

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Detection and Analysis of Functional Specialization in Duplicated Genes

Owen Z. Woody and Brendan J. McConkey
University of Waterloo
Canada

1. Introduction

Gene duplication has long been recognized as a powerful mechanism facilitating development and evolution in genomes. Duplication events produce additional copies of genomic information, perhaps including one or more genes. While in some cases these duplicated elements may be of immediate benefit (i.e. increasing availability and effective dosage of a desired gene product), often they are initially at least somewhat redundant, and either neutral or mildly detrimental to the fitness of the organism. It is perhaps no surprise, then, that the majority of duplicated genes are quickly deactivated by mutations abolishing transcription or translation. Some duplicated genes, however, survive and persist, suggesting that their retention has some benefit. Many of these genes seem to have acquired properties that distinguish them from their progenitors – they may be expressed in a novel tissue type, for example, or differ in their functional specificity. In these cases, it appears as though duplication has facilitated evolution, either by allowing specialization and refinement or, perhaps most intriguingly, generating genes free to mutate and acquire ‘novel’ functions. These retained duplicates form a family of genes related through common ancestry. As a result of their common origin, gene sequences within gene families are often quite similar, complicating the task of assigning them unique and specific functions. As such, there has been a significant effort to study and characterize the evolution of function in the aftermath of a duplication event. This chapter will briefly cover the various modes of gene duplication, and then will focus on the various functional outcomes of duplication. The theoretical models for functional specialization following a duplication event are discussed, as are practical techniques for applying these models to observed gene duplicates.

2. Mechanisms of duplication

There are several different mutational mechanisms through which gene duplicates can be produced. Depending on the type of event, the nature and scale of what is duplicated can differ significantly; single genes may be copied, with or without their peripheral regulatory elements, or entire genome can be duplicated. While each mechanism ultimately results in the duplication of one or more genes, the mechanisms differ in three key respects; how much regulatory information the duplicated genes retain, where the duplicates are integrated into the genome, and how many interaction partners are duplicated. Duplication

mechanisms can be broadly categorized into three groups – DNA/RNA-mediated transposition, unequal recombination, and genome doubling/hybridization. These mechanisms all produce paralogs -- homologous genes that are both present in and native to the same genome (in contrast to orthologs, where speciation acts as a 'duplication event' and the homologous genes are components of different genomes). Figure 1 provides a diagram depicting various modes of duplication.

2.1 DNA/RNA transposition

DNA/RNA transposition refer to mechanisms by which a specific short nucleotide sequence, either mRNA (as in retrotransposition) or DNA (e.g., transposon-mediated duplication) is copied from one location in the genome to another. The insertion location is essentially random; any compatible destination locus will do, and thus the produced duplicate need not necessarily be located near its progenitor template. RNA-mediated retrotransposition is unique in that it uses post-transcription sequence as a template for the nascent duplicate. Hence, upstream and downstream regulatory sequences lying outside the transcribed gene sequence are not preserved, and the newly produced gene will have most or all introns (and possibly some exons) spliced out. The new gene may also possess a genetically encoded poly-A tail. Since RNA-mediated retrotransposition does not preserve most non-coding elements, the duplicate gene must depend on the a-priori availability or acquisition of promoter/regulatory sequences in order to be transcribed. Absence of these elements effectively means the new gene duplicate is a pseudogene.

DNA-mediated duplications, as mediated by transposons, for example, often retain regulatory information and intron/exon structure. Nonetheless, they still operate on a very specific subsequence of DNA, and elements relocated by DNA-mediated transposition can be inserted in any eligible location in the genome.

2.2 Segmental duplication/unequal cross-over

Errors during homologous recombination can produce serial duplications of genetic sequence. Unequal crossing-over is an error stemming from the mis-alignment of homologous chromosomes during mitosis/meiosis. Ordinarily, homologous sequences are aligned and cross-over events result in balanced exchanges of sequence information across chromosomes. An abundance of repetitive sequences can, however, cause chromosomes to misalign, in which case a segment of one chromosome is inserted into its sister chromatid (thus producing a duplication and a reciprocal deletion).

Since multiple rounds of unequal crossing over tend to gradually inflate the number of candidate repeat regions, some genomic regions are hotbeds for sequence duplication and can give rise to a large number of duplicate genes in series. These serially arranged duplicates are referred to as “tandem duplicates”. These tandem gene arrays are highly localized in the genome, and tandem duplicates retain most or all of their intron/exon structure and peripheral non-coding elements. Unequal crossing-over also plays a role in the generation of copy number variations (Redon et al., 2006).

2.3 Whole genome duplication/allopolyploids

In some circumstances, errors during segregation can produce diploid gametes, and the fusion of these diploid gametes can result in a complete doubling of genomic content (all chromosomes present in duplicate). While very rare, these whole genome duplication

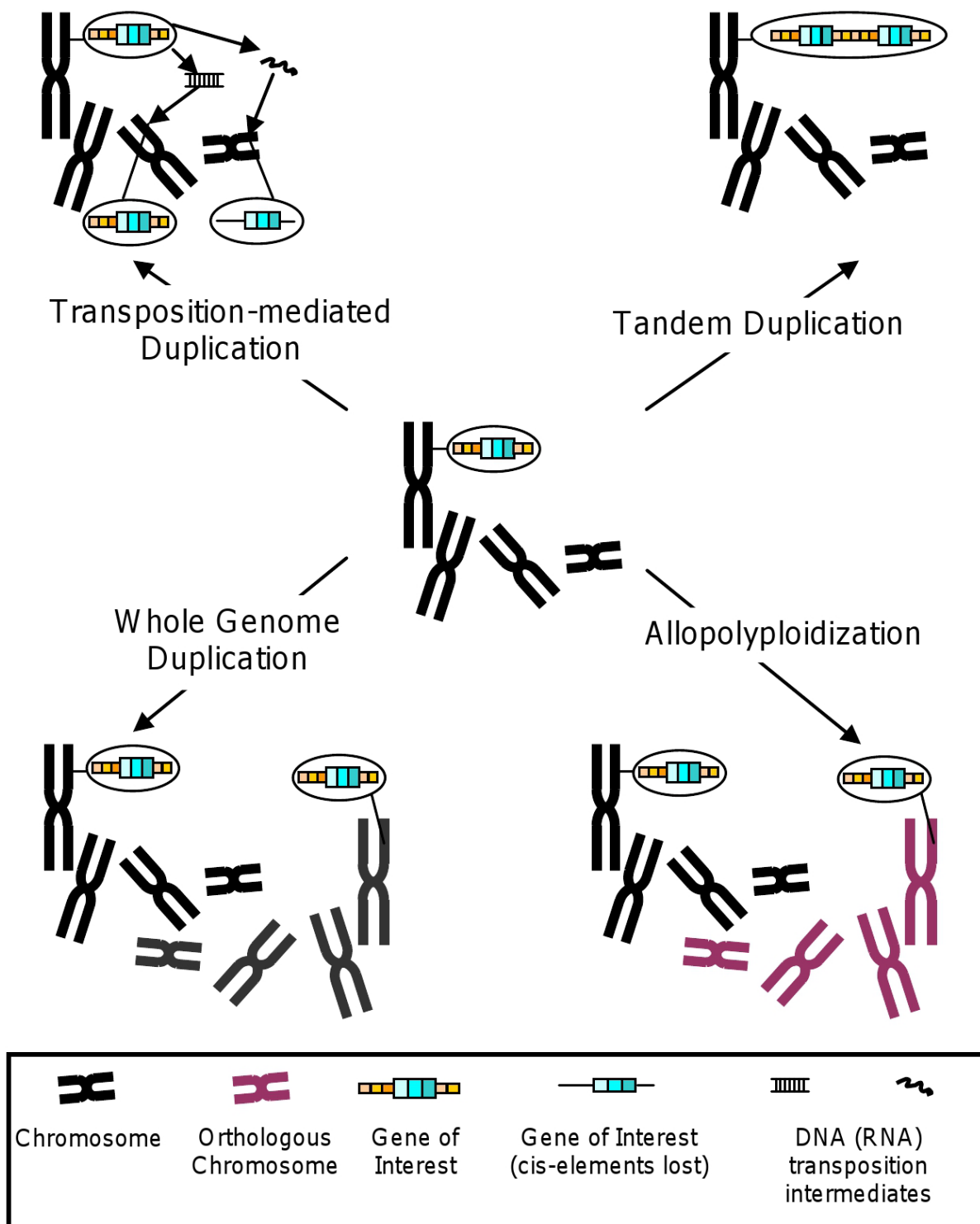


Fig. 1. Modes of gene duplication. Upper left: transposition mediated by either a DNA or RNA intermediate can produce gene locations at distant locations in the genome. RNA intermediates retain little of the regulatory sequence surrounding the parent gene. Upper right: Errors during homologous recombination can produce tandem arrays of genes, situated in series. Bottom Left: Doubling of all chromosomes will produce duplicates of all genes in the genome. Bottom Right: Allopolyploids contain genomes from two compatible species, with duplicate gene pairs from formerly orthologous genes.

(WGD) events have a dramatic impact on the content of the genome, and a number of WGD events have been hypothesized in the history of various lineages (Van de Peer et al., 2009). By their nature, WGD events result in the duplication of all loci, preserving non-coding elements, intron/exon structure, and even overall stoichiometry within gene/protein interaction networks. Interestingly, it has been observed that lineages that undergo separate, distinct WGD events (in this case, *Xenopus tropicalis* and zebrafish) often ultimately retain similar (i.e. orthologous) duplicates – that is to say, WGD duplicates that becomes fixed in one lineage, were also often fixed in the other (Semon & Wolfe, 2008). Pairs of duplicate genes that arose through WGD are sometimes referred to as Ohnologs (Turunen et al., 2009). WGD events are relatively common in plant lineages, which may have interesting implications for the evolution of gene regulation (Lockton & Gaut, 2005).

Allopolyploids are a variant of whole genome duplications in which the diploid gametes come from two different species. These genomic hybrids contain two formerly independent complete genomes. The most commonly studied allopolyploids are plants, though a number of examples have been documented elsewhere in the animal kingdom (including the model organism *Xenopus Laevis*). Duplicates produced through allopolyploidy (i.e. formerly orthologous genes now present in the same organism) are often referred to as “homeologs” (Flagel et al., 2008).

3. Defining gene function

One significant hurdle to the discussion of duplicate functional specialization is defining gene function. Gene function may be broken into two broad categories – regulation and gene product (MacCarthy & Bergman, 2007). Regulation encompasses the “when, where, why, and how much” aspects of a gene’s transcription – non-coding elements around a gene (such as enhancers and signaling sequences) can direct when a gene should be expressed, and in what quantity. These non-coding elements are responsive to various cellular and environmental triggers. Changes to regulation alone may be sufficient to bring about the specialization of a new duplicate.

Gene products, on the other hand, primarily dictate the “how” in a gene’s function (along with some regulatory and subcellular localization information present in the 5' and 3' untranslated regions (UTRs)). Studies of duplicated genes have focused on changes to various coding sequence properties, such as binding sites, eligible cofactors, indels, and catalytic residues (Turunen et al., 2009). It’s theoretically possible for a duplicate gene to become functionally specialized without any change to its regulation (Des Marais & Rausher, 2008).

It is worth noting that changes falling into these two categories can occur serially or in concert. For example, a change in tissue localization may precede structural mutations adapting a protein to a new environment.

4. Theoretical models for duplicate retention and functional specialization

A number of theoretical models have been proposed to describe how a parental gene’s functions can be partitioned between offspring, and how this partitioning affects the chances of these genes to avoid pseudogenization and eventual deletion.

Three archetypal outcomes – specifically, nonfunctionalization, subfunctionalization, and neofunctionalization, are based on concepts typically attributed to Ohno (1970).

Nonfunctionalization describes the situation where one duplicate's expression is abolished, making it invisible to natural selection and thus free to accumulate mutations. While it is technically possible for a nonfunctionalized gene to have its function restored, the vast majority become relics progressively crippled by the accumulation of disabling and deleterious mutations. There has been some interest in studying the impact losing a duplicate via nonfunctionalization has on sibling genes – for example, whole genome duplication events can lead to cases of “ohnologs gone missing”, where a WGD duplicate has been lost (Canestro et al., 2009). Reciprocal duplicate loss has been hypothesized as one means of speciation. Figure 2 depicts two hypothetical duplications and their respective functional specializations.

4.1 Neofunctionalization

Neofunctionalization refers to the scenario where one duplicate gene acquires mutations that allow it to acquire previously unexplored functions, either through changes in regulation (e.g. tissue localization) or coding sequence. Claims of neofunctionalization tend to focus on the generation of new functions, though it should be noted that these developments may also result in the loss of ancestral function(s) (Turunen et al., 2009).

A specific example of neofunctionalization can be found in a recent study of the MADS-box gene family in angiosperms. MADS-box genes are well-known for their role in developmental processes, but the functions of some gene family members have been difficult to determine. Viaene et al. (2010) provide evidence that a group of these genes, the AGL6 subfamily, can be neatly divided into two groups based on duplication history. One of these groups retains the ancestral function of guiding reproductive development, while the other seems to have acquired a novel role in regulating the growth of vegetative tissues.

A second example, describing the functional differentiation of two paralogs in maize, shows that ancestral functions can still be retained even when one duplicate acquires a novel function (Goettel & Messing, 2010). Two paralogs, named p1 and p2, both drive the synthesis of maysin, which in turn contributes to resistance against earworm. In addition, the p1 gene also has a secondary role in controlling the accumulation of red pigments. The authors propose of a series of recombination events that describe how these genes acquired their distinct characters.

4.2 Subfunctionalization

Subfunctionalization involves each gene taking upon a complementary subset of the parental gene's functions, such that neither is independently capable of fulfilling all the parental gene's roles. Subfunctionalization is conceptually synonymous with the Duplication, Degeneration, and Complementation (DDC) model. Regulatory subfunctionalization could result in non-overlapping tissue distributions for the nascent duplicates, with the union of the expression profiles matching the parental gene's range.

Jarinova et al. (2008) describe an instance of subfunctionalization the Hox genes of zebrafish. Through a careful analysis of peripheral non-coding elements, the authors show how the two *hoxb* complexes in zebrafish, *hoxb5a* and *hoxb5b*, acquired non-overlapping expression profiles. In particular, the experimental removal of one regulatory element unique to *hoxb5a* resulted in the two paralogs (re)acquiring a similar expression profile.

The idea of structural subfunctionalization is perhaps best captured in the "Escape from Adaptive Conflict" (EAC) hypothesis. Consider a hypothetical gene product with multiple

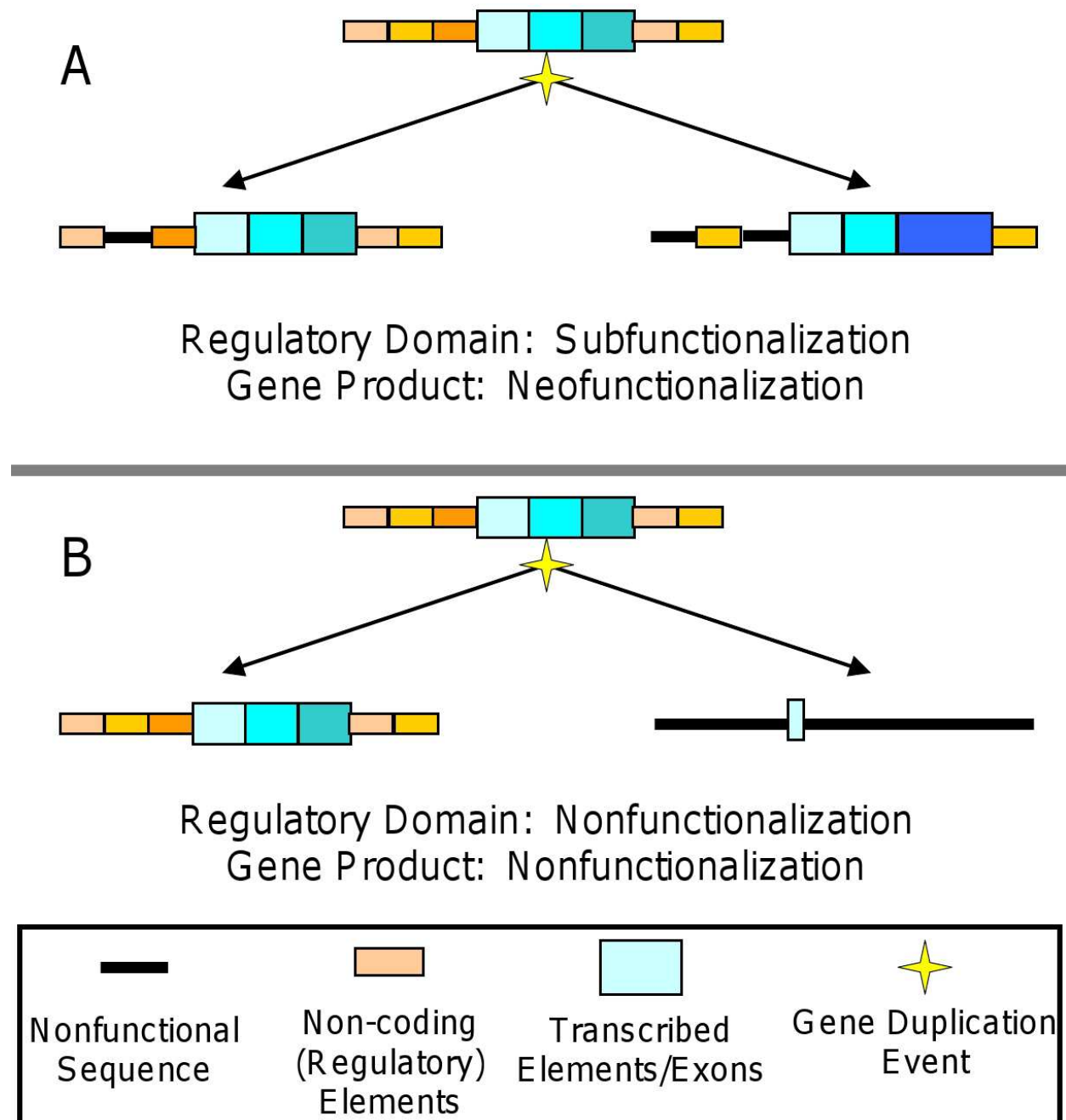


Fig. 2. Possible functional specializations following duplication. Two hypothetical examples showing how retention models can apply either to regulatory regions or gene products. A) Duplicated genes subfunctionalize at the regulatory level, partitioning their parental regulatory domains and suggesting subdivided roles. The gene product, however, has acquired a novel element (i.e. new exon), suggesting neofunctionalization at the coding sequence level. B) Following duplication, one gene loses its regulatory domains and is interrupted by an early stop codon, reflecting nonfunctionalization both at the regulatory and gene product levels.

interaction partners (e.g. an enzyme with two possible substrates). Selection for bifunctionality in this enzyme may limit the binding/catalytic efficiency of either specific reaction -- mutations that improve one may inhibit the other, hence the "adaptive conflict".

Should this gene be duplicated, however, each offspring gene could be free to acquire mutations that optimize binding to one specific substrate, thus escaping the conflict without a loss of functionality. The EAC model essentially describes this process, where a single enzyme with multiple interaction partners gives rise to duplicate genes with more specific but enhanced functionality.

EAC is interesting in that it lies somewhere on the boundary between subfunctionalization and neofunctionalization. Three claims are required to invoke the model: that i) both duplicates accumulate adaptive changes post mutation, that ii) these mutations enhance ancestral functions, and lastly that iii) the ancestral gene was constrained from improving functions (Barkman & Zhang, 2009). The key difference (and challenge) lies in proving the ancestral form was bi-functional. Studies demonstrating the EAC model in action are still relatively uncommon. An early attempt to apply the model to the genes from the anthocyanin biosynthetic pathway of morning glories has come under criticism for not clearly providing these three veins of supporting evidence (Barkman & Zhang, 2009; Des Marais & Rausher, 2008).

The EAC process has also been invoked to describe the evolution of a novel anti-freeze protein in an Antarctic zoarcid fish (Deng et al., 2010). The authors demonstrate that an ancestral gene had a rudimentary ice-binding affinity in addition to its primary catalytic function (a sialic acid synthase), and that a duplication event allowed one copy of this gene to abolish this ancestral function and refine its ice-binding capability. The discussion includes a careful analysis of the three EAC criteria listed above.

While duplicated genes are generally relegated to one of the fates listed above, a number of case studies have shown that recent duplicates can maintain identical functional profiles. One possible explanation for this is that the duplicates have acquired mutations that have restored the “status quo” that was present prior to duplication. If mutations cause the sum of the duplicate genes' expression levels to equal the expression level of their ancestor, both genes could experience some level of selective pressure to maintain expression despite being fully redundant. Ganko et al. (2007) observe that a vast majority of duplicates, regardless of duplication mechanism, showed asymmetric expression, with one gene consistently showing higher levels of expression than its sibling across all tissues. This suggests that a limited form of subfunctionalization may play an initial role in the retention of duplicates. Asymmetrical expression divergence was also observed in a study of duplicated genes in the fly, with a tendency for the “parent” gene of the duplicate to have high expression levels (Langille & Clark, 2007). Interestingly, Qian et al. (2010) point out that many gene duplicates are synthetically lethal or deleterious, and they suggest that expression load may be shared by both genes after duplication.

4.3 Alternative models and odd cases

A number of other subtle variations have been proposed to augment these three primary fates. Subneofunctionalization, for example, is a model that argues that subfunctionalization followed by neofunctionalization is a common and sequential process. Subfunctionalization permits a relaxation of selection on various subregions of the gene, which in turn allows the (eventual) evolution of novel functions, suggesting that subfunctionalization is more of a midstep than an endpoint (Johnson & Thomas, 2007).

In addition, while subfunctionalization does not make any a priori claims about the proportion of functions lost by each duplicate, it appears that in some cases the losses are

highly asymmetrical. Panchin et al. (2010) show that in human duplicate genes one duplicate appears to remain totally unchanged, while its sibling accumulates the majority of functional (in this case, amino acid) mutations.

Contrary to expectations, many gene duplicate pairs appear to be retained despite total apparent functional redundancy. A relatively recent model has been proposed to explain this phenomenon. The theory, coined "originalization", uses arguments based on purifying selection and recombination to support the preservation of both duplicate copies (i.e. prevent non-functionalization) for an extended period of time (Xue & Fu, 2009; Xue et al., 2010).

It has also been suggested that models of duplicate retention are focusing on too small a unit, and that protein interaction networks (themselves composed of a number of co-expressed and functionally related proteins) provide a more coherent perspective on the size of perturbation required to have a phenotypically relevant effect (MacCarthy & Bergman, 2007). The authors argue that cases of regulatory subfunctionalization and neofunctionalization often have no phenotypic consequence on the output of a protein network, and are thus effectively neutral for longer than duplicate-oriented models would suggest. A number of studies have reported an unexplained level and duration of retention for redundant duplicates (Skamnioti et al., 2008).

The relative importance of these models to the retention of duplicates is a subject of continued interest. Whole genome duplication events, which effectively introduce a paralogous copy of every gene in the genome, present an opportunity to tally the cases for which each model applies best. For a review of the relative importance of these various retention models specifically as they pertain to duplicates produced in plant WGD events, see (Edger & Pires, 2009).

5. Factors affecting the rate and trajectory of duplicate gene divergence

5.1 Gene conversion

Gene conversion describes the process by which the sequence content of one genetic locus is used as a template to alter and "paste over" the genetic sequence at a distal location. Gene conversion has the potential to enforce similarity across duplicate loci, both in terms of regulation and structure. A recent study on duplicated segments in a pair of *Drosophila* species made noted several anomalies that were suggestive of gene conversion. Interestingly, the edges of duplicated regions accumulated distinguishing mutations faster than more central regions, suggesting that these regions were being maintained by gene conversion and that the size of the region being converted was gradually being reduced by sequence mutations near the borders. Furthermore, paralogs near the boundaries of duplicated segments showed more divergence than those located near the centre (Osada & Innan, 2008). The authors note that this phenomenon could result in misleading estimates of synonymous divergence, as the conversion process would periodically homogenize the two sequences.

The requirements for a neofunctionalized gene to escape gene conversion and achieve fixation have been studied from a population genetics perspective (Teshima & Innan, 2008). The fit of the model is tested on a pair of human opsins, which differ in their light sensitivity.

Additional evidence that gene conversion may play a role in duplicate divergence was found in a study of WGD duplicates in rice. Duplicates that contained subsequences of

particularly high sequence similarity also showed a tendency towards retaining similar regulation profiles. The authors suggest that this was a consequence of the promoter regions being propagated via gene conversion activity acting on the locally similar subsequences (Yim et al., 2009).

5.2 Alternative splicing

In general, multiple splice forms (and the potential for these splice variants to have distinct functions) have not received much attention in studies of gene duplication and functional divergence. In a first step towards addressing this oversight, Zhan et al. (2011) studied the potential for alternative splicing in *Drosophila* duplicates. New genes tended to show lower levels of alternative splicing, and the subset of duplicates that retained the potential for multiple spliceforms were expressed in fewer tissues, at lower levels, and had had their expression breadth shifted towards preferential expression in testes. The authors also noted that a duplicate's alternative splicing potential depended on duplication mode, with retrotransposed genes being copied with a specific and frozen configuration of exons/introns.

5.3 Properties of genes that influence duplicate specialization

The rate at which duplicated genes acquire novel functions is of great interest. Studies have been done to compare the standard metrics of gene evolution (synonymous distance, non-synonymous distance) to measures of functional differentiation across duplicate genes. While initial studies demonstrated only a weak correlation between expression divergence and sequence divergence, subsequent studies have drawn attention to a number of gene parameters that strongly influence the rate and extent of functional differentiation across duplicates.

The mode of duplication has been cited in multiple instances as an important determinant of eventually retention/functional specialization. In a study comparing the functional evolution of genes duplicated through different mechanisms, Ganko et al. (2007) found that WGD-derived duplicates tended to be expressed at higher levels and were more broadly expressed (in contrast to duplicates derived in smaller scale duplications). Wang et al. (2010) found that tandem gene duplicates tended to have conserved function, whereas retrotransposed genes were more likely to have undergone EAC.

Duplicate genes can differ in their tissue distribution, and certain tissues seem to have a greater propensity to adopt genes with novel function than others. In particular, novel duplicates show a tendency towards expression in the testes. Langille and Clark (2007) found that retrotransposed duplicates in particular showed testis-biased expression. Mikhaylova et al. (2008) also found that duplicated genes expressed in the testes tended to show particularly divergent expression across species. Gallach (2010) illustrate a trend for mitochondrial-associated proteins to preferentially fixate in autosomes (i.e. avoiding the X chromosome), and to have a strong tissue bias towards testes expression.

Han et al. (2009) revealed an interesting trend for duplicate genes that had been relocated (i.e. created by transposition). In instances where these duplicated genes showed asymmetric sequence evolution, in more than 80% of cases it was the relocated gene that showed stronger support for positive selection. This suggests an important role for chromosomal distribution in the evolution of gene function and duplicate divergence. A study by Tsankov et al. (2010) also showed that local chromatin organization (i.e.

nucleosome positions) has a strong effect on gene expression, which suggests that translocated duplicates may show expression divergence by virtue of chromosomal position alone. In addition, Ren et al. (2005) found that tandem duplicates that shared expression domains tended to have dissimilar sequence-based functions. Shoja et al. (2007) noted that tandem gene duplicates tended to show a relationship between expression divergence and chromosomal distance.

In their work on the possible action of gene conversion on the evolution of duplicated segments in *Drosophila*, Osada and Innan (2008) noted that duplications lying near the edges of duplicated segments showed more sequence divergence, suggesting that sublocation within a duplicated segment is an additional factor to consider in studies of duplicate divergence.

The broad functional category to which a gene belongs can also influence its freedom to explore divergent functions. In an analysis of genes in the rice genome produced through a specific WGD, Yim et al. (2009) found that duplicate genes with divergent functions showed a significant enrichment towards metabolism-related activity. Langille and Clark (2007) showed that “cell physiological process” genes were particularly amenable to duplication via transposition. Perhaps reflecting similar functional pressures, Li et al. (2010) found that subcellular localization also influenced the divergence of expression between duplicate genes.

The mode of retention may also depend on the amount of selective pressure acting on its coding sequence. Semon and Wolfe (2008) showed that duplicates undergoing slow rates of sequence evolution seemed particularly prone to regulatory subfunctionalization. This observation is echoed in Arnaiz et al. (2010), who find that duplicate pairs in *Tetraurelia* with divergent expression profiles were unlikely to undergo sequence subfunctionalization. Li et al. (2009) found that the mode of duplication had a substantial effect on the degree of expression divergence between duplicates, based on microarray expression profiles of rice tissues.

Nielsen et al. (2010) suggest that genes under strong selective pressure produce duplicates that are quickly nonfunctionalized, suggesting low tolerance for (poisonous) isoforms of essential products. Thus, a gene’s essentiality and, by consequence, age, may both determine the extent to which gene duplicates may be retained.

6. Tools for measuring gene regulation

While a gene’s regulatory control is partly controlled by its complement of non-coding elements (as well as its genomic location, e.g., proximity to histones/heterochromatin), efforts to predict regulation from sequence alone have met with limited success, owing to non-linear interactions between various regulatory domains (Jarinova et al., 2008). A separate study found that transcription factor binding site turnover was insufficient to explain cis-regulatory evolution across orthologs (Venkataram & Fay, 2010).

Since accurate predictions of gene regulation based on genomic context and peripheral regulatory elements remain elusive, most studies of gene regulation depend on empirical measurements of gene products (i.e. mRNA or protein) as evidence for a gene’s expression under given conditions. Tools for quantifying the abundance of specific mRNA and/or protein species, such as PCR and Western blots, are standard laboratory techniques.

Within the past decade, however, a number of high-throughput technologies have become available that allow the localization and abundance of gene products to be measured empirically on a genomic/proteomic scale. At present, the most widely used platform is the microarray, an assay with a very large number of transcript-specific probes. Each probe is specific to a known transcript, allowing the potential for complete coverage of all known and predicted genes in a known genome sequence. Custom arrays can also be built from cDNA libraries when working with non-model organisms. Databases replete with microarray data are now publicly available for data mining, allowing a gene's expression (or lack thereof) to be profiled across tissues, timepoints, and stimuli. This aggregate gene behaviour is referred to as an "expression profile", and can serve as an empirical proxy of overall gene function. As more microarray data becomes available, the quality of this proxy will improve.

Expression measurement technologies measure gene activation directly, and are agnostic to the regulatory inputs/mechanisms that lead to transcription. In some cases, cis-regulatory regions can undergo substantial changes/shuffling without having much effect on the ultimate transcription behaviour of a gene -- transcription measurement technologies can help distinguish these cases from those that have actually changed a gene's expression phenotype (Comelli & Gonzalez, 2009).

In addition to general purpose (i.e. gene, exon) microarrays, several arrays have been designed to be maximally sensitive to differences between closely related genes. Microarrays use probes that measure targets by hybridizing to nucleotides directly via base complementation. Studies have previously demonstrated that the nucleotides at the center of the probe have the most influence on binding strength. In order to minimize the potential for cross-hybridization, some researchers have designed microarrays for comparing closely related genes (e.g. homeologs) by using probes that feature a known distinguishing SNP at the central position in a probe (Chaudhary et al., 2009; Fligel & Wendel, 2010; Fligel et al., 2008; Udall et al., 2006). This design should minimize cross-hybridization, though it should be noted that previous studies have found that cross-hybridization is only of concern when target sequences are >90-95% identical (Rajashekar et al., 2007). For duplicate genes that have highly similar sequences, alternative measurement technologies like deep sequencing can be used to obtain unbiased paralog-specific expression profiles.

Quantitative proteomics techniques such as iTRAQ (Burkhart et al., 2011) or 2D differential in-gel electrophoresis provide a similarly high-throughput platform for the quantitation of protein abundance. The data differs from microarray data in two respects -- the identities of quantified proteins are often not known in advance, and the coverage of the proteome is not complete and is sensitive to experimental parameters. However, protein abundances may be a more accurate reflection of gene action, as proteins are the active products of genes in most cases and mRNA abundance doesn't always correlate with protein abundance.

Gibson and Goldberg (2009) conducted a study on yeast duplicates using a novel metric of functional differentiation -- number and type of protein interactions. The authors used both affinity-precipitation mass spectrometry and yeast-2-hybrid assays to construct networks of protein interactions, and then sought to test whether the patterns of functional differentiation better fit models of subfunctionalization or neofunctionalization. Their work expands on previous studies that describe the functional evolution of the genome/proteome in terms of the growth of (novel) protein interactions. They illustrate how existing methods

overlook self-self interactions in the parents/progeny, and propose means of avoiding this bias. In general, they found that subfunctionalization was the prevalent driver of protein interaction network evolution.

Recently, sequencing and mass spectrometry have both achieved levels of throughput that make it possible to survey the transcriptome or proteome directly. While these technologies have considerable promise as a source for expression data, at present there are less data available from these platforms (but see Harhay et al., 2010). However, the essential idea of the expression profile holds constant, irrespective of the specific sort (and indeed, mixture) of data that is mined.

7. Models of parental gene function

Since available sequence data is generally restricted to present-day organisms, it is not directly possible to measure a gene's function pre- and post-duplication. As such, when presented with a pair of paralogs, it is not often clear which genes have retained the "ancestral function", if any. In this section, a number of proposed techniques for estimating pre-duplication function are described. These techniques can be broadly broken into two categories - techniques which seek to find an appropriate reference organism elsewhere in nature (i.e. a sister species with a somewhat divergent duplication history), and techniques which attempt to estimate/reconstruct ancestral function from those observed in extant species.

If gene information is available for two closely related species, it is often possible to find a number of examples where a gene duplication event has only occurred in one of the two species. In these cases, paralogs in one genome will correspond to a single ortholog in another. By comparing the functions of paralogs to an unduplicated ortholog, it may be possible to infer which of the two paralogs has undergone more dramatic functional changes. Unfortunately, this approach is restricted to genes present in a 2-to-1 fashion, and even in these cases caution must be taken to ensure the duplication event truly post-dates the speciation event.

One interesting variant of this strategy is to use distantly related members of the same gene family from within the same species (Panchin et al., 2010). Since most genes belong to families with several members, recent duplicates can take advantage of ancient and highly diverged gene family members to serve as a proxy for an orthologous outgroup. This process is useful for calculating rates of sequence evolution between recent duplicates.

Comparisons between lineages which have and have not undergone WGD events can also shed light on the evolution of function post-duplication. For example, Kassahn et al. (2009) used mouse orthologs as reference points for evaluating the post-WGD expression divergence in duplicate genes from five teleost fish species, suggesting that this approach is viable even when the organisms being compared are distant relatives.

Allopolyploids present a unique opportunity for studying gene evolution in the aftermath of widespread duplication. Allopolyploids are hybrids of distinct species, and in many cases the unhybridized lineages have persisted alongside their allopolyploid cousins to the present day. In these cases, the history of gene functional evolution can be inferred by examining gene expression behaviour at four stages: pre-hybridization (the two present day diploid parental strains), day zero hybridization (a cross of the two modern day parentals to

create an “F1” allopolyploid), and post-hybridization (the present day allopolyploid). Thus, the functions of both parental genes can be compared to novel and mature hybrids, revealing the immediate effects upon and eventual trajectory of functional evolution.

The utility of this approach can be seen in Chaudhary et al. (2009), where the functional profiles of homeologous genes could be succinctly depicted as two-component pie charts. The dominance of one genome's homelog over another can be visualized as an unequal partitioning in the pie, and changes to this partitioning following the transition from diploidy to allopolyploidy mark possible instances of functional specialization.

However, in many cases there are no suitable extant orthologs available to serve as models for ancestral gene function. In these cases, there are a number of algorithms for estimating ancestral gene function based directly on the functions of descendent (and other related) genes. Estimation methods can try to infer both gene regulation and gene sequence/structure from present day data.

Microarray-based gene expression profiles have been used in several efforts to estimate ancestral gene function. In a study of stress response genes in *Arabidopsis*, estimates of ancestral gene function were constructed using BayesTraits (Pagel & Meade, 2006), with the present day response profiles used as primary data. For each extant stress response gene, responses to various stresses were coded based on expression level changes (up-regulation, down-regulation, no response). By adjusting the parameters of the Bayestraits program, the authors were able to select a model for gain/loss of response behaviour. This information, when combined with phylogenetic trees mapping out the sequence relationships for each gene family, allowed estimates of the stress response behaviour of ancestral genes (internal nodes on the tree) (Zou et al., 2009).

Another microarray-based approach was explored in Doxey et al. (2007). The study examined the beta-(1,3)-glucanase gene family in *Arabidopsis*, using expression profiles constructed from microarray measurements on tissue and stress response patterns. The expression data for all genes in the family were grouped using hierarchical clustering, such that genes with similar (correlated) expression profiles were grouped together. Based on this clustering, genes were assigned labels according to their functional groups, and these labels were then used as primary data for the reconstruction of ancestral states on the gene family phylogenetic tree via parsimony. Using this approach, the expression profile of ancestral, pre-duplication sequences could be estimated from on the values reconstructed on the tree.

This approach of reconstructing gene functions as characters on a gene phylogenetic tree has a lot of potential, as it allows all members of a gene family to contribute information about the functional breadth explored in a gene family. The exact quantity reconstructed on the tree can vary from simple binary tissue presence/absence (Karanth et al., 2009) to the exact expression abundance as measured by a high-throughput assay (Guo et al., 2007; Li et al., 2005; Oakley et al., 2005).

There have also been efforts to reconstruct ancestral gene sequences, with the hope of reconstructing gene function. Working from the extant variety of fluorescent proteins, Field and Matz (2010) modeled the evolution of fluorescence color in the family by estimating and producing gene sequences at the internal nodes of the fluorescent protein family phylogenetic tree. By producing proteins based on the estimated ancestral sequences, the authors were able to estimate the fluorescence colors of evolutionary intermediates in the family.

8. Making a case for duplicate specialization

The most important techniques for studying functional specialization focus on different aspects of gene function, but are all generally associated with the task of distinguishing the roles and fates of duplicate genes. Figure 3 provides a diagram summarizing the various aspects of gene function that are amenable to these techniques.

8.1 Biochemical function and analysis

Unequivocal evidence for functional specialization can be drawn from studies of enzyme kinetics. By measuring the substrate affinity and catalytic rates of enzymes, for example, it is possible to quantitatively measure differences in performance between duplicate genes. Biochemical approaches are very labor intensive and only readily applicable to certain classes of genes, but the evidence they provide is direct and readily interpreted.

In a study highlighting the potential importance of the EAC model, Des Marais and Rauscher (2008) used enzyme affinity assays to demonstrate the enzymatic function of paralogous anthocyanin biosynthetic pathway genes in morning glories. Enzyme kinetics were compared across different species that differed by a duplication of a specific enzyme, with the unduplicated ortholog acting as a proxy for the ancestral function.

A recent innovative approach to studying gene function used directed (in lab) evolution to try and encourage a derived gene to revert to an hypothesized ancestral function (Bershtein & Tawfik, 2008). The authors studied the rate of 'reversion' and how this rate varied when various degrees of selective pressure (selecting for the ancestral function, the current function, or both) were applied. By studying the transitional states the gene underwent as its function shifted, the authors found evidence that best fit the subfunctionalization model of duplicate gene evolution.

8.2 Expression profiling and reconstruction

Expression profiles (mined from expression assays like high-throughput sequencing) can provide immediate evidence of functional differentiation between duplicated genes (based on divergent, non-overlapping expression behaviour). For example, Yasukawa et al. (2010) use RNA in-situ and reporter analysis to determine precise expression localization and timing of duplicates in *Drosophila*. Rajashekar (2007) used microarray expression profiles to analyze the similarity of duplicates in the hydrophobin gene family in a fungus, *Paxillus involutus*.

By augmenting comparisons of gene expression profiles with reconstructed estimates of parental gene expression (via ancestral character estimation projected onto phylogenetic trees, see section 7), it is further possible to estimate how each gene progeny specialized from its parent following duplication. Case examples are the studies by Doxey et al. (2007) and Zou et al. (2009), both mentioned previously. Zou et al. (2009) reconstructed the expression behaviour in stress response genes in *Arabidopsis*, and with the additional information made available in the estimates of ancestral behaviour, the authors were able to infer patterns of subfunctionalization and neofunctionalization leading to the expression behaviour in extant duplicates. Karanth et al. (2009) reconstructed the ancestral tissue expression patterns of fatty acid binding proteins in zebrafish, revealing

an apparent neofunctionalization event followed by subfunctionalization in a subsequent duplication.

8.3 Comparing with a non-duplicated ortholog

As discussed earlier, one effective technique for estimating the function of the ancestor of a pair of duplicate genes is to refer to a related species where the locus is unduplicated. In this case, the assumption is that the orthologous gene is behaving in the related genome as the parental gene was behaving prior to the duplication event. This point of reference makes it possible to distinguish between models of duplicate retention, lending to support towards subfunctionalization versus neofunctionalization, for example.

In a study of zebrafish-specific WGD-produced duplicates, Kassahn et al. (2009) use unduplicated mouse orthologs as a reference, despite the considerable distance separating these two organisms. Multiple gene properties were compared between paralogs and their mouse ortholog, including sequence, structure, and expression information. The authors found support for neofunctionalization in a number of duplicates, and that regulatory changes were far more common than changes to gene products.

In a study of human genes, Panchin et al. (2010) chose to use distantly related gene family members as proxies for ancestors of recent paralogs. They demonstrated that, in many cases, the recent duplicates are evolving asymmetrically, with one duplicate accumulating sequence mutations much faster than its sibling.

Semon and Wolfe (2008) conducted a study comparing the fate of WGD duplicates in *X.laevis*, an allopolyploid, to *X. tropicalis*, a related species that did not undergo any WGD. Expression patterns were compared across 11 tissue types, and related losses of tissue breadth to possible subfunctionalization. In addition to this, the authors also compared the fate of duplicated genes produced through two different large-scale duplication mechanisms by comparing *X.laevis* to zebrafish, a species with a well studied WGD that did not stem from allopolyploidy. They find that duplicates retained in the *X.laevis* duplication were also frequently retained in duplicate in zebrafish, suggesting common influences on the duplicability of these gene varieties.

Another example of a well-studied allopolyploid, cotton, has been discussed in previous sections (Chaudhary et al., 2009; Flagel et al., 2008; Flagel & Wendel, 2010). One unique observation made possible in this system is the phenomenon of transgressive segregation, where the expression profiles of homeologous genes eventually evolve to resemble neither of the parental strains, suggesting a unique adaptation to the presence of two essentially complete genomes within a single cell (Flagel & Wendel, 2010).

8.4 Comparing gene product properties

While not as easily assayed as gene expression, the transcribed content of genes (i.e. proteins) can also suggest the gain and loss of functions. As a simple example, the rate of protein sequence evolution can be compared between duplicates by comparing their respective rates of synonymous and non-synonymous mutation. While not necessarily illustrative of the nature of the difference, this method can provide evidence for asymmetrical selection, suggesting one duplicate is acquiring amino acid altering mutations faster than the other (Ganko et al., 2007). Working from a list of 15 of the most asymmetrically diverged WGD-derived protein sequences in *S. cerevisiae*, Turunen et al. (2009) noted substantial indels in addition to changes in important catalytic residues and

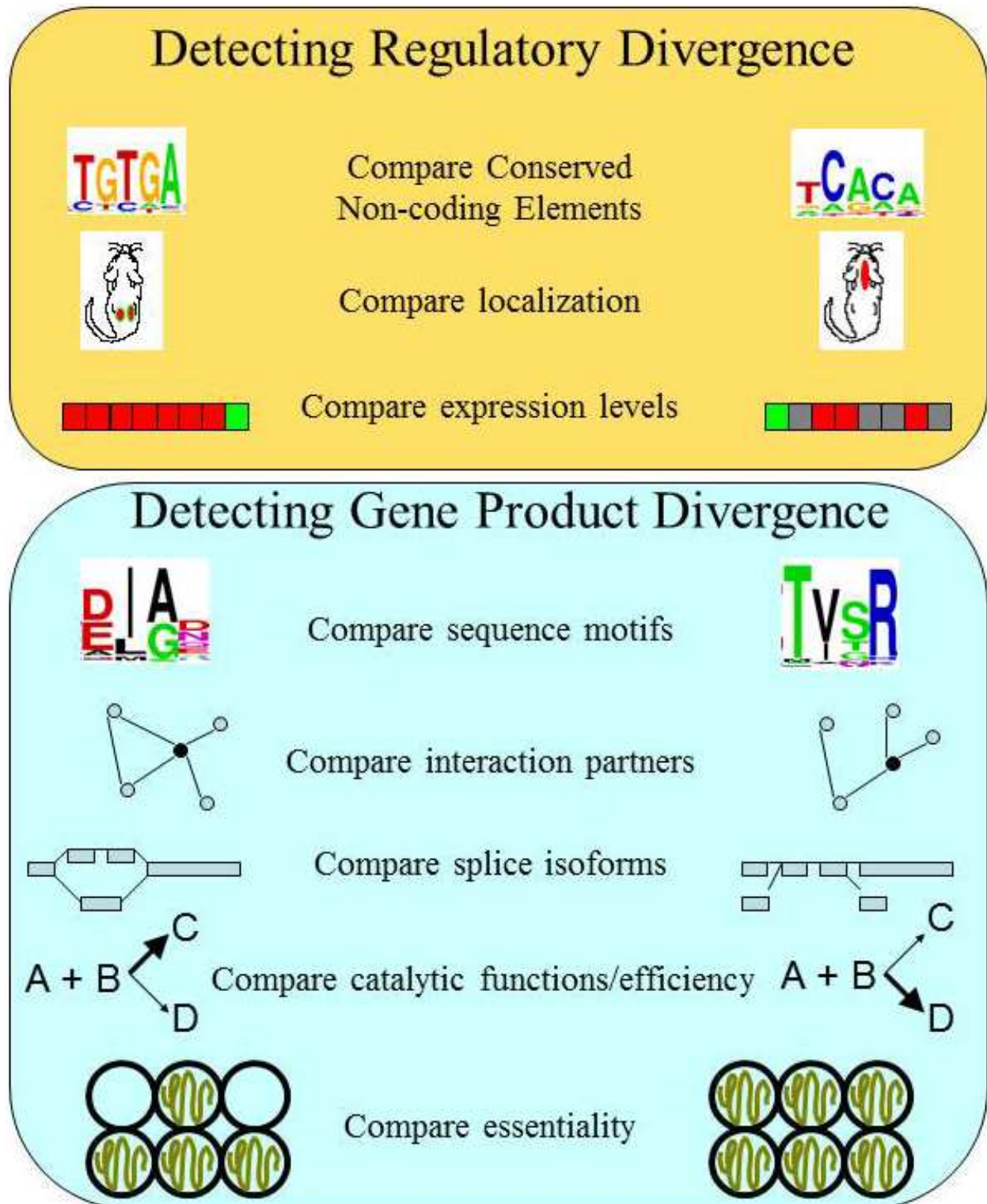


Fig. 3. Gene properties that can be examined for evidence of functional specialization. The top set (orange) are approaches that check for differences in gene regulation; expression levels reflect measurements of transcription in tissues in response to a series of stresses (e.g. as obtained from microarrays). The bottom set (blue) are aspects of the gene product that may differ between duplicates. Sequence logos may be generated using the WebLogo software (Crooks et al., 2004).

active/cofactor binding sites. A literature search seemed to support that many of these highly modified duplicates had acquired novel functions.

Other aspects of gene function, such as the position and number of introns or methylation sites, have also been used to characterize divergence between duplicated genes. For example, Xiong et al. (2010) include intron position in a study of the expansion of the ABC transporter gene family in the ciliate *Tetrahymena thermophila*. In addition to comparing the expression profiles constructed by clustering gene expression data, the authors also compare intron positions to group the family members into functional subcategories (Xiong et al., 2010). A similar study has examined differential splicing forms in duplicate genes of *Drosophila* (Zhan et al., 2011).

Yang et al. (2006) compared the “DNA-binding with one finger” (DOF) gene family across three plant species – rice, *Arabidopsis thaliana*, and poplar. Their multifaceted approach to describing gene function included an analysis of protein motif gain/loss and changes to methylation patterns. Combined with information about gene regulation drawn from microarrays, PCR and massively parallel signature sequencing, the authors compared the relative applicability of various duplicate retention models to the DOF family.

When the information is available, the protein-protein interaction partners of duplicates can also be compared to study duplicate specialization. Nielsen et al. (2010) compared a set of residues in the tail ends of tubulin genes in fruit flies, noting divergence in these regions which may reflect changes in protein-protein interaction partners. Studying the applicability of models like subfunctionalization and neofunctionalization at the level of gene networks has helped integrate duplicate specialization into a broader systems biology context (Gibson & Goldberg, 2009; MacCarthy & Bergman, 2007).

9. Conclusion

Studies of the evolution of duplicate genes are pushing the field towards more exacting standards and definitions for gene function. Since the rate and extent of duplicate gene specialization is dependent on so many factors, and since novel functions can emerge in so many different ways, integrative approaches will be of paramount importance to understanding this key aspect of genomic evolution. Future studies can benefit in particular from the inclusion of data from gene families as a whole, as this additional information helps both with estimating ancestral gene functions and with evaluating the breadth of function previously covered by related genes.

While empirical evidence of differential catalytic function remains the gold standard for proving functional specialization of duplicated genes, high-throughput studies exploiting the vast quantities of minable expression data provide a cheap and effective means for studying functional specialization at the level of whole gene families. Genomes with annotations beyond expression profiles (such as gene-by-gene interaction profiles and essentiality data) should be helpful for determining the extent to which functional changes at the regulatory level actually impact phenotype.

10. References

Arnaiz, O., Gout, J. F., Betermier, M., Bouhouche, K., Cohen, J., Duret, L., Kapusta, A., Meyer, E. & Sperling, L. (2010). Gene expression in a paleopolyploid: a

- transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics*, Vol. 11, No. 547
- Barkman, T. & Zhang, J. (2009). Evidence for escape from adaptive conflict? *Nature*, Vol. 462, No. 7274
- Bershtein, S. & Tawfik, D. S. (2008). Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. *Mol. Biol. Evol.*, Vol. 25, No. 11, pp. 2311-2318
- Burkhardt, J. M., Vaudel, M., Zahedi, R. P., Martens, L. & Sickmann, A. (2011). iTRAQ protein quantification: A quality-controlled workflow. *Proteomics*, Vol. 11, No. 6, pp. 1125-1134
- Canestro, C., Catchen, J. M., Rodriguez-Mari, A., Yokoi, H. & Postlethwait, J. H. (2009). Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet.*, Vol. 5, No. 5
- Chaudhary, B., Flagel, L., Stupar, R. M., Udall, J. A., Verma, N., Springer, N. M. & Wendel, J. F. (2009). Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics*, Vol. 182, No. 2, pp. 503-517
- Comelli, R. N. & Gonzalez, D. H. (2009). Divergent regulatory mechanisms in the response of respiratory chain component genes to carbohydrates suggests a model for gene evolution after duplication. *Plant. Signal. Behav.*, Vol. 4, No. 12, pp. 1179-1181
- Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.*, Vol. 14, No. 6, pp. 1188-1190
- Deng, C., Cheng, C. H., Ye, H., He, X. & Chen, L. (2010). Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. *Proc. Natl Acad. Sci. U.S.A.*, Vol. 107, No. 50, pp. 21593-21598
- Des Marais, D. L. & Rausher, M. D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature*, Vol. 454, No. 7205, pp. 762-765
- Doxey, A. C., Yaish, M. W., Moffatt, B. A., Griffith, M. & McConkey, B. J. (2007). Functional divergence in the Arabidopsis beta-1,3-glucanase gene family inferred by phylogenetic reconstruction of expression states. *Mol. Biol. Evol.*, Vol. 24, No. 4, pp. 1045-1055
- Edger, P. P. & Pires, J. C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.*, Vol. 17, No. 5, pp. 699-717
- Field, S. F. & Matz, M. V. (2010). Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals. *Mol. Biol. Evol.*, Vol. 27, No. 2, pp. 225-233
- Flagel, L., Udall, J., Nettleton, D. & Wendel, J. (2008). Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol.*, Vol. 6, No. 16
- Flagel, L. E. & Wendel, J. F. (2010). Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.*, Vol. 186, No. 1, pp. 184-193

- Gallach, M., Chandrasekaran, C. & Betran, E. (2010). Analyses of nuclearly encoded mitochondrial genes suggest gene duplication as a mechanism for resolving intralocus sexually antagonistic conflict in *Drosophila*. *Genome Biol. Evol.*, Vol. 2, pp. 835-850
- Ganko, E. W., Meyers, B. C. & Vision, T. J. (2007). Divergence in expression between duplicated genes in *Arabidopsis*. *Mol. Biol. Evol.*, Vol. 24, No. 10, pp. 2298-2309
- Gibson, T. A. & Goldberg, D. S. (2009). Questioning the ubiquity of neofunctionalization. *PLoS Comput. Biol.*, Vol. 5, No. 1
- Goettel, W. & Messing, J. (2010). Divergence of gene regulation through chromosomal rearrangements. *BMC Genomics*, Vol. 11, No. 678
- Guo, H., Weiss, R. E., Gu, X. & Suchard, M. A. (2007). Time squared: repeated measures on phylogenies. *Mol. Biol. Evol.*, Vol. 24, No. 2, pp. 352-362
- Han, M. V., Demuth, J. P., McGrath, C. L., Casola, C. & Hahn, M. W. (2009). Adaptive evolution of young gene duplicates in mammals. *Genome Res.*, Vol. 19, No. 5, pp. 859-867
- Harhay, G. P., Smith, T. P., Alexander, L. J., Haudenschild, C. D., Keele, J. W., Matukumalli, L. K., Schroeder, S. G., Van Tassell, C. P., Gresham, C. R., Bridges, S. M., Burgess, S. C. & Sonstegard, T. S. (2010). An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. *Genome Biol.*, Vol. 11, No. 10
- Jarinova, O., Hatch, G., Poitras, L., Prudhomme, C., Grzyb, M., Aubin, J., Berube-Simard, F. A., Jeannotte, L. & Ekker, M. (2008). Functional resolution of duplicated *hoxb5* genes in teleosts. *Development*, Vol. 135, No. 21, pp. 3543-3553
- Johnson, D. A. & Thomas, M. A. (2007). The monosaccharide transporter gene family in *Arabidopsis* and rice: a history of duplications, adaptive evolution, and functional divergence. *Mol. Biol. Evol.*, Vol. 24, No. 11, pp. 2412-2423
- Karant, S., Denovan-Wright, E. M., Thisse, C., Thisse, B. & Wright, J. M. (2009). Tandem duplication of the *fabp1b* gene and subsequent divergence of the tissue-specific distribution of *fabp1b.1* and *fabp1b.2* transcripts in zebrafish (*Danio rerio*). *Genome*, Vol. 52, No. 12, pp. 985-992
- Kassahn, K. S., Dang, V. T., Wilkins, S. J., Perkins, A. C. & Ragan, M. A. (2009). Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res.*, Vol. 19, No. 8, pp. 1404-1418
- Langille, M. G. & Clark, D. V. (2007). Parent genes of retrotransposition-generated gene duplicates in *Drosophila melanogaster* have distinct expression profiles. *Genomics*, Vol. 90, No. 3, pp. 334-343
- Li, Q., Liu, X., He, Q., Hu, L., Ling, Y., Wu, Y., Yang, X. & Yu, L. (2011). Systematic analysis of gene expression level with tissue-specificity, function and protein subcellular localization in human transcriptome. *Mol. Biol. Rep.*, Vol. 38, No. 4, pp. 2597-2602.
- Li, Z., Zhang, H., Ge, S., Gu, X., Gao, G. & Luo, J. (2009). Expression pattern divergence of duplicated genes in rice. *BMC Bioinformatics*, Vol. 10 Suppl 6
- Li, Z., Liu, Q., Song, M., Zheng, Y., Nan, P., Cao, Y., Chen, G., Li, Y. & Zhong, Y. (2005). Detecting correlation between sequence and expression divergences

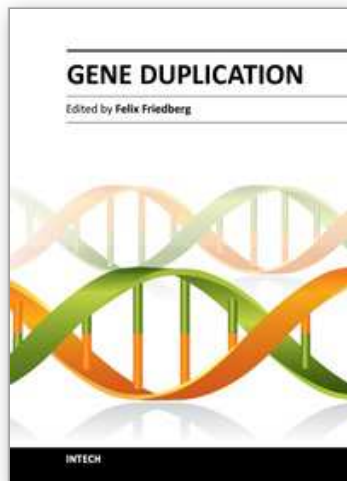
- in a comparative analysis of human serpin genes. *BioSystems*, Vol. 82, No. 3, pp. 226-234
- Lockton, S. & Gaut, B. S. (2005). Plant conserved non-coding sequences and paralogue evolution. *Trends Genet.*, Vol. 21, No. 1, pp. 60-65
- MacCarthy, T. & Bergman, A. (2007). The limits of subfunctionalization. *BMC Evol.Biol.*, Vol. 7, No. 213
- Mikhaylova, L. M., Nguyen, K. & Nurminsky, D. I. (2008). Analysis of the *Drosophila melanogaster* testes transcriptome reveals coordinate regulation of paralogous genes. *Genetics*, Vol. 179, No. 1, pp. 305-315
- Nielsen, M. G., Gadagkar, S. R. & Gutzwiller, L. (2010). Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. *BMC Evol.Biol.*, Vol. 10, No. 113
- Oakley, T. H., Gu, Z., Abouheif, E., Patel, N. H. & Li, W. H. (2005). Comparative methods for the analysis of gene-expression evolution: an example using yeast functional genomic data. *Mol.Biol.Evol.*, Vol. 22, No. 1, pp. 40-50
- Ohno, S. (1970). *Evolution by Gene Duplication*, Springer-Verlag, 0-04-575015-7, New York
- Osada, N. & Innan, H. (2008). Duplication and gene conversion in the *Drosophila melanogaster* genome. *PLoS Genet.*, Vol. 4, No. 12
- Pagel, M. & Meade, A. (2006). Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. *Am. Nat.*, Vol. 167, No. 6
- Panchin, A. Y., Gelfand, M. S., Ramensky, V. E. & Artamonova, I. I. (2010). Asymmetric and non-uniform evolution of recently duplicated human genes. *Biol. Direct*, Vol. 5
- Qian, W., Liao, B. Y., Chang, A. Y. & Zhang, J. (2010). Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet.*, Vol. 26, No. 10, pp. 425-430
- Rajashekar, B., Samson, P., Johansson, T. & Tunlid, A. (2007). Evolution of nucleotide sequences and expression patterns of hydrophobin genes in the ectomycorrhizal fungus *Paxillus involutus*. *New Phytol.*, Vol. 174, No. 2, pp. 399-411
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. *Nature*, Vol. 444, No. 7118, pp. 444-454
- Ren, X. Y., Fiers, M. W., Stiekema, W. J. & Nap, J. P. (2005). Local coexpression domains of two to four genes in the genome of *Arabidopsis*. *Plant Physiol.*, Vol. 138, No. 2, pp. 923-934
- Semon, M. & Wolfe, K. H. (2008). Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. *Proc. Natl Acad. Sci. U.S.A.*, Vol. 105, No. 24, pp. 8333-8338
- Shoja, V., Murali, T. M. & Zhang, L. (2007). Expression divergence of tandemly arrayed genes in human and mouse. *Comp. Funct. Genomics*, 60964
- Skamnioti, P., Furlong, R. F. & Gurr, S. J. (2008). The fate of gene duplicates in the genomes of fungal pathogens. *Commun. Integr. Biol.*, Vol. 1, No. 2, pp. 196-198

- Teshima, K. M. & Innan, H. (2008). Neofunctionalization of duplicated genes under the pressure of gene conversion. *Genetics*, Vol. 178, No. 3, pp. 1385-1398
- Tsankov, A. M., Thompson, D. A., Socha, A., Regev, A. & Rando, O. J. (2010). The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol.*, Vol. 8, No. 7
- Turunen, O., Seelke, R. & Macosko, J. (2009). In silico evidence for functional specialization after genome duplication in yeast. *FEMS Yeast Res.*, Vol. 9, No. 1, pp. 16-31
- Udall, J. A., Swanson, J. M., Nettleton, D., Percifield, R. J. & Wendel, J. F. (2006). A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics*, Vol. 173, No. 3, pp. 1823-1827
- Van de Peer, Y., Maere, S. & Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.*, Vol. 10, No. 10, pp. 725-732
- Venkataram, S. & Fay, J. C. (2010). Is transcription factor binding site turnover a sufficient explanation for cis-regulatory sequence divergence? *Genome Biol. Evol.*, Vol. 2, pp. 851-858
- Viaene, T., Vekemans, D., Becker, A., Melzer, S. & Geuten, K. (2010). Expression divergence of the AGL6 MADS domain transcription factor lineage after a core eudicot duplication suggests functional diversification. *BMC Plant. Biol.*, Vol. 10, No. 148
- Wang, Z., Dong, X., Ding, G. & Li, Y. (2010). Comparing the retention mechanisms of tandem duplicates and retrogenes in human and mouse genomes. *Genet. Sel. Evol.*, Vol. 42, pp. 24
- Xiong, J., Feng, L., Yuan, D., Fu, C. & Miao, W. (2010). Genome-wide identification and evolution of ATP-binding cassette transporters in the ciliate *Tetrahymena thermophila*: A case of functional divergence in a multigene family. *BMC Evol. Biol.*, Vol. 10, No. 330
- Xue, C., Huang, R., Liu, S. Q. & Fu, Y. X. (2010). Recombination facilitates neofunctionalization of duplicate genes via originalization. *BMC Genet.*, Vol. 11, No. 46
- Xue, C. & Fu, Y. (2009). Preservation of duplicate genes by originalization. *Genetica*, Vol. 136, No. 1, pp. 69-78
- Yang, X., Tuskan, G. A. & Cheng, M. Z. (2006). Divergence of the Dof gene families in poplar, *Arabidopsis*, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiol.*, Vol. 142, No. 3, pp. 820-830
- Yasukawa, J., Tomioka, S., Aigaki, T. & Matsuo, T. (2010). Evolution of expression patterns of two odorant-binding protein genes, Obp57d and Obp57e, in *Drosophila*. *Gene*, Vol. 467, No. 1-2, pp. 25-34
- Yim, W. C., Lee, B. M. & Jang, C. S. (2009). Expression diversity and evolutionary dynamics of rice duplicate genes. *Mol. Genet. Genomics*, Vol. 281, No. 5, pp. 483-493, 1617-4623
- Zhan, Z., Ren, J., Zhang, Y., Zhao, R., Yang, S. & Wang, W. (2011). Evolution of alternative splicing in newly evolved genes of *Drosophila*. *Gene*, Vol. 470, No. 1-2, pp. 1-6

Zou, C., Lehti-Shiu, M. D., Thomashow, M. & Shiu, S. H. (2009). Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*. *PLoS Genet.*, Vol. 5, No. 7, e1000581

IntechOpen

IntechOpen



Gene Duplication

Edited by Prof. Felix Friedberg

ISBN 978-953-307-387-3

Hard cover, 400 pages

Publisher InTech

Published online 17, October, 2011

Published in print edition October, 2011

The book Gene Duplication consists of 21 chapters divided in 3 parts: General Aspects, A Look at Some Gene Families and Examining Bundles of Genes. The importance of the study of Gene Duplication stems from the realization that the dynamic process of duplication is the "sine qua non" underlying the evolution of all living matter. Genes may be altered before or after the duplication process thereby undergoing neofunctionalization, thus creating in time new organisms which populate the Earth.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Owen Z. Woody and Brendan J. McConkey (2011). Detection and Analysis of Functional Specialization in Duplicated Genes, Gene Duplication, Prof. Felix Friedberg (Ed.), ISBN: 978-953-307-387-3, InTech, Available from: <http://www.intechopen.com/books/gene-duplication/detection-and-analysis-of-functional-specialization-in-duplicated-genes>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen