

**Genetic Redundancy**  
**The Ultimate Evidence of the Design of Life**

Peter Borger

Corresponding address: The Independent Research Institute on Origins

Peter Borger, PhD

e-mail: [peterborger@hotmail.com](mailto:peterborger@hotmail.com)

*Summary.*

The fundamental premise of the intelligent design concept is that biological systems are the result of intelligent design. Question is however, how the design of such systems can be unequivocally recognized. It has been argued that the irreducible complexity of biological systems can only be explained invoking intelligence and as such it is often presented as evidence of intelligent design. The Darwinians interpretation of irreducible complex systems and their apparent design is that they are the ultimate accumulated result of the natural selection of slightly advantageous traits. Straightforward unambiguous evidence for the design of biological systems should come from observed biological systems that cannot be the result of natural selection. If we would find a biological system to exist for which we must assume neutral selection, then we would not only have the evidence for intelligent design, but the system would also qualify as a falsifier of Darwinian Theory. Over the past two decades scientists have observed the peculiar biological phenomenon of genetic redundancy, which pertains to genes or genetic systems that seem to have no obvious function. Indeed, genetic redundancy is now defined as the situation in which the disruption of a gene is selective neutral. Genetic redundancy is the resultant of cooperating scale free genetic networks that provide robustness to organisms. One of the biggest surprises of modern biology genetic redundancy terminates Darwin's era of natural selection.

*More of the same*

To introduce the topic of genetic redundancy I would first like to focus on a related, but better known genetic concept: *high abundance genes*. Based on their functional output two types of genes can be distinguished: *single-step* output genes and *two-step* output genes. Single-step output genes produce RNA molecules with structural or regulatory functions, whereas the two-step output genes produce proteins. Gene output is always restricted. Physical limits determine how much output – RNA or protein – can be generated. For single-step genes the primary output is always an RNA molecule, and the only physical limit is the rate of gene transcription. For two-step genes it is a bit more complicated. Here, like single-step genes, the first step also produces an RNA molecule – now called the messenger RNA which is an exact template of the coded information

present in the DNA. The output of a protein-coding gene is further limited at the level of translation, when the messenger RNA is translated into the actual protein. Transcription and translation thus determine the amount of protein that can be formed. Because of these two rate-limiting factors, a single protein-coding gene can never exceed the production of a certain amount of protein. A valid question is – How do organisms cope with the situation in which gene output is required in such high abundance that a single copy of the gene would not be sufficient? For instance, during cell divisions the proteins that wrap up the DNA molecule in packages or the protein-machines that synthesise proteins are required in such huge amounts that one gene would not be enough. The biological solution turned out to be many, many copies of identical genes. These genes are known as *high abundance genes*. The most dramatic abundance of genes is found in single-step output genes that specify structural RNA molecules. Ribosomal RNA, transfer RNA and small nuclear RNA are the final output of a single-step gene and mediate the flow of information from DNA into functional proteins. A multitude of identical RNA genes are required to fulfil the organism's demand for protein synthesis. For example, the human genome contains approximately 500 copies of identical genes of the ribosomal RNA molecules. Even more impressive are the 20'000 5S-ribosomal copies present in the genome of the frog *Xenopus laevis* [1].

#### *The paradox of multigene families*

It isn't hard to understand why mutations in single copy genes may subject an organism to a selective disadvantage. Single copy genes may specify proteins that fulfil essential roles in the organism. Mutations that change the protein will be either incompatible with life or subject the organism to a reproductive disadvantage. It is much harder to imagine that a mutation in one of the 500 copies of the ribosomal RNA genes would lead to a reproductive disadvantage. The ribosomal RNA genes are backed up by a multitude of copies and losing a few – or even a dozen – would not harm the organism. This observation raises a clear-cut problem for Darwinian Theory:

It is very difficult to imagine how a mutation in any single gene copy would endanger the survival of the organism. In fact, our biological

intuition rejects the notion that a defect in one out of 500 gene copies could have any significant effect on the fitness of the organism. Transcription of the remaining 499 unmutated copies would still produce 99.8 percent of the normal level of ribosomal RNA required to assemble functional ribosomes. Besides, if one mutated copy is harmless, why should two mutated copies or even 20 mutated copies, leaving 480 normal copies, be harmful? [1].

This excerpt from a molecular biology textbook hits the problem right on the head – How could high abundance gene families be retained in the genome without natural selection acting on the individual genes? Large multigene families would theoretically quickly accumulate mutations because the many copies of wild-type genes would protect the organism from deleterious changes in any particular gene copy. Individual genes of huge multigene families are expected to have accumulated mutations and a lot of distinct diverging gene copies should be found. In sharp contrast to this expectation, the multiple copies of most multigene families are identical. For instance, the 500 copies of the ribosomal RNA genes are virtually the same. The discrepancy between observations that genes within a multigene family are usually identical and the theoretical expectation that they should be distinct is known as the *paradox of multigene families*.

To solve the paradox, organisms must have a way for maintaining homogeneity among the members of repeated DNA sequence families. Evidence for a gene homogenisation mechanism was demonstrated for tandemly repeated genes encoding ribosomal RNA of the Baker's yeast, *Saccharomyces cerevisiae*. Tandemly repeated genes are like a row of identical beads that make a chain. Between each individual gene there is an intervening sequence that serves as a localised stimulator of recombination proteins – in other words hotspots for recombination. The precise mechanisms by which the intervening sequences induces homogenisation is unclear, but it is conceivable that the recombination hotspots ensure homogenisation of the repeated genes by the promotion of genetic conversions, unequal chromatid exchange, or through controlled gene amplification [2]. Multigene families that are not tandemly repeated are also found homogenous, however. In yeast

eight identical genes for the transfer RNA specific for tyrosine were identified. The eight genes are located on eight different chromosomes. The coding part of these genes and the 14 nucleotide intervening sequence are identical except at one position at each locus, although the flanking regions of the genes display no obvious sequence similarities. It is very hard to explain such localised homogeneity by a Darwinian selection mechanism.

[O]ur biological intuition would anticipate that the eight individual genes in this small multigene family would have diverged slightly from each other through the accumulation of various minor or neutral mutations. The absence of such divergence indicates that these eight genes are coevolving; that is, their sequences all change together [3].

Co-evolution implies that the genes must know what they look like, and what their identical twin-genes look like. How does a gene know that? And how does a gene know the other gene changed? Genes on different chromosomes that change simultaneously have to communicate with each other trans-chromosomally. It is more likely, that (recombination) proteins in conjunction with special RNA molecules specially designed for this task guard the sequence identity of the eight genes. The mechanism could be similar to those in used in software development for optical character recognition methods, which are known as *voting*. In this process a handful of engines are asked to express their opinion of what a particular word is. If a majority of engines agree on what the word is, it is concluded that they are right. A similar mechanism could be operating in the DNA, where all eight genes are constantly compared. If seven of the eight identical genes all had the same value, but not number eight, then the latter would be in error, and could therefore be set back to rights by making it the same as the other seven genes. If such mechanisms really exist, identical sequences change simultaneously as the result of non-random mutations<sup>1</sup>. It should be noted that gene homogenisation mechanisms prevent the increase of genetic information, and thus microbe-to-microbiologist evolution to proceed. Therefore, redemption of the paradox for multigene families through gene

---

<sup>1</sup> To test whether such mechanism exists it would be sufficient to induce point mutations into one of the genes and see if the mutations are repaired.

homogenisation mechanisms immediately introduces another paradox: *the paradox for the increase of information*. If homogenization mechanisms do not exist, high abundance genes must be of recent origin, otherwise they would have accumulated neutral mutations.

### *Genetic redundancy*

The discovery of the primary rules governing biology in the second half of the 20th century paved the way for a more fundamental understanding of the complexity of life. One of the spin-offs of this knowledge has been the development of sophisticated techniques to elucidate the function of proteins. When molecular biologists want to know the function of a particular human protein they genetically modify a laboratory mouse in such way that it lacks the equivalent gene. Theoretically, the phenotype of a mouse lacking specific genetic information could provide essential information about the function of the gene. To create such mice, molecular biologists put a selectable marker in the gene of interest in embryonic stem cells that have also been derived from the mouse<sup>2</sup>. The marker interrupts the gene of interest and a functional protein cannot be produced. Then, they inject the manipulated embryonic stem cell into a mouse's oocyte (egg cell), transplant the resulting zygote back into the uterus of pseudo-pregnant mouse and hope for the best. Among the offspring there may be individuals that carry one copy of the selective marker – and those have the gene of interest interrupted. Individuals carrying the interrupted gene can be sorted out through screening them for the presence of the selection marker. It is now fairly easy to obtain animals in which both copies are interrupted through selective breeding. Mendel's genetic laws ensure that crossbreeding two *marked* littermates (which have one interrupted gene) will produce offspring of which a quarter has two interrupted genes. This is because the reproductive cells – sperm and eggs – have only one copy of each gene, while the cells of the body have two. Only half the number of sperm cells contains the interrupted gene and this also holds for the

---

<sup>2</sup> To replace genes scientists make use of homologous recombination to exchange the genetic marker with the gene of interest. Homologous recombination is a protein-mediated exchange program that commonly occurs in the stem cells of the reproductive organs. It ensures that the daughter chromosomes exchange DNA in an orderly fashion to produce gametes. All gametes produced by this mechanism have reshuffled genetic information and are slightly different.

oocytes. Selective breeding is required to bring the two interrupted genes together in one animal. Mice that have both genes interrupted cannot produce the corresponding protein – they are called *knockouts*. Over the years, thousands of knockouts have been generated. The knockout-strategy revealed the functions of hundreds of genes and has contributed immensely to our biological knowledge. There was, however, one unexpected surprise – the *no-phenotype knockout*. If all genes have a selectable value, as it is believed by Darwinians, then all knockouts should have measurable, detectable phenotypes. Not so! Many knockouts were just doing fine. They did not show any phenotypic abnormalities and their reproductive success was not different from that of the wild-type. The no-phenotypes knockouts demonstrated that genes can be disrupted without any detectable effects on the phenotype. Many genes seem to have no measurable function. This is known as *genetic redundancy*.

#### *Molecular switches*

An intriguing example of genetic redundancy is found in the *SRC-gene* family. The family comprises a group of eight genes that code for eight distinct proteins with a function that is technically known as *tyrosine kinase activity*. The SRC proteins attach phosphate groups to other proteins that contain the amino acid tyrosine in a specific amino acid context. The result of this attachment is that the protein becomes activated; it is switched on, and can hence pass down information in a signalling cascade. Four closely related members of the family are named SRC, YES, FYN and FGR, the other related members are known as BLK, HCK, LCK and LYN. Both families are so-called nuclear receptors, and transmit signals from the exterior of the cell to the nucleus, the operation centre where the information present in the genes is transcribed into messenger RNA. The proteins of the *SRC-gene* family operate as molecular switches that regulate growth and differentiation of cells. When a cell is triggered to proliferate, tyrosine kinase proteins are transiently switched on, and then immediately switched off. The *SRC-gene* family is among the most notorious gene families known to man, since they cause cancer as a consequence of single point mutations. A point mutation is a change in a DNA sequence that alters only one single nucleotide – a DNA letter – of the entire gene. When the point mutation is not on a silent position, it will cause the organism's protein-making

machines to incorporate a wrong amino acid. The consequence of the point mutation is that the organism now produces a protein that cannot be switched off. Mutated *SRC*-genes are of particular danger because they will permanently activate signalling cascades that induce cell proliferation. The signal that tells cells to divide is permanently switched on. The result is uncontrolled proliferation of cells – cancer. Importantly, the cancer promoting point mutations cannot be overcome by allelic compensation, because a normal protein cannot help to switch off the mutated protein. Therefore, there are essentially *no point mutations allowed* in *SRC* proteins without severe phenotypic consequences. Amino acid changing point mutations in most, presumably all, of the *SRC*-genes can lead to uncontrolled cellular replication. To obtain more insight in the function of the other members of the *SCR*-gene family, knockout mice models have been generated for all of them. With unexpected results – four out of eight knockout models did not have a detectable phenotype [4]. Despite their cancer inducing properties half of the *SRC*-genes appear to be redundant. Without natural selection acting on them, the existence of the redundancy observed within the *SRC*-gene family has been explained as follows:

”In the redundant gene family of *SRC*-like proteins, many, perhaps almost all point mutations that damage the protein also cause deleterious phenotypes and kill the organism. The genetic redundancy cannot decay away through the accumulation of point mutations” [4].

This scenario implies that the *SRC*-genes are destined to reside in the genome forever. Point mutations that immediately kill, raises an intriguing origin question. If the *SRC*-genes are really so potently harmful that point mutations induce cancer, how could this extended gene family come into existence through gene duplication and diversify through mutations at the first place? After the first duplication the genes are not allowed to change because it will invoke a lethal phenotype and kill the organism through cancer. Amino acid changing mutations in the *SCR*-genes will permanently be selected against. The same holds true for the third, fourth and additional gene duplication. New gene copies are only allowed to mutate at neutral sites that do not replace amino acid in the

protein. Otherwise the organism will demise through tumours. Due to this purifying selection mechanism the duplicates should stay as they are. Yet, the proteins of the *SRC*-family are distinctly different, only sharing 60-80 percent of their sequences.

### *Histone H1*

When the cells of eukaryotes<sup>3</sup> divide, the DNA must first be packaged into ordered units to form the rod-shaped chromosomes we often see presented in journals. For packaging the DNA histone proteins are indispensable – at least, that used to be the view. Histone H1, for instance, is a linker protein that connects the nucleosomes and efficiently packages the DNA when eukaryote cells prepare for cell division. It surprised me that scientists had succeeded in generating viable histone H1 knockouts in several species of fungi, because in these organisms it is present as a single copy gene. Interruption of the H1 gene in the fungus *Ascolobus* did not affect the fungus' growth rate and viability during the first two weeks – neither did it affect reproductive success [6]. The only measurable effect of the interruption was a decreased life span. Investigators who produced a histone H1 knockout in the fungus *Aspergillus* reported that they were unable to detect any deleterious effect on the organism's sexual reproduction cycle; neither did the inactivation negatively influence the life span of this fungus. These data show that an inactive histone H1 gene does not immediately influence the fungus' fitness and is empirical evidence of the redundant character of the H1 gene.

Mammals have eight histone H1 genes. Here too the H1 proteins are used to link the nucleosomes and they comes in 8 different varieties, designated H1a, H1b, H1c, H1d, H1e, H1t, H1oo, H1o. H1a through H1e are only expressed in somatic cells, whereas H1 through H1o is merely expressed in germ cells. Scientists of the Albert Einstein College of Medicine, New York, interrupted two of them in a mouse model [7]. The experiment showed that single knockouts (mice only missing a functional H1o gene) developed normally – but remarkably the double knockouts also did not demonstrate signs of reduced fitness in the forms H1o(-)/H1c(-), H1o(-)/H1d(-), H1o(-)/H1e(-). The latter was surprising because the H1c, H1d and H1e normally substituted when the H1o was absent.

---

<sup>3</sup> Eukaryotes are organisms built of nucleated cells.

The double knockouts lack two functional H1 genes, but still they do not have detectable phenotypes. None of the H1 double knockouts exhibited phenotypic or histological abnormalities, and the packaging of DNA showed no significant changes. The double knockouts also did not show a reduced fertility or reproductive abnormalities. These results suggest that any individual H1 subtype is dispensable for mouse development and that loss of even two subtypes is tolerated if a normal H1-to-nucleosome stoichiometry is maintained [7]. Apparently, the distinct H1 genes form a backup of homologous proteins that are equally capable of doing the same job. These data beg the question how such system could be retained during millions of years of alleged evolutionary time without natural selection to act on it.

### *Natural Knockouts*

Angiogenin is a small protein that stimulates the formation of blood vessels in vertebrates; also in humans and mice. For a while, angiogenin had considerable attention in cancer research since it could be a target to treat tumours. The idea was as follows. Tumours need blood vessels to grow and for that purpose they rely on angiogenin. If we can stop the angiogenin supply, the tumour can't form vessels and dies because of lack of oxygen and nutrients. To get a feeling for the function of a certain gene in humans, the equivalent gene is usually interrupted in a mouse model, and then, by assuming one-to-one gene equity the function of a human gene is deduced from the phenotype of the knockout. To understand how angiogenin functions in a living system, cancer researchers would like to have a knockout model. Unfortunately, the genome of the mouse contains three copies of the angiogenin gene. Because the three copies form a natural back up, disrupting one of them will most probably not give an informative phenotype. To obtain an informative angiogenin-knockout mouse the three copies have to be interrupted all at once – very tough business. For making informative knockouts scientists prefer an organism that has only one copy of the gene of interest. Primates are such organisms. They only have one angiogenin gene, which makes them a much easier knockout model. Scientists do not have to make the angiogenin knockout in primates; however, it is already there.

The Douc Langur (*Pygathrix nemaeus*) is an Asian leaf-eating Colobine monkey and has only one copy of the angiogenin gene. The monkey's gene cannot produce a functional protein, however, because it has a one-nucleotide deletion mutation in the sixth codon of the mature peptide. This mutation generates a premature stop signal and causes that only truncated functionless angiogenin proteins can be produced. The Douc Langur is the natural knockout for angiogenin. The investigators who reported this in 2003 in the scientific journal *Gene* also reported that the same nucleotide deletion is found in five unrelated individuals, suggesting the inactivation event was recent and unique to the Douc Langure lineage. They concluded that this natural knockout experiment suggests that primate angiogenin is dispensable – even in the wild [8]. In other words, angiogenin is a redundant gene. How can a single copy gene be dispensable? If angiogenin is responsible for making new blood vessel, how can there be still Douc Langurs? It must appear that building new blood vessels is not only dependent on angiogenin; there must be additional proteins have the same or similar function as angiogenin that can also direct vessel formation. If the angiogenin gene is dispensable even in the wild, how on earth did the angiogenin gene evolve without natural selection?

#### *What do we need CCR5 for?*

Acquired immunodeficiency syndrome (AIDS) is caused by a small selfish molecule known as human immunodeficiency virus (HIV) and a world-wide pest that impacts the life of millions of people worldwide. But not all humans are threatened by HIV. Millions of Europeans are completely resistant to HIV infections and even more are partially protected. The immunity stems from their genetic make-up – the protected Europeans lack a functional gene that normally specifies a chemokine receptor known as CCR5. Biology is for a huge part about sending and receiving messages, so many proteins are messengers and many others are receptors to receive in the message. The receptors are usually located on the outer membrane of cells so that the messengers don't have to enter the cell; they can simply leave the message at the entrance where the receptors will take it over the message. Viruses exploit receptors on the cell surface to gain entry into cells. HIV uses the CCR5 receptor to enter two cells of the human immune system, T cells and macrophages. As soon as HIV is inside, it hijacks the cell's biochemistry to use it for its

own reproduction, which then leads to the destruction of the host cell. The mutation that protects Europeans from HIV infections is a deletion mutation that chopped off an essential part of the CCR5 receptor. Individuals who have inherited the chopped-off-receptor from both of their parents – the natural knockouts – completely miss the CCR5 on their cell surfaces and that is the reason why HIV cannot enter the immune cells. The natural knockouts are completely resistant to HIV infections.

The CCR5 mutation has been estimated to be no more than 4'800 years of age [9]. This is very young; in particular so, because the chopped-off gene variant makes up ten percent of all CCR genes in Europe. It can be calculated that a neutral mutation – a genetic change that has no selective advantage – would require 127 thousand years of dispersal to reach the high frequency in Europe [9]. The high incidence of the CCR5 variant therefore tells us that the CCR5 mutation must render an individual a reproductive advantage<sup>4</sup>. It couldn't be HIV, so what was it? Some scientists have advocated the idea that the infamous bubonic plague – the Black Death – pandemic is responsible for the high incidence, since it killed around one third of Europe's population between 1346 and 1352 and the Great Plague of 1665-1666, which killed around twenty percent of all Europeans. Others argue that, while the plague was responsible for heavy mortality, it did not generate sufficient natural selection to drive the chopped-off-gene to a frequency of 10 percent. Instead, it is more likely that a disease that persisted since the incipience of the mutation drove the increase in frequency. Researchers of the Yale School of Medicine argue that, like HIV, susceptibility to smallpox virus may have hinged on the presence of the CCR5 receptor. Smallpox epidemics occurred frequently, and the cumulative number of deaths during the last 700 years was greater than from plague. And, unlike plague, small pox disproportional infected children, and thus a typical small pox death removed greater reproductive potential than the average plague victim. Pan-European smallpox epidemics date back at least 2'000 years, enough to account for the high incidence of the chopped-off-gene. Small pox killed about every third person during millennia and a protective mutation would rapidly increase in frequency [9]. Whether plague or small pox

---

<sup>4</sup> Darwinians will argue that the CCR5 mutation has a selective advantage or renders the carrier increased fitness.

was the driving force behind the rapid increase of the inactive CCR5 is still a matter of debate and can only be ended in the laboratory when scientists test whether the CCR5 molecule provides a way of entering the host cells for either small pox or bubonic plague, or both.

But there is more to the story. A large number of CCR5 variants can be found in human populations; it demonstrates that this gene easily accumulates mutations and is not under stringent selective constraint. On the contrary – people expressing a functional protein are in danger of being infected in environments with HIV and other viruses. A primate species known as the Sooty or red-capped Mangabey (*Cercocebus torquatus*) that dwells the African rainforests are the natural hosts of simian immunodeficiency virus (SIV). SIV is a virus very similar to HIV and also requires the CCR5 receptor to gain entry to the host cells. Like a subset of Europeans, the Mangabeys never suffer from SIV infections; and, like the subset of Europeans, the Mangabeys carry a chopped-off CCR5 gene that renders them immune to SIV infections. Millions of Europeans and the whole population of African Mangabeys are natural knockouts that demonstrate that species have genes they do not immediately require for survival – genes that can easily be lost from the genome.

It should be noted that genetic redundancy is not confined to the higher organisms, and that the presented examples are only the top of the iceberg. The vast majority of the several hundreds of knockouts obtained in *Arabidopsis* do not give rise to visible, directly informative phenotypes. In a recent study, molecular biologists demonstrated that fewer than 2% of approximately 200 *Arabidopsis* knockouts displayed significant phenotypic alterations. Many of the knockouts do not affect plant morphology even in the presence of severe physiological defects [10]. Bacteria also keep many more genes in their genome than absolutely required. Scientists have constructed a strain of *Escherichia coli* bacteria that lacks eight percent of the wild type genome and reduced the number of genes by more than nine percent [11]. This huge deletion of several hundreds of genes does not affect the bacterium's reproduction rate on minimal medium. Hundreds of genes can

simply be removed from the bacterium's genome without an effect on its *fitness*. It confirms the non-essential, redundant nature of these genes.

### *Paradigm Lost*

The presented examples of genetic redundancy raises a few intriguing questions. A first important question to be addressed is – can we understand genetic redundancy from Darwin's natural selection perspective? If a gene can be knocked out without reproductive consequences for the organism, there is no reason to believe natural selection acts on it. Such genes qualify as neutral genes. Genetic redundancy can thus be defined as the situation in which the disruption of a gene is selectively neutral. Logically, redundant genes should rapidly evolve because there is no selective force that demands the maintenance of a particular DNA sequence. Darwin's selection hypothesis thus predicts a rapid divergence of redundant genes between isolated populations and between species. Likewise, selection hypothesis must predict that essential genes – genes that prove lethal when they are inactivated – should change more slowly. If protein evolution is due in large part to neutral and slightly deleterious amino acid substitutions, then the incidence of such mutations should be greater in proteins that contribute less to individual reproductive success. The rationale for this prediction is that relatively dispensable proteins should be subject to weaker purifying selection, and should accumulate mildly deleterious substitutions more rapidly. This argument was forwarded over twenty years ago, is fundamental to many theoretical applications of evolutionary theory, but despite intense scientific scrutiny the prediction has not been confirmed. The prediction has, however, been found to be untrue in many organisms. A systematic analysis of mouse genes showed that essential genes do not evolve more slowly than non-essential ones [12]. Likewise, *E. coli* proteins that operate in huge redundant networks can tolerate just as many mutations as single-copy unique proteins [13].

A second question raised by genetic redundancy is – how did organisms obtain genes that are not subject to natural selection? First, let's have a look how it is thought genetic redundancies arise. If genes duplicate fairly often, it is reasonable to expect some level of redundancy in most genomes. Susumo Ohno influential 1970 book *Evolution by Gene*

*Duplication* dealt with this idea [14]. As he saw it, gene and genome duplications are the principal forces that drove the increasing complexity during the evolution from microbes to microbiologists. Ohno proposed that duplications of genetic material provide the genetic redundancies that are completely free to accumulate mutations. Duplicated DNA elements are not subject to natural selection and free to transform into functional novelties. With time, a duplicate would accumulate mutations and diverge in its conditions of expression due to mutations in regulatory DNA elements, or in the specialized function of the specified protein due to mutations that change the properties of the protein. Duplicates that specify novel protein functions with a selective advantage would certainly be favoured by natural selection. Meanwhile, the genetic redundancy would protect old functions as new ones arise, hence reducing the lethality of mutations. Originally proposed by Ohno, diversification of duplicated genetic material is now the accepted standard evolutionary idea on how genomes gain useful information. Ohno's idea of evolution through duplication also serves as explanation for the no-phenotype knockouts, because duplicated genes provide an organism with back-up genes. Ohno's theory predicts an association between genetic redundancy and gene duplication.

Some biologists have looked into this matter specifically using the wealth of genetic data available for *Saccharomyces cerevisiae* – the common baker's yeast. The yeast's genome was one of the first to be entirely analysed and it was found that a surprising sixty percent of its genes could be inactivated without producing a phenotype – these genes are redundant. In 1999, Winzeler and co-workers reported in *Science* that only nine percent of the non-essential [=redundant] genes of *Saccharomyces* have sequence similarities with other genes present in the yeast's genome and could thus be the result of duplication events [15]. Over ninety percent of the non-essential genes of *Saccharomyces* is thus not related to genes in the yeast's genome and demonstrate that genetic duplications do not explain genetic redundancy. In a 2000 study, Andreas Wagner included all data available for the baker's yeast, including the completely genome sequence, all gene expression data, and the results from systematic knockout studies. His accurate analysis recapitulated Winzeler's original findings that weak or no-effect [i.e. redundant, PB] genes are no more likely to have paralogous [i.e. duplicated, PB] genes within the yeast genome than genes

that do result in a defined phenotype when they are knocked out. Wagner concluded that robustness of mutant strains cannot be caused by gene duplication and redundancy, but is more likely to be caused by interactions between unrelated genes [16]. Additional evidence for this view comes from the *Caenorhabditis elegans* genome project. The address the function of single- and multicopy genes in relation to robustness to mutations, the worm's genes were temporarily inactivated – not all genes at once, but one at a time. A surprising 89% of single-copy and 96% of duplicate genes showed no detectable phenotypic effect; the major part of the genes is redundant [17]. More recent studies have now confirmed that cooperating networks of unrelated genes contribute significantly more to robustness than gene copy number [18]. Together, these data show the predicted association between genetic redundancy and gene duplication are non-existing. In other words, Ohno's evolution by gene duplication is wrong.

To circumvent the problem that redundant genes are being lost rather than being transformed into new genes, the computer models that simulate Darwinian evolution require some degree of *symmetry breaking* between duplicated genes. In order to preserve duplicated genes the strict equality among the copies must be disrupted: one of the copies must relocate to a different part in the genome, preferably to a distinct chromosome or to the opposite end of a chromosomal arm. To maintain functional equality one must assume mutational inequality, while mutational equality necessitates divergent efficacy. In the very simplest mathematical models, identical genes must mutate at different rates and perform with different efficacies [19]. If mutations are merely introduced at random, as the Darwinians believe, different mutation rates of duplicated genes are not very realistic. An unequal crossover of chromosome arms results in two identical genes that are located immediately next to each other. One of the genes is now likely to become inactivated, unless mutation rates in genes are of a non-random character, meaning that one of both genes is more likely to accumulate mutations. This has been observed for genes that are located in different parts of the genome, for instance when one of the genes had a centromeric location and the other a telomeric, or when they are located on distinct chromosomes. In order to be preserved duplicated genes must move. The high abundance genes of the previous chapter showed that duplicates are unlikely to be preserved,

because of gene homogenisation mechanisms. Only by ignoring these mechanisms and assuming that weak selection acts on duplicated genes, evolutionary theoreticians are able to demonstrate that redundant genes might become fixed in a population [20]. These restrictions are not very realistic and the models are merely required to explain genetic redundancy in terms of natural selection, i. e. a duplicated gene must provide a direct reproductive advantage which could be favoured by selection. The remarkable observation on redundant genes was, however, that they are not subject to natural selection and as such can per definition not be addressed by natural selection.

### *Disbelief*

Darwinians are selectionists. All traits and characteristics of organisms are explained as the product of natural selection. The genetics that determine these traits must, therefore, also be explained through natural selection. From a Darwinian perspective genes cannot have evolved without natural selection, and truly redundant genes are impossible paradoxes. In their opinion, natural selection cannot prevent the inactivation of redundant genes, truly redundant genes cannot exist. It is impossible for a Darwinian to imagine genetic redundancy to be evolutionary stable and truly redundant genes should rapidly 'evolve'. Yet, redundant genes do not change faster than essential genes. And now several studies have demonstrated that genetic redundancy is *not* associated with gene duplications, predictions from standard evolutionary theory addressing genetic redundancy are clearly not in fact born out. Genetic redundancy is a Darwinian paradox and evolutionary theory is in need of a new paradigm. One thing we can be certain about; evolutionary philosophy is not going to abstain from Darwin's magic – natural selection. Diethard Tautz has already argued that weak selective forces are sufficient to shaping genomes. The *no-phenotype knockouts* can be taken as biological evidence for weak selective forces, he says, otherwise the no-phenotype knockouts would not exist [21]. It may be true that things are the way they are, but Tautz' tautology is begging the question. In fact, Tautz is trying to introduce the biological equivalent of the *anthropic principle*, a once popular paradigm in physics. It holds that the universe is as we see it, because if it were otherwise, we might not be here to observe it. In physics the anthropic principle has met with a lot of scepticism. Why should we accept a similar tautology in biology?

The no-phenotype knockouts are the reason why an increasing disbelief is taking hold of biologists that explain biological complexity from selection hypothesis. These biologists are Darwinians and unable to understand the existence of genes without natural selection. Now the new biology data are starting to reveal that many genes can simply be erased from the genome, Darwinians do not reject Darwin's principle of natural selection as a major evolutionary force, but rather doubt their observations:

"Genetic redundancies only reflect our inability to uncover conditions that would reveal the need for the redundant [genes]" [22].

The inability of integrating the phenomenon of redundancy into standard evolutionary theory is also strikingly reflected by embryologist and defender of the integrity of science Louis Wolpert's response to why we are not able to detect diminished fitness in knockouts:

"But, did you take it to the opera?" [22].

From a Darwinian point of view it can simply not be understood how genes can reside in the genome without selection. This is also clear from a statement in Nature by Mario Capecchi, a pioneer in developing the knockout technology:

"I don't believe that there is a single [knockout] mouse that does not have a phenotype. We just aren't asking the right questions" [22].

The right question here to be asked is: Was Darwin wrong? My answer to that question is: Yes, he was. Selection theory does not explain the observations of the new biology. Genetic redundancy that produces the robustness of living systems does not have a solution in the Darwinian framework.

When predictions from evolutionary theory are not in fact borne out, it is usually saved from falsification by the addition of *ad hoc* hypotheses that make it compatible with the facts. By this means a theory that was initially meant genuine scientific degenerates into pseudo-scientific dogma. We are at the dawn of a new era, where, to solve the problems thrown up by the new biology worrisome models are being introduced. Theoreticians include *symmetry-breaking*, *genetic uncertainty*, and *weak selection* to keep up Darwinian dogma. In actuality, *neutral selection* must be introduced to explain genetic redundancy in Darwinian terms. The utter mathematical models that account for genetic redundancy should be encountered with a lot of caution, however. In mathematics there is the so-called *sceptic stance*, which says that a mathematical result cannot just be translated into reality. What we need in biology is new fundamental concepts to understand the observations, not new mathematical models that accommodate these observations in the Neodarwinian framework. Knockout experiments will continue to throw up organisms that do not have a phenotype, and the reported no-phenotype knockouts are only the tip of the notorious iceberg. Many knockout organisms in which no phenotype could be traced ever see the light of day:

“[A] lot of those things [i.e. no-phenotype knockouts, PB] you don’t hear about. No-phenotype knockouts are negative results, and as such they are usually not reported in scientific journals; because they do not have news value. To illustrate the problem, the journal *Molecular and Cellular Biology* has since 1999 a section given over to knockout and other mutant mice that seem perfectly normal.” [22].

The no-phenotype knockouts may not be hot news, they certainly have added to our knowledge. They tell us that the selection hypothesis is false. Darwin’s theory which strongly relies on selection hypothesis is thus scientifically untenable.

### *Scale free networks*

The evident surprise of many biologists when confronted with the unexpected no-phenotype knockouts stems from, I believe, a profound lack of appreciation of non-linear

effects in biochemical systems. It is ironic that standard wall charts of biochemical reactions show hundreds of coupled reaction working together in networks, while graduate students are tacitly encouraged to think in terms of linear cause and effect. Linear cause-and-effect thinking roots in ancient Greek Philosophy, was adopted by nineteenth century European scholars, and is still dominating most fields of science, including biology. We cannot understand genetic redundancy and biological robustness in linear terms of single causality, where A causes B causes C causes D causes E. Rather, biological systems are designed as redundant scale-free networks. In a scale-free network the distribution of node linkage follows a power law, in that it contains many nodes with a low number of links, few nodes with many links and very few nodes with a high number of links. The internet is an example of a scale-free robust network: the major part of the websites makes only a few links, less make an intermediate number of links, whereas a minor part makes the majority of links. Usually hundreds of routers routinely malfunction on the Internet at any moment, but the network rarely suffers major disruptions. As many as 80 percent of randomly selected Internet routers can fail and the remaining ones will still form a compact cluster in which there will still be a path between any two nodes [23]. Individual links are not essential for letting the system function as a whole. We rarely notice the consequences of the errors that routinely occur in our cells, because proteins operate in scale-free networks. The protein networks found in biology are the result of hundreds of cooperating genes, which ensure that would one of them become inactivated by a mutation, essential pathways are not shut down immediately. The scale-free networks of biology have the ability to by-pass the functions of individual components of those networks and due to this buffering capacity provide robustness to organisms. Scale-free networks govern intracellular signaling and homeostasis. They also govern and fine-tune gene expression, regulate the maintenance of genomes and provide regulatory feedback on gene expression. Biological systems rely on scale-free networks that can incorporate small failures in order to withstand larger failures. Natural selection does not act on the individual nodes of a scale-free network and this is why we observe genetic redundancy. Genetic redundancy not associated with genetic duplications provides independent evidence for a Grand Omnipotent Designer –

the great engineer that constructed multipurpose genomes with a high level of buffering capacity to make sure life could propagate in time.

## References

- [1] Quoted from: Molecular biology of the gene. The Benjamin/Cummings Publishing Company, Inc, California, 1987, page 659. ISBN 0-8053-9614-4.
- [2] Keil RL, Roeder GS. Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of *S. cerevisiae*. Cell 1984, volume 39, pages 377-386.
- [3] Quoted from: Molecular biology of the gene. The Benjamin/Cummings Publishing Company, Inc, California, 1987, page 660. ISBN 0-8053-9614-4.
- [4] Toby JG, Spring J. Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. Trends Genet 1998, Volume 14, pages 46-9.
- [5] Martin A. Is tetralogy true? Lack of support for the “One-to-Four-Rule”. Mol Biol Evol 2001, volume 18, pages 89-93.
- [6] Ausio J. Are linker histones (histone H1) dispensable for survival? BioAssays 2000, volume 22, pages 873-7.
- [7] Fan Y, Sirotkin A, Russell RG, Ayalla J, and Scoultchi AI. Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. Mol Cell Biol 2001, volume 21, pages 7933-43.
- [8] Zhang J, Zhang YP. Pseudogenization of the tumor-growth promoter angiogenin in a leaf-eating monkey. Gene 2003, volume 308, pages 95-101.
- [9] Galvani AP, Novembre J. The evolutionary history of the CCR5-Delta32 HIV-resistance mutation. Microbes Infect 2005, volume 7, pages 302-9.

- [10] Bouche N, Bouchez D. Arabidopsis gene knockout: phenotypes wanted. *Curr Opin Plant Biol* 2001, volume 4, pages 111-7.
- [11] Kolisnychenko V, Plunkett G, Herring CD, Feher T, Posfai J, Blattner FR, Posfai G. Engineering a reduced Escherichia coli genome. *Genome Res* 2002, volume 12, pages 640-7.
- [12] Hurst LD, Smith NGC. Do essential genes evolve slowly? *Curr Biol* 1999, volume 9, pages 747-50.
- [13] Hahn MW, Conant GC, Wagner A. Molecular evolution in large genetic networks: Does connectivity equal constraint? *J Mol Biol* 2004, volume 58, pages 203-11.
- [14] Ohno S. *Evolution by Gene Duplication*. Springer, New York, 1970.
- [15] Winzeler EA *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 1999, volume 285, pages 901-6.
- [16] Wagner A. Robustness against mutations in genetic networks of yeast. *Nat Genet* 2000, volume 24, pages 355-61.
- [17] Conant GC, Wagner A. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc Biol Sci* 2004, volume 271, pages 89-96.
- [18] Kitami T, Nadeau JH. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nat Genet* 2002, volume 32, pages 191-4.
- [19] Krakauer DC, Nowak MA. Evolutionary preservation of redundant duplicated genes. *Semin Cell Dev Biol* 1999, volume 10, pages 555-9.

[20] Krakauer DC, Plotkin JB. Redundancy, antiredundancy, and the robustness of genomes. Proc Natl Acad Sci USA 2002, volume 99, pages 1405-9.

[21] In: Tautz D. A genetic uncertainty problem. Trends Genet 2000, volume 16, pages 475-7.

[22] Quoted from: Pearson H. Surviving a knockout blow. Nature 2002, volume 415: pages 8-9.

[23] Barabasi A, Bonabeau LE. Scale-free Networks, Sci Am 2003, volume 288, pages 60-9.