Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Animal

Genomic mechanisms of gene expression regulation in the *Squalius alburnoides* hybrid complex: insights from microRNAs and DNA methylation

Joana Gonçalves Fernandes Pinho

Mestrado Biologia Evolutiva e do Desenvolvimento

2011

Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Animal

Genomic mechanisms of gene expression regulation in the *Squalius alburnoides* hybrid complex: insights from microRNAs and DNA methylation

Joana Gonçalves Fernandes Pinho

Dissertação orientada pela Doutora Maria Ângela Inácio

Mestrado Biologia Evolutiva e do Desenvolvimento

2011

Contents

Agradecimentos

À Doutora Maria Ângela Inácio pela paciência e por me ter guiado durante a elaboração desta Tese de Mestrado.

À Professor Maria Manuela Coelho pelo constante interesse na evolução deste trabalho.

À Maria Ana, Ana Rita e Mónica Silva por estarem disponíveis para a discussões de ideias e pela partilha de conhecimento do trabalho em laboratório.

Ao Max pelas discussões sobre o controverso mundo da quantificação de DNA.

Ao grupo *S. alburnoides*, Tiago Jesus, Miguel Machado, Miguel Santos e Isa Matos, pela útil partilha de conhecimentos, discussões e sugestões, pela boa disposição e animadas saídas de campo. Ao Tiago pelas sessões de "brainstorming" no autocarro, ao Miguel M. pelas suas acesas discussões e úteis limpezas, ao Miguel S. pelos seus ensinamentos de citometria de fluxo e tormento da vida das pessoas e finalmente à Isa Matos pela valiosa discussão e entusiasmo contagioso no estudo do *S. alburnoides*.

À Joana Mateus pelo apoio, incentivo e companheirismo nos melhores e nos piores momentos.

Aos colegas do Laboratório de Genética pelo apoio e compreensão nas horas mais difíceis.

Ao grupo de Biologia do Desenvolvimento pela constante preocupação e amizade demonstradas.

À Dona Branca pela sua disponibilidade e boa disposição.

À Doutora Patrícia Pereira do CESAM pelo precioso auxílio prestado na análise de microarrays e ao Doutor Luca Comai pelas dicas sobre RNA-seq.

Ao Doutor Francisco Enguita pela disponibilização do Bioanalyzer e à Doutora Ana Tenreiro pela disponibilização do leitor de fluorescência de microplacas.

Aos meus amigos pela força que me deram mesmo quando ao faltava aos eventos sociais por motivos de força maior.

Aos meus pais pelo afecto incondicional em todos os momentos, especialmente os mais complicados quando tiveram de aturar o meu mau humor.

À minha família pelo apoio constante nesta complicada etapa.

Ao Tiago Rio por estar a meu lado quando eu mais precisei.

A mim por ter conseguido superar esta fase sem endoidecer.

RESUMO

Organismos poliplóides são organismos caracterizados pelo aumento do número de cópias genómicas. Estes funcionam como reservas de variação latente que ao se tornarem visíveis à selecção lhes confere vantagem evolutiva. A evolução das plantas deve-se grandemente a este fenómeno, uma vez que as novas combinações geram novos fenótipos e portanto rápida evolução adaptativa. Mesmo em animais, embora mais raros há casos de taxa poliplóides.

Apesar disso, estes organismos estão sujeitos a uma maior instabilidade genómica, seja por aumento da carga genética de um só genoma devido à sua endoduplicação (autopoliplóides) ou por aumento desta carga através da hibridação de genomas distintos (alopoliplóides). Desta forma, para que o organismo se mantenha equilibrado a nível funcional esta instabilidade tem de ser ultrapassada.

Por outro lado, tem-se verificado em plantas poliplóides que esta amplificação não é linear, ou seja, a quantidade de genoma não é proporcional à sua expressão génica. Por exemplo, foi demonstrado em milho poliplóide a existência de um efeito de dosagem na expressão génica. Actualmente sabe-se que este efeito depende grandemente da re-organização do genoma. Esta organização pode ser não só genética, mas também epigenética.

Um excelente organismo para o estudo dos efeitos da poliploidização do genoma em animais é o complexo *Squalius alburnoides*. Esta espécie, pertencente à Família Cyprinidae, foi descrita inicialmente como tendo três diferentes ploidias no mesmo complexo: diplóide, triplóide e tetraplóide.

Actualmente sabe-se que este complexo tem uma origem híbrida resultante de um cruzamento entre um ancestral materno de *S. pyrenaicus* (genoma P) e de um ancestral paterno semelhante ao actual *Anaecypris hispanica* (genoma A). *S. alburnoides* encontra-se actualmente distribuído por toda a Península Ibérica. Em Portugal, a distribuição do complexo ocorre nos rios de Norte ao Sul do país, mas enquanto no Sul há incorporação do genoma P por interacção com a espécie *Squalius pyrenaicus*, no Norte o complexo incorpora o genoma C, uma vez que interage com a espécie *Squalius carolitertii*. Esta incorporação resulta das diferentes formas de reprodução do complexo que dão origem a diferentes combinações de genomas e ploidias.

Em vertebrados, Pala *et al.*, (2008) descobriu pela primeira vez que nas formas triplóides do complexo *S. alburnoides* existia silenciamento génico em alguns alelos envolvendo também um mecanismo de compensação de dose na expressão génica de alguns dos genes investigados. Consequentemente colocou-se a questão se o silenciamento seria num haploma inteiro ou não. Observou-se que nos indivíduos híbridos triplóides havia contribuição diferencial de genomas, ou seja, em diferentes genes verificou-se a ocorrência de duas situações diferentes: exclusiva expressão do genoma A ou expressão bialélica (PA). Não havendo portanto silenciamento de uma cópia de genoma inteira. Adicionalmente, ao comparar vários genes diferentes percebeu-se

que havia preferência para o silenciamento das cópias de genoma P, verificando-se que o silenciamento não ocorria completamente ao acaso. É possível que nos híbridos de *S. alburnoides* este mecanismo actue de forma plástica, podendo haver alelos que escapem aos vários mecanismos de regulação. De uma forma geral, em *S. alburnoides* os constrangimentos genéticos em poliplóides são ultrapassados mimetizando um contexto diplóide. No entanto, permanecem desconhecidos que mecanismos estão subjacentes à regulação da expressão e ao silenciamento diferencial dos alelos.

Em plantas, já foram estudados variados processos de regulação génica. Entre estes encontram-se a metilação ao nível do DNA e modificação das histonas que dependendo da sua localização desempenham um importante papel na regulação da estrutura da cromatina. A metilação envolve modificações covalentes no material genético das células. Eucariotas superiores possuem este mecanismo de regulação que actua adicionando um grupo metil às citosinas dos dinucleótidos CG, com a finalidade de impedir a transcrição. Uma vez adquirido um determinado padrão de metilação, os tecidos mantêm-no nas suas células ao longo de vários ciclos celulares devido à actividade da metiltransferase que metila as novas cadeias de DNA sintetizadas. Inclusive, pelo menos em plantas, sabe-se que esse padrão pode atravessar diversas gerações.

Estudos anteriores mostram uma relação entre plantas alotetraplóides, tanto sintéticas como naturais, e modificações por metilação no genoma. Por outro lado, diversos estudos evidenciam que, embora bastante conservados, os microRNAs variam a sua expressão mesmo em espécies próximas. Além disso, mostrou-se estarem envolvidos nas variações da regulação génica de plantas alopoliploides, não havendo expressão aditiva. Os *small* RNAs foram inicialmente identificados em *Caenorhabditis elegans* e fazem parte de uma classe de RNAs não codificantes tendo normalmente ≈22 nucleótidos de comprimento. Actualmente, já estão descritos diversos tipos de RNA não codificantes para além dos microRNA que possuem também uma função regulatória. Embora estas classificações ainda sejam relativamente recentes, pode-se afirmar com clareza que ambos os miRNAs e os siRNAs (também designados por tasiRNAs) são processados pelo mesmo complexo com o intuito de regular o silenciamento génico. Desta forma, o produto final da sua biogénese é responsável tanto pela degradação do transcrito como pela repressão de tradução. Assim, poderemos questionar se nos organismos poliplóides, este ou outro tipo de RNAs não codificantes, poderão estar a controlar a expressão génica de forma a permitir a este tipo de organismos superar a instabilidade causada pelo aumento da dose génica, ou até pela junção de dois genomas diferentes.

Tendo em conta os mecanismos já descritos em plantas, torna-se imperativo tentar perceber de que forma estes terão igualmente um papel decisivo na regulação génica que ocorre em vertebrados, mais especificamente no complexo *Squalius alburnoides.* Para tal, usaram-se indivíduos AA, PA e PAA do complexo e ainda PPs que foram analisados para esclarecer duas

possibilidades: a regulação por microRNAs e a regulação por metilação de DNA. Enquanto que o primeiro requereu uma análise através de Sequenciação de Nova Geração e microarrays, a segunda abordagem usou endonucleases com diferentes sensibilidades à metilação e um ensaio imunológico com afinidade para DNA metilado.

Após um extensivo controlo de qualidade das bibliotecas sequenciadas, verificou-se que *S. pyrenaicus* possuía um perfil de expressão mais distante do que os indivíduos pertencentes ao complexo e que estes últimos possuíam os seus miRNA mais expressos do que PP, o que seria de esperar visto fazerem parte de um complexo alopoliplóide, necessitando portanto de maior regulação. A técnica de hibridação em microarrays revelou ter um padrão de expressão bastante semelhante ao obtido nos dados da sequenciação, apesar do seu elevado *background* e do facto da análise individual da expressão de miRNAs na maioria dos casos não se mostrar reprodutível com os resultados obtidos através da sequenciação.

Além disso encontraram-se ainda sequências de 33 nucleótidos mais expressas em triplóides (PAA) provavelmente sendo algum tipo de *smallRNA* ainda não identificado, mas que poderá ter um papel importante da compensação de dosagem.

Por outro lado, a metilação de DNA mostrou ser significativamente diferente e elevada nos indivíduos triplóides em comparação aos outros indivíduos analisados do complexo. Adicionalmente, verificou-se o expectável para as diferenças nos níveis de metilação de DNA entre órgãos tendo em conta a diversidade de genes a serem expressos em cada caso, sendo o fígado o tecido menos metilado e as células sanguíneas o mais metilado.

No geral, os resultados deste estudos apontam para a existência de um mecanismo regulatório pré-transcricional por metilação de DNA nas formas triplóides, ao invés de um controlo pós-transcricional massivo. De uma perspectiva evolutiva, não faz sentido transcrever genes que irão depois ser regulados por um qualquer mecanismo pós-transcricional. Tendo isto em conta, faz sentido que os triplóides PAA regulem a sua expressão génica de forma a obterem uma expressão semelhante à de um diplóide, preferencialmente através do mecanismo de metilação de genoma e usando apenas os miRNA para pequenos ajustes de expressão dirigidos a específicos transcritos alvo.

Palavras-chave

Squalius alburnoides, allopoliploidia, regulação da expressão génica, metilação de DNA, microRNAs

Abstract

The cyprinidae *Squalius alburnoides* is an allopolyploid complex that inhabits the Iberian Peninsula rivers. Its hybrid origin and the vast variety of reproductive mechanisms generate different genomic compositions, including diploids (AA, PA) and triploids (PAA) forms.

Using *S. alburnoides* as her model, Pala *et al*. (2008) showed for the first time in vertebrates that, as a response to the increase in the genome dosage, triploids could regulate gene expression to a diploid state. Besides, several alleles were showed to be silenced in the triploids. However until now, the mechanism of gene regulation was not described.

Previous studies in plants demonstrate that silencing in allopolyploids could be explained by several epigenetic mechanisms. For instance, DNA methylation plays a major role in the chromatin remodelling and microRNAs seem to regulate gene expression post-trasncriptionaly. Therefore, in order to find out if these mechanisms have an important function in the regulation of gene expression in the *S. alburnoides* allopolyploid complex, they were investigated. MicroRNAs profiles were analysed by next generation sequencing and microarray hybridization while genomic DNA methylation was analysed by an immunoassay protocol and using methyl sensitive restriction analyses. Thus, results showed that the post-transcriptional regulation by non-coding microRNAs does not seem to play a major role in gene expression regulation of the complex. On the contrary, this regulation seemed to be mainly accomplished by the pre-transcriptional mechanism of DNA methylation since this work demonstrated that it is predominant in the complex, mainly in the triploid form PAA.

Thus, methylation should function as a primary step of regulation in order to buffer the polyploidy effect and, only then, microRNAs should be used as a tool to fine-tuning the allopolyploid expression of some essential genes is particular.

Keywords

Squalius alburnoides; allopolyploidy; gene expression regulation; DNA methylation; small RNAs

1. INTRODUCTION

1. 1. Polyploids and Hybrids

Evolution resulted occasionally due to changes occurred in chromosome number and genetic material. Polyploidy is characterized by the multiplication of the entire genome and provides a source of variation visible to natural selection that offers a rapid adaptative evolutionary potential and thus, higher evolutionary success.

Polyploids can be classified in two groups, autopolyploids and allopolyploids, depending on the origin of their chromosomes and the mechanism involved in their formation. The first, possess multiple chromosome sets derived from the same species once it emerged from a natural genome doubling either from the same individual or from different individuals of the same species. The second arise when hybridization occurs, usually between different species, producing a polyploid composed by chromosomes derived from different species (reviewed in Otto, 2007).

This phenomenon occurs mainly in a large range of plants, since sex determination and gene dosage imbalance restricts the existence of polyploidy in animals (Orr, 1990). Though less common, stable and well adapted polyploids have also been identified in some animal taxa such as amphibians and reptiles (reviewed in Otto & Whitton, 2000), fish (Leggat & Iwama, 2003) and even mammals (Gallardo *et. al*., 2006).

Natural hybridization was reported to play an important role in evolution, through introgression and hybrid speciation (Mallet, 2007). This phenomenon is particularly distributed through fish and positively correlated with polyploidy (Le Comber & Smith, 2004). Former studies, namely in *Cyprinidae* and in *Cobitidae* families (Gromicho & Collares-Pereira, 2007), reported allopolyploidization as the prevalent state of polyploidization presented in fish.

Many disadvantages and advantages of polyploidy have been extensively debated. For example, the emergence of asexual reproduction could give rise to less variability since it precludes the occurrence of recombination, however, asexual reproduction makes the encounter of sexual mates become needless which enhance the probability of the emergence of a next generation in a shorter period of time (reviewed in Comai, 2005). Additionally, heterosis, which is the maintenance of heterozygosity, originates hybrid vigour and possesses a positive effect in polyploidy success. Besides, beyond masking recessive (deleterious or not) alleles, gene redundancy also allows genes to undergo subfuncionalization or neofuncionalization.

Nevertheless, regarding the functionality of stable polyploids, a whole new level of complexity emerges where all the cellular machinery has to be adjusted to the new environment within the cell, particularly when dealing with allopolyploids. So, in order to achieve that and overcome the genetic constrains, polyploidization is accompanied by genetic as well as epigenetic alterations (Riddle & Birchler, 2003).

2.Genetics and epigenetics in polyploids

The response to the increase in gene dosage or the gathering of two heterologous genomes which contribute unequally to the regulatory networks, is an indispensable step to achieve a functional cell. Without an appropriate genome reorganization of the resulting polyploid, it could become unviable due to genomic instability. Innumerous studies about several genetic and epigenetic arrangements in polyploid plants have been described. Among others, the genetic consequences of polyploidization could be the mobilization of transposable elements (TE), already verified in plants, such as in the hybridization between rice and wild rice (*Shan et al*., 2005) or in the three hybrid species of sunflower (*Ungerer et al*., 2006). In animals, transposable elements activation was also verified, namely in interspecific hybrid between two species of Australian wallaby (*O'neill et al*., 1998). Loss of other genetic material (*Liu et al*., 1988), as well as gene conversion (*Wendel et al*., 1995) are other known genetic consequences of polyploidization.

On the other hand, it has been already reported that the genome amplification verified in polyploidy plants is not proportional (non-additive) to its gene expression (reviewed in Comai, 2005). For instance, a study from Guo (1996) described a dosage effect in the gene expression of wheat polyploids, referring to a dosage compensation mechanism. In fact, in plants it was observed that an allopolyploid should be capable of maintaining a diploid-like behavior (Ma & Gustafson, 2005). Despite the presence of multiple copies, gene expression has to be silenced somehow. Silencing patterns, on the other hand, are dependent on the gene, ones being silenced independently and repeatedly during polyploidization and others being affected stochastically (Adams, 2007).

Interestingly, many microRNA (miRNA) and their targets were not additively expressed in the *Arabidopsis* allotetraploids (Wang *et al*., 2006b; Ha *et al*., 2009), suggesting a miRNA's functional role. Beyond miRNAs, several other types of non-coding small RNAs have been found in animals, plants, and fungi, namely endogenous small interfering RNA (siRNA) (Reinhart and Bartel, 2002; Ambros *et al*., 2003; Baulcomb, 2004), transacting siRNA (tasiRNA) (Vasquez et. al, 2004), Piwi-interacting RNAs (piRNAs) (Aravin *et al*., 2007), among others. TasiRNAs were identified as mediators of post-transcriptional regulation, as well as of RNA-directed DNA methylation and chromatin remodeling (Vasquez *et. al*, 2004). The latest, piRNA, usually with 24-30nt length are derived from transposons and other repeated sequence elements and have been reported not only to be crucial to germ line development and fertility (O'Donnelland & Boeke, 2007), but also to transposon repression, which became extremely active after an event of allopolyploidization.Regarding the epigenetic regulation of gene expression, it occurs instantly by consequence of polyploidisation or hybridization. One of the main modifications is related to methylation patterns and gene silencing, which, in spite of being more frequent in plants (Xiong *et al*, 1999; Adams *et al*., 2004), was also described in animals (Salmon *et al*, 2005). Indeed, with the disruption of heterochromatin activation of TE (Josefsson *et al*., 2006), biased expression of homeologs (Udall *et al*., 2006), abnormal imprinting (Vrana, 2000), among others, could occur. As previously described, epigenetic and genetic modifications are closely related (Lu & Chen, 2011). Moreover, epigenetic regulation could be reversible which makes it advantageous for selection and adaptation during evolution and development.

In addition, some studies have also reported that these alterations are tissue specific (Adam *et al*, 2003). Besides, allopolyploids suffer a more critical alteration, when comparing to autopolyploids, meaning that hybridization causes the majority of modifications in gene expression, rather than ploidy changes (Albertin *et al*., 2006; Salmon *et al*., 2005).

2.1MicroRNAs biogenesis and cellular function

Transcriptome is defined as the complete collection of transcribed elements in a genome (Velculescu *et al*., 1997). It was considered to mainly consist of ribosomal RNA (rRNA) (80-90%), transfer RNA (tRNA) (5-15%), messenger RNA (mRNA) (2-4%) and a small fraction of intragenic (i.e. intronic) and intergenic noncoding RNA (ncRNA) (1%) with undefined regulatory functions (Lindberg & Lundberg, 2010). But, more recently it was described that ncRNA varies with the complexity of the genome, ranging from 0,25% in prokaryotes and 98,8% in humans (Taft *et al*., 2007). Several types of these ncRNAs have been already described. However, their functions are not clarified yet, thought huge efforts have been made in that direction.

The first ncRNA characterized was the short lin-4 RNA, in 1993, that was reported to directly down-regulate the lin-14 gene product by binding to specific repetitive sequences on the 3'UTR of the lin-14 messenger, controlling the time of development of the worm *Caenorhabditis elegans* (Lee *et al*., 1993). After, the first miRNA (let-7) was discovered (Reinhart *et al*., 2000), triggering the beginning of the comprehension of the silencing of specific genes by small non-coding RNAs. miRNAs have since then been found in plants, green algae, viruses, and more deeply branching animals (Griffiths-Jones *et al*., 2008), being verified that they are very conserved (Zhang *et al*, 2006). Particularly, miRNAs are post-transcriptional gene expression modulators, that are closely related with human disease and for that reason have been the most described of all small ncRNA. It was estimated that each animal miRNA regulates hundreds of different mRNAs, suggesting that a large proportion of the transcriptome is subjected to miRNA regulation (Bartel *et al.*, 2009; Voinnet, O *et al.*, 2009). Thus, they control a wide range of biological processes such as cell proliferation, differentiation, apoptosis and metabolism (*Bartel et al.*, 2009; Carthew & Sontheimer, 2009; Voinnet, O *et al.*, 2009; Krol, J. *et al.*, 2010).

MicroRNAs are composed by ≈22 nucleotides in length and derived from a complex biogenesis composed by several steps. At first, the capped and polyadenylated miRNAs precursors (pri-miRNAs) are transcribed, usually by RNA Pol II and fold back on themselves to form a long hairpin-like structure, which contains an imperfectly base-pairing stem. Once pri-miRNAs are formed, they are processed by the RNase III-like endonuclease Drosha which cuts them into ≈70 nucleotides hairpins designated as pre-miRNAs with 5' phosphate and 2 nucleotides 3' overhang. These molecules are transported to the cytoplasm by Ran-GTP through exportin5 and newly processed, this time by a protein complex (TAR RNA binding protein) including RNAse III type Dicer which is responsible for the cleavage of pre-miRNAs into ≈22 nucleotides duplexes. Finally, one of the strands is degraded and the other (mature miRNAs) becomes loaded into the Argonaute protein in miRNP or miRISC (miRNA-ribonucleoprotein or miRNA-induced silencing complex, respectively) where the miRNA will be complementary to target transcripts avoiding translation either by repression or endonucleolytic cleavage (*Filipowiczet al*, 2008).

The targeting issue still causes a lot of discussion. In plants, the complementarity between miRNA and mRNA is nearly perfect, however when talking about metazoans an imperfect base-pairing succeed which makes the miRNA harder to predict its targets transcripts (Winter *et al.*, 2009). Nevertheless, miRNA seed region, at 5'end (2-8 nucleotides) is required to have a perfect matching with its respective target mRNA transcript (Lewis *et al.*, 2005) being the secondary structure on the 3' UTR region of the mRNA also a relevant matching criteria (Brennecke *et al.*, 2005).

Until very recently it was thought that, in animals, miRNA repressed translation with little or no influence on mRNA abundance, but conversely, in plants, miRNAs were reported to degrade the transcript target. However, over the past few years, it has become clear that microRNAs are capable to induce mRNA degradation in animals and also translational repression in plants (Huntzinger & Izaurralde, 2011).

In overall, it seems that genetic constrains generated in polyploid organisms could be, at least partially overcome through the regulation of miRNAs. Indeed, studies have reported extensively that miRNAs regulate gene expression in allopolyploid plants (Ha *et al.*, 2008), being evolved into genetic regulation of polyploid plants. Thus, once gene expression is not additive, the functionally of the cell could be recovered avoiding a genomic shock.

2.2Next Generation Sequencing Technologies

Next generation sequencing (NGS) is a current toll to study transcriptomes and has been used to study small RNA profiles. It is a recent sequencing technology that has changed the world of molecular biology. The basic principle of these technologies is the miniaturized reactions which require DNA molecules at limiting dilutions, in a way that there is only a single DNA molecule firstly amplified and then sequenced. To do that, DNA has to be randomly broken into smaller sizes and then immobilized by primer or template to a solid surface, or even indirectly by linking a polymerase to the support (Mardis, 2008; Shendure *&* Ji, 2008; Metzker, 2010). Thus, the new sequencing platforms do not require the cloning of the fragments, which decreases considerably the cost and time of all the process when comparing with the older technologies. Before this, in 1977, the pioneer developments of Maxam and Gilbert (1977) and Sanger *et al.* started what is traditionally called Sanger sequencing. The massively parallel sequencing platforms only started with Roche (454) Genome Sequencer (GS) (2004), that was capable of producing several hundred thousand reads of 100 base pairs (bp) each. Today, GS FLX Titanium generates greater than 1 million reads composed by 400bp. In 2006, Illumina Genome Analyser (GA) was able to produce tens of millions of 32bp reads. Currently, Illumina HiSeq 2000 produces up to six hundred billion of reads of 100bps. These technologies were then followed by the most recent platforms: Applied Biosystems based on Sequencing by Oligo Ligation and Detection (SOLiD) and Helicos BioScience HeliScope. Whereas the first produces 400 million 50bp reads, the second is the first single-molecule sequencer that are able to produce 400 million though with 25-35bp reads (Costa *et al*, 2010).

These platforms are very different in what concerns sample preparation, chemistry, type, volume of raw data, and data formats, giving rise to their own characteristic error profiles. For example, in Illumina technology there is a greater probability of accumulating sequence errors at the end of the read and it produces short reads requiring specific alignment algorithms (Costa *et al*, 2010).

RNA-Seq and Microarray platform

Transcriptome analysis through RNA-seq allows identifying, characterizing and cataloguing all the transcripts expressed within a specific cell or tissue. Hence, a huge potential to address many biological-related issues could be answered by determining the expression or splicing patterns either gene or allele-specific (Wang *et al*., 2009). For example, in terms of disease-related studies, when comparing different samples, these data could give relevant biological information.

The older transcriptome techniques were based on hybridization or tag sequence. The widespread microarrays hybridization platforms typically suffer from cross-hybridization and undesirable background. Besides, transcripts hybridise only if it is included in the array design (Irizarry *et al*., 2005) which in RNA-seq does not occur. Also, a very significant issue has to do with the fact that we are dealing with an indirect method, which do not detect transcripts from repeated sequences or do not have the sensibility to detect very subtle changes in gene expression levels (Wang *et al*., 2009; Hoen *et al*., 2008; Bloom *et al*., 2009). On the other hand, when comparing to the older methodologies, the advantages of RNA-seq are evident. Besides, production of enormous volume of data with incredibly low background, it allows the discovery and quantification of new rare transcripts and Single Nucleotide Polymorphisms (SNPs); last, but not least, RNA-seq possesses higher technical and biological reproducibility (Metzker, 2010; Costa *et al*, 2010).

Still, studies about the comparison between RNA-seq and microarrays platforms are controversial. Based on comparison studies with biological data, it was described that, even if similarly correlated, microarrays produced less reliable absolute quantitative measurements and sequencing provides a better approximation of the real transcript content, when comparing microarrays with Illumina sequencing (Marioni *et al*., 2008; 't Hoen *et al.*, 2008) or other platforms (Coughlan *et al*., 2004; Chen *et al*., 2007; Liu *et al*., 2007). However, another study using pre-known synthetic samples demonstrated that microarray measurements are better correlated with RNA content than measurements from sequencing data. Additionally, reproducibility seems equivalent in both methods and microarrays sensibility showed to be higher than previous demonstrated (Willenbrock *et al.*, 2009). Moreover, a 2010 study using biological data, demonstrated that RNA-seq is more sensitive than microarrays (Xiong *et al*., 2010). Finally, Malone and Oliver (2011) affirmed that microarrays should be considered an accurate and useful tool, and RNA-seq should be used as complemented measurement.

In practice, what is necessary to do a trancriptome analysis through RNA-seq is: library preparation, sequencing and imaging, genome alignment and assembly (Metzker, 2010; Costa *et al*, 2010). The most common analysis of transcriptome is the quantitative expression profiling. To do so, it is necessary to trim the data and proceed to the crucial and complex step of genome alignment using algorithms like SOAP (Li *et al*., 2008), QPLAMA (Bona *et al*., 2008), TopHat (Trapnell *et al*., 2009), G.Mo.R-Se (Denoeud *et al*., 2008) or PASS (Campagna *et al.*, 2009), making use of a reference genome. When higher sequencing errors are expected in higher polymorphisms, insertion/deletion, complex exon-exon junctions, miRNA and small ncRNA, a more sophisticated alignment strategy is required to them. Particularly in the case of miRNAs, the first step (trimming) has to be efficiently performed, otherwise many observed miRNA would be derived from other sequences, since mature miRNA have paralogs with highly similar sequences (*Bartel et al*., 2004; *Guo et al*., 2009). Then, several approaches for the quantification analysis could be performed: (1) number of reads per each annotated element or (2) sum of the number of reads covering each base position of the annotated element, also known as base coverage (Costa *et al*., 2010). Afterwards, when comparing at least two samples sequenced independently, normalization should be done in order to take into account the sample size effect. Transcripts Per Million (TPM) is a simple and common approach which considers the raw abundance of each signature in each library. More recently, Reads Per Kilobase per Million (RPKM) was proposed by Mortazavi (2008) and is based on an algorithm that uses the molar concentration of a transcript by normalizing for the RNA length and the number of mapped reads. It is used to quantify the following comparisons: both different genes within the same sample and differences of expression across biological conditions (Li *et al*., 2009). The main problem is that these types of measurements do not reflect precisely the accurate RNA abundance into a library due to several reasons: integrity of the input RNA, extent of ribosomal RNA remaining in the sample, size selection steps, accuracy of the gene models used and non-uniform sequence coverage (Costa *et al*., 2010). Efforts have been made in order to obtain novel efficient approaches. For example, NEUMA, an algorithm recently developed by Lee *et al.* (2010), demonstrates superior accuracy over other recently developed methods.

Differential expression is not only characterized by transcripts exhibiting the most spectacular differential expression. Some studies have concluded that small changes as 2:1 or 2:3 can cause drastic effects for example in disease states caused by haploinsufficiency or trisomy (Audic & Claverie *et al*., 1997).

2.3DNA methylation, a pre-transcription mechanism

Methylation of DNA through addition of a methyl group to a cytosine is one of the most common inherited covalent modifications (Richards, 2006; Klose and Bird, 2006; Kim *et al*., 2009),. In complex multicellular eukaryotes, methylation is localized at cytosines within CG dinucleotides. In every cellular cycle, after DNA replication, DNA methyltransferase from Dnmt1 family (Goll & Bestor, 2005) fills in the new synthetized strand with group methyl in order to recover and maintain the original pattern. At least in plants, this process is described to be maintained through multiple generations (Chan *et al*., 2005, Zilberman *et al*., 2007). Today there is no doubt about its central role in many aspects of biology. The studies describing DNA methylation importance are extensive since it plays an important role in cellular activities such as differential gene expression, cell differentiation, chromatin inactivation, genomic imprinting, development and even carcinogenesis (Gonzalgo & Jones, 1997a). In animals, for example, once Dnmt1 in *Danio rerio* is depleted, the terminal differentiation of the intestine, exocrine pancreas and retina are abnormally formed (Rai *et al*., 2006). Knockout mutations on genes encoding DNA methyltransferares in mouse are lethal (Goll & Bestor, 2005). Additionally, in female mammals, DNA methylation also actively participates in X- chromosome inactivation in response to dosage compensation (Heard & Disteche, 2006). In plants, alterations in development are also evident, such as in *Arabidopsis thaliana*, where the loss-of-function of the ortholog of *Dnmt1* (*MET1*) generated a delayed flowering and reduced fertility, aggravated by additional mutations in other methyltransferases genes (*CMT* and/or *DRM2*). Other studies reported significant differences in DNA methylation among diverse organs in plants, such as rice (Dhar *et al*., 1990), tomato (Messeguer *et al*., 1991) and maize (Lund *et al*., 1995), suggesting a role in tissue or organ formation. Besides, it was also verified that methylation act as an important response to environmental changes, for example in the study where alteration of DNA methylation was induced by salt stress (Zhong *et al*., 2009).

Once formed, several allopolyploid plants are frequently accompanied by DNA methylation modifications which play a critical role in chromatin remodeling and are associated with gene silencing (Madlung *et al*, 2002; Liu &Wendel, 2003). So, several studies have been made to conclude about the functional role of DNA methylation patterns in polyploids. When treated with an inhibitor of DNA methyltransferase (5-aza-2'-deoxycytidine), DNA of synthetic allotetraploids suffered demethylation and altered morphologies are developed (Madlung *et al*., 2002). Curiously, the natural allotetraploid (*Arabidopsis suecica*) possess silenced genes due to epigenetic regulation (methylation) (Lee & Chen, 2001). *Tritivumaestivum* (Ozkan *et al*., 2001; Shaked *et al*., 2001), a wheat synthetic allopolyploid, and a rice hybrid (Xiong *et al*., 1999) also showed alteration in genome methylation. All these studies showed that allopolyploids are subjected to epigenetic alterations, namely DNA methylation. Overall, sequences that are actively transcribed are often less methylated than silent genes, either in promoter or coding region (Finnegan *et al*., 1993). Hence, epigenetic regulation through DNA methylation may suppress gene expression, an essential achievement for allopolyploids success, whereas respective demethylation may cause their reactivating.

Until now, a huge range of methods in order to detect DNA methylation have been developed. The classical techniques are based on bisulfite conversion, methylation-sensitive restriction enzymes or affinity purification of methylated DNA. The basic principle of bissulfite conversion is the treatment of DNA with sodium bisulfite, which converts every unmethylated cytosine to uracil and then to thymine by PCR. Once sequences are obtained and analyzed, methylation polymorphism is possible to uncover. While the last method is more frequently used to analyze methylation polymorphism, the others enable to analyze the whole genome methylation. Affinity enrichment of methylated DNA uses an antibody or a methyl-binding domain (MBD) protein to obtain, by attaching to a column, the methylated DNA through affinity purification. In addition, some kits use specially treated strip wells that have a high DNA affinity. The methylated fraction of DNA is detected using capture and detection antibodies and then quantified colorimetrically/fluorimetricaly by reading the absorbance in a microplate spectrophotometer. The amount of methylated DNA is, then, proportional to the OD intensity measured. On the other hand, methyl-sensitive restriction enzymes have been used on several techniques *Hpa* II and *Msp* I recognize the same restriction site (CCGG) but possess different sensibilities to certain methylation states of cytosine, whereas *Hpa* II does not cut if one of the two adjacent cytosine are fully methylated (double strand), *Msp* I does not cleave if the external cytosine is fully or hemi methylated (single strand) (McClelland *et al*., 1994). In other words, the fully-methylation of the internal cytosine or hemi-methylation of the external cytosine at CCGG sites can be unequivocally distinguished by these isoschizomers. Although they do not distinguish all methylated stated of CCGG sequences in the genome, they are a very good indicator of the methylation relative content. However, it is important to have in mind that methylation percentages are lower than the total absolute values (Mhanni & MacGowan, 2004).

3.The *Squalious alburnoides* complex

The *Squalius alburnoides* (Steindachner) complex belongs to the Cypriniae family and is considered an endemic freshwater fish of the Iberian Peninsula, sympatric with other *Squalius* species, *S. pyrenaicus*, *S. caroliterti* and *S. aradensis*. Molecular evidences showed that its origin resulted from the unidirectional hybridization between *Squalius pyrenaicus* (Günther) (P genome) (*Alves et al*., 1997b) as maternal ancestor and a probably extinct paternal ancestor related to *Anaeccypris hispanica* species (Steindachner) (A genome) (Alves *et al*., 2001; Crespo-López *et al*., 2006; Gromicho *et al*., 2006; Robalo *et al*., 2006).

S. alburnoides was described for the first time in 1983 by Collares-Pereira, who identified three different ploidies within the complex, diploid (2n = 50), triploid (3n = 75) and tetraploid (4n = 100). Particularly in Portugal, the distribution of the complex occurs from North to South basins, where *Squalius* species actively contribute with genetic material to the complex. Thus, while in southern populations this hybridogenetic complex incorporates the P genome due to the *S. pyrenaicus* interaction, in the North, *S. alburnoides* interacts with *S. carolitertii* and consequently incorporates the C genome. Additionally, in the restricted southern population of Quarteira drainage, *S. arandendis* also contributes to the complex's genetic diversity by introgression of its Q genome (reviewed in Collares-Pereira & Coelho, 2010). For this reason, *S. alburnoides* is considered an allopolyploid complex.

Besides, one of the characteristic features of the *S. alburnoides* complex is its high diversity of reproductive modes that promote an intricate network of genetic exchange and continuous shifting between different forms. Indeed, it was described that the allopolyploid complex possess mechanisms of sexual and asexual reproduction, since depending on its genetic composition, it produces gametes with different genomes. Thus, hybridogenesis, meiosis and meiotic hybridogenesis are some of the reproductive modes already described (Alves *et al*., 1998, 1999, 2001, 2004; Gromicho & Collares-Pereira, 2004; Pala & Coelho, 2005; Crespo-López *et al.*, 2006; Sousa-Santos et al., 2007b). Consequently, due to its hybrid origin and the variety of reproductive mechanisms, each hybrid could be composed by distinct combinations of parental genomes, resulting in the southern populations diploids PA and AA, triploids PAA and PPA or tetraploids PPAA, being PPA rarely found. In this complex, the constant shift of the genome composition in consecutive generations involving a change on the ploidy level, promotes repetitive situations of potential genomic shock. So, and since this hybridogenetic complex shows high evolutionary success, it is an excellent organism to investigate the allopolyploidization effects (Alves *et al.*, 2001; Crespo-López *et al.*, 2006). Furthermore, the fact that hybrid diploid (PA) and tripoid (PAA), are morphologically undistinguished, suggest that gene expression is non-addictive, and so genetic or epigenetic silencing mechanisms should be acting.

3.1Gene-copy silencing and dosage compensation

The allopolyploid *S. alburnoides* complex must then supplant the expected previously referred genetic constrains in order to achieve a functional gene expression programme. Indeed, for the first time in triploid vertebrates, the presence of a silencing mechanism was suggested by Pala *et al*. (2008) by studying this complex, specifically in individuals from southern populations. It seemed that there was a response to the increase of gene dosage since the PAA triploid form had some alleles with reduced transcript levels to a diploid state. Thus, since there were no significantly differences in the expression between diploids and triploids, a dosage compensation mechanism was proposed (Pala *et al.*, 2008; Pala *et al.*, 2010).

Consequently, the question about whether the silencing occurs preferentially in a species specific genome emerged. To answer that it was necessary to proceed to a genome-specific allele expression of the triploid hybrids. In each analyzed housekeeping or tissue-specific gene, two different situations were verified depending on the tissue: exclusive expression of A genome or both genomes being expressed. These results revealed the existence of silencing in the P allele or in one of the A alleles, in order to mimic the diploid expression. Additionally, when comparing different genes and observing each tissue, the more frequent situation is the silencing of the alleles from the minority genome (P), instead of being random (Pala *et al.*, 2008).

However, this pattern was recently described to vary within the complex, according to the geographical origin and consequently, according to the genomes involved in hybridization process. Thus, contrarily to southern populations, in northern population polyploids exhibit preferential biallelic gene expression patterns, irrespective of genomic composition (Pala *et al.* 2010). In spite of knowing that gene expression in the hybrid triploid is non-additive, the mechanisms underlying expression regulation and allele differential silencing are still undefined in this species.

2. AIMS

This study aims to investigate the genomic mechanisms involved in gene expression regulation operating in the *Squalius alburnoides* complex. Therefore, two possibilities were analysed: regulation by microRNAs and DNA methylation. Hence, regarding not only with the emergence of hybridization but also the increase of the ploidy level this work intended to focus on the specific goals, detect changes in the microRNA expression profile and Changes in the genome wide-DNA methylation

Thus, to accomplish these aims it is essential to use in this study not only the hybrid triploids (PAA) and diploids (PA), but also the non-hybrid form of the complex (AA) and the interacting species in the region of the study, *Squalius pyrenaicus* (PP). The first goal were achieved through the construction of four small RNA libraries of the genomic compositions mentioned above and posterior sequencing by next generation technologies or microarray hybridization technology. The second goal was achieved by approaches using endonucleases with different sensibility to methylation and an immuno assay. Hereupon, what this study pretends is to gather data about these pre- and post-transcriptional mechanisms, allowing us to comprehend whether they play a role and if they do, how much they contribute to a wider regulation or a finer tuning of the constant shifting between the forms of the *S. alburnoides* complex.

3. MATERIALS AND METHODS

3.1. Sampling and genomic constitution determination

Squalius alburnoides specimens were collected from Almargem river basin, in the south of Portugal and *Squalius pyrenaicus* individuals were sampled in Colares stream. Fish was captured by electrofishing and had approximately 10 cm long.

After settle all the fish in the fish facility, blood samples were collected in freezing solution and immediately frozen at -80ºC. Ploidy level was obtained by flow cytometry according to Próspero and Collares-Pereira (2000). Each genome contribution was determined according to Inácio *et al*., 2010.

3.2. RNA extraction

Fish were acclimatized in captivity for two weeks to ensure that their microRNA expression was not affected by external factors such as stress. After, captured fish were sacrificed by overdose of anaesthetic MS222 and the respective organs collected and preserved in RNAlatter® (Ambion) at -20ºC.

Total RNA was extracted from liver, muscle and brain using the Tri-reagent (Ambion) and following the suppliers' instructions. Contaminant DNA was eliminated by the addition of TURBO™ DNase (Ambion) and further purification with fenol/chloroform. Ethanol, Glycogen and Sodium Acetate (NaOAc) were used to achieve RNA precipitation.

Quality evaluation of the extracted RNA was performed in Nanodrop 1000 (Thermo Scientific) and in 2100 Bioanalyzer (Agilent Technologies). The concentrations were also registered.

3.3. Library construction and sequencing analysis

To construct the smallRNA libraries, only RNA samples with a RIN greater than 8 (Bioanalyzer) were considered. Three different organs (muscle, brain and liver) from about three individuals of the same genomic composition were pooled together for library construction. Four libraries were prepared one from *S. pyrenaicus* (PP) and three from the different main forms of *S. alburnoides*: AA, PA and PAA. Each library was made following the Illumina protocol *Small RNA v1.5*. It consists on the ligation of 5' and 3' RNA adaptors to the total RNA extracted, which were then reverse transcribed with primers complementary to the adaptors and size-fractionated on a 6% PAGE gel to collect the small RNAs of 22-30nt. Libraries were shipped in dry ice to Beijing Genomics Institute (BGI), Hong Kong, where they passed the Bioanalyzer and real-time PCR quality controls. Then, the same amount of each library was sequenced by the Illumina technology - Solexa. Bioinformatic analysis was also performed in BGI.

After trimming the low quality reads, contaminants, adaptors at both 5' and 3'and reads shorter than 18nt, cDNA sequences were mapped to *Danio rerio* by SOAP programme. Annotation was made not only using the miRBase 15.0 database (*http://www.mirbase.org/*) for the precursor/mature miRNA, but also using the GenBank (*http://www.ncbi.nlm.nih.gov/*) and Rfam 9.1 (*http://www.sanger.ac.uk/software/Rfam*) databases, since many other sequences of exons, introns, tRNA, rRNAs, snoRNAs and snRNA were obtained. In order to avoid the classification of each unique sRNA into several categories priority was given to rRNA, miRNA, exon and intron classes, by this order. In addition, GenBank had priority over Rfam database. All the clean tags were grouped so that each unique sequence from each category had its associated umber of reads (counts). Tags that could not be annotated to any category were used to predict novel miRNA, by the BGI's software *Mireap* (**Fig 1**).

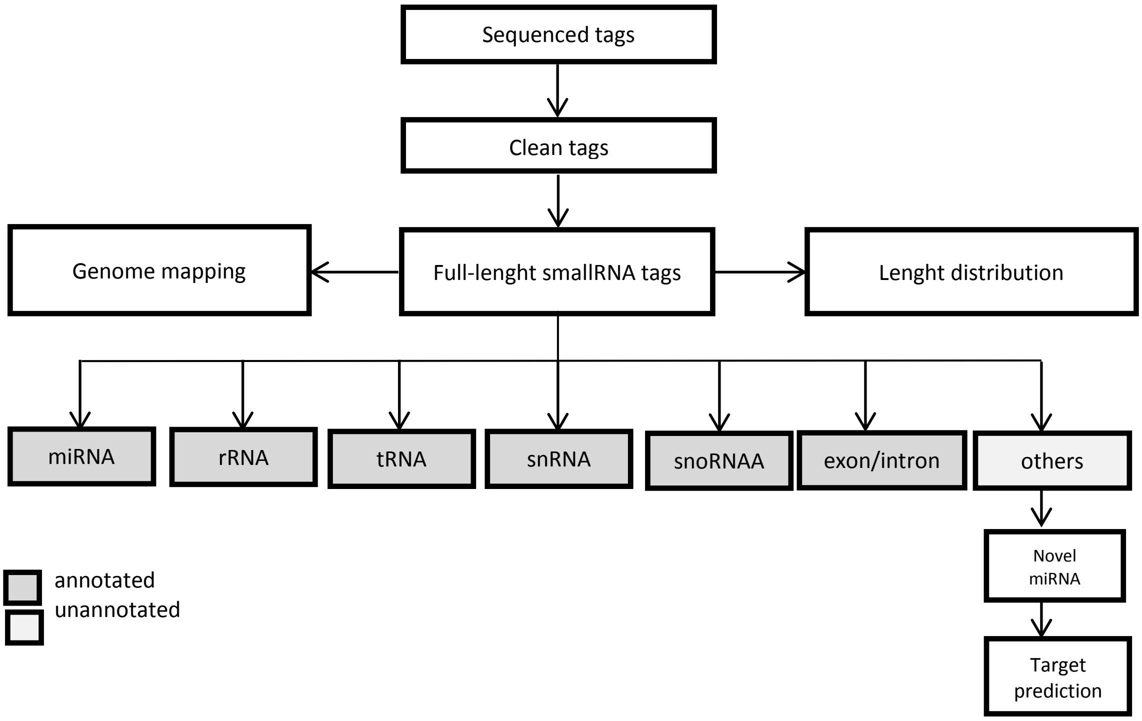


Fig 1 Scheme representing standard bioinformatics analysis applied to 35nt Solexa sequences after the data trimming.

3.4. Hybridization in microarray chip and data treatment

The same RNA pools that were used to construct the sequencing libraries were also subject to microarray analysis in a miRNAChip_MS_V1 produced by National Facility for DNA Microarrays (University of Aveiro). For that, RNA was reverse transcribed and cDNA was labeled using the miRNA labeling kit from *Kreatech*. Two µg of cDNA were incubated with Cy3-ULS (1 µl of Cy-ULS for 1 µg of cDNA) for 15min at 85ºC and then purified to remove non-reacted Cy-ULS. Dye incorporation was monitored by UV-visible spectroscopy. Hybridizations were performed at 42ºC for 16h and later, slides were washed according to the manufacture's recommendations and scanned using an Agilent microarray scanner. Resulting images were analyzed using *QuantArray* to extract microarray data. Cy3 median pixel intensity values were background subtracted, normalized and subjected to further analysis. Data points were removed when intensity values were below 200% of background. A specific normalization of microarray data was applied using *BRB ArrayTool*. This part of the work was performed in collaboration with Cesam, Universidade de Aveiro.

3.5. DNA extraction and 5-mC Immuno assay

Genomic DNA was extracted from fin, muscle liver and blood following standard phenol/chlorophorm extraction protocol of the digested tissue with SDS (Miller *et al*. 1988). In order to avoid incorrect gDNA quantification due to RNA contamination, RNase (*Sigma*) digestion was performed. Quality evaluation and concentration were carefully assessed by agarose gel electrophoresis and Nanodrop 2000 (Thermo Scientific). *MethylFlash™ Methylated DNA Quantification Kit – Fluorometric* (*Epigentek*) was used to quantify genome-wide levels of 5-methylcytosine (5-mC) in each organ of each genomic composition. The optimal recommended quantity of 100 ng was used in all reactions including negative and positive controls. 5-mC was retained in the respective well and capture antibody followed by detection antibody was applied. Samples fluorescence were measured in *Zenyth 3100* (*Anthos*) using the required excitation filter. Then, the relative fluorescence units (RFU) data was treated according to the following formula:

$$5\text{-}mC\,\% = \frac{(x - Cn) \div S}{(Cp - Cn) \times 2 \div P} \times 100\%$$

$5\text{-}mC\,\%$ - genome wide levels of 5-methylcytosine *per* individual; x - fluorescence measured for each sample; Cn - fluorescence measured for negative control, DNA non methylated; Cp - fluorescence measured for positive control, DNA composed by 50% of 5-methylcytosine; S - amount of input DNA for each sample in *ng*; P - amount of input positive control in ng; 2- factor to normalize positive control to 100%

As there were only few reactions available, samples were selected according to their quality. In this case, although muscle's DNA were from three individuals of the same genomic composition (except for AA -2 individuals), fin and liver's DNA were obtained only from 2 PAA, 2 PA, 1 AA and 1 PP. In addition, blood's DNA was from 3 PAA, 2 PA, 1 AA, 1 PP and in this special case, also 1 PPA.

3.6. *Msp* I / *Hpa* II assay

DNA previously extracted from the fin used in this assay were the same as in the previous method, however in this case it was possible to use a greater number of individuals. Thus, seven individuals of each genomic composition PA, PAA, AA and PP and even three PPA individuals were analysed. One µg of each sample was digested with either M*sp* I or H*pa* II restriction endonucleases at 37ºC for 16h with manufacturer-supplied buffers. This digestion time was the maximum time recommended by the manufacturer which was confirmed by the achievement of complete digestion of the control plasmid pBR322. Efficiency of the enzymes was also tested by comparing with the negative control, where no enzyme was added to samples. Three independent electrophoresis, comprising digestions with both enzymes in the same 1% agarose gel, were performed for each sample as technical replicates. Each gel, stained with RedSafe™, was visualized using a transilluminator (Uvitec) and images were captured digitally by Kodak DC290 and analysed in Image J 1.44p software. Image densitometry analysis comprised the measure of the mean intensity of the visible smear obtained in the respective lane, for both enzymes. The region scanned corresponded to mid and low molecular weight DNA restriction fragments, avoiding the regions of high and very low weight due to the saturation of the signal. Then, gDNA methylation percentage was calculated as follows:

$$M_{\%} = 1 - \left(\frac{I_{HpaI}}{I_{MspII}}\right) \times 100$$

$M_{\%}$ - genome wide DNA methylation for each individual; I_{HpaI} and I_{MspII} - mean intensity values of independent experiments of the digestion with *Hpa I* and *Msp II* endonucleases, respectively.

4. RESULTS

4.1. MicroRNAs

4.1.1. High-throughput sequencing

Small RNA libraries

In order to evaluate whether expression profiles of, miRNAs were influenced by the genetic shock created by the constant shifting in the genomic compositions, four small RNA libraries from three genomic compositions AA, PA and PAA of *S. alburnoides* as well as one from *Squalius pyrenaicus* (PP) were sequenced using Illumina technology. For each library, total RNA extracted from muscle, brain and liver was pooled together. Each run produced 68.926.168 (AA), 70.090.456 (PP), 69.740.296 (PA) and 66.362.907 (PAA) reads. After cleaning the data (discarding low quality reads, removing the adaptors and contaminants generated by ligation and sequences shorter than 18 nucleotides), 44.384.528 (64,39% out of the total reads), 43.137.518 (61,85%), 43.461.092 (65,49%) and 42.549.707 (60,71%) clean sequences remained from AA, PA PAA and PP libraries, respectively. In whole, this has generated ≈ 173.5 million of clean small RNA reads. Moreover, their length distribution presents a similar pattern among the libraries as well as an overall greater amount of sequences composed by 20-23 nucleotides in length, being the featured peak the one representing the 22 nucleotides length (**Fig 2**). In general, the distribution obtained by sequencing corroborates the sRNAs libraries quality since the traditional length of one of the most predominant class - miRNA is 20-23 nt (Lu *et al.* ., 2010). Interestingly, a contrast among the libraries is observed in the 31nt length sequences, in which the analysed triploid form, PAA, presents approximately 2-fold (10,77%) comparing not only to the other diploids of the complex (AA and PA) but also to PP, 5,52%, 4,42% and 5,24%, respectively. A similar pattern was seen in the 33nt length sequences.

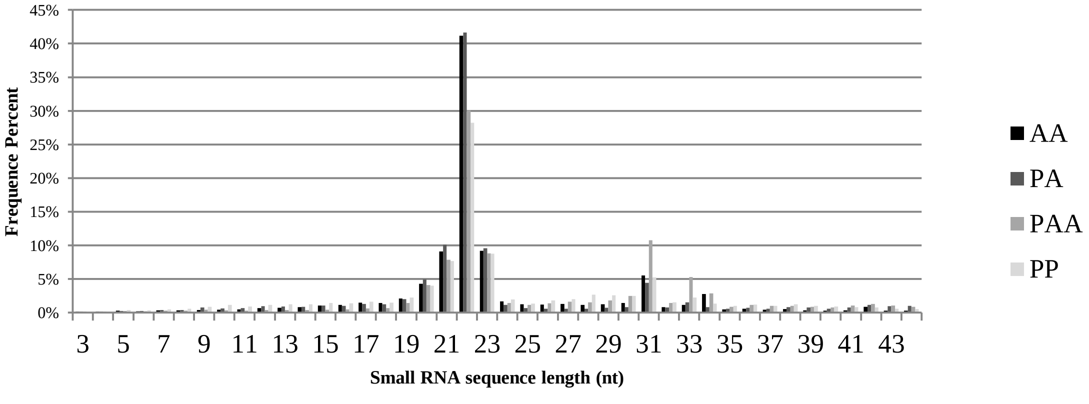


Fig 2 Length distribution of the sequences in all the libraries.

When grouped in unique non-redundant small RNA sequences, 2.133.078 (AA), 1.691.614 (PA), 2.358.444 (PAA), 2.607.075 (PP) distinct sequences were obtained.

After analyzing the common and specific sequences of each pair-wise comparison, it is clear that although most of the total sequences are common among libraries, varying between ≈92,5% and ≈95,2% (AA-PP 92,48%; PA-PAA 95,17%; PA-AA 94,59%; AA-PAA 93,81%; PA-PP 93,39%; PAA-PP 92,47%), when non-redundant sequences were analyzed and PAA was compared with the other forms of the complex it revealed more percentage of exclusive sequences (53,23% - PAA vs. 34,80% - PA; 47,93% - PAA vs. 42,42% - AA) (**Tab 1A**). However, both nuclear hybrids (PA and PAA) presented more sequences in common (≈12%) (**Tab 1B**). On the other hand, PP shared less common sequences with all the forms of the complex, ≈7-8% (**Tab 1B**).

Tab 1 Pair-wise comparisons of all unique sequences between every genomic composition. **A** Percentage of unique sequences specific from the compared genomic compositions. **B** Percentage of unique sequences common among the compared genomic compositions.

A

	PAA	PA	AA
PA	53,23% 34,80%		
AA	47,93% 42,42%	38,83% 51,49%	
PP	43,13% 48,55%	34,21% 57,31%	40,74% 51,51%

B

	PAA	PA	AA
PA	11,97		
AA	9,65	9,69	
PP	8,32	8,47	7,75

Small RNA annotation and length distribution

Using the ensemble database, clean reads were mapped to *Danio rerio* genome, the only *Cyprinidae* that already has its genome sequenced. Whereas AA, PA and PAA possess 29.869.369 (67,30%), 28.774.125 (66,70%) and 22.612.648 (52,03%) aligned total reads, respectively, PP has 24.344.581(57,21%). The percentage of aligned reads was in accordance to another *Cyprinidae* study (Chi *et al*., 2011). After, in order to classify the sRNA into different categories all the sRNA that matched or not the genome, were annotated using miRBase, GenBank and Rfam databases. While miRBase was used to annotate miRNAs, GenBank and Rfam were the databases chosen to annotate the remaining sequences. The most represented class was miRNA with 58,96% (AA), 62,46% (PA), 42,95% (PAA) and 44,22% (PP), proportions that are in accordance with the frequencies of annotated miRNA in the literature (Lu *et al.*, 2010; Chen *et al.*, 2005) and with results from **Fig 2**, considering the normal miRNA length. This was followed by unnanotated sequences, whose frequencies correspond to 28,75% (AA), 28,70% (PA), 41,76% (PAA) and 37,00% (PP). Additionally, degradation products from tRNAs, snRNA, snoRNA, rRNAs, exons and introns were also identified, even though underrepresented (**Table S1 and Fig 3**).

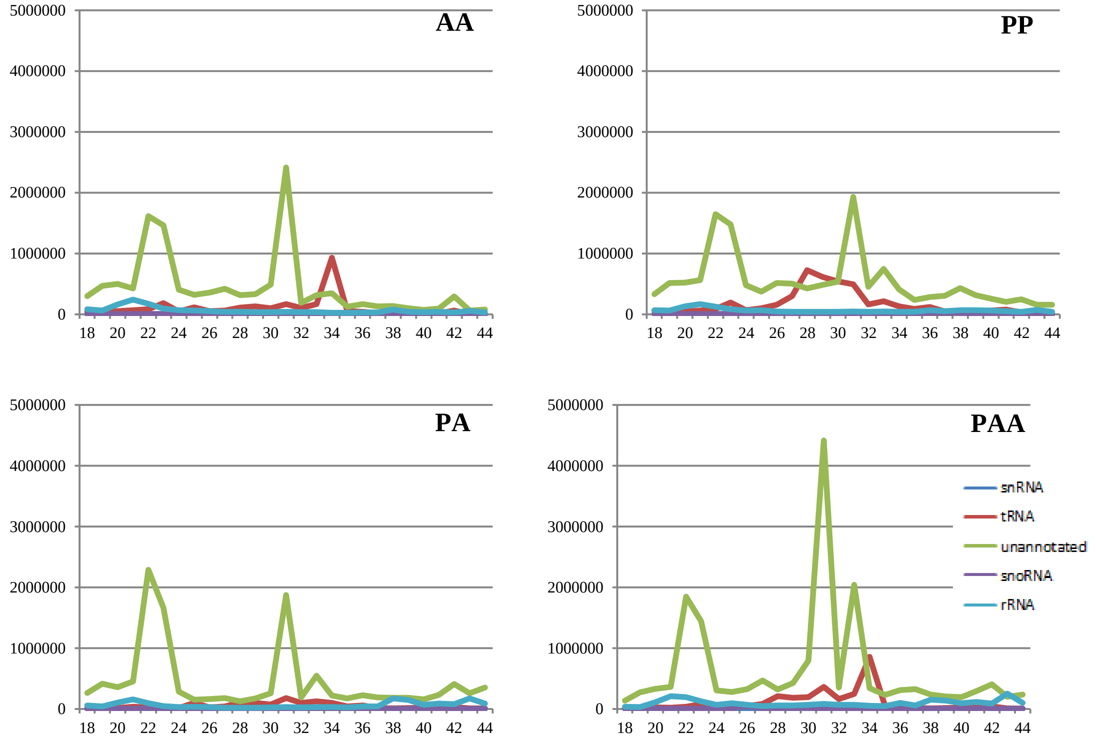


Fig3 Length distribution of the sequences in the four libraries AA, PA, PAA and PP of the following identified categories: snRNA, tRNA, unannotated, snoRNA and rRNA.

By analyzing the length distribution of all the RNA categories excluding miRNAs, (**Fig 3**), it was clear that the higher peak of 31nt (**Fig 2**) in the triploid PAA was from unannotated sequences (4.413.544 – 10,15% in PAA; 2.413.188 – 5,44% in AA; 1.873.619 – 4,34% in PA; 1.927.299 – 4,53% in PP). Curiously the same was observed for the higher 33nt length peak of PAA (2.040.515 – 4,69%). On the other hand, the unannotated sequences present in the 22nt length peak indicate the potential unknown miRNAs in all the libraries, whereas, as expected, known miRNAs contributed for the majority of the sequences of the 22nt peak (**Fig S1**).

miRNA expression profile

In order to untangle the microRNA expression profile, it was essential to align sRNA reads to each mature or precursor miRNA in miRBase, either from *Danio rerio* or from other organisms of the database, associating each unique miRNA to its respective count. However, to compare the miRNA expression of each miRNA the expression of each miRNA was presented in units of transcript per million (TPM) which considers the proportion of total clean reads in each library.

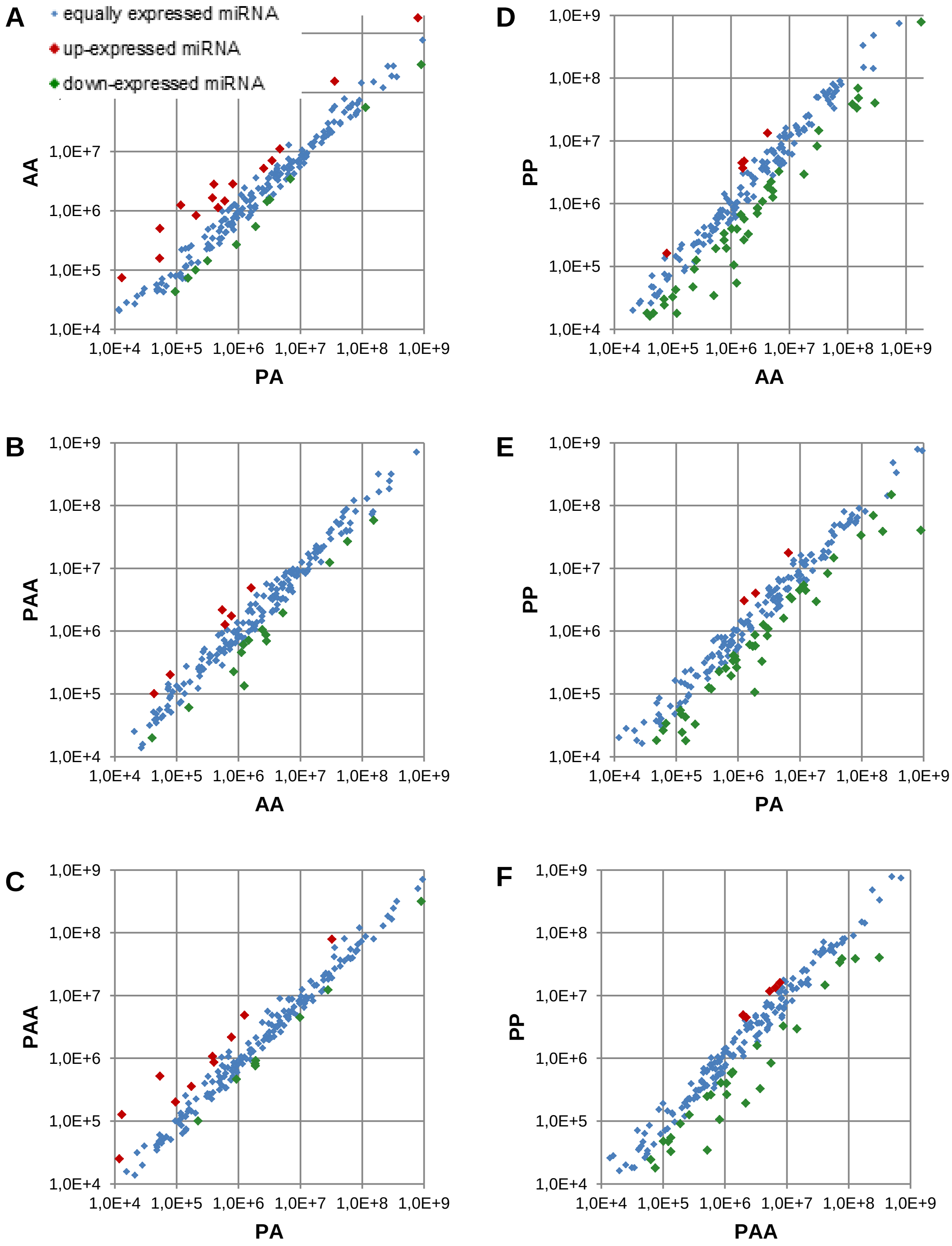


Fig4 Pair-wise comparisons in log2 plots of miRNA expression profiles between the genomic compositions AA, PA, PAA and PP, obtained through RNA-seq method. Genomic composition displayed on the x-axis was considered the control, while treatment is exhibited on the y-axis. **A, B, C** Pair-wise comparisons including only genomic compositions from *S. alburnoides* complex,. **D, E, F** Pair-wise comparisons including *S. pyrenaicus* and the genomic compositions of the complex.

Using TPM, pair-wise comparisons were performed, inferring the up- and down-expressed miRNA. In each comparison, two genomic compositions were presented as control vs. treatment. Two criteria to determine the up- and down-regulated miRNA were then established:

first, their fold-change and second, respective p-value (**Tab S2** and **Fig S2**). So, the significant up-expressed miRNAs were considered the ones presenting an expression superior to 2-fold (log2 (fold-change) > 1), whereas significant down-regulated miRNA showed half of the expression (log2 (fold-change) < -1). Additionally, p-value was obtained following Audic *et al.* (1997) which takes into account random fluctuations and sampling size. Hereupon, although the expression of some miRNA was significantly different (*p-value<0,01*), they were not considered significant up- or down-regulated, since they did not met the first criteria (**Table 2**). On the other hand, all the significant up- and down-regulated miRNA presented their p-value below 0,01 and were differently marked in the plots (**Table 2** and **Fig 4**).

Table2 Quantification of total and significant up- and down-expressed miRNA in each pair-wise comparison and respective percentage, in the RNA-seq method. Note that the values presented are for treatment (genomic composition represented in bold). Fold-change was calculated using all the differentially expressed miRNA, except when is demonstrated (w out), meaning that fold change was calculated without the one most distant outlier.

	PA-**AA**		AA-**PAA**		PA-**PAA**		AA-**PP**		PA-**PP**		PAA-**PP**	
	total	**	total	**	total	**	total	**	total	**	total	**
Up-expressed	68 (33.3%)	15 (55,6%)	96 (47.0%)	6 (26.0%)	50 (24.8%)	10 (52.6%)	76 (37.6%)	5 (10.0%)	140 (69.3%)	3 (6.5%)	82 (40.8%)	6 (16.7%)
Down-expressed	136 (66.7%)	12 (44,4%)	108 (53.0%)	17 (74.0%)	152 (75.2%)	9 (47.4%)	126 (62.4%)	45 (90.0%)	62 (30.7%)	43 (93.5%)	119 (59.2%)	30 (83.3%)
Total	**204**	**27**	**204**	**23**	**202**	**19**	**202**	**50**	**202**	**46**	**201**	**36**
Average ** **Log_2(fold-change)**	2.45 – 1.91 (w outl) 1.25 – 1.91		3.12 – 1.40		2.71 – 1.71 (w outl) 1.19 – 1.71		2.79 – 1.39		2.03 – 1.26		2.30 – 1.13	

** significant up- or down-expressed miRNA that were considered superior to 2-fold or half expression

The most evident different miRNA expression profile was from *S. pyrenaicus* (PP), which showed a general lower miRNA expression comparing to any of the other genomic compositions from *S. alburnoides* complex (**Fig 4 D, E, F**). Indeed, a chi-square test was performed to test if the quantity of up- or down-regulated miRNA were significantly different in each library. Whereas all the pair-wise comparisons containing PP obtained a low p-value (*p-value<0,01*), the remaining comparisons (**Fig 4 A, B, C**) between the genomic compositions of the complex, PA-AA (*p-value=0,846*), AA-PAA (*p-value=0,022*) and PA-PAA (*p-value=0,818*) seemed to have no significantly differences (p-value>0.01). However, considering not just the significant, but all the up and down expressed miRNAs in the libraries, all the comparisons are different, except for PAA vs. AA.

Regarding the fold change of the significant differently expressed miRNAs, the average of the up or down values was not considered significant different by the Mann-Whitney U test in none of the comparisons.

Novel miRNA

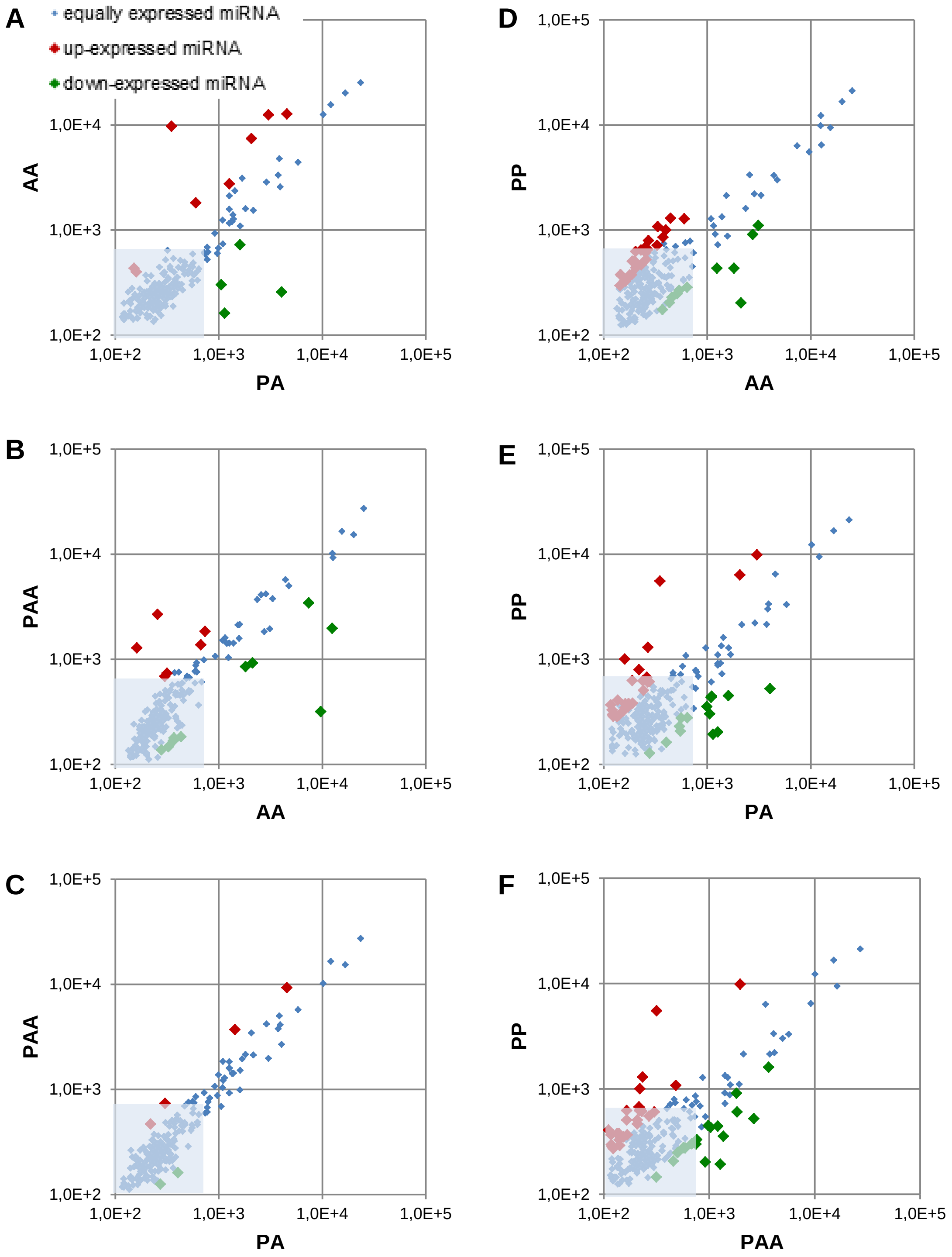
Afterwards, all unannotated but mapped to genome sequences were used to predict unknown miRNA, based on miRNA precursor characteristic hairpin structure. Secondary structure, Dicer cleavage site and minimum free energy were used as the main parameters to identify novel miRNAs, among others, by the *Mireap* software. In total, 80 (AA), 67 (PA), 56 (PAA) and 54 (PP) unique miRNA were predicted, being 21770, 25489, 19015 and 13954 the total counts respectively. From those, targets of all the novel miRNA were also predicted, except for PP in which only 4 from the 54 novel miRNA were possible to find. Thus, 442,040, 385,116, 304,152 and 18,073 targets were obtained for each library, respectively.

4.1.2. MicroArrays

miRNA expression profile

Using the same RNA's pools of sequencing libraries and in order to support the sequencing data, microarrays hybridization for miRNAs were conducted in chips designed including miRNAs from *Danio rerio* species. After data extraction, normalized miRNA profile expression was obtained. These data were plotted in log2ratio pair-wise plots and analyzed in order to understand whether the pattern obtained by this method was similar to the sequencing data.

In this analyses, the most different miRNA expression (here represented by +) was inferred by the same condition used in the RNA-seq (+ up-expressed (log2 (fold-change) > 1 or + down-expressed (log2 (fold-change) < -1).



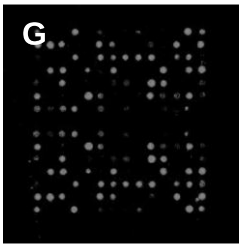


Fig 5 Pair-wise comparisons in log2 plots of miRNA expression profiles between the genomic compositions, AA, PA, PAA and PP, obtained through microarrays. Genomic composition displayed on the x-axis was considered as control, while treatment is exhibited on the y-axis. Note that the blue zone was considered the technical background **A, B, C** Pair-wise comparisons including only genomic compositions from *S. alburnoides* complex, AA, PA and PAA. **D, E, F** Pair-wise comparisons including *S. pyrenaicus* and the genomic compositions of the complex, AA, PA and PAA. **G** Example of a hybridization in a microarray chip.

Table3 Quantification of total and significant up- and down-expressed miRNA in each pair-wise comparison and respective percentage, in the microarray hybridization method. Note that the values presented are for treatment (genomic composition represented in bold). Fold-change was calculated using all the differentially expressed miRNA.

	PA-**AA**		AA-**PAA**		PA-**PAA**		AA-**PP**		PA-**PP**		PAA-**PP**	
	total	+	total	+	total	+	total	+	total	+	total	+
Up-expressed	111 (46.3%)	6 (60.0%)	118 (49.0%)	5 (50.0%)	108 (44.6%)	3 (100.0%)	117 (49.2%)	7 (58.3%)	120 (50.2%)	6 (35.8%)	130 (54.2%)	5 (29.4%)
Down-expressed	129 (53.7%)	4 (40.0%)	123 (51.0%)	5 (50.0%)	134 (55.4%)	0 (0.0%)	121 (50.8%)	5 (41.7%)	119 (49.8%)	8 (57.1%)	110 (45.8%)	12 (70.6%)
Total	**240**	**10**	**241**	**10**	**242**	**3**	**238**	**12**	**239**	**14**	**240**	**17**
Average + **Log_2(fold-change)**	2.44 – 2,15		2.20 – 1.98		0 – 1.22		2.01 – 1.36		2.00 – 2.35		1.62 – 2.45	

+ up- and down-expressed miRNA that were considered superior to 2-fold or half expression

Once more, the most evident different miRNA expression profile was from *S. pyrenaicus* (PP), which showed a general lower miRNA expression comparing to any of the other genomic compositions from *S. alburnoides* complex (**Fig 5 D, E, F**). Additionally, the chi-square test performed to test if the quantity of differently (+) up or down expressed miRNA were significantly different in the same library, never showed a low p-value (*p-value<0,01*), meaning that none of the miRNA profiles seemed to be different. The same result is obtained if we consider all the up- and down-expressed miRNAs in the libraries.

However, it is important to take into consideration that most of the miRNA showed an expression lower than the threshold that was considered background in the array.

4.1.3 Comparison of sequencing and microarray hybridization results

By means of comparing the sequencing and microarray data it was possible to verify the accuracy of the data obtained by both methods.

In fact, the overall expression profiles between the different genomic compositions seemed to be alike among methods, with similar patterns between the forms of the complex but *S. pyrenaicus* showing a more distant profile.

Nevertheless, individual analysis of each miRNA presented non reproducible results. Indeed the differently expressed miRNAs in both methods were mostly not the same (**Tab S2**, **S3**) and by analysing each same miRNA between methods, it was possible to verify that only eight out of forty seven miRNA exhibited a similar change of expression between genomic compositions (**Fig 6**).

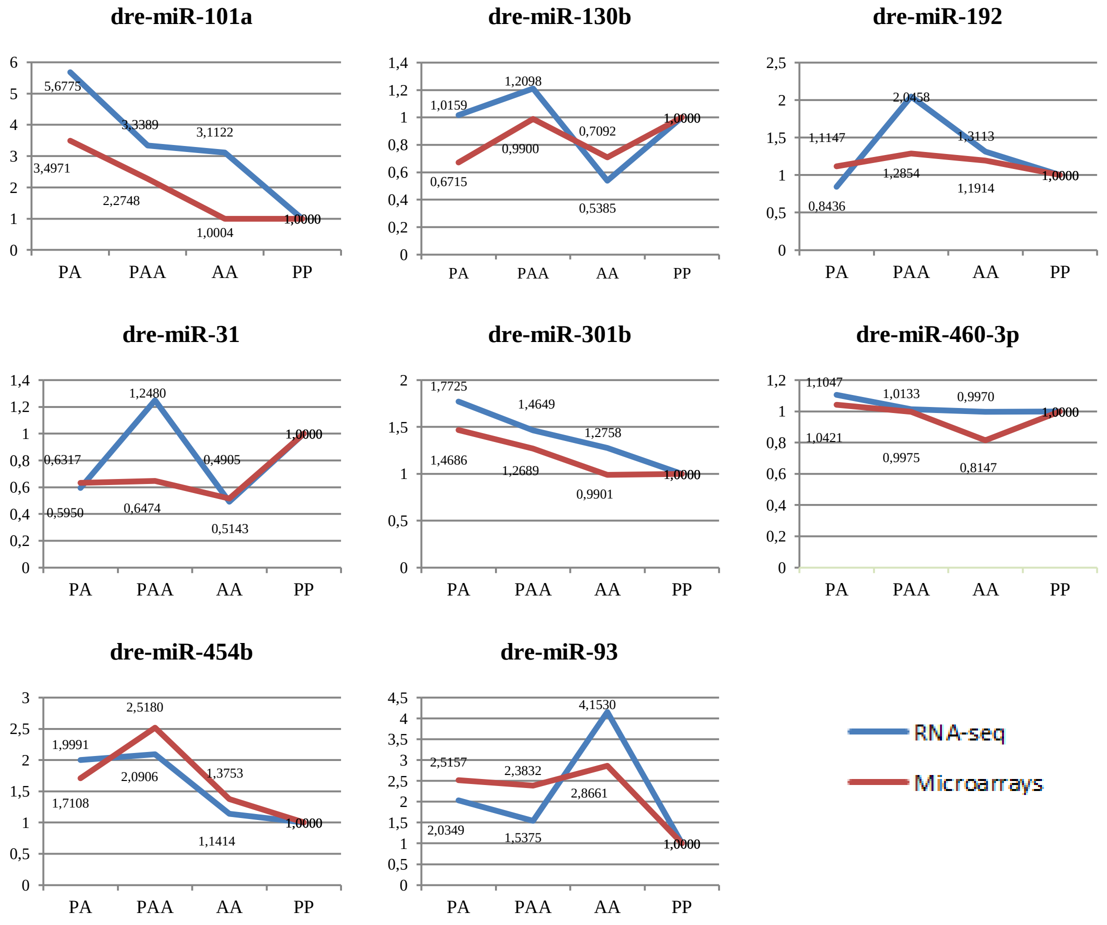


Fig 6 Fold-change of expression between the different genomic compositions of the miRNAs possessing similar expression pattern among high-throughput sequencing and microarray hybridization data. Differential expression was observed in the following comparisons. dre-miR-101a: seq PA-PP PAA-PP AA-PP, microarrays PA-PP PAA-PP AA-PAA PA-AA; dre-miR-130b seq: AA-PAA; dre-miR-192: seq PAA-PP PA-PAA; dre-miR-31 seq AA-PAA PA-PAA AA-PP; dre-miR301b; dre-miR-460-3p; dre-miR-454b seq PAA-PP, microarrays PAA-PP; dre-miR-93 seq AA-PAA PA-PAA AA-PP

4.2. Genome wide methylation

4.2.1 Immunoassay

The content of 5-mC was determined by a fluorimetric immunoassay which is based on the recognition of 5-methylcytosin by specific capture antibodies posteriorly linked to detection antibodies. Genomic compositions utilized in this method were the same as the used in the sequencing task, AA, PA, PAA and PP from several tissues. DNA was from blood, fin, muscle and liver and was extracted from 7 to 5 individuals from the complex, as well as 1 or 2 individuals from *S. pyrenaicus* species. DNA extracted from blood of one PPA was also used.

Given the few individuals considered, a statistical analysis of the data could not be performed. However, despite the low number of individuals, observing **Fig 7A** what stands out most in these data is the difference of the methylation content among the DNA from different types of cells. Blood cells showed the highest amount of 5-mC (mean 39.7%), followed by fin (19.0%), muscle (13.1%) and liver (11.0%). Since the results were in agreement with the expected given the low diversity of genes being expressed in red blood cells (mainly globins), confidence on the method was gain and it was decided to look closer to a tissue that allowed us to expand further the DNA methylation analysis (see next section). That tissue was fin since it is easiest to obtain and does not involve the fish to be sacrificed. So, although with no statistic power, we took into consideration the patterns of DNA methylation in each genomic composition in the fin (**Fig 7B**).

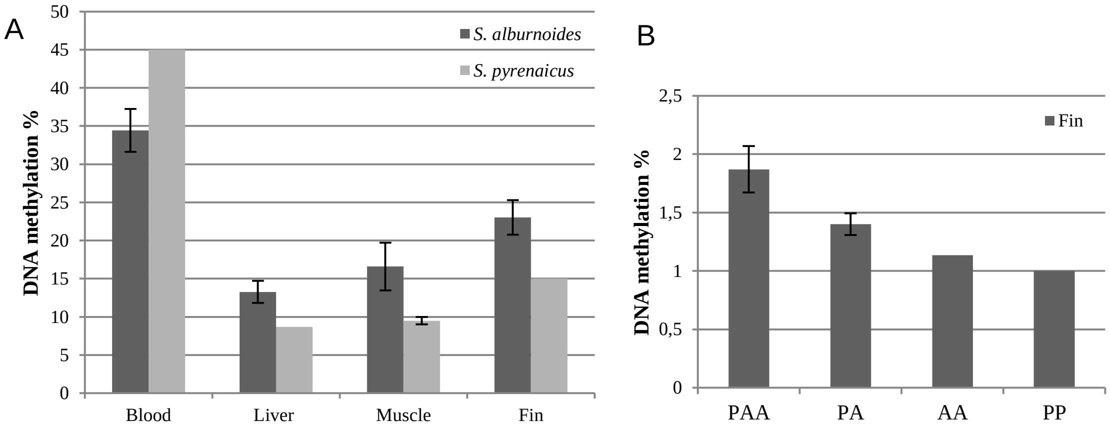


Fig 7 Quantification of 5-mC in genomic DNA. **A** 5-mC percentage of DNA from blood, liver, muscle and fin in the *S. alburnoides* complex, including diverse genomic compositions (PAA, PA, AA, and PPA) and in *S. pyrenaicus* **B** Fold-change of DNA 5-mC content from fin in PAA, PA, AA and PP. The data represent the average values ± standard deviation of the mean.

In particular, fin's DNA showed a pattern where non-hybrid genomic compositions AA and PP exhibited the least 5-mC content. Hybrid triploid PAA showed the highest percentage while PA intermediate values.

As mentioned before in order to expand the sample size, further fin's DNA methylation analysis was applied to *S. alburnoides* and *S. pyrenaicus* individuals using another method, the *Msp* I / *Hpa* II assay.

4.2.2. *Msp* I / *Hpa* II *assay*

In order to calculate DNA methylation, isoschizomers of restriction enzymes with different methylation sensibilities were used. This technique allows inferring the methylation content of a sample by measuring the densitometry intensities from images of smears obtained from electrophoresed gels with samples digested with one and another isoschizomer. Although recognising the same restriction site CCGG site, they show different sensibilities to certain methylation states of cytosine. Whereas *Hpa* II does not cut if one or the other cytosine are fully methylated (double strand), *Msp* I does not cleave if the external cytosine is fully or hemi methylated (single strand) (McClelland *et al*., 1994). So, even if not distinguishing all methylated states of CCGG sequences in the genome, they are a very good indicator of the methylation relative content, which could be calculated using the ratio between image densitometry results from both enzymes. Therefore, each DNA sample from fin was digested with either *Hpa* II or *Msp* I, run in two different lanes of an agarose gel and thereby the relative percentage of DNA methylation was calculated. Before, the efficiency of the restriction enzymes was confirmed by running in parallel the non-digested and the same DNA digested by *Hsp* II or *Msp* I (**Fig 8A**). An example of the *Msp* I/*Hpa* II assay is shown in **Fig 8B** where each sample was digested with either *Msp* I or *Hpa* II isoschizomers. In total, seven individuals from each genomic composition (PAA, PA, AA and PP) and three independent gels were analysed. Three PPA individuals were also studied. All the scan measurements were performed approximately between 1 and 4kb to guarantee that the saturation acquired in high and low molecular weight DNA was avoided.

After densitometry analysis, methylation degree was obtained and statistical analysis was performed in order to compare the different genomic compositions. Thus, results show that the hybrid triploid PAA presented the highest methylation content (23,2%), comparing to PA (7,9%), AA (10,2%), PP (8,7%) and PPA (13,5%) (**Fig 8C**). As well as in the previous technique the non-hybrid genomic compositions, namely AA and PP possessed their genome less methylated than the triploid PAA. On the other hand, AA also showed more methylation content than PP. Here, PA presented the lowest methylation degree, almost 3-fold lower than PAA and almost 2-fold lower than PPA. PPA seemed to possess intermediate methylation content. Hence, it seemed that part of the pattern obtained by this method remained preserved comparing to the general pattern achieved in immunoassay technique with PAA showing a higher methylation patterns than all the other genomic compositions.

In the statistical analysis, the dataset met the assumptions of homoscedasticity, however did not presented a normal distribution (assessed by Shapiro-Wilk test). Removal of one outlier that did not exhibit a normal distribution (one from PP which was confirmed by box-plot analysis) allowed the dataset to satisfy the ANOVA’s assumptions. Regardless of having datasets that fulfil or not ANOVA’s assumptions, two tests were performed, one including all data and another with the outlier removal. The first post hoc multiple comparisons Tukey test revealed that PAA possessed a significant higher methylation content comparing to the other genomic compositions PA (p=0.001), AA (p=0.006) and PP (p=0.002), but not to PPA, which did not differed significantly from any other genomic composition. AA and PP genomic compositions did not differ from each other as well as from diploid hybrid PA. Once PP outlier was removed the dataset interpretation remained the same although the significantly different presented changed p-values (p=0.000; p=0.002; p=0.000, respectively).

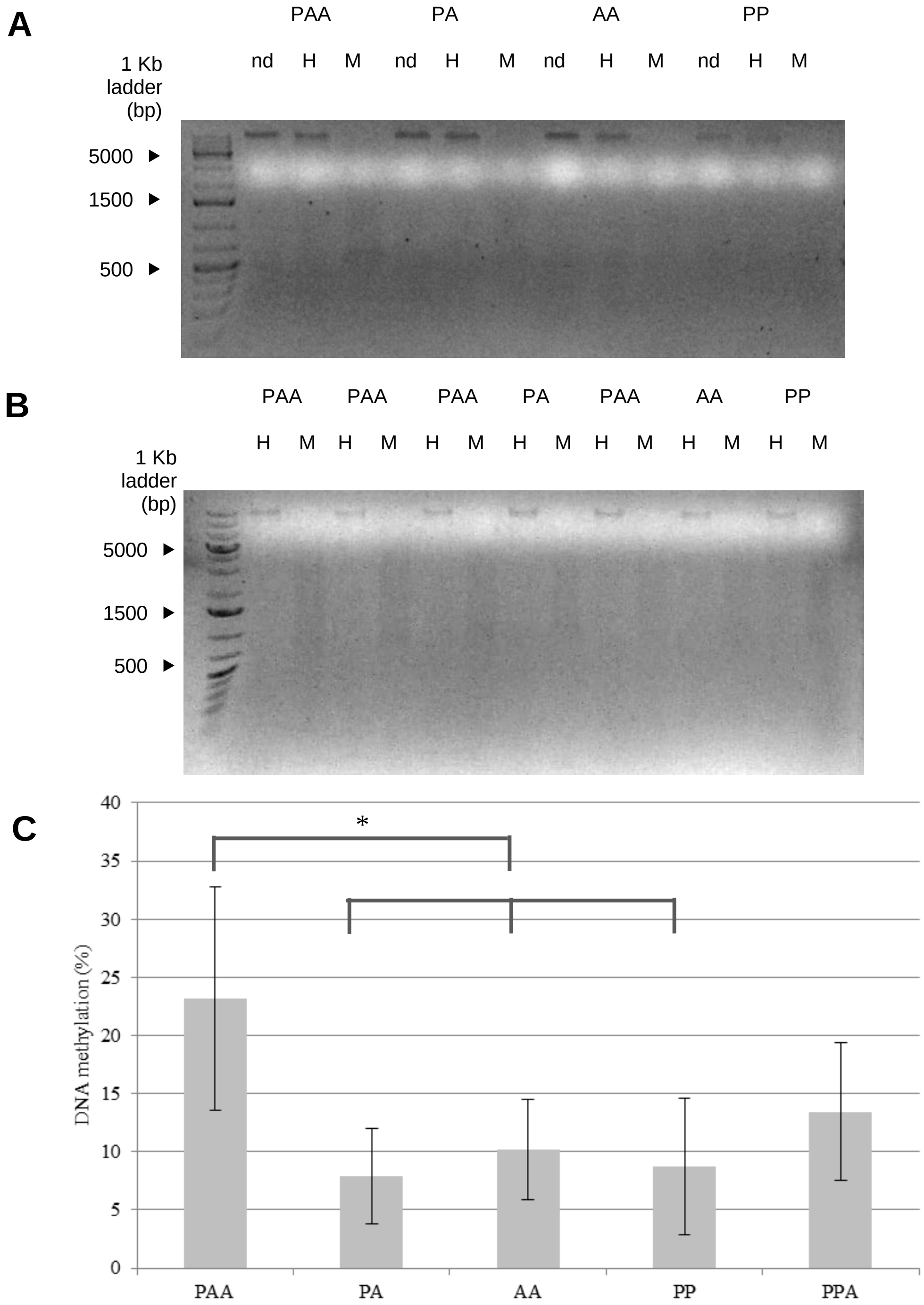


Fig 8 *Msp* I / *Hpa* II restriction assay. **A** Agarose gel presenting non-digested (nd), *Hpa* II (H) and *Msp* I (M) digested samples. **B** Example of an agarose gel of the *Msp* I/*Hpa* II assay, where lanes containing *Hpa* II and *Msp* I digested samples contained visible image densitometry variances. **C** DNA methylation percentage of each genomic composition analyzed (PAA, PA, AA, PP and PPA) and respective error bars. Statistical results (ANOVA) are also represented, * meaning statistically different with p<0,006

Additionally, since the assumptions were not fulfilled for the whole dataset, non-parametric tests were also accomplished. Thus, Kruskal-Wallis test pointed to the PAA having significantly higher methylation degree comparing to PA (p=0.016) and to PP (p=0.010). Here, however, the hybrid diploid PA do not show a significant lower methylation degree than PAA (p=0,210). All other comparisons were, as in parametric tests, not significantly different.

5. Discussion

5.1 MicroRNAs expression as a post-transcriptional regulation mechanism

MicroRNA expression profile of four libraries was assessed using NGS and microarrays hybridization. Since the proposal of this study was to compare different genomic compositions including different ploidy levels, the use of the same amount for each case, either in the NGS or in microarrays was attempt in order to facilitate the comparison among samples.

In the first approach, the platform used was Illumina technology which allowed obtaining a high volume of data. Thus, in this study, an overwhelming number of reads per library was provided, in comparison with the recent studies using the same platform. Whereas, on those, each run produced an average of only ≈ 6-10 million of high quality reads (Lu *et al.*, 2010; Xu *et al.*, 2010), the current work yielded ≈ 43 million, approximately four to seven times more.

Apart from the expected majority of 22nt reads containing miRNAs, it is intriguing that, by distributing the sequenced clean reads in a length plot, an evident peak of sequences with 31nt appears in the PAA genomic composition. Although it could be a technical artifact, it is not likely that this 31nt length sequences appeared randomly, since it is more evident on the triploid forms and it was not demonstrated any proportional increase in the other length classes. These triploid's sequences were two times more frequent than in the other libraries, suggesting a higher expression of approximatly 31nt length particular sequences. Once reads were annotated, this peak was, in fact, mainly composed by unannotated sequences. So, it is possible that some classes of non-coding small RNA, other than miRNA are being expressed in triploid forms rather than in the other diploid genomic compositions analysed (AA, PA and PP). Given its length, it might be that this greater peak was caused by the increase of piRNA, ranging between 24 and 30nt. Identified in germ cells (Siomi *et al*., 2011) and also somatic cells (Malone *et al*., 2009), piRNAs are directly involved in transposon repression, which would justify its increase in the triploid form. However, further investigation should be accomplished. In addition, the same was verified to sequences composed by 33nt length.

The majority of clean reads could be aligned to *Danio rerio* genome, since miRNAs are extremely conserved. Indeed, the obtained mapped reads, were higher than 60%, which is in accordance with former studies. Being this step a limiting phase to the whole analysis process, it was considered that the mapped reads allowed to greatly overcoming this step. Together, the high number of reads obtained, the length distribution indicating a high quantity of 22nt sequences and the percentage of reads mapped to the reference genome in all the libraries, indicate the great quality of the samples. Further analysis could, then, be performed.

An overall specific and common reads in each library, before the mapping, were also assessed.

Is was interesting to verify that PP presented more exclusive sequences, as expected from another species, while within the complex it was the PAA presenting the higher percentage, which could be related with the greater amount of the unannotated sequences of 31 and 33 nts in the PAA library.

After the genome alignment, sequences were annotated with several databases. MicroRNA proportions of each library were in accordance with other studies (Lu *et al.*, 2010; Chen *et al.*, 2005). Reads from this class of non-coding small RNA was then converted to TPM to make the libraries comparable with each other. This type of conversion is very useful since the total reads were different among libraries. However, this measurement does not consider, for example, the extent of ribosomal RNA remaining in the sample which could vary between the libraries. Consequently, it does not reflect the real accurate abundance into the libraries.

MicroRNA expression pattern

Observing both methods, RNA-seq and microarrays, it was clear that miRNAs are mostly differentially expressed between *Squalius pyrenaicus* and the genomic compositions of *Squalius alburnoides* complex. This can be explained by the fact that PP is a different independent species with its own microRNA expression pattern. Additionally, it was verified that PP, had always its own microRNA expression down-regulated in comparison with the individuals from the complex. Since *S. pyrenaicus* is only a sperm donor to the complex, its expression is not affected by any event of polyploidization or hybridization. On the other hand, it makes sense that miRNAs in *S. alburnoides* were more expressed than in *S. pyrenaicus*, since the constant shifting in the forms could create genomic shock, sometimes overcome by gene silencing mechanisms.

On the other hand, when comparing samples within the complex, the miRNA expression is expected to be different since constant new genome combinations are being formed causing genome instability. However, this did not seem to happen. In RNA-seq, when comparing the different forms of the complex by taking into account the number of miRNA significantly differently expressed and considering the number of up *vs.* down regulated as well as their fold-change, no significant differences were observed. In the microarray technology, the pattern of the comparison within the complex seems to be the similar, with all the forms presenting similar profiles, but the PP showing a slightly distinct pattern.

So in general the patterns between both methods are identical and thus, in spite of triploid hybrid (PAA) were differentially up-expressing a few miRNAs if compared to hybrid (PA), the results suggest that microRNAs do not have a major role in the genetic regulation of the complex. The better technique to evaluate a transcriptome is still controversial. Hence, individual analysis of each miRNA was performed to comprehend if the differential up and down-regulated miRNA and their pattern were the same. Despite de similarity between the patterns obtained in both techniques, individual analysis did not show a reproducible result, contrarily to what was postulated by Willenbrock *et al.* (2009). Indeed, it was verified that the differently expressed miRNAs in both methods were mostly not the same.

This could be have been originated by several reasons. Microarray analysis is always subjected to cross hybridization or background. Also, the indirect measurement through fluorescence, not having the sensibility to detect slightly differences of miRNA expression, is another huge limitation of this method that could have influenced these data. Besides, despite the conservation of miRNA among taxa, the using of *Danio rerio* genome for hybridization or mapping of the sequences could also be a limiting step to the reproducibility of this study. Since, both methods present different sensibilities, reproducibility might even be diminished due to the presence of low miRNA expression differences between the libraries.

Interestingly, at least in RNA-seq, where almost no outliers were verified, the comparisons between triploid and diploids within the complex (PAA *vs*. PA and PAA *vs*. AA), were identified as the most similar. Besides, interesting to note was the fact that once considering all miRNAs, including the no significant differently expressed, the only comparison showing no significantly difference in the amount of up *vs*. down regulated miRNAs was PAA *vs*. AA. Those results are in accordance with the preferentially silencing of P alleles reported for mRNAs, suggesting that the miRNA is not playing a role in gene silencing, but rather adapting to the environment of a preferentially expression of just A alleles, possible potentiate by another regulatory mechanism .

5.2 DNA methylation as a pre-transcriptional regulation mechanism

As microRNA were not significantly different expressed between the genomic compositions in *S. alburnoides*, a pre-transcriptional mechanism might be operating in order to overcome the genome instability created by the constant shifting of genomes within the complex.

DNA methylation was then investigated using the same genomic compositions of the previous approach from several organs. Besides, some rare PPA could also be included in this analysis. On the other hand, two independent methodologies were used.

In the first methodology, 5-methylcytosine relative percentage content was assessed in *S. alburnoides* complex and in *S. pyrenaicus* by organ. In spite of the exception observed in blood, in the majority of organs (liver, muscle and fin) the complex showed a confident higher methylation percentage than in *S. pyrenaicus*. Therefore, this suggests that this pre-transcriptional mechanism is operating in the allopolyploid complex, at least in some of the tissues.

It is quite obvious the great differences among the analyzed tissues. Our results showed that blood cells possessed a higher 5-mC. Blood was found to be more methylated according to the few genes under expression (mainly globins). Contrarily, liver was described to possess a high variety of gene expression (Shen *et al*., 2006) and in our data, the tissue showing the lowest methylation was precisely the liver, which might be related with a higher gene expression in this tissue. Although the real sensibility of this method is unknown, the fold difference of 4 between blood and liver should be considered. Moreover, the pattern observed between tissues is constant in both species.

In the second methodology, endonucleases with different sensibilities to methylation were used in order to answer the same question. Therefore, although the isoschizomers used do not distinguish all the cytosine states in CCGG sequences, it was considered an informative method to obtain the relative methylation content (Mhanni & McGowan *et al*., 2004). In fact, when methylation of fin was accessed by this method the pattern obtained was similar to the one obtained specifically in fin in the previous approach. Here, triploid hybrids showed a significantly higher methylation content than any of the diploids (PA, AA, PP), that did not show significantly different DNA methylation content between each other. Moreover, PPA which presented an *intermediate* DNA methylation did not exhibited any statistical differences from all the other genomic compositions.

As by the conjugation of both methods, the DNA methylation content in the fin of hybrid diploid PA is uncertain, it only can be suggested that there is difference between PAA and PA, being the first more methylated than the last.

Results point to the existence of a pre-transcriptional regulatory mechanism by DNA methylation in the triploid forms, rather than a massive post-transcriptional control. In an evolutionary point of view it does not make sense to waste energy transcribing genes that will then be regulated by miRNA through blocking the translation or destroying the transcripts. Thus, methylation should function as a primary step of regulation in order to buffer the polyploidy effect and, only then, microRNAs should be used as a tool to fine-tuning the allopolyploid expression of some essential genes is particular. The lack of a clear effect on DNA methylation in the PPA form can be related with its poor statistical power (3 samples only). The fact that there are less PPAs in nature could be also related to a weak efficiency of DNA methylation, compromising the gene expression regulation and thus the normal development of the fish, eventually not surviving as much as the others forms.

6. References

Adams KL (2007) Evolution of duplicated gene expression in polyploidy and hybrid plants. *Journal of Heredity* 98, 136-141.

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences* 100, 4649-4654.

Adams KL, Percifield R, Wendel JF (2004) Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. *Genetics* 168, 2217-2226.

Albertin W, Balliau T, Brabant P, Chevre A, Eber F, Malosse C, Thiellement H (2006) Numerous and rapid nonstochastic modifications of gene products in newly synthesized brassica napus allotetraploids. Genetics 173, 1101-1113.

Alves MJ, Coelho MM, Collares-Pereira MJ (1998) Diversity in the reproductive modes of females of the *Rutilus alburnoides* complex (Teleostei, Cyprinidae): a way to avoid the genetic constraints of uniparentalism. *Molecular Biology and Evolution* 15, 1233-1242.

Alves MJ, Coelho MM, Collares-Pereira MJ (2001) Evolution in action through hybridization and polyploidy in an Iberian freshwater fish: a genetic review. *Genetica* 111, 375-385.

Alves MJ, Coelho MM, Collares-Pereira MJ, Dowling TE (1997b) Maternal ancestry of the *Rutilus alburnoides* complex (*teleostei*, *Cyprinidae*) as determined by analysis of cytochrome b sequences. *Evolution* 51, 1584-1592.

Alves MJ, Coelho MM, Collares-Pereira MK (2001) Evolution in action through hybridization and polyploidy in an Iberian fresh water fish: a genetic review. *Genetica* 111, 375–385.

Alves MJ, Gromicho M, Collares-Pereira MJ, Crespo-Lopez E, Coelho MM (2004) Simultaneous production of triploid and haploid eggs by triploid *Squalius alburnoides* (Teleostei, Cyprinidae). *Journal of Experimental Zoology* 301A, 552-558.

Alves MJ, Coelho MM, Collares-Pereira MJ (2001) Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. *Genetica* 111, 375-385.

Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D (2003) MicroRNAs and other tiny endogenous RNAs in *Caenorhabditis elegans*. *Current Biology* 13, 807-818.

Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761-764.

Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761-764.

Audic S, Claverie J (1997) The significance of digital gene expression profiles. Genome Research l7, 986-995.

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136, 215-223.

Baulcombe D. (2006) Short silencing RNA: the dark matter of genetics? *Cold Spring Harbor Symposia on Quantitive Biology* LXXI, 13-20.

Bloom JS, Khan Z, Kruglyak L, Singh M, Caudy AA (2009) Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics* 10, 221.

Bona F, Ossowski S, Schneeberger K, Rätsch G (2008) Optimal spliced alignments of short sequence reads. *BMC Bioinformatics* 24, 174-180

Brennecke J, Stark A, Russel RB, Cohen SM (2005) Principles of microRNA – target recognition. PLoS Biology 3, e85.

Buckingham S (2003) The major world of microRNAs. *Nature* (published online), http://www.nature.com/horizon/rna/background/microRNAs.html.

C.'t Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, Menezes RX, Boer JM, van Ommen GJB, Dunnen (2008) Deep sequencing based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Research* 36, e141.

Campagna D, Albiero A, Bilardi A, Caniato E, Forcato C, Manavski S, Vitulo N, Valle G (2009) PASS: a program to align short sequences. *Bioinformatics* 25, 967-968.

Carthew RW, Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. Cell 136, 642-655.

Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Reviews Genetics* 6, 351-360.

Chen C, Durand E, Forbes F (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Molecular Ecology 7, 747-756.

Chen H, Nakamura A, Sugimoto J, Sakumoto N, Oda R, Jinno Y, Okazaki Y (2006) Tissue specificity of methylation and expression of human genes coding for neuropeptides and theirs receptors, ando f a human endogenous retrovírus K family. *Journal of Human Genetics* 51, 440-450.

Chen PY, Manninga H, Slanchev K, Chien M, Russo JJ, Ju J, Sheridan R, John B, Marks D, Gaidatzis D, Sander C, Zavolan M, Tuschl T (2005) The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes & Development* 19, 1288-1293

Chi W, Tong C, Gan X, He S (2011) Characterization and Comparative Profiling of MiRNA, *PloS ONE* 6, e23549.

Collares-Pereira MJ (1983) Estudo sistemático e citogenético dos pequenos ciprinídeos ibéricos pertencentes aos géneros *Chondrostoma agassiz*, 1835, *Rutilus rafinesque*, 1820 e *anaecypris*. PhD. Thesis, University of Lisbon, Lisbon, Portugal.

Collares-Pereira MJ, Coelho MM (2010) Reconfirming the hybrid origin and generic status of the iberian cyprinid complex *Squalius alburnoides*. *Journal of Fish Biology* 76, 707-715.

Comai L (2005) Advantages and disadvantages of polyploidy. *Nature Reviews Genetics* 6, 836-846

Costa V, Angelini C, D'Apice L, Mutarelli M, Casamassimi A, Casamassimi A, Sommese L, Gallo MA, Aprile M, Esposito R, Leone L, Donizetti A, Crispi S, Rienzo M, Sarubbi B, Calabro R, Picardi M, Salvatore P, Infante T, De Berardinis P, Napoli C, Ciccodiola A (2011) Massive-Scale RNA-Seq Analysis of Non Ribosomal Transcriptome in Human Trisomy 21. PLoS ONE 6, e18493.

Costa V, Angelini C, De Feis I, Ciccodicola A (2010) Uncovering the complexity of the transcriptomes with RNA-seq. *Journal of Biomedicine Biotechnology* 2010, ID 853916.

Coughlan SJ, Agrawal V, Meyers B (2004) A comparison of global gene expression measurement technologies in *Arabidopsis thaliana*. *Comparative and Functional Genomics* 5, 245-252.

Crespo-López ME, Duarte T, Dowling T, Coelho MM (2006). Modes of reproduction of the hybridogenetic fish *Squalius alburnoides* in the Tejo and Guadiana rivers: an approach with microsatellites. *Zoology* 109, 277-286.

Dahiya N, Sherman-Baust CA, Wang TL, Davidson B, Shin IM, Zhang Y, Wood W 3rd, Becker KG, Morin PJ (2008) MicroRNA expression and identification of putative miRNA targets in ovarian cancer. *PLOS One* 3, e2436.

Denoeud F, Aury J, Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C, Jaillon O, Artiguenave F (2008) Annotating genomes with massive-scale RNA sequencing. *Genome Biology* 9, i12

Dhar MS, Pethe VV, Gupta VS, Ranjekar PK (1990) Predominance and tissue specificity of adenine methylation in rice. *Theoretical and Applied Genetics* 80, 402-408.

Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature Review Genetics 9, 102-114.

Finnegan EJ, Brettell RI, Dennis ES (1993) The role of DNA methylation in the regulation of plant gene expression. In *DNA Methylation: Molecular Biology and Biological Significance* (ed. J.P. Jost and H.P. Saluz), 218–261.

Gallardo MH, Gonzalez CA, Cebrian I (2006) Molecular cytogenetics and allotetraploidy in the red vizca cha rat, *Tympanoctomys barrerae* (*Rodentia, Octodontidae*). *Genomics* 88, 214-221.

Goll MG, Bestor TH (2005) Eukaryotic methylcytosine methyltransferases. *Annual Review of Biochemistry* 74, 481-514.

Gonzalgo ML, Jones PA (1997a) Mutagenic and epigenetic effects of DNA methylation. *MutationResearch* 386, 107-118.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Research 36, 154-158.

Gromicho M, Coelho MM, Alves MJ, Collares-Pereira MJ (2006). Cytogenetic analysis of A*naecypris hispanica* and the relationship with the paternal ancestor of the diploid-polyploid *Squalius alburnoides* complex. *Genome* 49, 1621-1627.

Gromicho M, Collares-Pereira MJ (2004) Polymorphism of major ribosomal gene chromosomal sites (NOR-phenotypes) in the hybridogenetic fish *Squalius alburnoides* complex (Cyprinidae) assessed through crossing experiments. *Genetica* 122, 291-302.

Gromicho M, Collares-Pereira MJ (2007) The evolutionary role of hybridization and polyploidy in an Iberian cyprinid fish – a cytogenetic review. In: *Fish Cytogenetics* (eds. Pisano E, Ozouf-Costaz C, Foresti F, Kapoor BG), pp. 41-67. Science Publishers, Enfield.

Guo J, Miao Y, Xiao B, Huan R, Jiang Z, Meng D, Wang Y (2009) Differential expression of microRNA species in human gastric cancer versus non-tumorous tissues. *Journal of Gastroenterology and Hepatology* 24, 652-657.

Guo M, Davis D, Birchler JA (1996) Dosage effects on gene expression in a maize ploidy series. *Genetics* 142, 1349-1355.

Ha MS, Pang M, Agarwal V, Chen ZJ (2008) Interspecies regulation of microRNAs and their targets. Biochim Biophys Acta 1779, 735-742.

Ha MS, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ (2009) Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences* 106, 17835–17840.

Hale MC, McCormick CR, Jackson JR, DeWoody JA (2009) Next-generation pyrosequencing of gonad transcriptomes in the polyploidy lake sturgeon (*Acipenser fulvescens*): The relative merits of normalization and rarefaction in gene discovery. *BMC Genomics* 10, 203.

Heard E, Disteche CM (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome, *Genes Development* 20, 1848-1867.

Heard E, Disteche CM (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes & Development* 20, 1848-1867.

Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nature Review Genetics 12, 99-110.

Inácio A, Matos I, Machado M, Coelho MM (2010) An easy way to characterize the genomic composition of allopolyploids: the example of the Iberian cyprinid *Squalius alburnoides*. J*ournal of Fish Biology* 76, 1995–2001.

Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JGN, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martínez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye SQ, Yu W (2005) Multiple-laboratory comparison of microarray platforms. *Nature Methods* 2, 345-349.

Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martínez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye SQ, Yu W (2005) Multiple laboratory comparison of microarray platforms. *Nature Methods* 2, 345-350.

Josefsson C, Dilkes B, Comai L (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology* 16, 1322-1328.

Kim J, Kim JY, Issa JPJ(2009) Aging and DNA Methylation. *Current Chemical Biology* 3, 321-329.

Kim VN (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Molecular Cell Biology: Nature Reviews* 6, 376-385.

Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *TRENDS in Biochemical Sciences* 31, 89-97.

Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. Nature Review Genetics 11, 597-610.

Le Comber SC, Smith C (2004) Polyploidy in fishes: patterns and processes. *Biological Journal of the Linnean Society* 82, 431-442.

Lee RC, Feinbaum RL, Ambros V (1993) The *Caenorhabditis elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843-854.

Lee S, Seo CH, Lim B, Yang JO, Oh J, Kim M, Lee S, Lee B, Kang C, Lee S (2010) Accurate quantification of transcriptome from RNA-seq data by effective length normalization. *Nucleic Acids Research* 39, e9.

Lee HS, Chen ZF (2001) Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proceedings of the National Academy of Sciences* 98, 6753-6758.

Leggatt RA, Iwama GK (2003) Occurrence of polyploidy in the fishes. *Reviews in Fish Biology and Fish eries* 13, 237-246.

Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15-20.

Li B, Ruotti V, Stewart RM, Thomson JM, Dewey CN (2009) RNA-seq gene expression estimation with read mapping uncertaintly. *Bioinformatics* 26, 453-500.

Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* 24, 713-714

Liu B, Vega JM, Feldman M (1998) Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences. *Genome* 41, 535-542.

Liu B, Wendel JF (2003) Epigenetic phenomena and the evolution of plant allopolyploids. *Molecular Phylogenetics and Evolution* 29, 365-379

Liu F, Jenssen TK, Trimarchi J, Punzo C, Cepko CL, Ohno-Machado L, Hovig E, Kuo WP (2007) Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates. *BMC Genomics* 8, 153.

Lu J, Chen ZJ (2011) Small RNA inheritance in hybrids and allopolyploids. *RNA Technologies,* 91-106.

Lu Y, Li C, Zhang K, Sun H, Tao D, Liu Y, Zhang S, Ma Y (2010) Identification of piRNA in Hela cells by massive parallel sequencing. *Biochemestry and Molecular Biology Reports* 43, 635-641

Lund G, Messing J, Viotti A (1995) Endosperm-specific demethylation and activation of specific alleles of a-tubulin genes of *Zea mays* l. *Molecular and General Genetics* 246, 716-722.

Ma X, Gustafson JP (2005) Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenetic and Genome Research* 109, 236-249.

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davidson J, Comai L (2002) Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* Allotetraploids", *Plant Physiology* 129, 733-746.

Mallet J (2007) Hybrid speciation. Nature 445, 279-283.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ (2009) Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137, 522-535.

Malone JH, Oliver B (2011) Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biology 9, 1-9.

Mardis ER (2008) Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics 9, 387-402.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNAseq: An assessment of technical reproducibility and comparisonwith gene expression arrays. Genome Research 18, 1509-1517.

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proceedings of the National Academy of Sciences 74, 560-564.

McClelland M, Nelson M, Raschke E (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Research* 22, 1640-3659.

McClelland M, Nelson M, Raschke E (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Research* 22, 3640-3659.

Messeguer R, Ganal MW, Steffens JC, Tanksley SD (1991) Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. *Plant Molecular Biology* 16, 753-770.

Metzker ML (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics* 11, 31-44.

Metzker ML (2010) Sequencing technologies – the Next Generation. Nature Reviews Genetics 11, 31-46

Mhanni AA, McGowan RA (2004) Global changes in genomic methylation levels during early development of the zebrafish embryo. *Development Genes and Evolution* 214, 412-417.

Miller SA., Dykes DD, Polesky H. F (1988) A simple salting out procedure for extracting DNA from human nucleated cells" Nucleic Acids Research 16, 1215.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5, 621-628.

O'Neill RJW, O'Neill MJ, Graves JAM (1998). Undermethylation associated with retroelement activation and chromosome remodeling in an interspecific mammalian hybrid. *Nature* 393, 68-72.

O'Donnell KA, Boeke JD (2007) Mighty Piwis defend the dermline against genome intruders. *Cell* 129, 37-44.

Ohno S (1970) Evolution by gene duplication. George Allen and Unwin, London

Orr HA (1990) "Why polyploidy is rarer in animals than in plants" revisited. *American Naturalist* 136, 75 9-770.

Otto SP (2007) The evolutionary consequences of polyploidy. *Cell* 131, 452-462.

Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annual Review of Genetics* 34, 401-437.

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (Aegilops-Triticum) group. *Plant Cell* 13, 1735–1747.

Pala I, Coelho MM (2005) Contrasting views over a hybrid complex: between speciation and evolutionary "dead-end". *Gene* 347, 283-294.

Pala I, Coelho MM, Schartl M (2008) Dosage compensation by gene-copy silencing in a triploid hybrid fish. *Current Biology* 18, 1344-1348.

Pala I, Schartl M, Brito M, Vacas JM, Coelho MM (2010) Gene expression regulation and lineage evolution: the North and the South tale of the hybrid polyploid squalius alburnoides complex. *Proceedings of the Royal Society* B.277, 3519-3525.

Próspero MJ, Collares-Pereira MJ (2000) Nuclear DNA content variation in the diploid-polyploid *Leuciscus alburnoides* complex (Teleostei, Cyprinidae) assessed by flow cytometry. *Folia Zoologica* 49, 53-58.

Qu W, Hashimoto S, Morishita S (2009) Efficient frequency-based de novo short-read clustering for erros trimming in next-generation sequencing. *Genome Reasearch* 19, 1309-1315.

Rai K, Nadauld LD, Chidester S, Manos EJ, James SR, Karpf AR, Cairns BR, Jones DA(2006) Zebrafish Dnmt1 and Suv39h1 regulate organ-specific terminal differentiation during development, *Molecular and Cellular Biology* 26,7077-7085.

Reinhart BJ, Bartel DP (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science* 297, 1831.

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-Nucleotide let-7 RNA regulates developmental timing in caenorhabditis elegans. Nature 403, 901-906.

Richards EJ (2006) Inherited epigenetic variation – revisiting soft inheritance. Nature Reviews Genetics 7, 395-401.

Riddle NC, Birchler JA (2003) Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybids. *Trends in Genetics* 19, 597-600.

Robalo JI, Sousa-Santos C, Levy A, Almada VC (2006) Molecular insights on the taxonomic position of the paternal ancestor of the *Squalius alburnoides* hybridogenetic complex. *Molecular Phylogenetics and Evolution* 39, 276-281.

Salmon A, Ainouche ML, Wendel JF (2005) Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (*poaceae*). *Molecular Ecology* 14, 1163-1175.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74, 5463-5467.

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13, 1749–1759.

Shan XH, Liu ZL, Dong ZY (2005) Mobilization of the active mite transposons mPing and Pong in rice by introgression from wild rice (*Zizania latifolia* griseb). *Molecular Biology and Evolution* 22, 976-990.

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nature Biotechnology 26, 1135-1145.

Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology* 12, 246-258.

Sousa-Santos C, Collares-Pereira MJ, Almada V (2007b) Fertile triploid males – an uncommon case among hybrid vertebrates. *Journal of Experimental Zoology* 307A, 220-225.

StatSoft I (2009) STATISTICA (data analysis software system). Available at: www.statsoft.com.

StatSoft I (2010) Electronic Statistics Textbook (Electronic Version). Available at: http://www.statsoft.com/textbook/.

Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays news and reviews in molecular cellular and developmental biology 29, 288-299.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-sequencing. *Bioinformatics* 25, 1105-1111.

Udall JA, Swanson JM, Nettleton D, Percifield RJ, Wendel JF (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* 173, 1823-1827.

Ungerer MC, Strakosh SC, Zhen Y (2006) Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Current Biology* 16, 872-873.

Vazquez F, Gasciolli V, Crete P, Vaucheret H (2004) The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but no posttranscriptional transgene silencing. *Current Biology* 14, 346-351.

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DEJ, Hieter P, Vogelstein B, Kinzler KW (1997) Characterization of the yeast transcriptome. Cell 88, 243-251.

Voinnet O (2009) Origin, biogenesis and activity of plant microRNAs. Cell 136, 669-687.

Vrana PB, Fossella JA, Matteson P, del Rio T, O'Neill MJ, Tilghman SM (2000) Genetic and epigenetic incompatibilities underlie hybrid sdysgenesis in peromyscus. *Nature Genetics* 25, 120-124.

Wang J, Tian L, Lee HS, Wei NE, Jiang H (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172, 507-517.

Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen J (2006b) Genomewide non-additive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172, 507–517.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10, 57-63.

Wendel JF, Schnable A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences* 92, 280-284.

Willenbrock H, Salomon J, Søkilde R, Barken KB, Hansen TN, Nielsen FC, Møller S, Litman T (2009) Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. *RNA* 15, 2028-2034.

Winter J, Jung S, Keller S, Gregory RI, Diederichs S (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. Nature Cell Biology 11, 228-234.

Xiong LZ, Xu CG, Saghai-Maroof MA, Zhang QF (1999) Paterns of cytosine methylation in an elite rice hybrid and its parental lines by a methylation-sensitive amplification polymorphism technique. *Molecular and General Genetics* 261: 439-446.

Xiong Y, Chen X, Chen Z, Wang X, Shi S, Wang X, Zhang J, He X (2010) RNA sequencing shows no dosage compensation of the active X-chromosome. *Nature Genetics* 42, 1043-1047.

Xu Q, Liu Y, Zhu A, Wu X, Ye J, Yu K, Guo W, Deng X (2010) Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type. *BMC Genomics* 11, 246

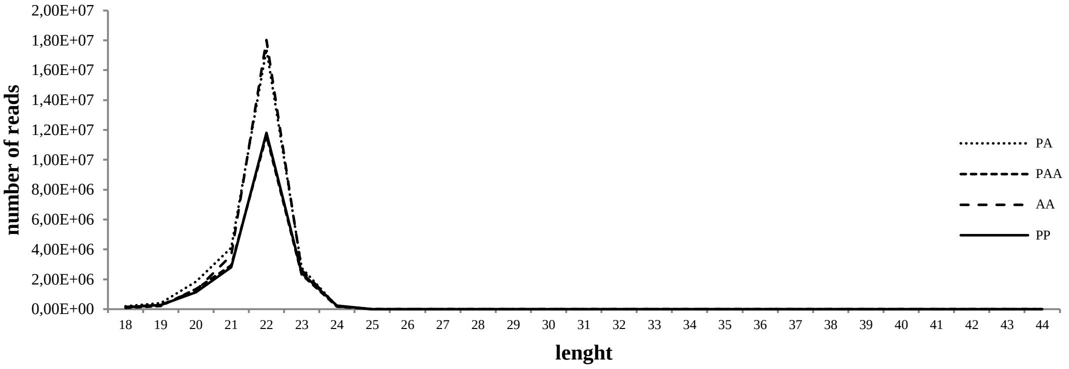
Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006) Conservation and divergence of plant microRNA genes. *The Plant Journal* 46, 243-259.

Zhong L, Xu Y, Wang J (2009) DNA-methylation changes induced by salt stress in wheat *Triticum aestivum*. *African Journal of Biotechnology* 8, 6201-6207.

Zilberman D, Gehring M, Tran RK, Ballinger T, Hanikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics* 39, 61-69.

Zilberman D, Henikoff S (2007) Genome-wide analysis of DNA methylation patterns. *Development* 134, 3959-3965.

7. APPENDIX



FigS1 Length analysis of sequences identified as miRNA in each library, PA, PAA, AA and PP.

Tab S1 Annotation of sequences of each library (AA, PP, PA and PAA)

	AA				PP				PA				PAA			
Category	**Unique sRNAs**	**Percent (%)**	**Total sRNAs**	**Percent (%)**	**Unique sRNAs**	**Percent (%)**	**Total sRNAs**	**Percent (%)**	**Unique sRNAs**	**Percent (%)**	**Total sRNAs**	**Percent (%)**	**Unique sRNAs**	**Percent (%)**	**Total sRNAs**	**Percent (%)**
miRNA	6300	0.30%	26168020	58.96%	6179	0.24%	18814080	44.22%	6595	0.39%	26938989	62.45%	6229	0.26%	18666251	42.95%
unann	1905709	89.34%	12758390	28.75%	2324670	89.17%	15743617	37.00%	1506461	89.05%	12381313	28.70%	2139007	90.70%	18149423	41.76%
rRNA	76055	3.57%	1661352	3.74%	67960	2.61%	1707323	4.01%	61014	3.61%	1687768	3.91%	74553	3.16%	2535056	5.83%
snRNA	9275	0.43%	54724	0.12%	11681	0.45%	128340	0.30%	8156	0.48%	59797	0.14%	11183	0.47%	93142	0.21%
snoRNA	5695	0.27%	67818	0.15%	7910	0.30%	91638	0.22%	7369	0.44%	99150	0.23%	9830	0.42%	163141	0.38%
tRNA	73725	3.46%	2672865	6.02%	104313	4.00%	4494841	10.56%	54130	3.20%	1330317	3.08%	63677	2.70%	2813046	6.47%
Exon antisense	9435	0.44%	106634	0.24%	8874	0.34%	240268	0.56%	7753	0.46%	70305	0.16%	8224	0.35%	97637	0.22%
Exon sense	34528	1.62%	700137	1.58%	60998	2.34%	1027195	2.41%	29763	1.76%	459180	1.06%	34489	1.46%	791073	1.82%
Intron antisense	6844	0.32%	79469	0.18%	7759	0.30%	187842	0.44%	6062	0.36%	42862	0.10%	6242	0.26%	61297	0.14%
Intron sense	5512	0.26%	115119	0.26%	6731	0.26%	114563	0.27%	4311	0.25%	67837	0.16%	5010	0.21%	91026	0.21%
Total	2133078	100%	44384528	100%	2607075	100%	42549707	100%	1691614	100%	43137518	100%	2358444	100%	43461092	100%

Tab S2 Significantly miRNA up- and down-regulated, their TPM, fold-change and respective p-value of the following comparisons PA-AA; AA-PAA; PA-PAA; AA-PP; PA-PP; PAA-PP

miR-name	PA (TPM)	AA (TPM)	fold-change (log2 AA/PA)	p-value
dre-let-7f	113.657.675	55.894.703	-1,023911635	0
dre-miR-1	905.279.251	295.271.587	-1,616320351	0
dre-miR-126*	2.902.810	1.433.382	-1,018027002	0
dre-miR-130a	153.926	73.224	-1,07184845	9,62E-16
dre-miR-133c	929.585	267.886	-1,794967603	0
dre-miR-150	3.228.512	1.547.161	-1,061246056	0
dre-miR-15a*	203.767	100.035	-1,026415572	2,72E-22
dre-miR-19b*	95.277	43.258	-1,139161023	8,07E-07
dre-miR-29b	6.936.421	3.460.665	-1,003142172	0
dre-miR-301a	315.966	143.969	-1,134011125	9,39E-51
dre-miR-499	1.907.157	545.460	-1,805878304	0
dre-miR-736	14.604	0,2704	-15,7209089	4,71E+04
dre-miR-107b	812.981	2.840.855	1,805031656	0
dre-miR-122	36.150.666	153.866.005	2,089580388	0
dre-miR-18c	207.476	835.201	2,009178978	0
dre-miR-194a	4.679.917	11.083.141	1,243811956	0
dre-miR-199	3.488.379	7.059.442	1,016997361	0
dre-miR-19c	603.419	1.466.502	1,28114701	0
dre-miR-19d	468.965	1.134.179	1,274096189	4,07E-258
dre-miR-202*	116.604	1.246.606	3,418316377	0
dre-miR-20b	401.275	2.802.103	2,803846805	0
dre-miR-22a	796.831.195	1.813.913.849	1,186759904	0
dre-miR-375	380.179	1.657.560	2,124310345	0
dre-miR-459	53.550	505.131	3,237699103	0
dre-miR-459*	12.982	74.125	2,513447538	1,10E-32
dre-miR-726	53.318	157.487	1,562538174	6,22E-39
dre-miR-93	2.566.212	5.237.411	1,029213464	0

miR-name	PA (TPM)	PAA (TPM)	fold-change (log2 PAA/PA)	p-value
dre-miR-1	905.279.251	314.949.288	-1,523243338	0
dre-miR-133a*	1.870.993	825.566	-1,1803487	0
dre-miR-133c	929.585	463.633	-1,003603533	1,17E-137
dre-miR-27b	9.833.899	4.479.179	-1,134529217	0
dre-miR-27d	1.884.439	759.990	-1,310082754	0
dre-miR-34c	221.849	100.320	-1,144968797	3,75E-32
dre-miR-499	1.907.157	920.133	-1,051509299	0
dre-miR-724	28.170.605	12.305.029	-1,19494249	0
dre-miR-736	14.604	0,5062	-14,81629654	5,57E+08
dre-miR-146b	1.257.606	4.865.041	1,951771958	0
dre-miR-20b	401.275	863.301	1,105272383	1,34E-150
dre-miR-223	173.167	357.101	1,044168148	3,04E-49
dre-miR-192	32.442.988	78.679.109	1,278073922	0
dre-miR-31	96.204	201.790	1,068685896	7,77E-25
dre-miR-375	380.179	1.072.914	1,496783692	0
dre-miR-429b	11.823	25.080	1,084941193	4,80E+08
dre-miR-459	53.550	519.775	3,278928765	0
dre-miR-459*	12.982	127.010	3,290357524	6,00E-89
dre-miR-96	769.632	2.170.447	1,495751504	0

miR-name	AA (TPM)	PAA (TPM)	fold-change (log2 PAA/AA)	p-value
dre-miR-103	58.423.287	26.749.903	-1,127009869	0
dre-miR-107	2.448.601	1.050.365	-1,221066957	0
dre-miR-107b	2.840.855	692.343	-2,036766338	0
dre-miR-122	153.866.005	58.154.544	-1,403710691	0
dre-miR-18c	835.201	224.569	-1,894964651	0
dre-miR-196a	21.629	0,7363	-14,84231374	1,69E+06
dre-miR-19c	1.466.502	710.981	-1,044496127	2,78E-243
dre-miR-202*	1.246.606	133.913	-3,218637635	0
dre-miR-203a	1.119.534	454.890	-1,299308718	4,62E-264
dre-miR-203b	1.210.782	604.909	-1,001149102	1,37E-184
dre-miR-20b	2.802.103	863.301	-1,698574423	0
dre-miR-2188*	40.329	19.788	-1,027191828	2,25E+06
dre-miR-22a	1.813.913.849	504.265.746	-1,846849804	0
dre-miR-457b	16.447	0,4142	-15,27713738	5,59E+05
dre-miR-724	29.989.054	12.305.029	-1,285187955	0
dre-miR-726	157.487	60.514	-1,379891888	1,84E-30
dre-miR-93	5.237.411	1.938.976	-1,433558877	0
dre-miR-130b	768.286	1.726.142	1,167835782	0
dre-miR-141	606.743	1.264.809	1,059762074	4,87E-214
dre-miR-146b	1.603.712	4.865.041	1,601036882	0
dre-miR-19b*	43.258	101.010	1,223459257	1,66E-11
dre-miR-31	79.307	201.790	1,347334566	1,08E-40
dre-miR-96	550.192	2.170.447	1,979985126	0

miR-name	PA (TMP)	PP (TPM)	fold-change (log2 PP/PA)	p-value
dre-miR-1	905.279.251	40.404.979	-4,4857579	0
dre-miR-101a	219.416.889	38.647.035	-2,505244937	0
dre-miR-10a	838.713	406.818	-1,043793704	2,42E-130
dre-miR-10b	1.732.367	567.806	-1,609274597	0
dre-miR-10d	142.799	17.861	-2,999101115	1,30E-90
dre-miR-133a*	1.870.993	106.229	-4,138554587	0
dre-miR-133b	11.504.603	5.398.862	-1,091483954	0
dre-miR-133c	929.585	346.888	-1,422116855	7,01E-245
dre-miR-1388*	121.240	47.474	-1,352656241	7,36E-19
dre-miR-142a-3p	7.364.819	3.251.726	-1,179444373	0
dre-miR-142a-5p	880.904	395.067	-1,156887461	1,36E-164
dre-miR-142b-5p	61.431	26.087	-1,235635824	2,46E-01
dre-miR-148	35.250.058	14.698.574	-1,261949434	0
dre-miR-152	18.381.215	2.943.381	-2,642685926	0
dre-miR-15a*	203.767	32.668	-2,640970388	4,91E-116
dre-miR-16b	154.730.506	69.253.826	-1,159791979	0
dre-miR-181a*	637.032	254.291	-1,324885443	5,04E-146
dre-miR-182	2.432.917	329.497	-2,884349868	0
dre-miR-183	2.981.859	842.544	-1,823388105	0
dre-miR-199*	3.056.968	1.090.019	-1,487748168	0
dre-miR-202*	116.604	54.524	-1,096653969	3,31E-10
dre-miR-206	98.111.579	33.359.102	-1,556342958	0
dre-miR-216a	5.450.939	1.594.371	-1,773517401	0
dre-miR-216b	1.540.422	608.230	-1,340636751	0
dre-miR-2184	336.366	125.970	-1,416951691	2,06E-80
dre-miR-2187*	141.640	42.538	-1,73540464	3,42E-40
dre-miR-218b	369.980	119.625	-1,62892836	6,83E-113
dre-miR-219	950.217	262.281	-1,857143717	0
dre-miR-27b	9.833.899	4.480.642	-1,134058077	0
dre-miR-27d	1.884.439	867.221	-1,119663498	0
dre-miR-29b	6.936.421	3.457.603	-1,004419235	0
dre-miR-338	12.396.633	4.443.744	-1,480100717	0
dre-miR-34	828.513	337.018	-1,297698687	4,91E-187
dre-miR-34b	125.181	24.442	-2,356581361	7,78E-57
dre-miR-455b	48.218	18.096	-1,413900977	7,44E-01
dre-miR-460-5p	68.618	33.843	-1,019729615	7,56E+01
dre-miR-499	1.907.157	584.258	-1,706746125	0
dre-miR-724	28.170.605	8.242.125	-1,773102299	0
dre-miR-725	494.465	224.913	-1,136501354	2,56E-84
dre-miR-736	14.604	0,2585	-15,78583977	4,48E+04
dre-miR-9	302.528.068	148.560.600	-1,026017452	0
dre-miR-93	2.566.212	1.261.113	-1,024942808	0
dre-miR-96	769.632	193.186	-1,994178238	0
dre-miR-146b	1.257.606	3.026.578	1,267007526	0
dre-miR-25	1.912.952	4.012.014	1,068525966	0
dre-miR-722	6.497.592	17.639.604	1,440841112	0

miR-name	n3 (TPM)	PP (TPM)	fold-change (log2 PP/n3)	p-value
dre-miR-1	314.949.288	40.404.979	-2,962514561	0
dre-miR-101a	129.040.246	38.647.035	-1,739391454	0
dre-miR-10a	861.690	406.818	-1,082785428	9,16E-144
dre-miR-10b	1.294.951	567.806	-1,189427509	3,63E-260
dre-miR-10d	74.780	17.861	-2,065839615	7,91E-23
dre-miR-133a*	825.566	106.229	-2,958205888	0
dre-miR-1388	190.745	90.717	-1,072200404	8,96E-23
dre-miR-1388*	121.718	47.474	-1,358333017	2,78E-19
dre-miR-141	1.264.809	584.493	-1,113661887	2,92E-228
dre-miR-142a-3p	8.774.285	3.251.726	-1,432075871	0
dre-miR-148	41.618.374	14.698.574	-1,501544404	0
dre-miR-152	14.611.000	2.943.381	-2,311508716	0
dre-miR-15a*	134.373	32.668	-2,040293244	7,43E-50
dre-miR-182	3.741.277	329.497	-3,50519355	0
dre-miR-183	5.595.580	842.544	-2,731463744	0
dre-miR-192	78.679.109	38.458.784	-1,032667472	0
dre-miR-19b*	101.010	47.944	-1,075075943	8,93E-06
dre-miR-200a	1.061.409	397.183	-1,418104907	5,84E-281
dre-miR-202*	133.913	54.524	-1,296332711	4,88E-20
dre-miR-206	72.478.391	33.359.102	-1,119470475	0
dre-miR-216a	3.330.105	1.594.371	-1,062580293	0
dre-miR-216b	1.347.642	608.230	-1,147748415	1,40E-255
dre-miR-2184	265.755	125.970	-1,077016641	1,62E-35
dre-miR-219	593.404	262.281	-1,177901351	8,21E-111
dre-miR-375	1.072.914	264.632	-2,019475009	0
dre-miR-454b	515.404	246.535	-1,063911257	1,64E-78
dre-miR-34b	63.505	24.442	-1,377507845	1,81E-04
dre-miR-459	519.775	34.548	-3,911213145	0
dre-miR-459*	127.010	0,6816	-17,50758532	2,82E-110
dre-miR-96	2.170.447	193.186	-3,489929742	0
dre-miR-129*	7.786.965	16.003.635	1,039266583	0
dre-miR-139	1.982.923	4.747.154	1,259434196	0
dre-miR-150	2.224.288	4.459.255	1,003459102	0
dre-miR-222	6.587.731	13.343.923	1,018329318	0
dre-miR-489	5.315.329	11.643.323	1,131271964	0
dre-miR-92b	1.992.357	4.900.621	1,298488398	0

miR-name	AA (TPM)	PP (TPM)	fold-change (log2 PP/AA)	p-value
dre-miR-1	295.271.587	40.404.979	-2,869437549	0
dre-miR-101a	120.276.597	38.647.035	-1,637926315	0
dre-miR-107b	2.840.855	857.115	-1,728764507	0
dre-miR-10b	1.671.979	567.806	-1,558086728	0
dre-miR-10d	117.834	17.861	-2,721871115	3,25E-65
dre-miR-122	153.866.005	48.185.996	-1,674988688	0
dre-miR-133a*	1.114.803	106.229	-3,391539216	0
dre-miR-1388	236.118	90.717	-1,38006319	5,17E-51
dre-miR-1388*	222.149	47.474	-2,22631813	2,34E-103
dre-miR-140*	4.882.107	2.232.683	-1,128725485	0
dre-miR-142a-3p	6.662.006	3.251.726	-1,034750957	0
dre-miR-142a-5p	1.260.800	395.067	-1,674170192	0
dre-miR-148	31.835.418	14.698.574	-1,11495651	0
dre-miR-152	17.876.274	2.943.381	-2,602499857	0
dre-miR-15a*	100.035	32.668	-1,614554817	1,56E-21
dre-miR-16b	150.319.499	69.253.826	-1,118066482	0
dre-miR-182	1.988.305	329.497	-2,593201844	0
dre-miR-183	2.855.049	842.544	-1,760691571	0
dre-miR-18c	835.201	195.536	-2,094689194	0
dre-miR-190b	36.499	18.096	-1,012186103	1,78E+07
dre-miR-199*	3.487.927	1.090.019	-1,678016563	0
dre-miR-196a	21.629	0,5875	-15,16801859	1,50E+04
dre-miR-19c	1.466.502	674.035	-1,121483627	4,38E-271
dre-miR-200a	1.021.302	397.183	-1,362533756	1,27E-255
dre-miR-202*	1.246.606	54.524	-4,514970346	0
dre-miR-206	144.591.602	33.359.102	-2,115831404	0
dre-miR-20b	2.802.103	700.592	-1,999863568	0
dre-miR-216a	5.237.861	1.594.371	-1,7159904	0
dre-miR-2184	255.044	125.970	-1,017665967	6,50E-30
dre-miR-2187*	112.652	42.538	-1,405048818	9,80E-19
dre-miR-2188*	40.329	16.216	-1,314399639	1,35E+03
dre-miR-219	759.499	262.281	-1,533934765	1,08E-225
dre-miR-22a	1.813.913.849	786.809.883	-1,205018953	0
dre-miR-34	763.104	337.018	-1,179054042	1,30E-147
dre-miR-34b	71.872	24.442	-1,556067492	1,04E-10
dre-miR-375	1.657.560	264.632	-2,647001662	0
dre-miR-455	71.196	30.317	-1,231669189	1,20E-03
dre-miR-455b	46.863	18.096	-1,372778479	4,69E+00
dre-miR-457b	16.447	0,1175	-17,09480417	3,38E-02
dre-miR-459	505.131	34.548	-3,869983483	0
dre-miR-459*	74.125	0,6816	-16,73067534	4,40E-49
dre-miR-462	4.280.771	1.847.016	-1,212674297	0
dre-miR-724	29.989.054	8.242.125	-1,863347765	0
dre-miR-93	5.237.411	1.261.113	-2,054156271	0
dre-miR-96	550.192	193.186	-1,509944615	3,04E-157
dre-miR-139	1.676.260	4.747.154	1,501816913	0
dre-miR-150	1.547.161	4.459.255	1,527179368	0
dre-miR-222	4.226.247	13.343.923	1,658733877	0
dre-miR-31	79.307	161.693	1,027737108	4,96E-15
dre-miR-458	1.585.913	3.686.277	1,216850854	0

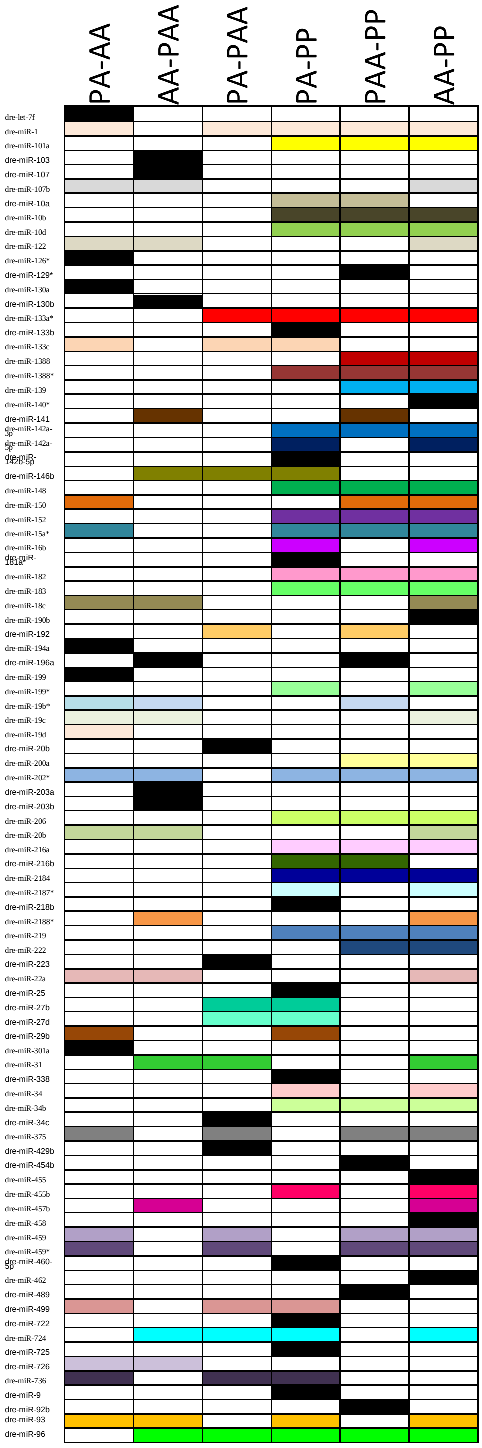


Fig S2 Graphical representation of each Significantly up- and down-regulated miRNA in each pair-wise comparison in high-throughput sequencing. Black squares represent the miRNA that are significantly up-or down-regulated in only one comparison, while colored squares different from black or white squares represent the ones significantly up- or down-regulated in more than one pair- wise comparison. White squares showed the inexistence of significantly up or down expression of the respective miRNA.

Tab S3 Significantly miRNA up- and down-regulated, their pixel intensity and respective fold-change of the following comparisons PA-PAA**;** AA-PAA**;** PA-AA**;** AA-PP**;** PA-PP**;** PAA-PP

PA-PAA			
miR-name	**PA**	**PAA**	**fold-change (log2 PAA/PA)**
dre-miR-138	1438,455571	3679,903729	1,355147
dre-miR-216a	219,6610442	466,5061274	1,086617
dre-miR-461	4558,437178	9235,208453	1,018605
gga-miR-16	302,634403	733,0500565	1,276336
dre-miR-18b	403,0683027	160,2894175	-1,33035
dre-miR-730	273,5457394	124,4309714	-1,13644

AA-PAA			
miR-name	**AA**	**PAA**	**fold-change (log2 PAA/AA)**
dre-miR-1	670,8754596	1364,706792	1,024474
dre-miR-731	256,6414897	2663,880165	3,375703
gga-miR-16	317,6265596	733,0500565	1,20658
hsa-miR-101	301,8310603	686,3624263	1,185229
mmu-miR-182	161,4975685	1275,568843	2,981556
dre-mir-2195	735,8703522	1838,73356	1,321189
dre-miR-133b*	327,0046938	146,151169	-1,16185
dre-miR-150	370,1015537	177,6471028	-1,05891
dre-miR-21	1812,790567	850,2621106	-1,09223
dre-miR-218a	7408,732279	3427,61997	-1,11202
dre-miR-220	12457,27626	1968,809222	-2,66159
dre-miR-92a	2118,457633	916,6390649	-1,20859
hsa-let-7c	9701,119889	317,4261084	-4,93366
mmu-miR-128a	278,9958266	136,6893095	-1,02934
mmu-miR-204	365,9374704	167,3167848	-1,12901
mmu-miR-30b	431,8255768	182,8911024	-1,23946

PA-AA			
miR-name	**PA**	**AA**	**fold-change (log2 AA/PA)**
dre-miR-21	599,8715926	1812,790567	1,595487
dre-miR-218a	2070,481914	7408,732279	1,83926
dre-miR-219	1267,67071	2750,17599	1,117344
dre-miR-220	3022,605213	12457,27626	2,043124
dre-miR-461	4558,437178	12714,96747	1,479917
dre-miR-736	159,3925713	398,6929839	1,322694
hsa-let-7c	349,4342798	9701,119889	4,795058
mmu-miR-30b	152,4010478	431,8255768	1,502576
dre-miR-430j	1606,445471	721,7334086	-1,15433
dre-miR-731	4058,182532	256,6414897	-3,98301
hsa-miR-101	1055,155267	301,8310603	-1,80564
mmu-miR-182	1141,927365	161,4975685	-2,82189

AA-PP			
miR-name	**AA**	**PP**	**fold-change (log2 PP/AA)**
dre-let-7i	159,142843	327,4284296	1,040857
dre-miR-10b	228,9624689	645,8372041	1,496059
dre-miR-10c	252,8313137	622,5497683	1,300014
dre-miR-10d	267,8024464	650,5633634	1,28052
dre-miR-133c	204,3541472	437,4421004	1,098021
dre-miR-152	203,9544825	623,0186754	1,611028
dre-miR-18c	145,5630749	376,3897888	1,370583
dre-miR-19a*	231,2052271	465,2634815	1,008874
dre-miR-19d	198,4410834	507,1893273	1,353814
dre-miR-203b	254,0827841	610,7246723	1,265224
dre-miR-216a	271,1875335	794,7285439	1,353814
dre-miR-216b	149,1759311	367,0698505	1,265224
dre-miR-301c	189,4219526	505,7132706	1,551171
dre-miR-430a	600,0683073	1281,558605	1,29904
dre-miR-729	178,9672727	372,7389443	1,29904
dre-miR-731	256,6414897	522,3280636	-1,59678
dre-miR-736	398,6929839	1001,792206	-1,16399
dre-miR-738	442,4871206	1292,41234	1,416716
mmu-miR-125b	328,8487648	713,9371963	1,094701
dre-mir-2191	162,6651403	326,9928051	1,05847
dre-mir-2192	332,7944554	1078,03281	1,025202
dre-mir-2194	186,9493369	375,9941443	1,329233
dre-mir-2197	373,8842152	855,3062769	1,546359
dre-mir-126b	247,8363054	674,1322243	1,05847
dre-mir-429b	143,5718312	296,0953497	-3,38732
dre-miR-150	370,1015537	175,0629255	-1,08005
dre-miR-153c	3116,071494	1102,89264	-1,49844
dre-miR-21	1812,790567	433,7710233	-2,06321
dre-miR-219	2750,17599	909,248173	-1,59678
dre-miR-301a	638,7544143	285,0613192	-1,16399
dre-miR-455	536,0667789	267,1498353	-1,00476
dre-miR-92a	2118,457633	202,4578879	-3,38732
dre-miR-93	1240,150483	432,6950831	-1,51909
hsa-miR-222	514,174752	247,7269738	-1,05351
hsa-miR-24	458,4164918	228,5695167	-1,00403
mmu-miR-30b	431,8255768	203,5634713	-1,08497

PA-PP			
miR-name	**PA**	**PP**	**fold-change (log2 PP/PA)**
dre-let-7d	136,4516524	406,1454148	1,573606
dre-let-7g	266,5696426	605,3692407	1,183303
dre-let-7i	120,6948859	327,4284296	1,439815
dre-miR-10b	253,6338095	645,8372041	1,348423
dre-miR-10c	239,4676561	622,5497683	1,378358
dre-miR-146a	170,6363373	362,9753868	1,088947
dre-miR-152	189,0797369	623,0186754	1,720281
dre-miR-16b	189,6920456	380,8804699	1,005679
dre-miR-18c	161,4518834	376,3897888	1,221123
dre-miR-202*	124,0962397	285,1451856	1,200237
dre-miR-203b	276,2361606	610,7246723	1,14462
dre-miR-203b*	134,3638403	283,0156844	1,074737
dre-miR-216a	219,6610442	794,7285439	1,855183
dre-miR-216b	117,4001073	367,0698505	1,644621
dre-miR-218a	2070,481914	6336,472263	1,613713
dre-miR-220	3022,605213	9807,7695	1,698133
dre-miR-301c	242,8555231	505,7132706	1,058221
dre-miR-729	176,1723912	372,7389443	1,081178
dre-miR-734	171,729555	350,9909059	1,031295
dre-miR-736	159,3925713	1001,792206	2,651927
dre-miR-738	267,7511134	1292,41234	2,271102
hsa-let-7c	349,4342798	5515,166511	3,980311
hsa-miR-19b	139,6205186	289,9594861	1,05434
dre-mir-2193	160,545137	344,9340487	1,103342
dre-mir-2194	172,737604	375,9941443	1,122128
dre-mir-2198	158,1858937	326,7529619	1,046579
dre-mir-126b	261,3305088	674,1322243	1,367156
dre-mir-429b	121,1906804	296,0953497	1,288783
dre-miR-1	995,5780691	354,1455923	-1,49119
dre-miR-10a*	561,534992	274,2219506	-1,03403
dre-miR-132*	277,4149923	126,7518689	-1,13004
dre-miR-18b	403,0683027	161,7663361	-1,31711
dre-miR-20b	1106,533693	443,284902	-1,31974
dre-miR-430j	1606,445471	448,1451358	-1,84183
dre-miR-731	4058,182532	522,3280636	-2,95781
dre-miR-92a	1268,648168	202,4578879	-2,6476
dre-miR-93	1088,51548	432,6950831	-1,33094
hsa-miR-101	1055,155267	301,7223407	-1,23776
hsa-miR-24	539,0417034	228,5695167	-1,21966
hsa-miR-25	643,8886886	276,4751076	-2,56958
mmu-miR-182	1141,927365	192,3614067	-1,42111
mmu-miR-451	552,2364494	206,219223	-1,23776

PAA-PP			
miR-name	**PAA**	**PP**	**fold-change (log2 PP/PAA)**
dre-let-7d	111,4047089	406,1454148	1,866186
dre-let-7g	302,341468	605,3692407	1,001637
dre-let-7i	139,990832	327,4284296	1,225847
dre-miR-10b	222,8644357	645,8372041	1,535004
dre-miR-10c	218,8896262	622,5497683	1,507986
dre-miR-152	165,4944843	623,0186754	1,912492
dre-miR-18c	139,4113059	376,3897888	1,43288
dre-miR-194a	123,551188	271,9044635	1,137991
dre-miR-19a*	208,6598164	465,2634815	1,156895
dre-miR-19d	208,2293621	507,1893273	1,284351
dre-miR-202*	121,023283	285,1451856	1,236412
dre-miR-203b	230,3059411	610,7246723	1,406971
dre-miR-203b*	120,6223843	283,0156844	1,230384
dre-miR-216b	117,693046	367,0698505	1,641026
dre-miR-220	1968,809222	9807,7695	2,316602
dre-miR-301c	165,289818	505,7132706	1,613322
dre-miR-375	268,8741838	549,6946433	1,031699
dre-miR-729	136,3337675	372,7389443	1,451023
dre-miR-734	154,4614388	350,9909059	1,184187
dre-miR-736	220,0463175	1001,792206	2,186704
dre-miR-738	234,6685546	1292,41234	2,46137
hsa-let-7c	317,4261084	5515,166511	4,118912
hsa-miR-19b	142,2276003	289,9594861	1,02765
mmu-miR-204	167,3167848	365,0934951	1,125684
ZF_miR_10	121,9171369	275,5895935	1,17662
dre-mir-2192	484,6278325	1078,03281	1,153452
dre-mir-2194	133,4670171	375,9941443	1,494227
dre-mir-126b	216,1637042	674,1322243	1,640907
dre-mir-429b	118,2283136	296,0953497	1,324486
dre-miR-1	1364,706792	354,1455923	-1,94618
dre-miR-10a*	592,3386896	274,2219506	-1,11108
dre-miR-138	3679,903729	1601,051501	-1,20065
dre-miR-20b	1212,259286	443,284902	-1,45139
dre-miR-219	1824,404592	909,248173	-1,00468
dre-miR-27c*	316,0475878	145,5427252	-1,1187
dre-miR-430j	981,328233	448,1451358	-1,13077
dre-miR-454b	752,8606933	298,9947071	-1,33226
dre-miR-731	2663,880165	522,3280636	-2,3505
dre-miR-92a	916,6390649	202,4578879	-2,17873
dre-miR-93	1031,21535	432,6950831	-1,25292
hsa-let-7b	763,5127941	330,6040025	-1,20755
hsa-miR-101	686,3624263	301,7223407	-1,18575
hsa-miR-222	501,0934837	247,7269738	-1,01633
hsa-miR-25	572,2903711	276,4751076	-1,0496
mmu-miR-182	1275,568843	192,3614067	-2,72925
mmu-miR-451	457,1860258	206,219223	-1,1486
dre-mir-2195	1838,73356	605,3180209	-1,60295