CellPress

Opinion

Paralogue Interference Affects the Dynamics after Gene Duplication

Elisabeth Kaltenegger[1,*] and Dietrich Ober[1]

Proteins tend to form homomeric complexes of identical subunits, which act as functional units. By definition, the subunits are encoded from a single genetic locus. When such a gene is duplicated, the gene products are suggested initially to cross-interact when coexpressed, thus resulting in the phenomenon of paralogue interference. In this opinion article, we explore how paralogue interference can shape the fate of a duplicated gene. One important outcome is a prolonged time window in which both copies remain under selection increasing the chance to accumulate mutations and to develop new properties. Thereby, paralogue interference can mediate the coevolution of duplicates and here we illustrate the potential of this phenomenon in light of recent new studies.

Trends

The biologically active units of proteins are often homomeric complexes of two or more subunits, for example, transcription factors binding DNA as dimers.

After duplication of genes that encode such self-interacting proteins, the gene products of both duplicates are able to cross-interact and form 'paralogous heteromers'.

A physical and functional link in the paralogous heteromer mediates a dominant-negative effect of detrimental mutations and thereby contributes to shape the fate of the duplicates, a phenomenon called 'paralogue interference'.

The consequences of paralogue interference on the fate of duplicated genes are diverse and further studies will reveal if this phenomenon plays a vital role in the period while a duplicated gene struggles to survive.

Birth and Evolutionary Fate of Duplicates

Gene duplication is considered a prominent mechanism that provides potentially redundant genetic material that is less constrained by selection [1–5]. Based on analyses of genomic databases of several eukaryotic species, the average rate by which new gene duplicates arise was estimated to be in the order of 0.01 per gene per million years [6]. Thus, the frequency of gene duplications is of the same order of magnitude as the rate of mutation per nucleotide site [6]. Gene duplicates are estimated to account for greater than 60% of the genes in eukaryotic genomes [7,8]. Starting from Ohno's classic model published in 1970 [9], diverse models have been suggested to predict the fate of a duplicated gene and to describe the underlying mechanism (Box 1, reviewed in [2]). The three most discussed outcomes for duplicated genes are: (i) the additional copy gains a new function (neofunctionalisation); (ii) after duplication of an at least bifunctional ancestor gene, the functions are distributed to its descendants, each of which performs a subset of the ancestral gene function (subfunctionalisation); and (iii) the additional gene copy accumulates degenerative mutations and is lost over time (nonfunctionalisation). However, when a protein-encoding gene is duplicated, one prominent feature of proteins is largely neglected in most gene duplication models, that is, the tendency of proteins to form symmetrical homomers, which is a widespread phenomenon [10–15].

Homomeric proteins (see Glossary) can be divided into two subgroups: (i) **facultative oligomers**: the physical interactions in the oligomer are nonobligate for function [16]; and (ii) **obligate oligomers**: oligomerisation is a structural need for function as it is the case if either the active site of a homomeric enzyme requires residues from adjacent subunits ('shared active site') or interaction of subunits is important for conformation stabilisation [17]. Further examples include multimeric ion channel proteins and transcription factors that bind DNA only in their multimeric form. In the following section, we will focus on duplications of genes that encode obligate homomeric proteins and discuss the possible effects of this organisation on the fate of duplicates.

[1]Department of Biochemical Ecology and Molecular Evolution, Botanical Institute, Christian-Albrechts-University, 24098 Kiel, Germany

*Correspondence: ekaltenegger@bot.uni-kiel.de (E. Kaltenegger).

 http://dx.doi.org/10.1016/j.tplants.2015.10.003

Box 1. Gene Duplication Models

Duplication–Degeneration–Complementation
This model is one of the first subfunctionalisation models [52]. It hypothesises that paralogues are initially functionally redundant. Through neutral mutation processes, the copies will accumulate degenerative mutations that will affect the function of each, with the consequence that the ancestral function is split to the paralogues. Selective constraints immediately after the duplication are predicted to be relaxed until gene function is divided so that both copies must be maintained by purifying selection.

Escape from Adaptive Conflict
According to this subfunctionalisation model, duplication resolves a biochemical conflict of a multifunctional protein [53]. It assumes that optimisation of one of the functions in the ancestral protein would have detrimental effects on the other functions. Via gene duplication, the functions are split and can be individually optimised. In this scenario, both duplicates are predicted to bear the signature of positive selection as a sign of functional adaptation.

Neofunctionalisation
This model postulates that one paralogue will be selected to preserve the ancestral function, while the other is free to evolve a new function [9]. Thus, one paralogue will continue to evolve under purifying selection, while the other will experience relaxed selective constraints until its recruitment to a new function. Positive selection will indicate the period of adaption to the new function.

Innovation–Amplification–Divergence
This model assumes an ancestral gene that encodes a protein with various side activities. In a changing environment, a side activity may become valuable [54]. Selection will then favour its increase, which can be conferred by duplication of the parent gene. In the duplicated gene, the side activity will be optimised by selective forces. This model predicts like the neofunctionalisation model purifying selection in the copy that maintains the ancestral function, but positive selection in the second copy immediately after the duplication event.

Paralogue Interference
After duplication, the two copies are functionally linked when they encode proteins whose active biological unit is a homooligomer. The gene products of the paralogues will interact if coexpressed and assemble into paralogous heteromers. These heteromers provide the functional link between the copies. Paralogue interference predicts that both copies inherit the selective constraints of their ancestor. Immediately after the duplication, either coevolution of the paralogues will be favoured and positive selection will optimise the paralogues, while the complex maintains its overall function, or purifying selection will prevent change of the individual components of the complex. In the latter case, interference between paralogues can be escaped, for example, via regulatory divergence.

Duplication of Homomeric Enzymes

Homomeric protein organisation is predicted to affect duplicates in a direct way: assuming the whole gene including its regulatory elements is duplicated, the two paralogous copies are identical immediately after the duplication event. The gene products should be coexpressed and be able to cross-interact and to form **paralogous heteromers** (Figure 1A) [18]. Because of this interaction, the two copies are physically linked. If the protein is only active in the oligomeric state, the two copies are also functionally linked. Degenerative mutations in only one copy can 'poison' the protein complex and affect the ancestral function [19]. For instance, if a gene encoding an obligate dimeric protein is duplicated, a deleterious mutation in one of the paralogue genes should reduce the activity not only of the mutated homomer (A′A′) but also of the paralogous heteromer (AA′). This principle is similar to the 'dominant-negative effect between heterozygous alleles of multimeric proteins' suggested by Herskowitz [20] (Box 2). As a consequence of this effect, mutations in only one paralogue might not be phenotypically silent and the duplicates not as independent as suggested in most gene duplication models. On the contrary, both copies inherit the selective constraints of their ancestor and can mutually influence each other, as long as they can form paralogous heteromers. The situation is different if a facultative oligomer is duplicated. In this case, the duplicates are physically linked in the paralogous heteromer but are expected to not functionally interfere with each other. Indeed, the multimeric state of such facultative oligomers was described to vary stochastically on an evolutionary time scale [15].

Paralogous Interference over Time

After duplication of obligate oligomers, the capability of forming multimeric complexes will be maintained in at least one copy to guarantee the ancestral function. However, what about the

Glossary

Duplication-resistant genes: genes that are convergently restored to single copy status after multiple genome-wide and smaller-scale duplication events.
Facultative oligomer: proteins that do not require oligomerisation for activity but that assemble into oligomeric complexes. Examples are enzymes that bear the active site within a single subunit.
Homomeric proteins: proteins that are encoded by a single gene locus and that self-interact with each other and form oligomeric complexes composed of identical subunits.
Isoenzymes: show identical catalytic properties but are encoded at different loci in the genome. Gene duplications are the most common mechanism through which isoenzymes arise.
Obligate oligomer: oligomeric proteins that require the formation of an oligomeric complex for their activity. Examples are enzymes of which the active site is formed within the interface of the interacting subunits.
Oligomeric proteins: proteins composed of two or more associated subunits (= polypeptide chains).
Paralogues (or paralogous genes): genes that result from a duplication from a common ancestor gene.
Paralogous heteromer: the gene products of two paralogues interact and form an oligomeric protein, with the paralogue proteins being subunits of the oligomeric complex.
Positive selection: selection for an advantageous allele (mutations) in populations.
Purifying selection (negative selection): a selection regime that reduces the frequency of deleterious alleles (mutations) in populations.
ω (or d_N/d_S or K_a/K_s) ratio: this ratio can be used to analyse selective pressure in protein-coding sequences. It is calculated as the ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site. Synonymous (silent) mutations are largely invisible to natural selection, whereas nonsynonymous (amino acid-altering) can experience strong selective pressure if functionally relevant. Thus, the fixation rates of these two types