Review Article

Novel Genes from Formation to Function

Rita Ponce,[1] Lene Martinsen,[2, 3] Luís M. Vicente,[4] and Daniel L. Hartl[3]

[1] *Centro de Biologia Ambiental (CBA), Faculdade de Ciências da Universidade de Lisboa, 1749 Lisboa, Portugal*
[2] *Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, Blindern, 0316 Oslo, Norway*
[3] *Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA*
[4] *Centre for Environmental and Marine Studies (CESAM), Faculdade de Ciências da Universidade de Lisboa, 1749 Lisboa, Portugal*

Correspondence should be addressed to Rita Ponce, arponce@fc.ul.pt

The study of the evolution of novel genes generally focuses on the formation of new coding sequences. However, equally important in the evolution of novel functional genes are the formation of regulatory regions that allow the expression of the genes and the effects of the new genes in the organism as well. Herein, we discuss the current knowledge on the evolution of novel functional genes, and we examine in more detail the youngest genes discovered. We examine the existing data on a very recent and rapidly evolving cluster of duplicated genes, the *Sdic* gene cluster. This cluster of genes is an excellent model for the evolution of novel genes, as it is very recent and may still be in the process of evolving.

1. Introduction

The availability of complete genome sequences allows the comparison of genomes, thus revealing the differences in gene complement and demonstrating the nature of the changes that occur in the evolution of genomes. In particular, genomic analysis showed that more than one-third of the eukaryotic genome is composed of gene duplications and gene families (e.g., [1–4]), highlighting the prominence of duplications in the evolution of genomes. The study of whole genomes also allows an analysis of the rates and dynamics of duplications as well as the divergence and silencing of duplicates (e.g., [5–7]). Moreover, evidence indicates that the rates of duplication are extremely high and of the same order of magnitude as the rate of mutation per nucleotide; for example, the frequency of duplication for coding genes was found to be 0.01 per gene per million years [8] and for internal duplications of gene segments, 0.001–0.013 duplications/gene per million years [9].

The fates of gene duplicates can be very different according to the mutations they undergo and the selective pressures they are under. After a gene is duplicated, the most common fate seems to be the loss of function of one copy by the acquisition of degenerative mutations, while the other copy retains the original function. The originally identical copies can also both be maintained in the genome, allowing a higher production of the corresponding RNA or protein. Subsequent mutation by retrotransposon insertion into one of the copies can affect adaptive evolution because the effect on the phenotype can in some cases be beneficial to the organism [10]. Partial gene duplications can also occur, and if the duplicated part involves a structural or functional domain the new gene can increase the functional complexity of the encoded protein. It has also been proposed that if two copies of a gene acquire mutations in distinct subfunctions, the copies may undergo complementary loss of function in such a way that both copies are required to perform the original function; both copies would be maintained by selection [11]. Probably the most infrequent fate for a gene duplicate, and the one with most evolutionary impact, is for a duplicate copy to gain a new function by acquisition of mutations and to be maintained subsequently by selection while the old copy retains the original function. This last phenomenon leads to novel cellular functions and is the basis of the formation of divergent gene families. For a recent review of models of the evolution of gene duplications and the predictions for each model, see Innan and Kondrashov [12].

Gene duplication has long been thought to be a primary source of material for the evolution of new genes and new functions. The potential of gene duplication was first

recognized by Haldane [13], Bridges [14] and Muller [15]. In 1970, these ideas resurfaced with Ohno [16] who maintained that new genes and novel functions arise only by duplication and who presented evidence for the role of genome duplications in the evolution of multicellular organisms. Since then, the importance of duplications and of the divergence of duplicated copies has been widely supported (for a review see, e.g., [17–20]).

However, other evidence has accumulated showing that there are also other mechanisms responsable for the evolution of novel functions, such as exon shuffling, retroposition, gene fusion, and gene fission, and that several mechanisms can act together ([21–23] and recently reviewed in Kaessmann [24]). Another issue that has received attention is the formation of new regulatory regions, as these can be responsible for the development of novel expression patterns (for review see, e.g., [25–28]). In this paper, we will focus on the origin of very recent genes, which are of special interest since they may have not yet lost the signature of the events that took place during their formation, therefore conveying valuable information about those events.

2. Mechanisms of Origin of Novel Genes

The origin of novel genes has been extensively studied. In Table 1 is summarized the information on a sample of novel genes that reveals a picture of the events involved in their origin and their function. The sample is limited to genes that originated less than 50 million years ago to avoid confounding events of origination with effects of later evolution; this cutoff is also based on the dynamics of formation, preservation, and decay of duplicates and chimeras found by Rogers et al. [29]. Many of the novel genes described and characterized are from *Drosophila* species. The extensive work carried out on the evolution of new genes in *Drosophila* (see e.g., [30]) reflects the fact that *D. melanogaster* is a well-studied model organism at the molecular level that also has a long history of evolutionary studies.

Overall, there is no single mechanism of the molecular events involved in the formation and maintenance of novel genes, but rather several mechanisms. While gene duplication is conceded to have an important role in creating genetic novelties, an alternative mechanism for the origin of new genes is the shuffling of existing genes or functions to form chimeric genes. However, many chimeric genes derive from previously duplicated material, as shown by the presence of the parental genes in the genome and the organization of the chimeras [29, 31]. A chimeric gene would hardly be recognized as such were the parental genes not present. A second type of chimeric gene arises not from tandem duplication, but from duplication through retrotransposition. In addition, it has been shown that both nonhomologous recombination and nonallelic homologous recombination can lead to the formation of chimeric genes which both can be facilitated by repetitive elements (transposable elements or satellite DNA sequences) [32]. Yang et al. [32] identified 17 chimeric genes formed by ectopic recombination within the last 12 Mys in the *Drosophila melanogaster*.

The contribution of chimeric genes to the evolution of genetic novelties has been revealed by recent studies. Rogers and colleagues [29, 31] compare the fate of simple genetic duplicates versus genetic duplicates that underwent fusion to form chimeras in *D. melanogaster*. The results are very interesting: gene duplications are formed at a rate of about 80.4 duplicates per million years, but only 4.1% are preserved; chimeras are formed at a rate of about 11.4 per million years and show a similar rate of decay (with 1.4% preservation) [29]. Some chimeric genes were also implicated in selective sweeps [31], revealing the impact of this kind of molecular event. This analysis of the *D. melanogaster* genome is interesting, but raises the question whether the same pattern would be observed in species that do not have such a high rate of DNA loss as *D. melanogaster* as found by Petrov and colleagues [33–35].

Equally relevant, and conceptually on the opposite side of the spectrum, is the origination of new genes "from scratch" or *de novo*, the case when coding sequences are derived from noncoding DNA, for example, introns and other untranslated regions [36–39]. This reutilization of noncoding sequences may have important consequences in the fate of such genes.

3. Tissues Where Novel Genes Are Expressed

Among the examples of new genes with very recent origin, many are involved in male reproduction, in particular in spermatogenesis. It is the case of *Jingwei*, which appears to be expressed in the testis [43]; *Odysseus*, which contains rapidly evolving homeodomains involved in sperm function [56]; *Dntf-2r*, which has male-specific expression [50]; *K81*, which is expressed in primary spermatocytes [58]; *Sdic*, whose product is incorporated into the sperm tail [47]. The prominence of expression in male reproductive tissue has led to the hypothesis that spermatogenesis is more prone to the cooption of novel genes and functions, while other developmental processes may be under a higher level of selective constraint [63]. As debated in Nielsen et al. [64], the testes undergo intense selection pressures because the cells are subject to genetic conflict, sperm competition, reproductive isolation and are exposed to germline pathogens and mutations that cause segregation distortion. An interesting observation in this regard is that newly retroposed duplicated genes in *Drosophila* are dominated by genes going from the X chromosome to the autosomes, not the opposite direction, and most of these have evolved a testis-specific expression pattern [65]. A possible explanation is that this predominantly one-way pattern is driven by positive selection because the retroposed genes escape X inactivation during spermatogenesis and are therefore free to develop expression patterns in the testis [65].

Kaessmann [24] emphasizes the importance of transcription during spermatogenesis, during which a permissive state of the chromatin may allow a widespread transcription of genes that would not otherwise be expressed.

Additionally, during spermatogenesis, there are numerous cell divisons, affording an opportunity for new

TABLE 1: Novel genes described (less than 50 million years (mys) old): Their formation, expression and function. The genes are ordered by age.

Gene	Mechanisms	Functions	Expression tissue	Species, genome location	Age (mys)	References
Monkey-king (mkg)	Duplication by retroposition and gene fission	Not determined	D. simulans/D. sechellia only males, D. mauritania males and females	D. simulans, D. sechellia, D. mauritania, X chromosome (chr.)	1-2	[23]
Quetzalcoatl	Duplication, gene fusion	Not determined	Early pupae, adult females, and male testes	D. melanogaster	<2	[40]
Ifc-2h	Duplication by retroposition	Not determined	Whole body	D. simulans, D. sechellia, D. mauritiana	2	[41]
Hun	Exon shuffling by illegitimate recombination	Not determined	Testis	D. sechellia, D. mauritania, D. simulans	2-3	[42]
Jingwei	Exon shuffling, retroposition	Not determined	Testis	D. teissieri, D. yakuba, D. santomea	2-3	[43]
Sphinx	Exon shuffling, retroposition	Male courtship behaviour	Embryo, pupae, adult of males and females	D. melanogaster, chr. 4	2-3	[44, 45]
Poldi	De novo emergence from intergenic DNA	Involved in spermatogenesis, perhaps through chromatin modification pathways	Testis	Mus mus domesticus and four other species in the Mus genus.	2.5-3.5	[38]
Adh-Twain	Duplication, gene fusion	Not determined	Larvae, adult males and females	D. subobscura, D. guanche, D. madeirensis	3	[22, 46]
CG32582, CG32712, CG15323, CG32690, CG31909	From non-coding sequences	Not determined	Testis (enriched in primary spermatocytes)	D. melanogaster; D. simulans	2–5	[37]
Sdic	Duplication, gene fusion	Sperm competition	Testis	D. melanogaster, X chr.	<5.4	[47–49]
Dntf-2r	Duplication by retroposition	Not determined	Testis	D. melanogaster, D. simulans, D. sechellia, D. mauritiana, 2nd chr.	3–12	[50]
Nsr (kep1 gene family)	Duplication	Spermatogenesis	Testis (enriched in primary spermatocytes)	D. melanogaster subgroup, 2nd chr.	5.4–12	[51]
CG3927 (kep1 gene family)	Duplication	Spermatogenesis	Testis (enriched in primary spermatocytes)	D. melanogaster subgroup, 2nd chr.	5.4–12	[51]
CG4021 (kep1 gene family)	Duplication	Spermatogenesis	Testis (enriched in primary spermatocytes)	D. melanogaster subgroup, 2nd chr.	5.4–12	[51]
Sflc	Duplication	Male and female fertility, life expectancy	Not specified	D. melanogaster subgroup	6–11	[52]
Hydra	De novo emergence from intergenic DNA	Not determined	Testis	D. melanogaster subgroup	<13	[53]
PIPSL	Exon shuffling, retrotransposition	Ubiquitin binding	Testis	Humans Chimpanzees	15–19	[54]
Adh-Finnegan	Duplication, gene fusion	Not determined	Pupae and adults	D. buzzatii, D. hydei, D. metleri, D. mojavensis, D. mulleri (repleta group)	20–30	[45, 46, 55]

TABLE 1: Continued.

Gene	Mechanisms	Functions	Expression tissue	Species, genome location	Age (mys)	References
Odysseus (OdsH)	Duplication	Involved in sperm function	Mainly in male reproductive tissue, but also detected in larvae and embryo	D. melanogaster, D. simulans, D. sechellia, D. mauritiana, D. yakuba	<40 (after split of subgenus Sophophora and Drosophila	[56, 57]
K81	Duplication by retroposition	Telomere capping in post-meiotic male germ cells	Male specific expression (testis: primary spermatocytes), but weak expression also in wild-type females	D. melanogaster subgroup	<30	[58, 59]
Obp57d and Obp57e	Duplication	Taste sensation of octanoic acid	Taste sensilla on the legs	D. melanogaster subgroup	<30	[60, 61]
PGAM3	Exon shuffling, retroposition	Possibly phosphoglycerate mutase activity	White blood cells	Humans, chimpanzees, macaques, X chr.	>25	[45, 62]

mutations to arise, and those may affect sperm performance. Hence, spermatogenesis offers a large arena of competition in which cells and their descendants are under intensive selective pressure.

An important pattern observed in speciation, often considered the first step in the evolution of complete sterility and inviability, is Haldane's rule, which states that, in an interspecific cross, if one sex is sterile or inviable, this sex is the heterogametic sex [66]. If novel genes appear more often in male reproductive tissue than in female reproductive tissue—as observed in the summary of novel genes in Table 1—this could be an explanation for Haldane's rule when the heterogametic sterile F_1 hybrid is the male, but not in the cases when it is the female. In *Xenopus*, in which males are the homogametic sex, males are always sterile in interspecific crosses [67]. However, while *Xenopus* seems an exception to Haldane's rule, it may be a consequence of the fast evolution of male-specific genes. Strong sexual selection in males is reflected by rapid evolution of genes in male reproductive tissue, which then may cause hybrid sterility in recently diverged sister species even when males are homogametic. From this angle, tissue specificity of novel genes seems more important to hybrid sterility than heterogamy. It would be interesting to examine the tissue specificity of novel genes in species in which the female is the heterogametic sex, but this is beyond the scope of our paper.

4. The *Sdic* Gene Cluster

The *Sdic* gene is a recently evolved chimeric gene in *D. melanogaster*, discovered and described by Nurminsky and colleagues in 1998 [47, 68]. This gene possesses several unique features that provide an exceptional opportunity for the study of new gene functions, the fate of gene duplications, and the evolution of male reproductive traits.

Sequence analysis of the *Sdic* gene revealed that *Sdic* is a chimera of two genes that exist intact in the genome. *Sdic* is composed of parts of *AnnX*, which encodes an annexin protein, and *Cdic* (also referred to in the literature and FlyBase as *sw*), which encodes an intermediate polypeptide chain for the cytoplasmic dyneins [47]. The structure of *Sdic* along with the fact that *AnnX* and *Cdic* exist intact in the genome indicates that *Sdic* originated as a duplication and fusion of *AnnX* and *Cdic*, followed by small deletions and rearrangements. Its formation involved the creation of novel promoter elements (which provided testis-specific expression) from the fusion of portions of an *AnnX* exon and a *Cdic* intron. Its coding region, however, derived solely from *Cdic*. The comparison of the coding region of *Sdic* with *Cdic* shows that *Sdic* lacks the 3′ region of *Cdic* (which corresponds to 100 amino acids residues at the C-terminal part of the *Cdic* protein) and at its 5′ end underwent extensive refashioning by the occurrence of multiple mutations, deletions (including frameshift deletions), and insertions, culminating in a new 5′ exon that encodes a totally novel N-terminus for the protein [47, 69] (Figure 1).

There are several copies of *Sdic* located in tandem at the base of the X chromosome, in region 19 of the larval salivary gland polytene chromosomes, forming a gene cluster. This repeated region is flanked by the parental genes, on the 5′ side by *Cdic* and on the 3′ side by *AnnX*. According to the available genomic sequence of *D. melanogaster, the Sdic* gene is repeated four times in tandem between the genes *Cdic* and *AnnX* genes. Within this cluster there are also four dead-on-arrival retrotransposable elements of the *RT1C* family, one *RT1C* copy located upstream of each *Sdic* gene copy [70] (Figure 2).

Sdic is present in all wild-type strains of *D. melanogaster* [47], but it has not been found in any other species of the *D. melanogaster* subgroup. In the species closest to *D. melanogaster* (*D. simulans*, *D. mauritiana*, *D. yakuba*, *D. teissieri*, and *D. erecta*) *AnnX* and *Cdic* are adjacent to each other, without signs of an ancestral *Sdic* gene or *RT1C* element between them [71]. Furthermore, in these species, *AnnX* and *Cdic* do not show any signs of duplication and are reasonably conserved across species [71]. Consequently, the formation of the *Sdic* gene and the subsequent duplication that formed the entire cluster happened only in the lineage that gave rise to *D. melanogaster*; the original *Sdic* and the *Sdic* cluster were formed within the last 2 million years, after the split of the lineage that formed *D. melanogaster* and its sibling species *D. simulans*.

A model for the evolution of this cluster was developed based on the available genomic data [70], and the divergence of the *RT1C* copies, which are expected to be evolving neutrally, was used to date the duplications [48]. In the ancestral situation *AnnX* and *Cdic* must have been adjacent to each other. An initial event duplicated *AnnX* and *Cdic*, and then one or more deletions fused one copy of *AnnX* and one copy of *Cdic* giving rise to an ancestral *Sdic*. During the early steps of the formation of the ancestral *Sdic*, a dead-on-arrival retrotransposable element from the *RT1C* family was inserted upstream of the ancestral *Sdic*. The ancestral *Sdic* and its upstream region were duplicated in the last 232–463 thousand years, giving rise to two *Sdic* genes (and two *RT1C*, one upstream of each *Sdic*). Another duplication in the last 100–180 thousand years gave rise to four *Sdic* genes (and four *RT1C*, one upstream of each *Sdic* copy).

While the *Sdic* cluster has been isolated in the extremities of BAC clones, it is difficult to determine the overlap of the BAC sequences; moreover, there are nonassembled pieces of BAC clones containing *Sdic* portions that do not seem to match any of the assembled copies. The recent *in situ* hybridization work by Yeh et al. [49] indicates that all *Sdic* copies should be in region 19 of chromosome X. Given the young age of the cluster, the duplicates are predicted to be very recent and to have few or no differences in sequence, making it extremely difficult to confirm the exact number of genes by present sequencing and assembly techniques. The available sequences of assembled and nonassembled BAC clones could be explained either by four copies in tandem or eight copies that resulted from a duplication of preexisting four, yielding eight with the same pattern of similarities. The hypothesis of a cluster formed of four copies is the most parsimonious; however eight copies in tandem would be closer to the original Southern blot experimental results by Nurminsky et al. [47], which suggested as many as 10 duplicates.
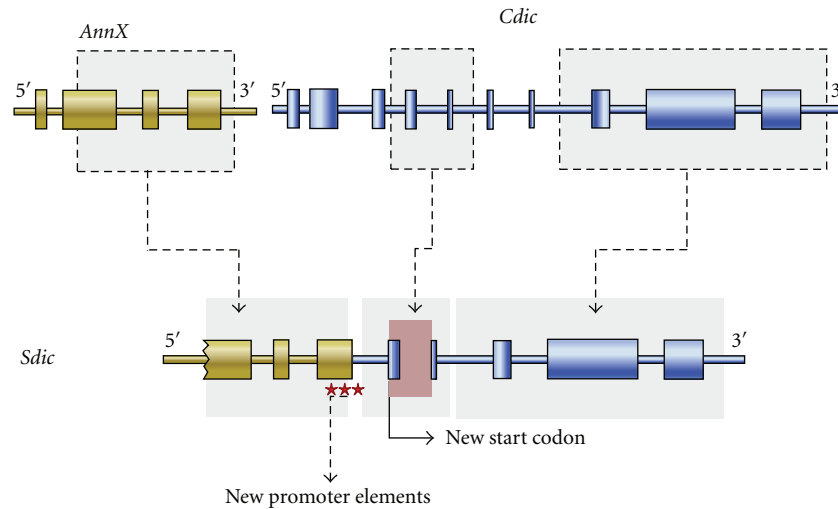
FIGURE 1: Formation of the *Sdic* gene from parts of the genes *AnnX* and *Cdic*. Introns are represented as thin cylinders and exons as thick cylinders. The stars represent *Sdic* promoter elements.
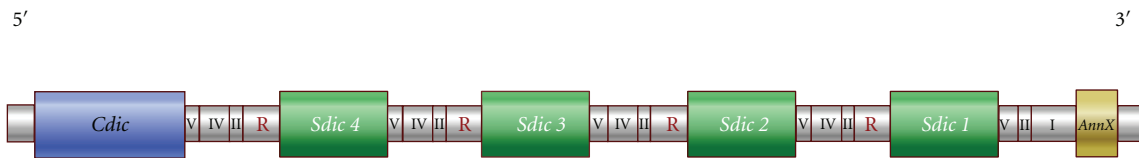

FIGURE 2: The *Sdic* gene cluster. The cluster is composed by four *Sdic* genes, with an *RT1C* retrotransposable element upstream of each Sdic gene. The cluster is located between the parental genes *Cdic* and *AnnX*. *Cdic* is represented in blue, *Sdic* genes in green and *AnnX* in yellow; intergenic regions are grey; R represents *RT1C* elements. These genes are located in the minus strand, so the order of genes in this figure is the opposite order of these genes in Flybase.

One question still unanswered is that of the expression of the genes that belong to this cluster. Although the cluster is composed of four almost identical *Sdic* genes, there is no evidence for expression of all copies. The RNA sequences described in Nurminsky et al. [47] and Kulathinal et al. [63] are similar to what would be the expected transcription of *Sdic1*. Later work found evidence of expression for both *Sdic1* and *Sdic3* [49, 72]. *Sdic1* is the oldest gene in the array and *Sdic3* was originated from a duplication of *Sdic1*, being the second oldest. The promoter sequences of all copies are identical, except for the promoter of *Sdic1*, which differs from the others by two nucleotides. Although these changes at the promoter level might not be relevant, there is experimental evidence that the promoter of *Sdic1*, at least, is a functional promoter: Nurminsky et al. [47] tested the *Sdic1* promoter *in vivo* and *in vitro* and found it to be functional.

The presence of extra genetic material in the *Sdic* region when one compares *D. melanogaster* with the other species of the *D. melanogaster* subgroup, extra material that contains a cluster of nearby identical genes and dead-on-arrival transposable elements is remarkable. *D. melanogaster* has a high rate of gene loss of unessential genetic material caused by deletions [33–35]. Additionally, according to existing models of the fate of gene duplicates and in light of the high rate of nonessential DNA loss in *D. melanogaster*, if not all copies of *Sdic* are functional then one would predict that the extra

copies of *Sdic* would be in the process of degeneration showing many deletions and mutations in the coding sequences as well as the *RT1C* elements. However, the *Sdic* copies and *RT1C* elements appear very similar to each other. It is possible that, given the young age of the cluster, degeneration is in such early stages that such mutations have not occurred or that those that may have occurred have been repaired by gene conversion from nonmutated copies. It is also possible that the *Sdic* cluster has important cellular functions yet to be discovered. Novel gene functions are frequently associated with rapid changes and show signs of positive selection [44, 50, 73], supporting the idea that novel genetic functions allow adaptive changes. There is limited evidence for positive selection of *Sdic* and at least one selective sweep in this region [47, 63, 74].

A key question concerning any novel gene and the functional role it fulfills is whether the function itself is novel or whether the function is redundant or overlapping with an already existing functional gene or genes. The function of *Sdic* protein was deduced from its sequence to be a sperm-specific dynein intermediate chain, and this deduction was supported by the finding of the *Sdic1* gene product in the sperm tail [47]. Is the Sdic protein redundant with another sperm dynein intermediate chain or has it supplanted a pre-existing gene? Has *Sdic* played a role in the speciation events in the split of the lineage of *D. melanogaster* from its sibling

species and has *Sdic* had an effect in adaptation or sexual selection in *D. melanogaster*?

These questions on the role of *Sdic* have been addressed very recently in experiments in which Yeh and colleagues [49] knocked out the *Sdic* region in *D. melanogaster* and looked for phenotypic effects on male reproduction. While no effects were detected in progeny size or sex ratio, males without *Sdic* had sperm that were less competitive in the female reproductive tract in being more easily displaced by sperm from subsequent males and, to a lesser extent, in being less able to displace sperm from previous males. These results support the role of the *Sdic* protein in sperm competition, but most importantly they answer the central question for new genes of "how did the sibling species manage without this function?" the work by Yeh et al. [49] shows that *Sdic* plays an important role in reproduction, and although it is not an essential gene, it gives an advantage in sperm competition, which could in turn be related to the positive selection.

It remains to be determined whether positive selection has acted upon all copies of *Sdic* and if the copies are subjected to the same selective pressures, but the importance of this gene cluster in reproduction has at least been demonstrated, and the results may help to shed light on the general issue of the function of novel genes

5. Concluding Remarks

Ultimately, differences observed between species are due to differences at the genome level. Genomic studies are revealing the extent of these differences—in gene number, in encoded functions, in expression—and are also revealing the mechanisms involved in the evolution of genomes. The analysis of particular newly evolved genes provides information in finer detail, which hopefully can be generalized and help to understand the evolution of new genes and new functions. Equally as important as the formation of new coding sequences is the formation of regulatory regions responsible for new patterns of expression as well as the processes leading to spread and maintenance of the novel gene in the population.

Bacterial genome studies have made very clear that, at least in bacterial species, a great part of the genes are not shared by all individuals of a species [75]. Different strains of the same species share a core genome containing genes present in all strains; however there is also a pan-genome consisting of genes present in only a subset of strains. As more complete genome sequences become available, we will be able to determine if similar patterns are observed in eukaryotes.

Author's Contribution

R. Ponce and L. Martinsen contributed equally to this work.

Acknowledgments

References

[1] S. Kaul, H. L. Koo, J. Jenkins et al., "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana," *Nature*, vol. 408, no. 6814, pp. 796–815, 2000.

[2] G. M. Rubin, M. D. Yandell, J. R. Wortman et al., "Comparative genomics of the eukaryotes," *Science*, vol. 287, no. 5461, pp. 2204–2215, 2000.

[3] J. C. . Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton et al., "The sequence of the human genome," *Science*, vol. 291, pp. 1304–1351, 2001.

[4] R. D. Morin, E. Chang, A. Petrescu et al., "Sequencing and analysis of 10,967 full-length cDNA clones from Xenopus laevis and Xenopus tropicalis reveals post-tetraploidization transcriptome remodeling," *Genome Research*, vol. 16, no. 7, pp. 947–803, 2006.

[5] Z. Gu, D. Nicolae, H. H. S. Lu, and W. H. Li, "Rapid divergence in expression between duplicate genes inferred from microarray data," *Trends in Genetics*, vol. 18, no. 12, pp. 609–613, 2002.

[6] P. Zhang, Z. Gu, and W. H. Li, "Different evolutionary patterns between young duplicate genes in the human genome," *Genome Biology*, vol. 4, no. 9, article R56, 2003.

[7] L. Bu, U. Bergthorsson, and V. Katju, "Local synteny and codon usage contribute to asymmetric sequence divergence of Saccharomyces cerevisiae gene duplicates," *BMC Evolutionary Biology*, vol. 11, article 279, 2011.

[8] M. Lynch and J. S. Conery, "The evolutionary fate and consequences of duplicate genes," *Science*, vol. 290, no. 5494, pp. 1151–1155, 2000.

[9] X. Gao and M. Lynch, "Ubiquitous internal gene duplication and intron creation in eukaryotes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 49, pp. 20818–20823, 2009.

[10] A. Kanazawa, B. Liu, F. Kong, S. Arase, and J. Abe, "Adaptive evolution involving gene duplication and insertion of a novel Ty1/copia-like retrotransposon in soybean," *Journal of Molecular Evolution*, vol. 69, no. 2, pp. 164–175, 2009.

[11] A. Force, M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait, "Preservation of duplicate genes by complementary, degenerate mutations," *Genetics*, vol. 151, no. 4, pp. 1531–1545, 1999.

[12] H. Innan and F. Kondrashov, "The evolution of gene duplications: classifying and distinguishing between models," *Nature Reviews Genetics*, vol. 11, no. 2, pp. 97–108, 2010.

[13] J. B. S. Haldane, *The Causes of Evolution*, Princeton University Press, 1990.

[14] C. B. Bridges, "The Bar "gene" a duplication," *Science*, vol. 83, no. 2148, pp. 210–211, 1936.

[15] H. J. Muller, "Bar duplication," *Science*, vol. 83, no. 2161, pp. 528–530, 1936.

[16] S. Ohno, *Evolution by Gene Duplication*, Springer, 1970.

[17] V. F. Irish and A. Litt, "Flower development and evolution: gene duplication, diversification and redeployment," *Current Opinion in Genetics and Development*, vol. 15, no. 4, pp. 454–460, 2005.

[18] T. Ohta, "Role of gene duplication in evolution," *Genome*, vol. 31, no. 1, pp. 304–310, 1989.

[19] J. S. Taylor and J. Raes, "Duplication and divergence: the evolution of new genes and old ideas," *Annual Review of Genetics*, vol. 38, pp. 615–643, 2004.

[20] J. Zhang, "Evolution by gene duplication: an update," *Trends in Ecology and Evolution*, vol. 18, no. 6, pp. 292–298, 2003.

[21] M. Long, E. Betrán, K. Thornton, and W. Wang, "The origin of new genes: glimpses from the young and old," *Nature Reviews Genetics*, vol. 4, no. 11, pp. 865–875, 2003.

[22] C. D. Jones, A. W. Custer, and D. J. Begun, "Origin and evolution of a chimeric fusion gene in Drosophila subobscura D. madeirensis and D. guanche," *Genetics*, vol. 170, no. 1, pp. 207–219, 2005.

[23] W. Wang, H. Yu, and M. Long, "Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species," *Nature Genetics*, vol. 36, no. 5, pp. 523–527, 2004.

[24] H. Kaessmann, "Origins, evolution, and phenotypic impact of new genes," *Genome Research*, vol. 20, no. 10, pp. 1313–1326, 2010.

[25] G. A. Wray, "The evolutionary significance of cis-regulatory mutations," *Nature Reviews Genetics*, vol. 8, no. 3, pp. 206–216, 2007.

[26] P. J. Wittkopp, "Evolution of cis-regulatory sequence and function in Diptera," *Heredity*, vol. 97, no. 3, pp. 139–147, 2006.

[27] S. B. Carroll, "Evolution at two levels: on genes and form," *PLoS Bology*, vol. 3, no. 7, p. e245, 2005.

[28] F. Rodriguez-Trelles, R. Tarríol, and F. J. Ayala, "Evolution of cis-regulatory regions versus codifying regions," *International Journal of Developmental Biology*, vol. 47, no. 7-8, pp. 665–673, 2003.

[29] R. L. Rogers, T. Bedford, and D. L. Hart, "Formation and longevity of chimeric and duplicate genes in Drosophila melanogaster," *Genetics*, vol. 181, no. 1, pp. 313–322, 2009.

[30] Q. Zhou, G. Zhang, Y. Zhang et al., "On the origin of new genes in Drosophila," *Genome Research*, vol. 18, no. 9, pp. 1446–1455, 2008.

[31] R. L. Rogers and D. L. Hartl, "Chimeric genes as a source of rapid evolution in Drosophila melanogaster," *Molecular Biology and Evolution*, vol. 29, no. 2, pp. 517–529, 2012.

[32] S. Yang, J. R. Arguello, X. Li et al., "Repetitive element-mediated recombination as a mechanism for new gene origination in Drosophila," *PLoS Genetics*, vol. 4, no. 1, article e3, 2008.

[33] D. A. Petrov and D. L. Hartl, "High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups," *Molecular Biology and Evolution*, vol. 15, no. 3, pp. 293–302, 1998.

[34] D. A. Petrov, E. R. Lozovskaya, and D. L. Harti, "High intrinsic rate of DNA loss in Drosophila," *Nature*, vol. 384, no. 6607, pp. 346–349, 1996.

[35] D. A. Petrov and D. L. Hartl, "Trash DNA is what gets thrown away: high rate of DNA loss in Drosophila," *Gene*, vol. 205, no. 1-2, pp. 279–289, 1997.

[36] M. E. Johnson, L. Viggiano, J. A. Bailey et al., "Positive selection of a gene family during the emergence of humans and African apes," *Nature*, vol. 413, no. 6855, pp. 514–519, 2001.

[37] M. T. Levine, C. D. Jones, A. D. Kern, H. A. Lindfors, and D. J. Begun, "Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 26, pp. 9935–9939, 2006.

[38] T. J. A. J. Heinen, F. Staubach, D. Häming, and D. Tautz, "Emergence of a New Gene from an Intergenic Region," *Current Biology*, vol. 19, no. 18, pp. 1527–1531, 2009.

[39] M. Toll-Riera, N. Bosch, N. Bellora et al., "Origin of primate orphan genes: a comparative genomics approach," *Molecular Biology and Evolution*, vol. 26, no. 3, pp. 603–612, 2009.

[40] R. L. Rogers, T. Bedford, A. M. Lyons, and D. L. Hartl, "Adaptive impact of the chimeric gene Quetzalcoatl in Drosophila melanogaster," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 24, pp. 10943–10948, 2010.

[41] C. Fan and M. Long, "A new retroposed gene in Drosophila heterochromatin detected by microarray-based comparative genomic hybridization," *Journal of Molecular Evolution*, vol. 64, no. 2, pp. 272–283, 2007.

[42] J. R. Arguello, Y. Chen, S. Yang, W. Wang, and M. Long, "Origination of an X-linked testes chimeric gene by illegitimate recombination in Drosophila," *PLoS Genetics*, vol. 2, no. 5, article e77, 2006.

[43] W. Wang, J. Zhang, C. Alvarez, A. Llopart, and M. Long, "The origin of the jingwei gene and the complex modular structure of its parental gene, yellow emperor, in Drosophila melanogaster," *Molecular Biology and Evolution*, vol. 17, no. 9, pp. 1294–1301, 2000.

[44] W. Wang, F. G. Brunet, E. Nevo, and M. Long, "Origin of sphinx, a young chimeric RNA gene in Drosophila melanogaster," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 7, pp. 4448–4453, 2002.

[45] M. Long, M. Deutsch, W. Wang, E. Betrán, F. G. Brunet, and J. Zhang, "Origin of new genes: evidence from experimental and computational analyses," *Genetica*, vol. 118, no. 2-3, pp. 171–182, 2003.

[46] C. D. Jones and D. J. Begun, "Parallel evolution of chimeric fusion genes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 32, pp. 11373–11378, 2005.

[47] D. I. Nurminsky, M. V. Nurminskaya, D. De Aguiar, and D. L. Hartl, "Selective sweep of a newly evolved sperm-specific gene in Drosophila," *Nature*, vol. 396, no. 6711, pp. 572–575, 1998.

[48] R. Ponce, "The use of a non-LTR element to date the formation of the Sdic gene cluster," *Genetica*, vol. 131, no. 3, pp. 315–324, 2007.

[49] S.-D. Yeh, T. Do, C. Chan et al., "Functional evidence that a recently evolved Drosophila sperm-specific gene boosts sperm competition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 6, pp. 2043–2048, 2012.

[50] E. Betrán and M. Long, "Dntf-2r, a young Drosophila retroposed gene with specific male expression under positive Darwinian selection," *Genetics*, vol. 164, no. 3, pp. 977–988, 2003.

[51] Y. Ding, L. Zhao, S. Yang et al., "A young Drosophila duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes," *PLoS Genetics*, vol. 6, no. 12, article e1001255, 2010.

[52] S. Chen, H. Yang, B. H. Krinsky, A. Zhang, and M. Long, "Roles of young serine-endopeptidase genes in survival and reproduction revealed rapid evolution of phenotypic effects at adult stages," *Fly*, vol. 5, pp. 345–351, 2011.

[53] S. T. Chen, H. C. Cheng, D. A. Barbash, and H. P. Yang, "Evolution of hydra, a recently evolved testis-expressed gene with nine alternative first exons in Drosophila melanogaster," *PLoS Genetics*, vol. 3, no. 7, article e107, 2007.

[54] D. V. Babushok, K. Ohshima, E. M. Ostertag et al., "A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids," *Genome Research*, vol. 17, no. 8, pp. 1129–1138, 2007.

[55] D. T. Sullivan, W. T. Starmer, S. W. Curtiss, M. Menotti-Raymond, and J. Yum, "Unusual molecular evolution of an Adh pseudogene in Drosophila," *Molecular Biology and Evolution*, vol. 11, no. 3, pp. 443–458, 1994.

[56] C. T. Ting, S. C. Tsaur, M. L. Wu, and C. I. Wu, "A rapidly evolving homeobox at the site of a hybrid sterility gene," *Science*, vol. 282, no. 5393, pp. 1501–1504, 1998.

[57] C. T. Ting, S. C. Tsaur, S. Sun et al., "Gene duplication and speciation in Drosophila: evidence from the Odysseus locus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 33, pp. 12232–12235, 2004.

[58] B. Loppin, D. Lepetit, S. Dorus, P. Couble, and T. L. Karr, "Origin and neofunctionalization of a Drosophila paternal effect gene essential for zygote viability," *Current Biology*, vol. 15, no. 2, pp. 87–93, 2005.

[59] R. Dubruille, G. A. Orsi, L. Delabaere et al., "Specialization of a drosophila capping protein essential for the protection of sperm telomeres," *Current Biology*, vol. 20, no. 23, pp. 2090–2099, 2010.

[60] T. Matsuo, "Rapid evolution of two odorant-binding protein genes, Obp57d and Obp57e, in the Drosophila melanogaster species group," *Genetics*, vol. 178, no. 2, pp. 1061–1072, 2008.

[61] E. Harada, J. Nakagawa, T. Asano et al., "Functional evolution of duplicated odorant-binding protein genes, Obp57d and Obp57e, in Drosophila," *PLoS ONE*, vol. 7, no. 1, article e29710, 2012.

[62] E. Betrán, W. Wang, L. Jin, and M. Long, "Evolution of the Phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene," *Molecular Biology and Evolution*, vol. 19, no. 5, pp. 654–663, 2002.

[63] R. J. Kulathinal, S. A. Sawyer, C. D. Bustamante, D. Nurminsky, R. Ponce, and J. M. Ranz, "Selective sweep in the evolution of a new sperm-specific gene in Drosophila," in *Selective Sweep*, D. Nurminsky, Ed., pp. 22–33, Springer, Boston, Mass, USA, 2005.

[64] R. Nielsen, C. Bustamante, A. G. Clark et al., "A scan for positively selected genes in the genomes of humans and chimpanzees," *PLoS Bology*, vol. 3, no. 6, article e170, 2005.

[65] E. Betrán, K. Thornton, and M. Long, "Retroposed new genes out of the X in Drosophila," *Genome Research*, vol. 12, no. 12, pp. 1854–1859, 2002.

[66] J. B. S. Haldane, *The Causes of Evolution*, Princeton University Press, 1990.

[67] M. J. Madison-Villar and P. Michalak, "Misexpression of testicular microRNA in sterile Xenopus hybrids points to tetrapod-specific microRNAs associated with male fertility," *Journal of Molecular Evolution*, vol. 73, pp. 316–324, 2011.

[68] D. I. Nurminsky, M. V. Nurminskaya, E. V. Benevolenskaya, Y. Y. Shevelyov, D. L. Hartl, and V. A. Gvozdev, "Cytoplasmic dynein intermediate-chain isoforms with different targeting properties created by tissue-specific alternative splicing," *Molecular and Cellular Biology*, vol. 18, no. 11, pp. 6816–6825, 1998.

[69] J. M. Ranz, A. R. Ponce, D. L. Hartl, and D. Nurminsky, "Origin and evolution of a new gene expressed in the Drosophila sperm axoneme," *Genetica*, vol. 118, no. 2-3, pp. 233–244, 2003.

[70] R. Ponce and D. L. Hartl, "The evolution of the novel Sdic gene cluster in Drosophila melanogaster," *Gene*, vol. 376, no. 2, pp. 174–183, 2006.

[71] R. Ponce, "The recent origin of the Sdic gene cluster in the melanogaster subgroup," *Genetica*, vol. 135, no. 3, pp. 415–418, 2009.

[72] B. R. Graveley, A. N. Brooks, J. W. Carlson et al., "The developmental transcriptome of Drosophila melanogaster," *Nature*, vol. 471, no. 7339, pp. 473–479, 2011.

[73] M. Long and C. H. Langley, "Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila," *Science*, vol. 259, no. 5104, pp. 91–95, 1993.

[74] D. Nurminsky, D. De Aguiar, C. D. Bustamante, and D. L. Hartl, "Chromosomal effects of rapid gene evolution in drosophila melanogaster," *Science*, vol. 291, no. 5501, pp. 128–130, 2001.

[75] M. Touchon, C. Hoede, O. Tenaillon et al., "Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths," *PLoS Genetics*, vol. 5, no. 1, article e1000344, 2009.