**Evolution by the Gene Duplication**

Gene duplication is a fundamental process in evolution for providing new genetic materials which contribute for the emergence of novel and adaptive traits. Classical evolutionary models proposed by Ohno (1970) stated that the emergence of novel functions of gene duplicate (paralogs) is due to relaxed selective constraint following duplication, in which that natural selection could not effectively remove mutations in paralogs due to functional redundancy with ancestral copies, thus the mutations accumulate through genetic drift. In this model, two alternative pathways are possible following period of relaxed selective constraint; (1) pseudogenization in which that degenerate mutations will cause loss of function on the most of paralogs and (2) neofunctionalization in which that change in selective pressure in environment confer mutations in few paralogs become beneficial and facilitate for the emergence of a new (and possibly unrelated) functions (Dyhkhuizen–Hartl model of adaptation). This model of gene duplication is exemplified in the evolution of antifreeze protein in Antartic icefish. The ancestral of this paralog is SAS (Sialic acid synthase, SAS gene) with minimal ability for ice binding. Following duplication, paralog acquire mutations which eliminate the enzymatic function of the protein and allowing the optimization of the antifreeze functionality [(Deng et al., 2010](http://www.pnas.org/content/107/50/21593.full.pdf)).

In other evolutionary model of gene duplication, termed as subfunctionalization, the ancestral functions are distributed among paralogs. This model differs in three aspect with neofunctionalization (Lynch, 2004); the functions is diversified before duplications (instead of emergence after duplication), all mutations is degenerative and there are equal evolutionary rates between ancestral and paralog. In this model, following duplication, degenerative mutations accumulate independently between ancestral and paralog and make the ancestral functions suboptimal in each locus, thus the ancestral and paralog is required to preserve the ancestral function and thus, the gene duplicates is retained. Thus, the duplicates undergo three phases before establishment; duplication-degeneration-complementation (DDC). Organ specialization of Pax6 duplicates (Pax6a in brain and Pax6 in pancreas) in Zebrafish is proposed due to sub-functionalization from of ubiquitous expression pattern of ancestral genes [(Kleinjan et al., 2008)](http://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.0040029&type=printable). Another mechanism of subfunctionalization proposed by Proulx SR, Phillips (2006) is that gene duplication is a way to avoid segregational load. When, loci is under heterozygote advantage, fitness of population will reduce when the individual is homozygous. Duplication could ensure population could attain permanent heterozygosity for the locus subject to the balancing selection.

Paralogs might also have similar functions with the ancestral copies (functional conservation) and this could be explained by redundancy and dosage. Gene duplication creates functional redundancy so that the paralogs could act as buffer for harmful mutations. Maintenance of paralog could also be caused by the paralogs are selected for increasing product for gene expression. Numerous paralogs in ribosomal RNA (which ~400 duplicates exist in human, [(Eickbush & Eickbush, 2007))](http://www.genetics.org/content/genetics/175/2/477.full.pdf) is thought to be retained as it promote more gene product.

Gene duplication could also resolve the selective constraint on the pleiotropic loci. If a locus is responsible for promiscuous functions, improvement of one function is often constrained on the detrimental effect on the other functions (called as antagonistic pleiotropy). Duplication causes the paralogs to specialize on the narrower function and thus, the functional optimization could be done independently among paralogs. Thus, the duplication is a way to escape from adaptive conflict in loci with numerous functions or termed as EAC model. [(Hittinger & Carroll, 2007)](https://www.nature.com/nature/journal/v449/n7163/full/nature06151.html) suggested that paralogs of GAL gene in *Saccharomyces cerevisease* evolve under this model. The ancestral protein of this gene performed bifunctional function; as regulatory protein and as an enzyme. Duplication results in the separation of function where the regulation is only performed by one duplicate (GAL1) and enzymatic function by another paralog (GAL3), allowing for tighter regulation of galactose metabolism.

**RNA interference pathways act as a defence against intra-genomic parasites and virus**

RNA interference is a collection of molecular pathways which unified by the use Argounate family protein to modify complementary nucleic acid using small RNA (approximately 20-30 nucleotides) as a guide. The basic pathway involves the recognition of double stranded RNA (dsRNA) by Dicer family protein (a type of RNAse III family protein) which then ‘diced’ dsRNA into small RNAs known as short-interfering RNAs (siRNAs). siRNAs are then incorporated into Argounate-containing complex and one RNA strand is released forming active RISC (RNA-induce silencing) complex. The loaded siRNAs is then used for sequence specific binding for degradation of RNA target. There is vast diversity of these pathways, which differ from the source of dsRNA (endogenous vs exogenous), fate of the target after complementary binding (transcriptional inhibition, heterochromatin formation or degradation) and its biological function. However, generally, these board arrays of pathways are divided into three major subpathways based on the source of the short RNAs; miRNA, viRNA and piRNA.

miRNA pathway is initiated by the expression of genome-encoded RNA, which can form a foldback hairpin structure. This structure is then recognized by Drosa-Pasha (Dicer family protein), which processed the RNA into short RNAs (22-23 nts). The short RNAs is loaded into RISC containing Argonaute 1 for regulating the gene expression and developmental processes. Meanwhile, the viRNA pathway is responsible for an antiviral immunity. This pathway is triggered by the presence of exogenous viral dsRNA which upon recognition by Dicer 2, it is processed to produce 21-24 nts long RNA and used as guide by Argounate2 to degrade viral genome in cytoplasm.

Regarding the piRNA pathway, the pathway is differentiated from the first two by the Dicer-independent short RNA processing which employ Aub-Ago3 to produce longer short RNAs (24-29 nts). The short RNAs is transcribed from non-functional Transposon element which trapped in the genomic region known as piRNA cluster. This small RNA is then processed through positive feedback loop, where the transcripts are cleaved alternately by Ago3-Aub (known as ping-pong cycle) and results in signal amplification. The antisense short RNAs strand from this ping-pong cycle is then loaded into Piwi which silence the gene target through heterochromatin formation. This piRNA pathway is only found in germ-line and associated somatic cells (with lack of ping-pong mechanism) and responsible for maintaining the host genomic stability from mobile genetic elements (e.g Transposon Element, TE).

**RNA interference pathway exhibits dynamic evolution across multiple taxa**

There is a conflict of interest between host and parasite relationship. Infection causes disease (and possibly reduces fitness) and host try to eliminate them, but the parasite intends to use the host for replication. RNA interference is one of the defence mechanism against virus, but the effectivity of this system lessen overtime as the virus develop mechanism to counter the defence. Virus is known able to produce VSR (viral suppressor of RNAi) which can inhibit host RNAi pathway in various stages. For example, P19 from Tombus virus could interfere with dsRNA loading into Argounate and prevent RISC assembly [((Park et al., 2004))](http://www.sciencedirect.com/science/article/pii/S0042682204001230?via%3Dihub) and PO from Poleroviruses facilitate degradation of Argounate through ubiquitin-proteosome pathway [((Baumberger et al., 2007)](http://ac.els-cdn.com/S0960982207018568/1-s2.0-S0960982207018568-main.pdf?_tid=be9b4b36-7048-11e7-b6f4-00000aab0f01&acdnat=1500884519_fd2797275f2b86bdedfefe493dc59a1b). This viral adaptation in turn will select for more effective RNA interference component. Thus, host and virus is locked in an reprocical adaptation which result in an evolutionary-arm race between host and parasite. This molecular arms race is reflected in the rapid adaptive evolution among the host RNAi machinery. ([Obbard et al., 2006)](http://www.cell.com/current-biology/abstract/S0960-9822%2806%2901208-5) identified that three components of viral-RNAi (Dcr2, R2D2 and Ago2) as the top 3% fastest evolving protein in *Drosophila* and further research shown that selective sweep has occurred in the 100 kb surrounding Ago2 genomic region [(Obbard et al., 2010)](https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/mbe/28/2/10.1093_molbev_msq280/2/msq280.pdf?Expires=1500979103&Signature=I569eC0kE6cBmazY3C7P6Rkxu1xIlWklpobx77QT1h8mvDRgFBT8qHRQjIKDDjpre2kGG2shovEu-nXlo8Oewq~gSnd3ccX65~vQGZl~KMzpFx4T-eMVsI0bT0QhnTR6xiUnXV5oTwQApvMpRvyYvHamBNOSaDMvbX4nbLjRwqmka5De~RY4UCpZY3OeXkCJLQ5SNK4nYND7Qofirk4uwc-bvMxPyxoyAhI9ZVijP1dqi-SloVIZpYEYL-F57F125VRfElncuE~2k2rXAjCnd9DjD8lAjYB8H97eZiTq~TiaZmnDE-~pg0bKNl~zkYlkaYFRPwEgIpklUEYirgYqgw__&Key-Pair-Id=APKAIUCZBIA4LVPAVW3Q). Interestingly, there is no evidence of rapid evolution among component with similar function in miRNA (which is not engage in coevolution with parasites), suggesting that the virus is the driver of the rapid evolution. Recent research by [(Palmer et al., 2017)](http://www.biorxiv.org/content/early/2017/06/23/154153.full.pdf+html) shown that this pattern is consistent across invertebrates.

Rapid adaptive evolution of RNAi is also reflected in the frequent gene gain, losses and functional divergence. For example, Piwi is present in the ancestral eukaryotes, but separately lost in fungi, nematode and plants. Loss of Piwi in Nematode is accompanied by extensive duplication of Argounate where 18 Argounates duplicates has been identified (known as WAGO, worm specific Agos) and associated with a novel 22 small RNA Pathway (22 WGs) with highly derived functions such as for environmental sensing and transgenerational epigenetic. Expansion of Piwi in Aedes also mediates the paralogs to acquire novel role as antiviral defense in somatic tissue, besides of the ancestral function for TE suppression in germline. In addition, regeneration in Planaria is partly regulated by two a Piwi paralogs (smedwi2-smedwi3), which give further evidence of the functional divergence following RNAi gene duplication.

RNA interference pathway might also be lost and compensated by other pathways with similar function. For example, interferon-based pathway might take role as antiviral protection following the loss of RNAi pathway in mammals, but recent research shown that the RNAi is retained in the cells that lack of interferon-mediated antiviral protection. Interestingly, ancestral pathway could compensate when the derived RNAi pathway is lost as exemplified in the several lineages of Nematodes which lost the piRNA pathway. In these lineages, transposon control is reverted to use the ancestral version of the pathway which employs Dicer-RdRP mediated DNA methylation.

**Expression profile and adaptive evolution as signatures of functional diversification**

Gene expression profile could be informative for the evolution of duplicate genes. Before duplication, the pattern of gene expression could indicate the potential of the gene duplication (duplicability). The genes with ubiquitous expression pattern tend to be duplicated less than the tissue-specific genes as observed in *Drosophila*. In the period of post-duplication, there is also asymmetrical pattern of expression between ancestral and paralog as there is tendency that the duplicates become specifically expressed in testis. For example, [((Betrán, et al., 2002.)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC187566/pdf/T06X.pdf) observed that the expression of 50% *Drosophila* genes that retro-transposed from X chromosome to autosomal were only found in testis . Interestingly, the testis specific expression is not permanent, in which that the expressions expand into other tissues as the paralogs age. [(Kaessmann, 2010)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2945180/) proposed a ‘out of testis’ hypothesis to explain this phenomenon. This hypothesis stated that testis is the catalysis for the emergence of the novel traits from paralogs. Open chromatin configuration and strong selective pressure in testis is thought able to prevent pseudogenization and thus, facilitate functional innovation of the paralogs.

**Pathway duplication could be indicated by synchronous duplication**

Interestingly, the piRNA pathway might have diverge among closely related species[. (Kelleher, Edelman, & Barbash, 2012)](http://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.1001428&type=printable) observed that the F1 hybrid from *Drosophila melanogaster* - *Drosophila simulans* displayed characteristic of piRNA effector protein mutant with an elevated transposon activity. The incompatibility of the piRNA pathway (even in sister species) indicates that the pathway evolves rapidly and that the piRNA machinery might have been adapted at species level.

Genes that have higher connectivity tend to be less duplicated than genes that do not interact. The reasoning might be that the interacting genes should maintain dosage-balance with other gene in the pathway and the duplication disrupts the dosage balance. However, when the force of selection for new functions is strong, the duplication of the component in the pathway could become the selective pressure for the rest of the pathway which results in pathway duplication.

**Dissertation aim**