Opinion TRENDS in Ecology and Evolution Vol.22 No.7

Protein subcellular relocalization: a new perspective on the origin of novel genes

S. Ashley Byun-McKay[1] and R. Geeta[2]

[1] Fairfield University, Department of Biology, Fairfield, CT 06824-5195, USA
[2] SUNY Stony Brook, Department of Ecology and Evolution, 640 Life Sciences Building, Stony Brook, NY 11794, USA

Gene duplication is considered to be the most important evolutionary process for generating novel genes. However, the mechanisms involved in the evolution of such genetic innovations remain unclear. There is compelling evidence to suggest that changing the subcellular location of a protein can also alter its function, and that diversity in subcellular targeting within gene families is common. Here, we introduce the idea that protein subcellular relocalization might be an important evolutionary mechanism for the origins of new genes.

Gene duplication: the making of a gene

For the past three decades, gene duplication has been considered to be the most important evolutionary process for the origin of new genes. By producing copies of preexisting genes, duplication provides the raw material that evolutionary processes can act upon to generate genes with novel functions [1–5]. With the recent availability of several sequenced genomes, such as thale cress *Arabidopsis thaliana*, humans *Homo sapiens* and the protozoan *Trichomonas vaginalis*, biologists have begun to think about gene duplication and the processes involved with retention and subsequent functional diversification with renewed vigor (e.g. Refs [6–15]). However, many of the key evolutionary events that ultimately enable the proteins produced by duplicate genes to acquire novel functions still remain unclear [16].

Analyses of diverse genomes suggest that gene duplication is a widespread feature in all three domains of life, Bacteria, Archaea and Eukarya (reviewed in Ref. [17]). Duplication can occur through a variety of processes, including unequal crossing over between homologous chromosomes, strand breakage and rejoining of non-homologous ends, and retroposition (see Glossary), the products of which can be complete or incomplete, tandem or non-tandem duplicates (Figure 1). Gene duplication can also occur through whole-chromosome or whole-genome duplications resulting from non-disjunction or polyploidization (reviewed in Ref. [18]). Here, we outline the evidence that subcellular relocalization of duplicate proteins might facilitate functional diversification and consequently precede the evolution of new genes following their duplication.

Corresponding author: Byun-McKay, S.A. (sbyun@mail.fairfield.edu).
Available online 15 May 2007.

The origin of new genes with novel functions

The origin of novel genes by duplication traditionally postulates that new genes arise as a consequence of a gradual accumulation of mutations. Because duplicates were conventionally thought to be functionally redundant, it was assumed that one of the duplicates was potentially dispensable and, as such, less hindered by the selective constraints that would normally limit functional exploration. Once a duplicate gene acquired a new advantageous function, it would be retained by selection. This process is commonly known as classic neofunctionalization (Box 1) [19]. Unfortunately, classic neofunctionalization cannot satisfactorily explain how a duplicate gene can escape the load of deleterious mutations that would probably accumulate before enough beneficial mutations could confer a new function. Although models such as subfunctionalization [20], epigenetic complementation [21] and dosage compensation (Box 2) [10,22] explain how duplicate genes might initially be retained in the genome, they generally do not include a clear mechanistic process that explains how duplicates evolve novel biochemical

Glossary

5′ end nucleotide sequence of a gene: part of the gene that is translated into the N terminal amino acid sequence of a protein.
C-terminal motifs: common elements found at the C terminus of a protein that can influence protein subcellular localization.
Cytosol: the fluid portion of the cytoplasm (cytoplasm minus the organelles).
Gene family: sets of genes that have originated through duplication from a single gene ancestor.
Non-tandem duplication: results in duplicates that are not in adjacent positions.
Novel (gene) function: definition is often subjective and contextual; here, we define novel gene function as a new biochemical function, rather than the change in the expression pattern of a gene.
N-terminal target peptide (NTP): short N-terminal amino acid sequence of a protein that directs it to a specific subcellular location Recent work suggests that the first 20 amino acids of the N terminus are particularly important predictors of protein subcellular location.
Positive (directional) selection: occurs when an allele has a greater fitness than others in a population, resulting in an increase in its frequency and, ultimately, its fixation.
Purifying (stabilizing or negative) selection: occurs when an allele is less fit than others in a population, resulting in a decrease in its frequency and, ultimately, its elimination.
Retroposition: the process in which DNA is transcribed into mRNA, which is then reversed transcribed into DNA and randomly inserted into the genome. The products of retroposition are duplicate genes that lack introns.
Subcellular: pertaining to membrane-bound compartments within the cell.
Tandem duplication: results in duplicates that are adjacent to one another.

  doi:10.1016/j.tree.2007.05.002

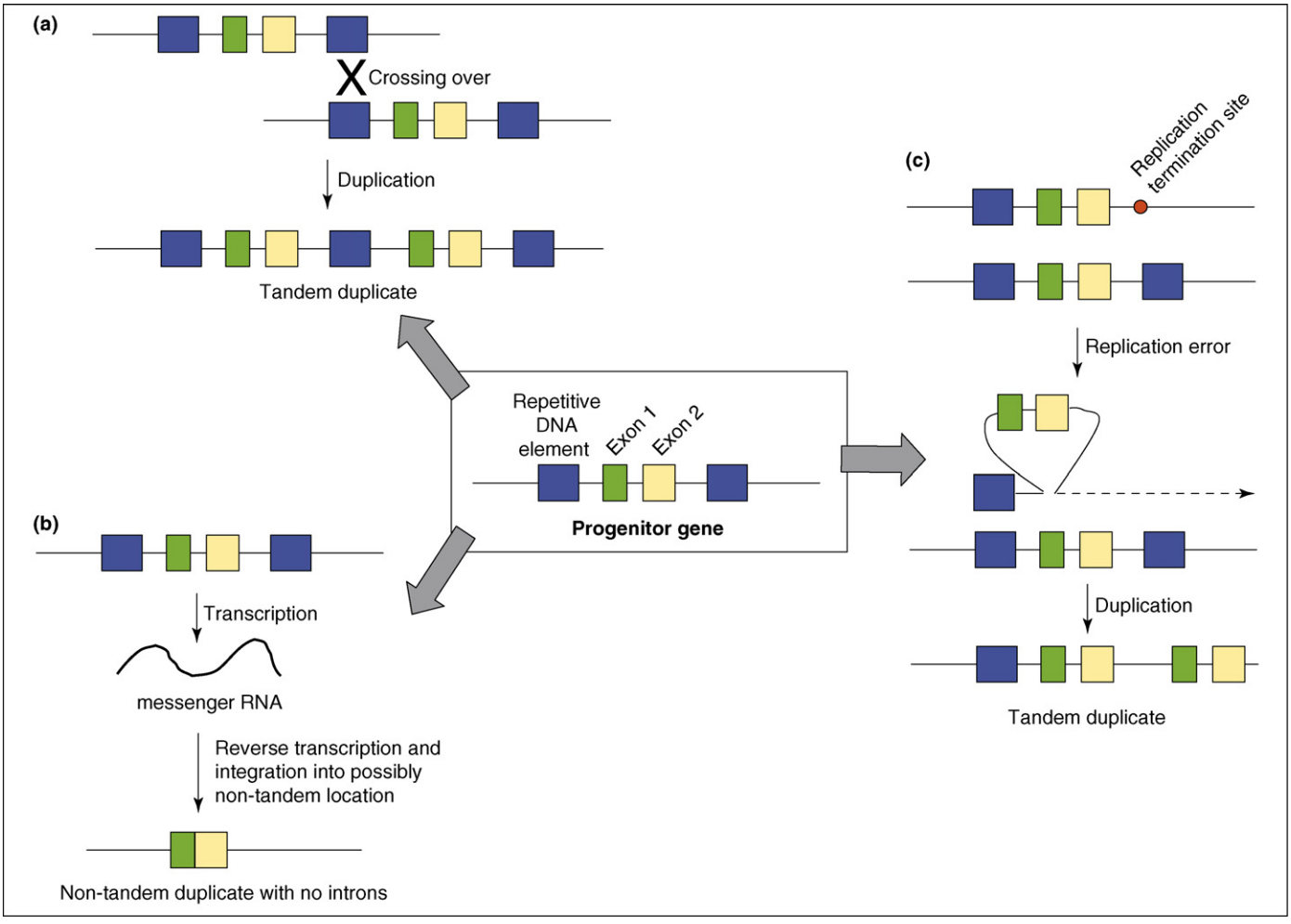

Figure 1. Three possible mechanisms of gene duplication: **(a)** unequal crossing over between homologous chromosomes; **(b)** non-homologous recombination through strand breakage and end joining; and **(c)** retroposition. Unequal crossing over often occurs in regions comprising repetitive DNA, resulting in mispairing between homologous strands and subsequent recombination; the product is a tandem duplication on one strand. Duplication can also occur between non-homologous strands. Replication-dependent breakage points can produce tandem duplications through strand breakage and recombination, resulting in a tandem duplication. Retroposition occurs when reverse transcribed RNAs are randomly integrated into the genome. Products of retroposition are duplicates that lack introns. Adapted with permission from Ref. [18].

functions, other than invoking diversification through the accumulation of point mutations.

Recently, more attention has been directed towards identifying and understanding the mechanics involved in the evolution of new genes and new gene functions [16,22,23]. One process suggests that the products of duplication, which can include imperfect or incomplete copies, have new functions at inception because they are not identical to their progenitor (Box 1) [23]. Although feasible, such a process cannot explain how duplication events that produce identical gene copies can ultimately result in different novel functions. A second process, the adaptive radiation model (Box 1), is based upon the premise that gene duplication can result in many gene copies. If these copies have some degree of multifunctionality and, as such, some measure of preadapation for different functions, then as a result of successive rounds of mutation and competition, only the copy with the highest level of functionality will be retained in the genome. The others would be removed by selection [16]. However, because adaptive radiation requires bursts of duplication, it is less applicable to smaller scale gene duplications.

Here, we propose that protein subcellular relocalization (PSR) could lead to the origin of new genes following perfect or imperfect, large- or small-scale duplications, and yet obviate the need for substantial accumulation of point mutations. We suggest that altering the N-terminal target peptide (NTP) and thus changing the subcellular location of a duplicate protein, might be one evolutionary process that can lead to new protein functions and, ultimately, the origin of novel genes.

Missing the target is easy to do

As many as half the different proteins synthesized within the cytosol of a typical eukaryotic cell are delivered to the cell membrane, secreted, or targeted to various membrane-bound organelles. Protein subcellular localization can involve several different processes, such as those that utilize nuclear localization signals, nuclear export signals and specific C-terminal motifs. Directing proteins to membrane-bound organelles, such as the endoplasmic reticulum (ER), mitochondria and chloroplasts is largely determined by short NTPs (including sequences such as the signal peptide), which are typically 13–36 amino acids

Box 1. Models of functional diversification mechanisms

The following are models or mechanisms mentioned here that could account for functional diversification. Once a new beneficial function evolves, it would be retained by selection.

Classic neofunctionalization
Following gene duplication, functional redundancy enables one of the duplicates to accumulate mutations. In rare cases, these mutations might be beneficial, resulting in the evolution of gene function [19]. A classic example of neofunctionalization is duplication of the *RNAase1* gene in the colobine monkey, which has resulted in two genes, *RNAse1a* and *RNAse1b*. Nine amino acid substitutions in the mature RNAse1b protein altered its optimal pH, enabling these monkeys to use leaves as a primary food source instead of the fruit and insects commonly used by most other monkeys (Ref. [1] and references therein).

Partial and chimeric duplications
The partial and chimeric duplications model is based upon observations from *Caenorhabditis elegans* that ~50% of all duplications do not produce copies that are identical to their progenitor. These imperfect duplications could be partial or chimeric (a mixture of introns and exons from two or more different genes). Because these products of duplication are not identical, they could have different or novel functions [23].

Neofunctionalization via adaptive radiation
Neofunctionalization via adaptive radiation assumes that a gene is multifunctional and has some measure of preadaptation for these different functions. If this gene is duplicated such that many copies exist, each of these copies can improve these various functions through point mutations, competing against one another so that most duplicates are lost through selection. Through the process of adaptive radiation, gene duplication can result in new genes with novel functions [16].

PSR
According to PSR, the products of gene duplication could acquire new functions and ultimately evolve into new genes by relocalizing their proteins to different locations within the cell. Although subcellular localization depends upon several different factors, we focus on the NTP. PSR via alteration of the NTP could occur either through an initial duplication error, which alters the 5′ end of the gene, or through point mutations.

Box 2. Models of duplicate gene retention

The following are models mentioned here that could account for duplicate gene retention. Once retained, novel functions could arise through mutations accumulating in the coding region of the gene.

Subfunctionalization
After gene duplication, functional redundancy enables deleterious mutations to accumulate in the regulatory elements of duplicate genes, altering their spatial and/or temporal expression pattern such that the original gene function becomes subdivided between the two duplicates. This is also referred to as the DDC (duplication, degeneration, complementation) model [20]. For example, in mouse and chicken, the single *eng1* is expressed in the hindbrain, neurons and pectoral appendages; however, in zebrafish, duplication of *eng1* has resulted in *eng1a*, which is expressed in the hindbrain and neurons, and *eng1b*, which is expressed in the pectoral appendage bud [20].

Epigenetic complementation
In epigenetic complementation, the specific positioning of a duplicate gene can result in epigenetic inactivation, which can be caused by methylation (addition of methyl groups to cytosine and occasionally adenosine residues), heterochromatinization, or homologous RNAi-mediated silencing. Owing to the specific localization of the duplicate within the genome, epigenetic complementation can change the temporal and/or the expression pattern of a gene (Ref. [17] and references therein).

Dosage compensation
Dosage compensation is based on the premise that duplicate genes are not necessarily neutral upon inception. Increased protein levels as a consequence of duplication can confer a selective advantage [11,22]. There are many examples of dosage compensation, including the duplication of genes in response to toxins, pesticides and antibiotics (Ref. [11] and references therein).

long. Proteins that do not have a NTP sequence remain within the cytosol as a default. After directing proteins to their appropriate place, NTPs are typically cleaved off, degraded and, thus, do not usually participate directly in protein function (reviewed in Ref. [24]; Ref. [25] and references therein) (Figure 2).

NTPs typically exhibit little sequence identity, as demonstrated by the numerous alternatives that can direct proteins to the same subcellular location [26,27]. For example, Kaiser *et al.* [26] showed that 20% of short random human genomic DNA sequences could all direct an invertase protein of *Saccharomyces cerevisiae* to the same subcellular location, the ER. Another illustration of NTP promiscuity is dual targeting, where the same NTP of a single protein can be directed to different subcellular targets depending upon the current state of the cell. The final location of the protein is determined by the outcome of subcellular competition (reviewed in Ref. [28]).

Despite the lack of a consensus sequence, the destination of a NTP can be predicted based on primary sequence features, such as the low abundance of acidic residues and the presence of α-helices. As such, a change in the subcellular destination of a NTP does not necessarily require a substantial change in sequence. In fact, a single amino acid substitution is potentially sufficient to redirect proteins to a new location. Because an amino acid change can be caused by a single nucleotide mutation, changing the subcellular target of a duplicate protein could occur through one base substitution (Table 1). Given that NTPs are estimated to evolve at least twice as fast as the rest of the protein [27], altering the subcellular target of a duplicate protein by mutations in the NTP might be expected to occur relatively rapidly, enabling a protein to acquire a new subcellular expression pattern and, possibly, a new biochemical function in a relatively short amount of time.

Table 1. Putative N terminal sequences and their subcellular targets

N terminal sequence[a]	Putative target[b]	Possible codon[c]
MSSSTLLHLLELSSLLFSCLPNAKPQQAED	Cytosol	E (GAG)
MSSSTLLHLL**L**LSSLLFSCLPNAKPQQAED	Endoplasmic reticulum	L (CUC)
MSSSTLLHLL**R**LSSLLFSCLPNAKPQQAED	Mitochondria	R (C**G**C)
MSSSTLLHLL**H**LSSLLFSCLPNAKPQQAED	Chloroplast	H (C**A**C)

[a]A single amino acid substitution (red) is sufficient to send a protein to different subcellular locations. In some cases, this substitution might result from a single nucleotide transition G to A (green).
[b]Prediction based on iPSORT.
[c]E, glutamic acid; L, leucine; R, arginine; H, histidine.

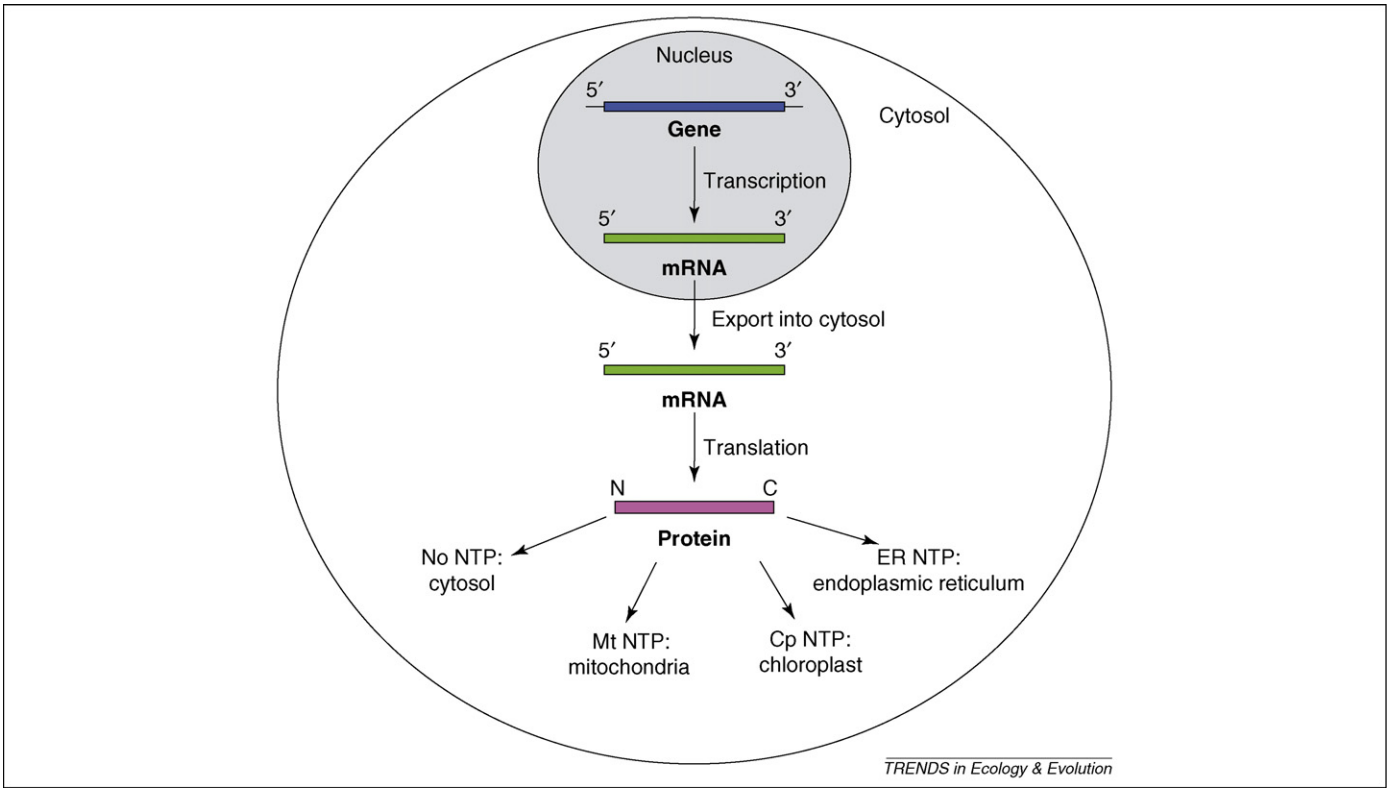

Figure 2. NTP protein subcellular localization. A gene is transcribed into messenger RNA (mRNA) within the nucleus. The mRNA is exported into the cytosol, where it is translated into a protein with N and C termini. The NTP is located at the N terminus; however, if there is no NTP, the protein remains within the cytosol. If the NTP sequence represents an Mt NTP, the protein is directed to the mitochondria. Alternatively, the NTP could be a chloroplast (Cp) NTP or endoplasmic reticulum (ER) NTP, which would direct the protein to the chloroplast or endoplasmic reticulum, respectively. Once the protein arrives in its correct subcellular location, the NTP is usually cleaved off and degraded.

Such changes could also occur immediately upon duplication if the duplication itself produces a partial or chimeric duplicate at the 5′ end of the gene. For example, the protein acetyl-CoA carboxylase (ACCase) from *Brassica napus* is organized into two groups, class I and class II. Class II is represented by a single ACCase gene (**Y10301**) that is likely to be localized to the cytosol based on sequence similarity to the *Arabidopsis* ACCase protein. Class I includes two ACCase genes (**X77576** and **Y10302**), both of which are characterized by an additional exon at the 5′ end that is not found in the Class II ACCase gene. The addition of the N-terminal exon in class I ACCase appears to result in a NTP that targets the protein to the chloroplast [29].

Protein function and subcellular location

An increasing amount of biochemical data suggest that protein function is strongly influenced by subcellular location and that altered protein targeting can have unpredictable effects on the role of a protein within the cell (e.g. Refs [30–34]). For example, insulin-degrading enzyme (IDE) is a widely expressed zinc-metallopeptidase that is largely localized to the cytosol. Its primary function is to regulate levels of cerebral amyloid β-peptide and plasma insulin and it is suspected of having a significant role in late-onset Alzheimer's disease and type 2 diabetes, respectively (Ref. [35] and references therein). An experimentally produced IDE targeted to the mitochondria was shown to degrade cleaved peptides, an activity that is unknown to the cytosolic form [35].

In some cases, a protein retargeted to a new subcellular location might change its overall protein efficiency, rather than its enzymatic function *per se*. For example, methylmercury lyase is a bacterial protein that can help bind and reduce the toxic effects of methylmercury, a serious environmental contaminant that accumulates in aquatic food chains (Ref. [36] and references therein). Introduction of bacterial mercury resistance genes such as *merB* into *Arabidopsis* resulted in enhanced mercury resistance (10–70 times higher specific activity) when the MerB protein was targeted to the ER instead of the cytoplasm. The conclusion was that the hydrophobic environment of the vesicular structures that MerB appeared to accumulate in provided better reaction conditions for their specific activity than did the conditions found in the cytoplasm [36].

Changes in protein function are typically associated with point mutations. For example, each of the members of the *Arabidopsis* desaturase gene family (ADS) catalyze the desaturation of fatty acids by inserting double bonds in specific locations along the fatty acid chain. Altering the placement of these double bonds would probably require at least two–six key mutations within the mature ADS enzyme [37–39]. However, it is possible to achieve similar results by subcellular relocalization. By experimentally relocalizing ADS members, ADS1 (**At1g06080**) and 2 (**At2g31360**) from the cytoplasm to the chloroplast and ADS3 (**At3g15850**) from the chloroplast to the cytoplasm, a functional change occurred in which these enzymes now catalyzed desaturation by inserting double bonds at

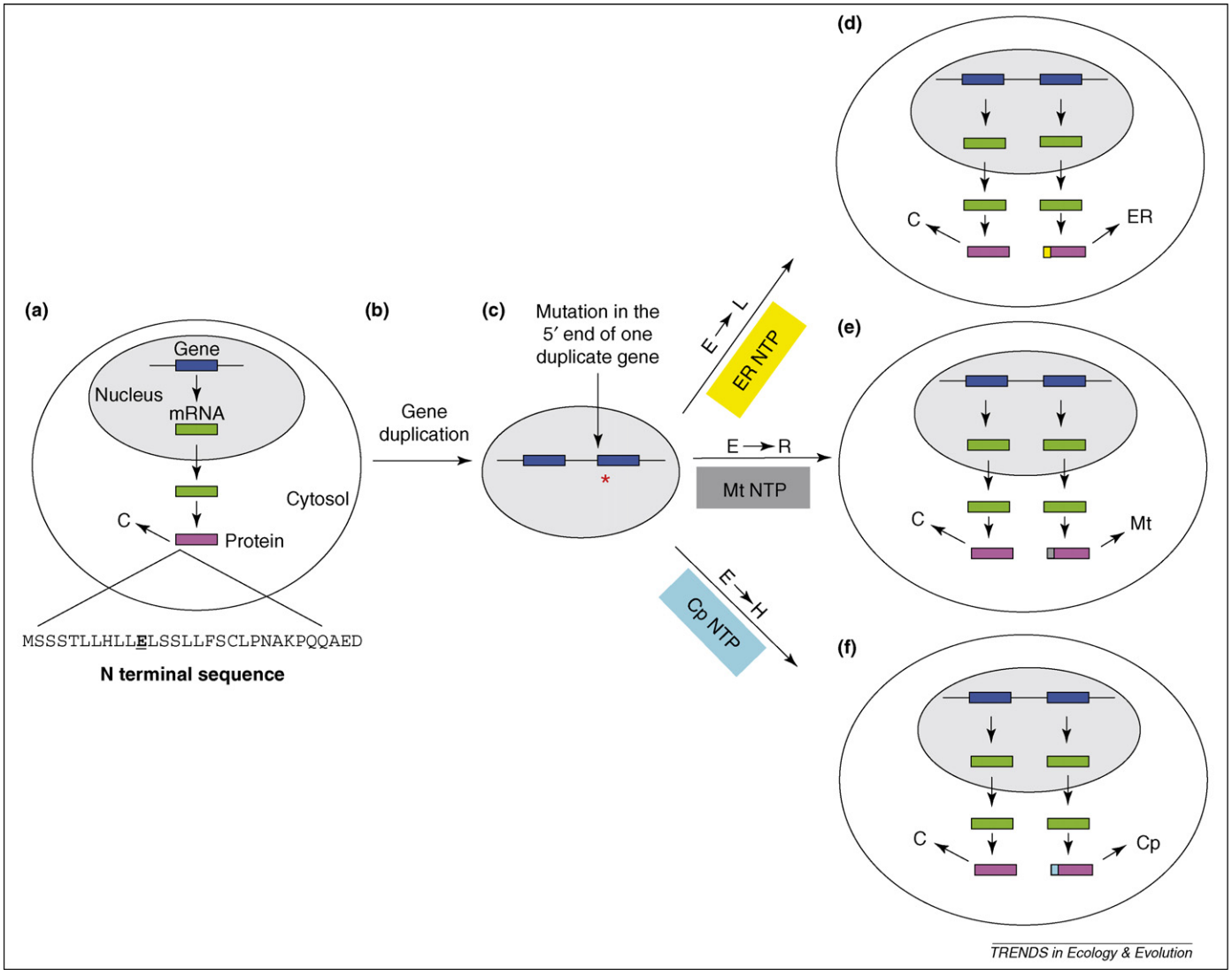

Figure 3. Origin of novel gene function by PSR. **(a)** A single gene (blue) produces a protein (pink) without a NTP, which therefore remains within the cytosol (C). The gene is then duplicated **(b)**. As a consequence of either imperfect duplication or subsequent mutation(s) in the 5′ end of the duplicate gene **(c)**, the resulting protein acquires a NTP and no longer remains within the cytosol; instead, it is targeted to the **(d)** endoplasmic reticulum (ER), **(e)** mitochondria (Mt) or **(f)** the chloroplast (Cp), depending upon the sequence features of the NTP (see Table 1, main text). If the duplicate protein has a negative or neutral effect on fitness, the probable fate of its gene will be pseudogenization or complete loss from the population. Neutral duplicates however, could also become fixed owing to drift. It is also possible that these duplicates could be relocalized by a subsequent mutation in the NTP. In some cases, the protein might acquire a new function in its new subcellular environment.

different positions in the fatty acid [37]. In this case, changing the subcellular location of the ADS enzymes sufficiently altered the surrounding biochemical milieu, enabling it to modify its function even in the absence of any mutations in the catalytic site [37].

Altered subcellular targeting within gene families

We suggested earlier that a duplicate protein could be redirected to a different subcellular location with relative ease. If relocalizing a duplicate protein via its NTP is relatively simple, then we should expect members of diverse gene families to show variability in subcellular locations. Recently, a genome-wide analysis of *Arabidopsis* demonstrated that at least 239 gene families have members that are potentially targeted to various subcellular locations [37]. Although most of these plant gene families are involved in secondary metabolism, variability in subcellular targeting was also found in gene families involved in other processes, such as transcription. Because transcription occurs in the nucleus, targeting transcription factors to other subcellular locations might seem surprising. However, transcription factors, similar to other proteins, can have different functions depending upon their cellular location. For example, some plant transcription factors, such as LEAFY, are capable of intercellular movement when localized to the cytoplasm and might be involved in cell–cell signaling [40,41].

Empirical observations have documented varied subcellular localization within many different gene families (e.g. Refs [42–46]). For example, phospholipase D (PLD) is a gene family found in yeast, plants and animals. The PLD proteins are believed to have a role in a variety of cellular processes, such as membrane trafficking, signal transduction and cell fate determination [45]. Within mammals, PLD has been empirically determined to occur in the plasma membrane, secretory granules, golgi apparatus,

ER, nucleus, mitochondria, caveolae, cytoskeleton and cytosol. Within each of these locations, PLD activity is stimulated by, or is dependent upon, different factors, including, but not limited to, oleate, Ca^{2+}, the G protein ADP–ribosylation factor, and phosphatidylinositol-4,5-bisphosphate (PIP_2), suggesting different physiological roles for PLD activity in each of these subcellular locations [45].

Evidence of altered subcellular targeting has also been observed in other eukaryotes. In a recent survey of the green alga, *Chlamydomonas,* FK506 binding proteins (FKBP) and cyclophilin, two gene families that function in protein folding, have been empirically and theoretically shown to have members that are directed to subcellular locations such as the nucleus, mitochondria, chloroplasts, and cytosol, as well as the ER [46], again illustrating how common altered subcellular targeting within gene families and across different eukaryotic kingdoms appears to be.

PSR: a novel mechanism for the evolution of new genes

Given observations that the biochemical function and efficiency of a protein is influenced by its subcellular location, the lack of a NTP consensus sequence, the apparent ease with which its destination can be changed, and the widespread diversity of subcellular locations within gene families, we propose that PSR might be an important first step in the evolution of new function following gene duplication (Figure 3).

During the early stages following duplication, changes in the 5′ end of the gene can result in relocalization of duplicate proteins. This relocalization can occur as a result of an initial imperfect duplication or with as little as a single point mutation. In the case where a dually targeted gene is duplicated, it is possible that subfunctionalization (i.e. degenerative mutations affecting the N terminus of the proteins) could result in two different genes being permanently targeted to separate locations.

Once a duplicate protein is relocalized, one of three outcomes is possible: (i) no overall effect on fitness; (ii) a negative effect on fitness; or (iii) a positive effect on fitness. In the first scenario, when there is no or little effect on fitness, the duplicate gene could drift towards fixation. Alternatively, subsequent mutations in the NTP or N-terminus could result in a protein being retargeted to another location, effectively giving the duplicate gene a second chance to acquire a new function. In the second scenario, a negative effect on fitness would probably result in the elimination of that duplicate gene by purifying selection. Again, the duplicate protein could be retargeted to a different location before it can be totally eliminated from the population. In the third case, where the duplicate protein acquires a novel biochemical function as a result of its new location, purifying selection could act on the duplicate gene to prevent the accumulation of any mutation that might destroy or minimize its function.

Positive selection could occur shortly after relocalization as the protein adjusts to its new subcellular environment. Proteins have been shown to have amino acids that are characteristic of their metabolic surroundings, especially at surfaces that are in direct contact with their environment. Presumably, proteins evolve so that the surface amino acid composition maximizes their efficiency in their specific subcellular location [47]. In fact, surface composition has been used to predict the subcellular location of a protein [47]. The potential impact that subcellular location has on protein evolution is further illustrated by the recent suggestion that subcellular location was the single largest factor affecting the evolutionary rate of mammalian proteins [48].

Concluding remarks

PSR suggests that changes in the N terminus of duplicate proteins result in altered subcellular targeting, which can then change protein function as a consequence of a new metabolic environment. PSR does not discount models such as subfunctionalization, dosage compensation, or epigenetic complementation, but rather might be one way in which these retained duplicates could acquire novel functions. PSR also complements models of novel gene evolution, such as partial or chimeric duplication [23], by describing how such duplications, which might be altered at the 5' end, could have new functions at inception by being relocalized to different subcellular locations. PSR could also have a role in adaptive radiation [16] as duplicate genes could improve their various functions by relocalizing their proteins to different places within the cell in addition to accumulating point mutations.

A positive feature of PSR is that it makes several testable predications. First, minor changes to the 5′ end of a duplicate gene can cause a relocalization of duplicate proteins within the cell. A comparison of N-terminal amino acid sequences within recently diverged gene families with empirically determined subcellular locations might provide some insight into the types and degree of change that have led to altered subcellular targeting between duplicate genes. Second, because functional divergence by PSR can occur relatively quickly, we might expect to see some level of standing variation in the subcellular location of duplicate proteins within gene families. We might also expect that the higher the levels of subcellular complexity (as measured by the number of different subcellular compartments), the greater the number of large gene families and perhaps the greater number of retained duplicates overall. A comparative survey of genomes from organisms encompassing a range of subcellular complexities, focusing on both large and small gene families, could provide pertinent information. Third, PSR suggests that once the retargeted protein reaches its new subcellular location and acquires a novel function, positive selection could occur shortly thereafter, as the protein adjusts to its new metabolic environment. In looking for signatures of positive selection, particular attention could be placed on these particular surface residues.

We have described how changes at the 5′ end of a duplicate gene can alter the subcellular location of its protein and potentially give rise to new genes without the mutational load typically associated with classic neofunctionalization. PSR represents one of the few attempts to understand the actual mechanics of novel gene evolution, describing not only how duplicates initially retained by epigenetic complementation, dosage compensation, or subfunctionalization could evolve new biochemical func-

tions, but also how duplicates could acquire novel functions with few point mutations. We suggest that PSR might be a key event in the evolution of new protein functions and as such, influential in determining whether a duplicate ultimately becomes a novel gene.

Acknowledgements

We thank: S. McKay, J. Flowers and J. True for their reading of and thoughtful suggestions for, this article; F. Kondrashov and two anonymous reviewers for their insight and comments; M. Bell for the advice, encouragement and the introduction that made this collaboration possible, and the genome discussion group at Stony Brook's Department of Ecology and Evolution for stimulating discussions that helped to consolidate our ideas. This work was supported in part by NSF/DEB01297 (to R.G.).

References

1 Li, W.H. and Gojobori, T. (1983) Rapid evolution of goat and sheep globin genes following gene duplication. *Mol. Biol. Evol.* 1, 94–108
2 Hughes, A.L. (1994) The evolution of functionally novel proteins after gene duplicationm. *Proc. R. Soc. B* 256, 119–124
3 Long, M. *et al.* (2003) The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* 4, 865–875
4 Hughes, A. (2005) Gene duplication and the origin of novel proteins. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8791–8792
5 Taylor, J.S. and Raes, J. (2004) Duplication and divergence: the evolution of new genes and old ideas. *Ann. Rev. Genet.* 28, 615–643
6 Kondrashov, F.A. *et al.* (2002) Selection in the evolution of gene duplications. *Genome Biol* DOI: 10.1186/gb-2002-3-2-research0008
7 Lynch, M. and Conery, J.S. (2003) The origins of evolutionary complexity. *Science* 302, 1401–1404
8 Papp, B. *et al.* (2003) Evolution of *cis*-regulatory elements in duplicated genes of yeast. *Trends Genet.* 19, 417–422
9 Kondrashov, F.A. and Kondrashov, A.S. (2006) Role of selection in gene duplications. *J. Theor. Biol.* 239, 141–151
10 He, X. and Zhang, J. (2005) Gene complexity and gene duplicability. *Curr. Biol.* 15, 1016–1021
11 Li, W.H. *et al.* (2005) Expression divergence between duplicate genes. *Trends Genet.* 21, 602–607
12 Moore, R.C. and Purugganan, M.D. (2005) Evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* 8, 122–128
13 Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
14 Moore, R.C. and Purugganan, M.D. (2003) The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15682–15687
15 Shui, S. *et al.* (2006) Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2232–2236
16 Francino, M.P. (2005) An adaptive radiation model for the origin of new gene functions. *Nat. Genet.* 37, 573–577
17 Zhang, J. (2003) Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298
18 Hurles, M. (2004) Gene duplication: the genomic trade in spare parts. *PloS Biol.* 2, e206
19 Ohno, S. (1970) *Evolution by Gene Duplication,* Springer-Verlag
20 Force, A. *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 15, 1531–1545
21 Rodin, S.N. *et al.* (2005) Repositioning dependent fate of duplicate genes. *DNA Cell Biol.* 24, 529–542
22 Papp, B. *et al.* (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424, 194–197
23 Katju, V. and Lynch, M. (2006) On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. *Mol. Biol. Evol.* 23, 1056–1067
24 Lodish, H. *et al.* (2000) *Molecular Cell Biology* (4th edn), W.H. Freeman
25 Bannai, H. *et al.* (2002) Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* 18, 298–305
26 Kaiser, C.A. *et al.* (1987) Many random sequences functionally replace the secretion signal sequence of yeast invertase. *Science* 235, 312–317
27 Williams, E.J.B. *et al.* (2000) The molecular evolution of signal peptides. *Gene* 253, 313–322
28 Karniely, S. and Pines, O. (2005) Single translation-dual destination: mechanisms of dual protein targeting in eukaryotes. *EMBO Rep.* 6, 420–425
29 Schulte, W. *et al.* (1997) Multi-functional acetyl-CoA carboxylase from *Brassica napus* is encoded by a multi-gene family: indication for plastidic localization of at least one isoform. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3465–3470
30 Sirover, M.A. (1999) New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochim. Biophys. Acta* 1432, 159–184
31 Björses, P. *et al.* (2000) Mutations in the AIRE Gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. *Am. J. Hum. Genet.* 66, 378–392
32 Domínguez, D. *et al.* (2003) Phosphorylation regulates the subcellular location and activity of the snail transcriptional repressor. *Mol. Cell. Biol.* 23, 5078–5089
33 Bijur, G.N. and Jope, R.S. (2003) Glycogen synthase kinase 3β is highly activated in nuclei and mitochondria. *Neurochem.* 14, 2415–2419
34 Simpson, J.C. and Pepperkok, R. (2006) The subcellular localization of the mammalian proteome comes a fraction closer. *Genome Biol.* 7, 222.1–222.3
35 Lessring, M.A. *et al.* (2004) Alternative translation initiation generates a novel isoform of insulin-degrading enzyme targeted to mitochondria. *Biochem. J.* DOI: 10.1042/BJ20041081
36 Bizily, S.P. *et al.* (2003) Subcellular targeting of methylmercury lyase enhances its specific activity for organic mercury detoxification in plants. *Plant Physiol.* 131, 463–471
37 Heilmann, I. *et al.* (2004) Switching desaturase enzyme specificity by alternate subcellular targeting. *Proc. Natl. Acad. Sci. U. S. A.* 101, 10266–10271
38 Broadwater, J.A. *et al.* (2002) Desaturation and hydroxylation. Residues 148 and 324 of *Arabidopsis* FAD2, in addition to substrate chain length, exert a major influence in partitioning of catalytic specificity. *J. Biol. Chem.* 277, 15613–15620
39 Taverna, D.M. and Goldstein, R.A. (2002) Why are proteins marginally stable? *Proteins* 46, 105–109
40 Sessions, A. *et al.* (2000) Cell–cell signaling and movement by the floral transcription factors LEAFY and APETALA1. *Science* 289, 779–781
41 Wu, X. *et al.* (2003) Modes of intercellular transcription factor movement in the *Arabidopsis* apex. *Development* 130, 3735–3745
42 Fan, L. *et al.* (1999) Subcellular distribution and tissue expression of phospholipase Dα, Dβ, and Dγ in *Arabidopsis*. *Plant Physiol.* 119, 1371–1378
43 Sung, D.Y. *et al.* (2001) Comprehensive expression profile analysis of the *Arabidopsis* Hsp70 gene family. *Plant Physiol.* 126, 789–800
44 Helms-Lackey, K. *et al.* (2003) Expression of 1L –myoinositol –1- phosphate synthase in organelles. *Plant Physiol.* 132, 2240–2247
45 Liscovitch, M. *et al.* (1999) Localization and possible functions of phospholipase D isozymes. *Biochim. Biophys. Acta* 1439, 245–263
46 Vallon, O. (2005) *Chlamydomonas immunophilins* and *paravulins*: survey and critical assessment of gene models. *Eukaryot. Cell* 4, 230–241
47 Andrade, M.A. *et al.* (1998) Adaptation of protein surfaces to subcellular location. *J. Mol. Biol.* 276, 517–525
48 Julinius, K. and Pederson, A.G. (2006) Protein evolution is faster outside the cell. *Mol. Biol. Evol.* 23, 2039–2048