J.F. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci.* **94**: 6809–6814.
- Goldschmidt, R.B. 1953. Evolution as viewed by one geneticist. *American Scientist* **40**: 84–98.
- Gould, S. 1988. Presidential address. Trends as changes in variance: A new slant on progress and directionality in evolution. *J. Paleontol.* **62**: 319–329.
- . 1995. The pattern of life's history. In *The Third Culture: Beyond the Scientific Revolution* (ed. J. Brockman), Chapter 2. Simon and Schuster, New York.
- Gould, S.J. and Lewontin, R.C. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proc. R. Soc. Lond. B Biol. Sci.* **205**: 581–598.
- Gu, Z., Nicolae, D., Lu, H.H., and Li, W.H. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* **18**: 609–613.
- Gu, Z., Rifkin, S.A., White, K.P., and Li, W.H. 2004. Duplicate genes increase gene expression diversity within and between species. *Nat. Genet.* **36**: 577–579.
- Haberer, G., Hindemitt, T., Meyers, B.C., and Mayer, K.F. 2004. Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of *Arabidopsis*. *Plant Physiol.* **136**: 3009–3022.
- Haldane, J.B.S. 1933. The part played by recurrent mutation in evolution. *Am. Nat.* **67**: 5–19.
- Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. 1999. From molecular to modular cell biology. *Nature* **402**: C47–C52.
- He, X. and Zhang, J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**: 1157–1164.
- Hurst, L.D., Pal, C., and Lercher, M.J. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**: 299–310.
- Inada, D.C., Bashir, A., Lee, C., Thomas, B.C., Ko, C., Goff, S.A., and Freeling, M. 2003. Conserved noncoding sequences in the grasses. *Genome Res.* **13**: 2030–2041.
- Irvine, K.D. and Rauskolb, C. 2001. Boundaries in development: Formation and function. *Annu. Rev. Cell Dev. Biol.* **17**: 189–214.
- Jareborg, N., Birney, E., and Durbin, R. 1999. Comparative analysis of noncoding regions of 77 mouse–human gene pairs. *Genome Res.* **9**: 815–824.
- Kellis, M., Birren, B.W., and Lander, E.S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617–624.
- Kellogg, E.A. 2001. Evolutionary history of the grasses. *Plant Physiol.* **125**: 1198–1205.
- Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., et al. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**: R7.
- Ku, H.M., Vision, T., Liu, J., and Tanksley, S.D. 2000. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl. Acad. Sci.* **97**: 9121–9126.
- Lercher, M.J., Urrutia, A.O., and Hurst, L.D. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.* **31**: 180–183.
- Levine, M. and Tjian, R. 2003. Transcription regulation and animal diversity. *Nature* **424**: 147–151.
- Lewis, E.B. 1951. Pseudoallelism and gene evolution. *Cold Spring Harb. Symp. Quant. Biol.* **16**: 159–174.
- Li, H., Pellegrini, M., and Eisenberg, D. 2005. Detection of parallel functional modules by comparative analysis of genome sequences. *Nat. Biotechnol.* **23**: 253–260.
- Lockton, S. and Gaut, B.S. 2005. Plant conserved non-coding sequences and paralogous evolution. *Trends Genet.* **21**: 60–65.
- Lynch, M. and Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.
- . 2003. The origins of genome complexity. *Science* **302**: 1401–1404.
- Lynch, M. and Force, A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473.
- Lyttle, T. 1993. Cheaters sometimes prosper: Distortion of Mendelian segregation by meiotic drive. *Trends Genet.* **9**: 205–210.
- Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and Van de Peer, Y. 2005. Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci.* **102**: 5454–5459.
- Makova, K.D. and Li, W.-H. 2003. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* **13**: 1638–1645.
- Maynard-Smith, J. and Szathmáry, E. 1995. *The major transitions in evolution*. W.H. Freeman, Oxford, UK.

- McLysaght, A., Hokamp, K., and Wolfe, K.H. 2002. Extensive genomic duplication during early chordate evolution. *Nat. Genet.* **31**: 200–204.
- McShea, D. 1996. Metazoan complexity and evolution: Is there a trend? *Evolution Int. J. Org. Evolution* **50**: 477–492.
- . 1998. Possible largest-scale trends in organismal evolution: Eight “live hypotheses.” *Annu. Rev. Ecol. Syst.* **29**: 293–310.
- Morgan, T.H. 1925. *Evolution and genetics*. **154**: 459–473.
- . 1932. *The scientific basis of evolution*. Princeton University Press, Princeton, NJ.
- Ni, M. 2005. Selectionism and neutralism in molecular evolution. *Mol. Biol. Evol.* **22**: 2318–2342.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, Berlin.
- Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.S., Comai, L., Madlung, A., Doerge, R.W., Colot, V., et al. 2003. Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* **19**: 141–147.
- Papp, B., Pal, C., and Hurst, L.D. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**: 194–197.
- Paterson, A.H., Bowers, J.E., Van de Peer, Y., and Vandepoele, K. 2005. Ancient duplication of cereal genomes. *New Phytol.* **165**: 658–661.
- Pereira-Leal, J.B. and Teichmann, S.A. 2005. Novel specificities emerge by stepwise duplication of functional modules. *Genome Res.* **15**: 552–559.
- Pierce, B.A. and Smith, H.M. 1979. Neoteny and paedogenesis. *Journal of Herpetology* **13**: 119–121.
- Prasad, V., Stromberg, C.A.E., Alimohammadian, H., and Sahi, A. 2005. Dinosaur coprolites and the early evolution of grasses and grazers. *Science* **310**: 1177–1180.
- Prince, V.E. and Pickett, F.B. 2002. Splitting pairs: The diverging fates of duplicated genes. *Nat. Rev. Genet.* **3**: 827–837.
- Raes, J. and Van de Peer, Y. 2003. Gene duplication, the evolution of novel gene functions, and detecting functional divergence of duplicates in silico. *Appl. Bioinformatics* **2**: 91–101.
- Ravasz, E., Somera, A.L., Monfru, D.A., Otvai, Z.N., and Barabasi, A.-L. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297**: 1551–1555.
- Rhoades, M. 1942. Preferential segregation in maize. *Genetics* **27**: 395–407.
- Roth, G., Nishikawa, K.C., and Wake, D.B. 1997. Genome size, secondary simplification and the evolution of the brain in salamanders. *Brain Behav. Evol.* **50**: 50–59.
- Schilthuizen, M. 2000. Dualism and conflicts in understanding speciation. *Bioessays* **22**: 1134–1141.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**: 501–506.
- Seoighe, C. and Gehring, C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet.* **20**: 461–464.
- Seoighe, C. and Wolfe, K.H. 1999. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2**: 548–554.
- Simillion, C., Vandepoele, K., Van Montagu, M.C., Zabeau, M., and Van de Peer, Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **99**: 13627–13632.
- Simillion, C., Vandepoele, K., Saeys, Y., and Van de Peer, Y. 2004. Building genomic profiles for uncovering segmental homology in the twilight zone. *Genome Res.* **14**: 1095–1106.
- Taylor, J.S. and Raes, J. 2004. Duplication and divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.* **38**: 615–643.
- Taylor, J., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* **13**: 382–390.
- Tian, C.G., Xiong, Y.Q., Liu, T.Y., Sun, S.H., Chen, L.B., and Chen, M.S. 2005. Evidence for an ancient whole-genome duplication event in rice and other cereals. *Acta Genetica Sinica* **32**: 519–527.
- True, J.R. and Carroll, S.B. 2002. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **18**: 53–80.
- Veitia, R.A. 2002. Exploring the etiology of haploinsufficiency. *Bioessays* **24**: 175–184.
- . 2003. Nonlinear effects in macromolecular assembly and dosage-sensitivity. *J. Theor. Biol.* **220**: 19–25.
- von Mering, C., Zdobnov, E.Z., Tsoka, S., Ciccarelli, F.D., Pereira-Leal, J.B., Ouzounis, C.A., and Bork, P. 2003. Genome evolution reveals biochemical networks and functional modules. *Proc. Natl. Acad. Sci.* **100**: 15428–15433.
- Wagner, A. 2002. Asymmetric functional divergence of duplicate genes in yeast. *Mol. Biol. Evol.* **19**: 1760–1768.
- Wellman, C.H., Osterloff, P.L., and Mohiuddin, U. 2003. Fragments of the earliest land plants. *Nature* **425**: 282–285.
- Wilkins, A.S. 2002. *The evolution of developmental pathways*. Sinauer, Sunderland, MA.
- Williams, E.J. and Bowles, D.J. 2004. Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res.* **14**: 1060–1067.
- Williams, E.J. and Hurst, L.D. 2002. Clustering of tissue-specific genes underlies much of the similarity in rates of protein evolution of linked genes. *J. Mol. Evol.* **54**: 511–518.
- Wolfe, K.H. and Shields, D.C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708–713.
- Yang, J., Lusk, R., and Li, W.-H. 2003. Organismal complexity, protein complexity, and gene duplicability. *Proc. Natl. Acad. Sci.* **100**: 15661–15665.
- Yu, J., Wang, J., Lin, W., Li, S., Li, H., Zhou, J., Ni, P., Dong, W., Hu, S., Zeng, C., et al. 2005. The genomes of *Oryza sativa*: A history of duplications. *PLoS Biol.* **3**: e38.
- Yuh, C.H., Bolouri, H., and Davidson, E.H. 2001. *Cis*-regulatory logic in the *endo16* gene: Switching from a specification to a differentiation mode of control. *Development* **128**: 617–629.