The Company of Biologists

RESEARCH ARTICLE

Gene copy silencing and DNA methylation in natural and artificially produced allopolyploid fish

Isa M. N. Matos[1,2,*], Maria M. Coelho[1] and Manfred Schartl[2,3,4]

ABSTRACT

Allelic silencing is an important mechanism for coping with gene dosage changes in polyploid organisms that is well known in allopolyploid plants. Only recently, it was shown in the allotriploid fish *Squalius alburnoides* that this process also occurs in vertebrates. However, it is still unknown whether this silencing mechanism is common to other allopolyploid fish, and which mechanisms might be responsible for allelic silencing. We addressed these questions in a comparative study between *Squalius alburnoides* and another allopolyploid complex, the Amazon molly (*Poecilia formosa*). We examined the allelic expression patterns for three target genes in four somatic tissues of natural allo-anorthoploids and laboratory-produced trigenomic hybrids of *S. alburnoides* and *P. formosa*. Also, for both complexes, we evaluated the correlation between total DNA methylation level and the ploidy status and genomic composition of the individuals. We found that allelic silencing also occurs in other allopolyploid organisms besides the single one that was previously known. We found and discuss disparities within and between the two considered complexes concerning the pattern of allele-specific expression and DNA methylation levels. Disparities might be due to intrinsic characteristics of each genome involved in the hybridization process. Our findings also support the idea that long-term evolutionary processes have an effect on the allele expression patterns and possibly also on DNA methylation levels.

KEY WORDS: Allelic silencing, Allopolyploidy, DNA methylation, Freshwater fish, *Poecilia formosa*, *Squalius alburnoides*

INTRODUCTION

In allopolyploid organisms, ancestral homologous alleles that diversified during evolution, designated 'homoeologs', are brought together again in one individual. Consequently, a successful allopolyploidization process requires the reconciliation of two or more sets of diverged genomes in the same nucleus (Feldman et al., 2012). Importantly, the regulatory interactions between genomes must be stabilized as the increased ploidy level and increased heterozygosity lead to gene redundancy, altered gene dosage and altered relationships within and between loci (Feldman et al., 2012; Yoo et al., 2013). These features make allopolyploid

plants and animals exciting objects for understanding the molecular mechanisms of gene regulation in an evolutionary context.

However, studies of the different aspects of allopolyploidy are strongly biased towards plant models (Mable, 2003; Stöck and Lamatsch, 2013). A few years ago, data on the mechanisms underlying gene expression regulation and the dynamics of genome-specific expression in vertebrate allopolyploids were almost absent. Pala et al. (2008) reported for the first time a regulation mechanism of 'functional diploidization' involving gene-copy silencing in an allopolyploid vertebrate, the *S. alburnoides* complex. *Squalius alburnoides* is a hybridogenetic fish that resulted from a cross of a *Squalius pyrenaicus* female (contributing the *p* genome) with an *Anaecypris*-like male (contributing the *a* genome) (Alves et al., 2001) (Fig. S1A). It emerged between 1.4 million years ago (MYA) (Cunha et al., 2004) and less than 0.7 MYA (Sousa-Santos et al., 2007). In present days the complex comprises several ploidy levels and genomic compositions distributed across the Iberian Peninsula (Alves et al., 2001; Collares-Pereira et al., 2013). Taking advantage of the hybrid status of *S. alburnoides*, genome-specific sequence differences were used to determine the contribution of each parental genome to the overall expression of loci individually analyzed in diploid and triploid hybrid individuals (Pala et al., 2008). Results showed that in most triploid *S. alburnoides* of *paa* genome composition, which is the most common form in Iberian southern river basins, for several loci and in different tissues the unpaired minority genome, the *p* haplome, was not contributing to the overall expression, whereas it was contributing to expression in other tissues. Also, the observed allelic expression patterns were different between genes and between different tissues for the same gene. This indicated a most extreme case of homoeolog expression bias (Grover et al., 2012), namely, allele silencing (AS). Therefore, in *S. alburnoides*, the problem of keeping the balance of the expression regulatory networks in an uneven-numbered genomic context might have been solved by AS. These observations were in accordance with gene regulation phenomena already reported in polyploid plants, which showed patterns of differential expression according to organs (Adams et al., 2003) and non-additiveness of expression following gene copy rise (Auger et al., 2005; Wang et al., 2006).

However, it remained unclear whether the silencing mechanism reported for triploid *S. alburnoides*, which is very frequent among both natural and synthesized allopolyploid plants (Adams et al., 2003), was also a common mechanism in allopolyploid vertebrates. A further restriction for generalization is that the allotriploid *S. alburnoides* analyzed so far were all carriers of a duplicated genomic set from one parental species and an unpaired genomic set from another parental species: *paa* and *ppa* in southern populations, and *cca* and *caa* in northern populations, where *S. pyrenaicus* is absent and is replaced by *Squalius carolitertii* (contributing the *c* genome) (Pala et al., 2008, 2010). This situation did not allow the exclusion of monoallelic expression in those cases where the minority genome was not expressed.

[1]Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa 1749-016, Portugal. [2]Department of Physiology Chemistry, Biocenter, University of Würzburg, Würzburg 97078, Germany. [3]Comprehensive Cancer Center Mainfranken, University Clinic Würzburg, Würzburg 97078, Germany. [4]Texas Institute for Advanced Study and Department of Biology, Texas A&M University, College Station, TX 77843, USA.

*Author for correspondence (immatos@fc.ul.pt)

I.M.N.M., 0000-0002-3725-6742

Journal of Experimental Biology

List of abbreviations

5-mC	5-methylcytosine
AS	allelic silencing
HK	housekeeping (gene)
SNP	single nucleotide polymorphism
TF	transcription factor
TFBS	transcription factor binding site
TGH	tri-genomic hybrid

So far, the molecular mechanism responsible for AS in the *S. alburnoides* complex is unknown. A reasonable explanation could be an epigenetic regulation. CpG methylation has long been recognized as a gene expression regulation mechanism by which genes can be silenced by methylation and turned on by demethylation (Martienssen and Colot, 2001). In allopolyploid plants, it is known that among the dramatic genome reconfigurations that can be induced by allopolyploidy, epigenetic changes can play a major role (Wang et al., 2014). However, epigenetic research in (allo)polyploid animals is scarce (Xiao et al., 2013; Covelo-Soto et al., 2015).

To answer these questions and contribute to a better understanding of gene expression regulation in a genomic context of raised ploidy and heterozygosity, we performed a comparative study between *S. alburnoides* and another allopolyploid complex, the Amazon molly (*Poecilia formosa*). *Poecilia formosa* is a unisexual all-female species that originated from a hybridization event between a *Poecilia mexicana limantouri* female (*m* genome) and a *Poecilia latipinna* male (*l* genome) (Lampert and Schartl, 2008) (Fig. S2A), and occurs in the Atlantic drainages, from Rio Tuxpan, Mexico, to South Texas, USA. It reproduces by gynogenesis, thus it depends on sperm from closely related gonochoristic (bisexual) species to trigger embryogenesis of their unreduced diploid eggs (Lampert and Schartl, 2008). Generally, paternal genes do not contribute to the next generation because the paternal pronucleus does not fuse with the unreduced diploid oocyte nucleus. Hence, the vast majority of *P. formosa* are diploid and genetically identical to their mothers. However, in rare cases, the exclusion mechanism fails and paternal introgression occurs (Lampert and Schartl, 2008). In one scenario, small parts of male genetic material are included as microchromosomes (Nanda et al., 2007). In other cases, the sperm nucleus fuses with the oocyte nucleus, resulting in triploid offspring. Such triploids are found in the wild and are true natural allopolyploids having an *mml* genomotype (Fig. S2B). They are fertile and produce all triploid offspring. It has, however, been demonstrated that the formation of such persisting triploid clones is an extremely rare event (Lampert et al., 2005; Schories et al., 2007). These allopolyploidizations were traced back to the evolutionary past of *P. formosa* and have to be considered as ancient events.

The naturally occurring old triploid *P. formosa* (*mml*) are gynogenetically maintained in nature and in the laboratory. On the contrary, triploids that are obtained *de novo* from diploid *P. formosa* as rare introgression cases in laboratory broods (Nanda et al., 1995) do not give rise to stable gynogenetic lines. These *de novo* triploids comprise different genomotypes depending on the parental species used for breeding, including tri-genomic hybrids (TGHs) with *mls* (*P. formosa*, *ml*, with introgressed genome from *P. salvatoris*, *s*) or *mlb* (*P. formosa*, *ml*, with introgressed genome from black molly, *b*) genomic composition (Lamatsch et al., 2010) (Fig. S2C). Such individuals are of great advantage for studying AS in allopolyploids

because they offer the opportunity to distinguish all three alleles and evaluate their expression contribution if diagnostic single nucleotide polymorphisms (SNPs) can be found.

To also obtain TGHs of the *S. alburnoides* complex, advantage was taken from the existence of another *Squalius* species, *Squalius aradensis* (*q* genome), which was reported to naturally hybridize with *S. alburnoides* (Sousa-Santos et al., 2006). Thus, triploid hybrids with the *pqa* genomotype can be produced and studied.

In this work, we examined the allelic expression patterns in several somatic organs of diploid and allotriploid *S. alburnoides* and *P. formosa* with particular analyses of TGHs. As a first step towards a mechanistic explanation, we also evaluated the correlation between levels of DNA methylation and the ploidy status and genomic composition of *S. alburnoides* and *P. formosa*.

We show that AS occurs both in *S. alburnoides* and in *P. formosa*. However, we found disparities within and between the two allopolyploid complexes concerning the pattern of allele-specific expression and DNA methylation levels. Our results indicate that long-term evolutionary processes affect allele expression patterns and DNA methylation levels. This study highlights that the relationships between polyploidy, hybridization, methylation and AS are far from linear, and underscores once more the need for further studies in this field.

MATERIALS AND METHODS
Fish samples

Squalius alburnoides (Steindachner 1866) and *S. pyrenaicus* (Günther 1868) were collected from the Almargem stream [29°S; 622,495.24 m E; 4,113,964.49 m N (UTM)], and *S. aradensis* (Coelho, Bogutskaya, Rodrigues & Collares-Pereira 1998) specimens were collected from Arade River basin [29°S; 545,693 m E; 4,133,136 m N (UTM)]. Fish were captured by electrofishing and brought alive to the animal facility of the Faculdade de Ciências da Universidade de Lisboa. Fish were maintained in high-quality glass tanks (30 litres capacity) equipped with filtration units, at 18°C and under a 14 h:10 h light:dark cycle. A *pa S. alburnoides* female and an *S. aradensis* male (previously genotyped) with evident sexual maturation and ready for breeding were used to perform an experimental cross in order to obtain a progeny specifically with *pqa* genotypes. Eggs and sperm were collected from the selected individuals applying gentle pressure to the abdomen and immediately mixed in a Petri dish with water. For 1 year, the progeny was reared constantly at 20°C. Several individuals were genotyped according to Sousa-Santos et al. (2005) in order to confirm the *pqa* genotype of the batch.

Poecilia mexicana limantouri (Jordan & Snyder 1899), *Poecilia latipinna* (Lesueur 1821), *Poecilia salvatoris* (Regan 1907), black molly and *Poecilia formosa* (Girard 1859) individuals were raised and maintained at standard conditions according to Kallman (1975), under a light cycle of 14 h:10 h light:dark. All fish were derived from laboratory stocks of the aquarium of the Biocenter at the University of Würzburg, Germany, that were originally established from fish collected in the wild, except for black molly, which is an ornamental variety of the *P. mexicana*/*P. sphenops* species complex. The strains used in this work are listed in Table S1.

Ethics statement
Fish were captured, handled and euthanized with the approval of the Portuguese National Forest Authority (AFN; fishing credentials nos 53/2013 and 51/2014) and the Biodiversity and Nature Conservation Institute (ICNB; license nos 235/2013/CAPT and 262/2014/CAPT), the Portuguese national authority and relevant

body concerned with protection of wildlife. The maintenance and use of animals in the animal facility of the Faculty of Science of University of Lisbon (FCUL) had the approval of the Portuguese Directorate-General of Veterinary (DGV), Directorate of Health Services and Animal Protection (DGV-DSSPA) (circular letter no. 99-0420/000/000-9/11/2009).

The selected populations for the fish captures were not imperiled, and sampling was done avoiding depletion of the natural stock. Fish were handled following the recommended ethical guidelines described in the 'Guidelines for the treatment of animals in behavioural research and teaching' (*Animal Behaviour*, 2006, 71, 245–253), and at all times, all efforts were made to minimize fish discomfort. Individuals were submitted to an overdose of the anesthetic MS222 before they were quickly decapitated. Only then were the organs harvested. Fish that were not used were later returned to the collecting site.

All *P. formosa* individuals and fish from parental species used in this study were raised under standard conditions in the aquarium facility of the Biozentrum at the University of Würzburg, where studies were approved by the Institutional Review Board.

Ploidy determination
Fin cells were stained with DAPI as described previously (Lamatsch et al., 2000). At least 10,000 cells were measured per sample. Chicken blood (2.5 pg of DNA per erythrocyte) was used as standard (Vinogradov, 1998).

DNA and RNA extraction
Total genomic DNA was obtained from dissected livers and muscle with a standard phenol/chloroform/isoamyl alcohol (25/24/1) protocol (Blin and Stafford, 1976). DNA was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA).

RNA was extracted from dissected livers, eyes, muscle and gills preserved in RNAlater (Ambion, Foster City, CA, USA) at −20°C. Total RNA was extracted using the Tri-Reagent (Ambion) following the supplier's instructions. Contaminant DNA was eliminated by the addition of TURBO DNase (Ambion) followed by purification with phenol/chloroform. Ethanol and glycogen were used to precipitate the RNA. RNA amount and quality evaluation was performed with Nanodrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA) and a 2100 Bioanalyser (Agilent Technologies, Santa Clara, CA, USA).

Sequence analysis and genome-specific expression
From the extracted RNA, first-strand cDNA was synthesized with the RevertAid First Strand cDNA Synthesis Kit (Fermentas, Thermo Fisher Scientific) with oligo dT primers. Primer sequences and amplification conditions for *actb*, *rpl8* and *gapdh* with *Squalius* and *Poecilia* samples are given in Table S2.

In *S. alburnoides*, SNPs between the P and A genomes for the three genes have already been reported (Pala et al., 2008; Matos et al., 2011). For the *S. aradensis* derived Q genome of the *S. alburnoides* complex and for all genomes present in allotriploid *P. formosa*, SNPs were identified in the present study.

PCR products were sequenced and sequences were aligned and compared with Sequencher ver.4 (Gene Codes Corporation, Ann Arbor, MI, USA). Within each of the fish complexes, polymorphic sites between the intervenient genomes were identified.

cDNA samples from adult liver, eye, gill and muscle of *S. alburnoides* and *P. formosa* diploid and triploid natural hybrids and TGHs were used as templates for independent amplifications

and direct sequencing of gene products of the three target genes (*actb*, *rpl8* and *gapdh*). Through sequence comparison, on the basis of the identified polymorphic sites between the involved genomes *p*, *a* and *q*, or *m*, *l*, *s* and *b*, the contribution of each genome-specific allele to the overall expression at each of the three target loci was determined.

Global DNA methylation quantification
The percentage of methylated DNA for the genomotypes of each one of the allopolyploid complexes was determined by colorimetric quantification of 5-methylcytosine (5-mC). Three to five specimens were sampled and analyzed independently for each genomotype. One hundred nanograms of DNA of each individual were loaded into each well of the MethylFlash Methylated DNA Quantification Kit (Epigentek, Farmingdale, NY, USA). The protocol and calculations were performed according to the manufacturer's instructions.

In addition, the observed mean methylation level for each genomotype in the hybrids (diploids and triploids) was compared with an expected methylation level, which was calculated by considering that each of the *p*, *a* and *q* genomes in the hybrids would be methylated at the same level as in the non-hybrid situation. The mean methylation level obtained for each parental diploid genomotype (*pp*, *aa* and *qq*) was used to calculate the expected methylation level for each hybrid genomotype [(*pp*/2)+(*aa*/2)+(*qq*/2)=additive expectation]. Expected additive values for *P. formosa* were calculated accordingly.

The mean observed methylation value (obs) for each hybrid genomotype was divided by its corresponding expected additive value (exp) (Table S3).

Comparative sequence analysis for promoter and CpG island predictions
Sequences for *P. formosa*, *P. mexicana* and *P. latipinna rpl8* (ID: 103134768, 106918910 and 106964237, respectively), *gapdh* (ID: 103136734, 106921370 and 106955760, respectively) and *actb* (ID: 103153440, 106927995 and 106956540, respectively) were obtained from GenBank. Ensemble84 Amazon molly gene annotations were used to identify exons, introns and untranslated regions. Putative promoter regions were defined as 2000 bp 5′ of the first nucleotide of the first exon (adapted from Farré et al., 2007).

For each gene, sequences were aligned and compared using Bioedit (Hall, 1999) with ClustalW multiple sequence alignment. The putative promoter regions served as templates for the design of degenerated primer pairs that were used to amplify the homoeologous DNA regions from *P. salvatoris* and black molly liver DNA samples. Primer sequences and amplification conditions are given in Table S2.

PCR products were sequenced and all sequences for each gene were aligned as previously.

Several tools were employed to analyze the nucleotide sequence of the putative promoter regions of *rpl8*, *gapdh* and *actb* between *mm*, *ll*, *bb* and *ss* genomes. Identity matrices were obtained with BioEdit. Promoter 2.0 Prediction Server (Knudsen, 1999) and the Gene Promoter Miner (Lee et al., 2012) were used to predict RNA polymerase II (Pol.II) promoters in *Poecilia* DNA sequences. With the Sequence Manipulation Suite – CpG Islands Sequence Analysis option (Stothard, 2000), the occurrence of CpG islands was predicted. Also, with DBCAT (Kuo et al., 2011) the occurrence of CpG islands was investigated as well as the number of CpG per 1 kb within the *mm*, *ll*, *bb* and *ss* sequences.

RESULTS

Analysis of allele-specific gene expression in triploid *S. alburnoides*

In *S. alburnoides* individuals we analyzed the qualitative pattern of allele-specific contribution for three genes, *actb*, *rpl8* and *gapdh*, in liver, muscle, eye and gill of naturally occurring allotriploids (*paa* genomotype) and laboratory-produced TGHs (*pqa* genomotype).

Several informative SNPs between *p* and *a* alleles for *actb*, *rpl8* and *gapdh* were previously reported (Pala et al., 2008; Matos et al., 2011) and used for this study. When *q* sequences were inspected and compared with *p* and *a* sequences, diagnostic SNPs between them were also identified.

The sequencing of reverse-transcribed PCR products of these three genes from all four organs once again confirmed that in *paa* individuals, AS of *p* is occurring (Table 1). Consistent with previous reports, monoallelic expression of the single *p* allele was not detected.

In contrast, in the TGH hybrids containing the *q* genome, sequencing of the reverse-transcribed PCR products of all three genes revealed no indication of silencing in any of the four analyzed tissues. The observed qualitative pattern of allele usage in the TGH individuals was consistently tri-allelic (Table 1).

Allele-specific expression in triploid *P. formosa*

For naturally occurring *P. formosa* allotriploids (*mml*) and laboratory-produced TGHs (*mlb* and *mls*), the qualitative pattern of allelic-specific contribution in *actb*, *rpl8* and *gapdh* in the liver, muscle, eye and gill was inspected (Table 2). Contrary to what was observed in *S. alburnoides*, in natural triploid *P. formosa* (*mml*) no evidence for AS was obtained.

We then looked at the laboratory-generated TGHs, either with *mlb* or *mls* genomic composition (Table 2). For both types of TGH we clearly detected allele-specific silencing. Moreover, in *mls* TGHs for *gapdh* and *rpl8*, even monoallelic gene expression (silencing of two alleles) was detected.

Global DNA methylation in *S. alburnoides* of different ploidy levels and genomic composition

Allele-specific silencing can be due to an epigenetic mechanism. Therefore, we determined the total amount of 5-mC in total DNA extracts from livers and muscle of natural allodiploid (*pa*), allotriploid (*paa*) and laboratory-produced TGH (*pqa*) *S. alburnoides*, as well as from the parental non-hybrids – *aa*, *pp* and *qq* (Fig. 1A,B). In both liver and muscle samples, there was a significantly higher amount of 5-mC in the *aa* diploids than in all other diploids. We found also that both natural triploids (*paa*) and the TGH triploids (*pqa*) have a similarly high level of 5-mC as the *aa* diploids, and again significantly higher (*t*-test for independent samples, $P>0.05$) than the *pp*, *qq* and *pa* diploids.

Global DNA methylation in *P. formosa* of different ploidy levels and genomic composition

For *P. formosa* we determined the global 5-mC levels in natural allodiploids (*ml*), allotriploids (*mml*), TGHs (*mls* and *mlb*) and in all the parental diploids (*mm*, *ll*, *bb* and *ss*) (Fig. 1C,D). For all *Poecilia* genotypes, the pattern of 5-mC was consistent between the two analyzed tissues. In both liver and muscle, higher levels of 5-mC were found in the natural diploid and triploid hybrids, while all diploid parental genomotypes (*mm*, *ll*, *bb* and *ss*) and the laboratory-produced TGH (*mlb* and *mls*) displayed a similar low methylation level.

Additivity of global DNA methylation in *S. alburnoides* and *P. formosa* allopolyploid complexes

For each hybrid genomotype we performed a simple relative comparison (ratio) between the mean observed methylation value and an expected methylation level in case of additivity (obs/exp) for a hybrid situation (Table S3). Results show that the genomotypes of both allopolyploid complexes can be separated into two distinct groups. One group is composed of *pa*, *paa*, *pqa*, *mlb* and *mls* genomotypes, with obs<exp, and a second group is composed of *ml* and *mml* genomotypes, with obs>exp (Fig. 2).

Promoter and CpG island prediction of *Poecilia* target genes

We used available genomic sequences of *P. mexicana*, *P. latipinna* and *P. formosa* as templates to isolate and characterize the homoeologous sequences in *P. salvatoris* and black molly. The selected target zones were the 2000 bp 5′ of the first nucleotide of the first exon of *rpl8*, *gapdh* and *actb*. We could amplify between 1100 and 1429 bp for *P. salvatoris* and black molly within these template regions. For each gene, we found, as expected for comparisons between species and/or strains, a high percentage (98–99% for *actb*, 93–99% for *gapdh* and 97–99% in *rpl8*) of positive similarity between *mm*, *ll*, *bb* and *ss* sequences (Table S4).

Within the selected sections for *mm*, *ll*, *bb* and *ss* we could predict for *gapdh* and for *actb* a highly likely promoter region – *gapdh*, from −592 to −294 bp of the first nucleotide of the first exon; and *actb*, from −611 to −296 bp of the first nucleotide of the first exon. Concerning CpG islands, for none of the individual genomes at any of the three genes were any CpG islands predicted with the DBCAT within the defined target zone, but with the Sequence Manipulation Suite a CpG island was found within the defined target zone for *rpl8* (from −498 to −283 bp of the first nucleotide of the first exon) and *actb* (from −1411 to −1204 bp of the first nucleotide of the first exon). Also, we quantified the number of CpG sites per 1 kb within the *mm*, *ll*, *bb* and *ss* sequences (Table S5), but no substantial differences were found between the genomotypes for each gene.

Table 1. Allelic expression pattern of *actb*, *rpl8* and *gapdh* in liver, eye, gill and muscle of *Squalius alburnoides*

Ploidy level	*n*	Genotype	Liver			Eye			Gill			Muscle		
			actb	*rpl8*	*gapdh*	*actb*	*rpl8*	*gapdh*	*actb*	*rpl8*	*gapdh*	*actb*	*rpl8*	*gapdh*
2n	2	*pa*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*
3n	5	*paa*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*
3n	1	*paa*	**a**	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*
3n	1	*paa*	**a**	*p+a*	*p+a*	*p+a*	*a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*	*p+a*
3n	6	*pqa*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*	*p+q+a*

Allelic silencing is highlighted in bold.

Table 2. Allelic expression pattern of *actb*, *rpl8* and *gapdh* in liver, eye, gill and muscle of *Poecilia formosa*

Ploidy level	n	Genotype	Liver			Eye			Gill			Muscle		
			actb	*rpl8*	*gapdh*	*actb*	*rpl8*	*gapdh*	*actb*	*rpl8*	*gapdh*	*actb*	*rpl8*	*gapdh*
2n	2	*ml*	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l
3n	9	*mml*	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l	m+l
3n	1	*mls*	m+l+(s)	_	**s**	_	_	**s**	m+l+(s)	_	_	m+l+(s)	_	_
3n	1	*mls*	**l+s**	m+l+(s)	**s**	m+l+(s)	m+l+(s)	_	**m+s**	_	**s**	**l+s**	m+l+(s)	**s**
3n	1	*mls*	m/s+l	m/s+l	**m+l**	m/s+l	m/s+l	m/s+l	m/s+l	m/s+l	m/s	m/s+l	m/s+l	m/s+l
3n	1	*mls*	m+l+(s)	m/s+l	l+s+(m)	m+l+(s)	m/s+l	l+s	m+l+(s)	m/s+l	m+l+s	m+l+(s)	m/s+l	m+l+s
3n	1	*mls*	m+l+(s)	**m/s**	m+s	m+l+(s)	m/s	m+s	m+l+(s)	m/s	m+s	m+l+(s)	m/s	m+l+s
3n	1	*mls*	**m+s**	m/s	m+s	m+s	m/s	m+s	m+s	m/s	m+s	m+s	**m**	m+s
3n	1	*mls*	m+l+(s)	**m/s**	m+s	m+l+(s)	_	m+s	m+l+(s)	m/s	m+s	m+l+(s)	**m**	m+s
3n	1	*mls*	m+l+(s)	m/s+l	m+l+s	m+l+(s)	m/s+l	m/s+l	m+l+(s)	m/s+l	m+l+s	m+l/s	m/s+l	m+l+s
3n	1	*mlb*	_	_	_	_	_	_	m/b+l	**m/b**	_	_	_	_
3n	1	*mlb*	m/b+l	**m/b**	**m/b**	m/b+l	m/b+l	**m/b**	m/b+l	m/b+l	m/b+l	m/b+l	**m/b**	m/b+l
3n	1	*mlb*	m/b+l	m/b+l	m/b+l	m/b+l	m/b+l	m/b+l	m/b+l	m/b+l	m/b+l	m/b+l	**m/b**	**m/b**
3n	1	*mlb*	m/b+l	_	_	m/b+l	m/b+l	_	m/b+l	m/b+l	_	m/b+l	_	_
3n	1	*mlb*	m/b+l	m/b+l	m+l+(b)	m/b+l	m/b+l	m/b+l	m/b+l	m/b+l	m+l+(b)	m/b+l	m/b+l	m+l+(b)
3n	1	*mlb*	m/b+l	m/b+l	m+l+(b)	m/b+l	m/b+l	m/b+l	m/b+l	m/b+l	m+l+(b)	m/b+l	m/b+l	m+l+(b)

Allelic silencing is highlighted in bold. /, to be read as 'either or'; (), presence or absence of the allele not unequivocally determined; _, allele expression pattern not assessed.

DISCUSSION

In this work we intended to answer three fundamental questions concerning the mechanism underlying gene expression regulation and the dynamics of genome-specific expression in vertebrate allopolyploids. First, we wanted to explore whether the silencing mechanism reported for natural triploid *S. alburnoides* was common to another allopolyploid vertebrate. Second, we wanted to investigate whether, in an allotriploid condition with increased heterozygosity, one of the three alleles is consistently silent, converting triploids into functional diploids. Third, it was our goal

to begin to identify possible mechanisms responsible for allele silencing. Specifically, we wanted to evaluate CpG methylation as a candidate mechanism, but other possibilities have been considered.

Allele-specific silencing in *P. formosa*

In TGH *P. formosa* triploids of *mlb* and *mls* genomic composition, AS was obvious and quite frequent. This shows for the first time that AS is indeed not a unique phenomenon in the *S. alburnoides* complex, but is more widespread. This is in line with earlier findings that the variation in pigmentation phenotypes between TGH of
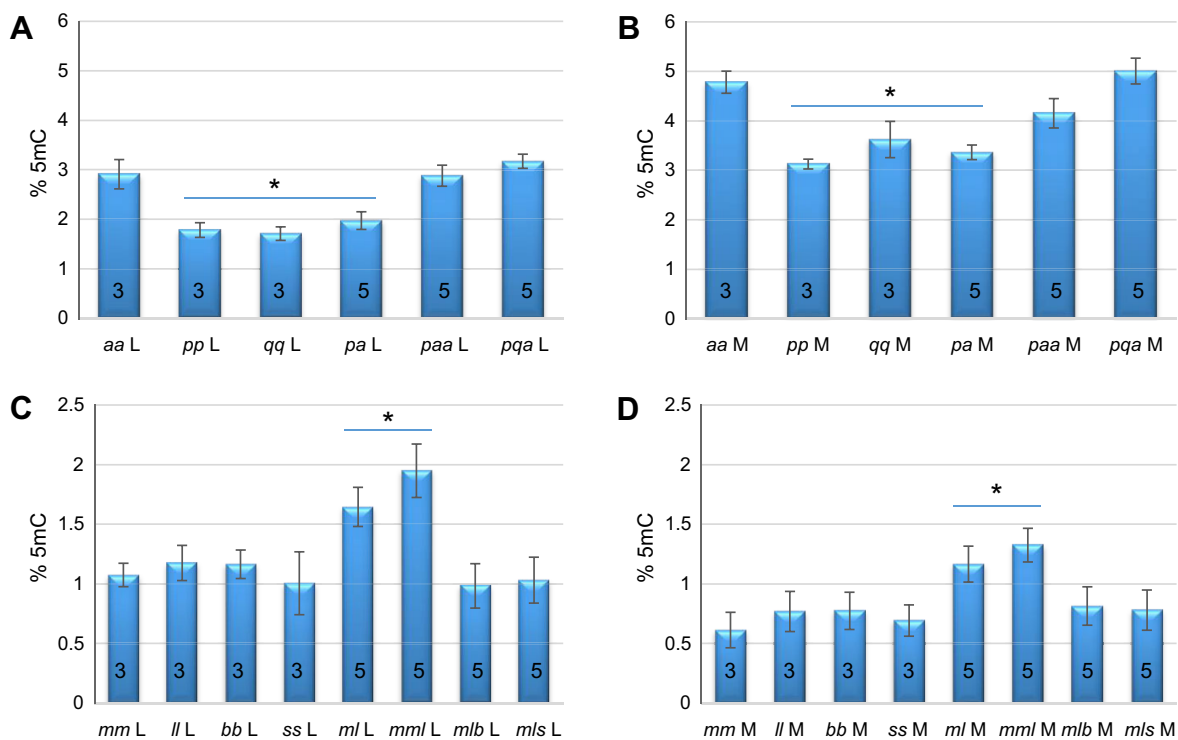

Fig. 1. Levels of global DNA methylation within the *Squalius alburnoides* and *Poecilia formosa* allopolyploid complexes. Global DNA methylation in (A) liver and (B) muscle tissue of several *S. alburnoides* complex intervenient genomotypes. Global DNA methylation in (C) liver and (D) muscle tissue of the *P. formosa* complex intervenient genomotypes. 5-mC, 5-methylcytosine; L, liver; M, muscle. Data are means±s.e.m. *Significant difference (*P*<0.05) between the underneath genomotypes and all other groups.
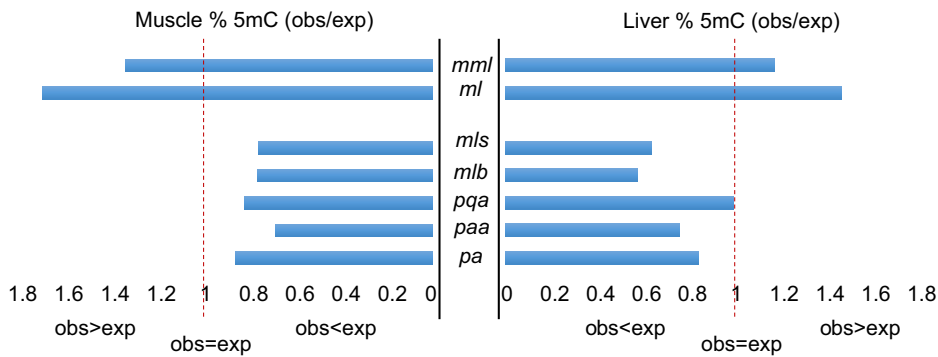
Fig. 2. Additivity of global DNA methylation in *S. alburnoides* and *P. formosa* allopolyploid complexes. Ratio between the mean observed methylation value and an expected methylation level in the case of additivity (obs/exp), in muscle and liver tissues. *pa*, *paa*, *pqa*, *mlb* and *mls* genomotypes present observed 5-mC levels<expected 5-mC levels, whereas *ml* and *mml* genomotypes have observed 5-mC levels>expected 5-mC levels. Dashed red line indicates the position at which observed 5-mC level=expected 5-mC level.

P. formosa individuals may be the consequence of differential contribution of genomes to overall expression (Lamatsch et al., 2010, 2011).

The failure then to detect AS also in the naturally occurring triploid of the *mml* genomic constitution was somehow unexpected as the naturally occurring triploid *P. formosa* were proposed earlier as good candidates where a comparable gene-copy silencing phenomenon like in *S. alburnoides* could occur (Pala et al., 2008). Comparison of expression levels at several allozyme loci between diploid and triploid *P. formosa* revealed them to be indistinguishable quantitatively (Turner et al., 1983), which could be a consequence of AS.

Our failure to detect AS in the naturally occurring *P. formosa* could be due to the following reasons. (1) AS is not random and it is always one of the 'm' alleles that is silenced. This phenomenon would escape our observation because our sequencing chromatograms did not allow for quantitation of peak heights at SNP positions. (2) AS does not occur on a full genomic scale and the three selected genes are not subjected to this phenomenon. However, if there were genome-wide occurrence of AS in triploid *P. formosa*, our study would most likely have been sufficient to detect it. Considering a parsimonious null hypothesis of random inactivation of one of the genomes (neither haplome nor tissue dependent), for each gene and per tissue, 2.7 instances of AS occurrences would be expected ($n=9$). We analyzed the allele expression pattern in four tissues, so in total per gene, approximately 11 (2.7+2.7+2.7+2.7) 'l' allele silencing occurrences should be seen in our evaluation if this phenomenon exists. If AS is not random and affects only a subset of genes or cell types, more genes and other organs need to be investigated in the future, preferably using transcriptome-wide approaches as recently described by Garcia et al. (2014). (3) AS does not occur at all in the *mml* genomotypes. Although this is a valid assumption in this context, as we did not find AS in naturally occurring allotriploid *P. formosa*, we cannot promptly discard that it does not occur at all. In fact, the occurrence of variegated skin phenotypes presented by some individuals is a strong contra-indicator of this third hypothesis.

The difference between the natural occurring *mml* and the TGH *P. formosa* triploids may be explained by different magnitudes of 'genomic shock'. 'Genomic shock' refers to a series of genomic perturbations at both genetic and epigenetic levels, and has been described in many plant allopolyploid systems (Wang et al., 2014). Some of its most frequent consequences are deviations from expected expression levels and allele specific expression patterns. Also, in plants it has been found that hybridization usually has a greater impact on gene silencing than does genome doubling (Chelaifa et al., 2010; Buggs et al., 2014). Despite both *P. formosa*

types having the same ploidy level, the increased diversity of genomes in the TGHs may lead to a higher level of 'genomic shock'. Compared with natural allotriploids, where only two distinct genomes have to be managed, the interactions and simultaneous regulation of three different genomic sets may pose additional challenges with different outcomes. In addition, it has to be considered that some intergenomic combinations are not well tolerated and can lead to hybrid incompatibilities and dysgenesis (Bomblies and Weigel, 2007; Ishikawa and Kinoshita, 2009; Walia et al., 2009; Malone and Hannon, 2009). So, immediate allele-specific expression adjustments in the TGH *P. formosa* may be a necessity to allow for the viability of these organisms.

Absence of AS in TGHs of *Squalius*

Contrary to what was observed in the naturally occurring allotriploid *S. alburnoides*, AS was not observed in any of the analyzed tissues in TGH individuals. It has been previously shown (Pala et al., 2010) that the patterns of gene expression in triploid *S. alburnoides* depend on the genomic contexts brought about by different parental contributions. For instance, the presence of *c* or *p* genomes in allopolyploid *S. alburnoides* biotypes results in substantial difference in genome-specific allele usage in either *paa* or *caa* genomic contexts (Pala et al., 2010). Because the effect of the *q* genome to the overall gene expression in natural occurring *S. alburnoides* of *qaa* and *qqa* genomotypes has never been assessed, the absence of AS in the TGH fish with one *q* haplome is difficult to assess, and the effects of the presence of the *q* genome are difficult to infer. However, we can at least say that the absence of AS in TGH *S. alburnoides* supports the previous conclusion that different genome combinations lead to different mechanisms of how to cope with genomic shock. In contrast, the absence of AS in TGH *Squalius* is not readily explained by the simple reasoning presented for AS occurrence in the TGH *P. formosa*, where we relate the higher genomic shock with the need for AS. This demonstrates the complexity of the phenomenon where two different deviations from normal come together, namely ploidy change and hybridization.

Despite our inability to show AS in the TGH *S. alburnoides*, its occurrence cannot be totally discarded, based on the same considerations presented for the naturally occurring *P. formosa*. So, to fully enlighten this matter, applying a transcriptome-wide approach to *S. alburnoides* would also be desirable.

However, despite new and promising tools that are constantly emerging (Shen et al., 2012a,b, 2013), assessing allele-specific gene expression on a large scale is still a technically challenging problem (Garcia et al., 2014), even more so in species with scarce genomic resources, and as in this case, higher levels of ploidy than diploidy.

Differences in global DNA methylation between genomotypes

DNA methylation modifications associated with ploidy changes have been studied extensively in plants (Diez et al., 2014). It has been shown that normal function and structure of newly formed polyploid genomes are intimately related with this epigenetic process (Matzke et al., 1999; Salmon et al., 2005; Chen and Ni, 2006; Wang et al., 2014). Also, it is known that methylation impacts directly on gene transcription (Wang et al., 2014; Sehrish et al., 2014). In general, it is assumed that methylated DNA sequences are transcriptionally inactive (Wang et al., 2014). So, one goal of this study was to relate AS occurrence in these fish to the degree of total DNA methylation.

We determined the total amount of DNA methylation in two tissue types (liver and muscle) for all the available genomotypes involved in both allopolyploid complexes. If the AS phenomenon was 5-mC mediated, our hypothesis was that the total methylation level would be higher in those triploid individuals where AS occurs. However, the pattern of global methylation in both the *S. alburnoides* and *P. formosa* allopolyploid complexes does not fit this initial expectation, nor does it help to clarify the different AS patterns between *S. alburnoides* and *P. formosa*. For instance, AS occurs in *P. formosa* TGH, where we identified low levels of methylation compared with naturally occurring diploids and triploids in which AS was not detected. Also, TGH *S. alburnoides*, where no AS was detected, presented similar high levels of methylation as the naturally occurring triploid *S. alburnoides* (*paa* genomotype), where AS has been encountered. So, global methylation levels do not seem to reflect the AS status. This is in line with findings in *Arabidopsis*, where for most of a pool of 77 analyzed genes, expression did not directly correlate with the methylation level (Shen et al., 2012a). In contrast, in *Tragopogon* it was shown that by DNA methylation one homeolog can be completely silenced (Sehrish et al., 2014).

We further observed that the levels of DNA methylation were non-linearly related to the ploidy level in each tested allopolyploid series. Higher ploidy level did not consistently correspond to higher or lower levels of DNA methylation in either of these allopolyploid complexes. Additionally, our results do not show a linear correspondence between higher levels of heterozygosity and higher or lower levels of DNA methylation.

Similar results have been found in an analysis of genomic DNA methylation in several annual herbaceous and woody perennial plants of several ploidy levels (Li et al., 2011). In addition, in a study that investigated DNA methylation changes associated with ploidy in *Salmo trutta*, no evidence of genome-wide methylation differences between diploid and triploid specimens was found (Covelo-Soto et al., 2015). However, in *Cyprinus carpio*×*Carassius auratus* hybrids it was found that hypermethylation was more prominent in the allotetraploids than in the diploid parental individuals (Xiao et al., 2013).

We have determined global methylation levels, but with this broad approach, underlying mechanisms of methylation as effectors at the single-locus scale are diluted. In this sense, investigating differences in 5-mC of promoters of genes presenting AS would be interesting. Methylation of promoters is canonically associated with stable, long-term transcriptional silencing, and one of the reasons is that a transcription factor (TF) is physically prevented from binding to its specific transcription factor binding site (TFBS) if the TFBS is methylated (Zhu et al., 2003; Defossez and Stancheva, 2011). A differential methylation status of CpG sites in the promoter and/or at its surroundings between the different alleles of a gene may lead to differential allelic expression (Kerkel et al., 2008; Sehrish et al., 2014). However, the three target genes focused on in the present study (*rpl8*, *gapdh* and *actb*) are housekeeping (HK) genes. HK genes are expressed in virtually all tissues and across developmental stages and are, in general, exempted from complex transcriptional programs as, for example, the transcriptional programs governing genes involved in responses to external stimuli or in cell differentiation (Farré et al., 2007). In principle, HK genes are activated by default; therefore, the CpG sites around or on the proximal promoter should be unmethylated. Also, contrary to what has been widely reported in other vertebrate organisms, it was found that in zebrafish, methylation and expression were most strongly correlated with regions 10,000 bp upstream and downstream from genes (McGaughey et al., 2014) and not at the proximal promoter sites. So, in the present case, for the specific gene targets on hand, a locus-specific approach did not offer much promise and it was not pursued.

Mechanisms other than DNA methylation may intervene or be responsible for allele expression bias

In any case, mechanisms other than DNA methylation may intervene or be responsible for allele expression bias and AS. For example, an miRNA-linked mechanism has been already identified as a good candidate in the *S. alburnoides* complex (Inácio et al., 2012) and should be similarly investigated for the *P. formosa* complex.

From another angle, in the analysis of the putative promoter regions of *rpl8*, *gapdh* and *actb* of *Poecilia* parental genomotypes, we found a high percentage of positive identity between the sequences. This is an expected result for comparisons within species and/or strains. However, as there is no perfect homology (less than 100% identity), it is conceivable that in the cells of the TGH individuals three different sequences are working simultaneously as promoter of each gene. Conversely, each of these different sequences can work more or less effectively as the docking site for polymerases and transcription factors originated from homoeolog genes. So, another mechanism that may intervene or be responsible for allele expression bias and AS is the strength of the promoter. A promoter can be classified from strong to weak according to its affinity for RNA polymerase and TFs (Li and Zhang, 2014). Thus, the strength of the promoter depends from how closely the promoter sequence resembles the ideal consensus sequences for the docking of polymerase and TFs (Li and Zhang, 2014). For example, in *Escherichia coli* it was observed that several non-consensus bases could have a positive effect on the promoter strength while certain consensus bases have a minimal effect (Kiryu et al., 2005). Also, it was demonstrated in yeasts that variations in the binding sites of TFs between three different strains were responsible for up to 50% of the observed differences in expression (Tirosh et al., 2008). Additionally, a more recent study showed that nucleotides in different regions of promoter sequence have different effects on promoter strength (Li and Zhang, 2014). So, we hypothesize that the conspicuous AS that we encountered in the *P. formosa* TGH may be due to different promoter strengths resulting from the different nucleotidic sequences detected. To support this assumption, a similar analysis for the *S. alburnoides* complex should be performed, and results should show higher levels of identity between the promoter sequences of the parental genomotypes. However, while large-scale annotated genomic data are available for the *P. formosa* complex, no reference genome has yet been produced for the *S. alburnoides* complex, so we could not perform the same analysis.

'Old' versus 'de novo' allopolyploids and the effects of long-term evolutionary processes

The analyzed laboratory-bred triploid *P. formosa* individuals with the *mml* genomotype were derived by gynogenesis from natural triploids. In these individuals, the original hybridization (*m×l*) and polyploidization (*ml+m*) events occurred a long time ago, and are merely clonally propagated at each generation (Lampert and Schartl, 2008). Therefore, we consider them as naturally occurring 'old triploids'. We also analyzed TGH *P. formosa* triploids of *mlb* and *mls* genomic composition that were experimentally produced through specific crosses between *Poecilia* strains and species (Lampert et al., 2007; Lamatsch et al., 2010). We can consider these individuals as '*de novo*' allotriploids, as increases in both ploidy and hybridity happen at the moment of production of each of these TGH individuals.

Inversely to what was observed in the 'old' *P. formosa* triploids, in the '*de novo*' triploids AS was quite frequent and evident. We hypothesize that AS may be an immediate mechanism to cope with the genomic shock. In fact, whenever AS has been detected in vertebrates, it was in individuals that could be considered '*de novo*' triploids. In *S. alburnoides* the reproductive complex is maintained through an intricate network of genetic exchanges and continuous *de novo* hybridizations. Hence, allopolyploidy is established '*de novo*' at the moment of each individual conception. Another example is the laboratory-produced TGH allotriploid medakas (*Oryzias latipes*), where it was found that allele suppression, despite not being abundant, consistently occurred (Garcia et al., 2014).

These examples support the hypothesis that AS may be an immediate mechanism to cope with genomic shock. Consecutively, refined mechanisms operate leading to a stable regulation of the three haplomes. However, we have not found AS in TGH *S. alburnoides*, which are also '*de novo*' allotriploids. This may indicate that AS is not a ubiquitous mechanism to cope with an abrupt increase of ploidy and heterozygosity in fish.

Several studies on allopolyploid plants have also revealed differences between 'old' and 'young' polyploids. The degree of non-additive expression was lower in recent allopolyploids compared with 'older' allopolyploid cotton and coffee genotypes (Flagel and Wendel, 2010). These results suggested that non-additive expression, that is due or related to AS, may increase over time, via selection and modulation of regulatory networks. In another study, results showed that in F1 hybrids and early allopolyploid *Tragopogon miscellus* plants there was activation of allele/homeolog expression in all tissues, eliminating the tissue-specific expression patterns observed in the parental diploids (Buggs et al., 2011). Tissue-specific expression patterns were then reestablished as generations succeeded (Buggs et al., 2011).

In this context, the differences in DNA methylation levels that we observed can also be interpreted. Comparing allotriploids of different evolutionary age, we observed a tendency towards higher DNA methylation levels than expected from additivity in the 'old' hybrids, whereas the opposite tendency was observed in the genomotypes of '*de novo*' hybrids.

In the *S. alburnoides* complex, we found other evidence that long-term evolutionary processes may influence DNA methylation levels. We observed that the percentage of methylated DNA is much higher in the *aa* genomotype than for the other two parental genomotypes (*pp* and *qq*). This may indicate that in individuals of the *aa* genomotype, more genes or alleles are downregulated or inactivated. These increased DNA methylation levels may be related to the fact that both *pp* and *qq* genomotypes exist as independent

species (*S. pyrenaicus* and *S aradensis*, respectively), having their own separate evolutionary paths, while an independent species with the *aa* genomotype does not exist. Individuals with the *aa* genomotype, called 'diploid nuclear non-hybrid males of the *S. alburnoides* complex' (Alves et al., 2001), perpetuate only inside the complex by mating with triploid hybrid females (*paa* or *qaa*) (Fig. S1). In each *aa* individual that arises, the nuclear hybrid status is lost and epigenetic changes are likely to occur.

In summary, our results imply that DNA methylation may play some role in the evolution of these vertebrate allopolyploids, probably somehow providing genome stability and reducing the degree of incompatibility that arises from multiple incongruous genomes within the same nucleus. Nevertheless, as in plants, the mechanisms by which all this happens at the whole genomic level (and also at specific sites) seem to be diverse and are still obscure.

Conclusions

With this work, we showed that in vertebrates, AS also occurs in allopolyploid situations besides the previously studied naturally occurring triploid *S. alburnoides*. In *P. formosa* AS was observed quite frequently in two distinct TGH genomic configurations.

We assume that AS is the result of genomic stress, induced by the presence of distinct genomes in the same nucleus. Of note, we found several disparities within and between the two complexes concerning the pattern of allele-specific expression and DNA methylation levels. These differences might be due to the intrinsic characteristics of each genome involved in the hybridization process. Expression silencing or downregulation can result from the interaction between divergent regulatory hierarchies (Riddle and Birchler, 2003) and differential capacity of interaction between proteins or complexes (Comai, 2000; Adams and Wendel, 2004). However, our results also point out that AS is not a ubiquitous mechanism to handle an abrupt increase in ploidy and heterozygosity in fish.

In addition, our findings support the notion that long-term evolutionary processes have an effect on the allele expression patterns and possibly also on DNA methylation levels. Our study highlights the complexity of allopolyploidy at the gene expression regulation level, and that attempts to find a common global mechanism or explanation that fits all allotriploid conditions might fail, as it might not exist.

to KX870956; KX870957 to KX870960; KX870961 to KX870968; KX871034 to KX871041; KX871114 to KX871121; KX870969 to KX870978; KX870979 to KX870988; KX870989 to KX870999; KX871000 to KX871009; KX871010 to KX871015; KX871016 to KX871021; KX871022 to KX871027; KX871028 to KX871033; KX871042 to KX871053; KX871054 to KX871064; KX871065 to KX871077; KX871078 to KX871089; KX871090 to KX871095; KX871096 to KX871101; KX871102 to KX871107, KX871108 to KX871113; KX871122 to KX871132; KX871133 to KX871142; KX871143 to KX871150; KX871151 to KX871160; KX871161 to KX871166; KX871167 to KX871172; KX871173 to KX871178; KX871179 to KX871184).

Supplementary information

Supplementary information available online at
http://jeb.biologists.org/lookup/doi/10.1242/jeb.140418.supplemental

References

Adams, K. L. and Wendel, J. F. (2004). Exploring the genomic mysteries of polyploidy in cotton. *Biol. J. Linn. Soc.* **82**, 573-582.

Adams, K. L., Cronn, R., Percifield, R. and Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl. Acad. Sci. USA* **100**, 4649-4654.

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

Auger, D. L., Gray, A. D., Ream, T. S., Kato, A., Coe, E. H., Jr and Birchler, J. A. (2005). Nonadditive gene expression in diploid and triploid hybrids of maize. *Genetics* **169**, 389-397.

Blin, N. and Stafford, D. W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.* **3**, 2303-2308.

Bomblies, K. and Weigel, D. (2007). Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat. Rev. Genet.* **8**, 382-393.

Buggs, R. J. A., Zhang, L., Miles, N., Tate, J. A., Gao, L., Wei, W., Schnable, P. S., Barbazuk, W. B., Soltis, P. S. and Soltis, D. E. (2011). Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr. Biol.* **21**, 551-556.

Buggs, R. J., Wendel, J. F., Doyle, J. J., Soltis, D. E., Soltis, P. S. and Coate, J. E. (2014). The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, pii:20130354.

Chelaifa, H., Monnier, A. and Ainouche, M. (2010). Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina×townsendii* and *Spartina anglica* (Poaceae). *New Phytol.* **186**, 161-174.

Chen, M. and Ni, M. (2006). RED AND FAR-RED INSENSITIVE 2, a RING-domain zinc finger protein, mediates phytochrome-controlled seedling deetiolation responses. *Plant Physiol.* **140**, 457-465.

Collares-Pereira, M. J., Matos, I., Morgado-Santos, M. and Coelho, M. M. (2013). Natural pathways towards polyploidy in animals: the *Squalius alburnoides* fish complex as a model system to study genome size and genome reorganization in polyploids. *Cytogenet. Genome Res.* **140**, 97-116.

Comai, L. (2000). Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol.* **43**, 387-399.

Covelo-Soto, L., Leunda, P. M., Pérez-Figueroa, A. and Morán, P. (2015). Genome-wide methylation study of diploid and triploid brown trout (*Salmo trutta* L.). *Anim. Genet.* **46**, 280-288.

Cunha, C., Coelho, M. M., Carmona, J. A. and Doadrio, I. (2004). Phylogeographical insights into the origins of the *Squalius alburnoides* complex via multiple hybridization events. *Mol. Ecol.* **13**, 2807-2817.

Defossez, P.-A. and Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. *Prog. Mol. Biol. Transl. Sci.* **101**, 377-398.

Diez, C. M., Roessler, K. and Gaut, B. S. (2014). Epigenetics and plant genome evolution. *Curr. Opin. Plant Biol.* **18**, 1-8.

Dreos, R., Ambrosini, G., Périer, R. and Bucher, P. (2015). The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res.* **43**, D92-D96.

Farré, D., Bellora, N., Mularoni, L., Messeguer, X. and Albà, M. M. (2007). Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biol.* **8**, R140.

Feldman, M., Levy, A. A., Fahima, T. and Korol, A. (2012). Genomic asymmetry in allopolyploid plants: wheat as a model. *J. Exp. Bot.* **63**, 5045-5059.

Flagel, L. E. and Wendel, J. F. (2010). Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* **186**, 184-193.

Garcia, T. I., Matos, I., Shen, Y., Pabuwal, V., Coelho, M. M., Wakamatsu, Y., Schartl, M. and Walter, R. B. (2014). Novel method for analysis of allele specific expression in triploid *Oryzias latipes* reveals consistent pattern of allele exclusion. *PLoS ONE* **9**, e100250.

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E. and Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* **196**, 966-971.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95-98.

Inácio, A., Pinho, J., Pereira, P. M., Comai, L. and Coelho, M. M. (2012). Global analysis of the small RNA transcriptome in different ploidies and genomic combinations of a vertebrate complex – the *Squalius alburnoides*. *PLoS ONE* **7**, e41158.

Ishikawa, R. and Kinoshita, T. (2009). Epigenetic programming: the challenge to species hybridization. *Mol. Plant* **2**, 589-599.

Kallman, K. D. (1975). The platyfish, *Xiphophorus maculates*. In *Handbook of Genetics* (ed. R. C. King), pp. 81-132. New York: Plenum Publishing.

Kerkel, K., Spadola, A., Yuan, E., Kosek, J., Jiang, L., Hod, E., Li, K., Murty, V. V., Schupf, N., Vilain, E. et al. (2008). Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* **40**, 904-908.

Kiryu, H., Oshima, T. and Asai, K. (2005). Extracting relations between promoter sequences and their strengths from microarray data. *Bioinformatics* **21**, 1062-1068.

Knudsen, S. (1999). Promoter 2.0: for the recognition of PolII promoter sequences. *Bioinformatics* **15**, 356-361.

Kuo, H. C., Lin, P. Y., Chung, T. C., Chao, C. M., Lai, L. C., Tsai, M. H. and Chuang, E. Y. (2011). DBCAT: database of CpG islands and analytical tools for identifying comprehensive methylation profiles in cancer cells. *J. Comput. Biol.* **18**, 1013-1017.

Lamatsch, D. K., Nanda, I., Epplen, J. T., Schmid, M. and Schartl, M. (2000). Unusual triploid males in a microchromosome-carrying clone of the Amazon molly, *Poecilia formosa*. *Cytogenet. Cell Genet.* **91**, 148-156.

Lamatsch, D. K., Stöck, M., Fuchs, R., Döbler, M., Wacker, R., Parzefall, J., Schlupp, I. and Schartl, M. (2010). Morphology, testes development and behaviour of unusual triploid males in microchromosome-carrying clones of *Poecilia formosa*. *J. Fish Biol.* **77**, 1459-1487.

Lamatsch, D. K., Trifonov, V., Schories, S., Epplen, J. T., Schmid, M. and Schartl, M. (2011). Isolation of a cancer-associated microchromosome in the sperm-dependent parthenogen *Poecilia formosa*. *Cytogenet. Genome Res.* **135**, 135-142.

Lampert, K. P. and Schartl, M. (2008). The origin and evolution of a unisexual hybrid: *Poecilia formosa*. *Philos. Trans. R Soc. B Biol. Sci.* **363**, 2901-2909.

Lampert, K. P., Lamatsch, D. K., Epplen, J. T. and Schartl, M. (2005). Evidence for a monophyletic origin of triploid clones of the Amazon molly, *Poecilia formosa*. *Evolution* **59**, 881-889.

Lampert, K. P., Lamatsch, D. K., Fischer, P., Epplen, J. T., Nanda, I., Schmid, M. and Schartl, M. (2007). Automictic reproduction in interspecific hybrids of poeciliid fish. *Curr. Biol.* **17**, 1948-1953.

Lee, T. Y., Chang, W. C., Hsu, J. B., Chang, T. H. and Shien, D. M. (2012). GPMiner: an integrated system for mining combinatorial cis-regulatory elements in mammalian gene group. *BMC Genomics* **13** Suppl 1, S3.

Li, J. and Zhang, Y. (2014). Relationship between promoter sequence and its strength in gene expression. *Eur. Phys. J. E Soft. Matter* **37**, 86.

Li, A., Hu, B.-Q., Xue, Z.-Y., Chen, L., Wang, W.-X., Song, W.-Q., Chen, C.-B. and Wang, C.-G. (2011). DNA methylation in genomes of several annual herbaceous and woody perennial plants of varying ploidy as detected by MSAP. *Plant Mol. Biol. Rep.* **29**, 784-793.

Mable, B. K. (2003). Breaking down taxonomic barriers in polyploidy research. *Trends Plant Sci.* **8**, 582-590.

Malone, C. D. and Hannon, G. J. (2009). Small RNAs as guardians of the genome. *Cell* **136**, 656-668.

Martienssen, R. A. and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070-1074.

Matos, I., Sucena, E., Machado, M. P., Gardner, R., Inácio, A., Schartl, M. and Coelho, M. M. (2011). Ploidy mosaicism and allele-specific gene expression differences in the allopolyploid *Squalius alburnoides*. *BMC Genet.* **12**, 101.

Matzke, M. A., Scheid, O. M. and Matzke, A. J. M. (1999). Rapid structural and epigenetic changes in polyploid and aneuploid genomes. *Bioessays* **21**, 761-767.

McGaughey, D. M., Abaan, H. O., Miller, R. M., Kropp, P. A. and Brody, L. C. (2014). Genomics of CpG methylation in developing and developed zebrafish. *G3* **4**, 861-869.

Nanda, I., Schartl, M., Feichtinger, W., Schlupp, I., Parzefall, J. and Schmid, M. (1995). Chromosomal evidence for laboratory synthesis of a triploid hybrid between the gynogenetic teleost *Poecilia formosa* and its host species. *J. Fish Biol* **47**, 619-623.

Nanda, I., Schlupp, I., Lamatsch, D. K., Lampert, K. P., Schmid, M. and Schartl, M. (2007). Stable inheritance of host species-derived microchromosomes in the gynogenetic fish *Poecilia formosa*. *Genetics* **177**, 917-926.

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

Pala, I., Schartl, M., Brito, M., Malta-Vacas, J. and Coelho, M. M. (2010). Gene expression regulation and lineage evolution: the North and South tale of the hybrid polyploid *Squalius alburnoides* complex. *Proc. R. Soc. B Biol. Sci.* **277**, 3519-3525.

Riddle, N. C. and Birchler, J. A. (2003). Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. *Trends Genet.* **19**, 597-600.

Salmon, A., Ainouche, M. L. and Wendel, J. F. (2005). Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Mol. Ecol.* **14**, 1163-1175.

Schories, S., Lampert, K. P., Lamatsch, D. K., García de León, F. J. and Schartl, M. (2007). Analysis of a possible independent origin of triploid *P. formosa* outside of the Río Purificación river system. *Front. Zool.* **4**, 13.

Sehrish, T., Symonds, V. V., Soltis, D. E., Soltis, P. S. and Tate, J. A. (2014). Gene silencing via DNA methylation in naturally occurring *Tragopogon miscellus* (Asteraceae) allopolyploids. *BMC Genomics* **15**, 701.

Shen, H., He, H., Li, J., Chen, W., Wang, X., Guo, L., Peng, Z., He, G., Zhong, S., Qi, Y. et al. (2012a). Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids. *Plant Cell* **24**, 875-892.

Shen, Y., Catchen, J., Garcia, T., Amores, A., Beldorth, I., Wagner, J., Zhang, Z., Postlethwait, J., Warren, W., Schartl, M. et al. (2012b). Identification of transcriptome SNPs between *Xiphophorus* lines and species for assessing allele specific gene expression within F_1 interspecies hybrids. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **155**, 102-108.

Shen, Y., Garcia, T., Pabuwal, V., Boswell, M., Pasquali, A., Beldorth, I., Warren, W., Schartl, M., Cresko, W. A. and Walter, R. B. (2013). Alternative strategies for development of a reference transcriptome for quantification of allele specific expression in organisms having sparse genomic resources. *Comp. Biochem. Physiol. D Genomics Proteomics* **8**, 11-16.

Sousa-Santos, C., Robalo, J. I., Collares-Pereira, M.-J. and Almada, V. C. (2005). Heterozygous indels as useful tools in the reconstruction of DNA sequences and in the assessment of ploidy level and genomic constitution of hybrid organisms. *DNA Seq.* **16**, 462-467.

Sousa-Santos, C., Collares-Pereira, M. J. and Almada, V. C. (2006). Evidence of extensive mitochondrial introgression with nearly complete substitution of the typical *Squalius pyrenaicus*-like mtDNA of the *Squalius alburnoides* complex (Cyprinidae) in an independent Iberian drainage. *J. Fish Biol.* **68**, 292-301.

Sousa-Santos, C., Collares-Pereira, M. J. and Almada, V. (2007). Reading the history of a hybrid fish complex from its molecular record. *Mol. Phylogenet. Evol.* **45**, 981-996.

Stöck, M. and Lamatsch, D. K. (2013). Why comparing polyploidy research in animals and plants? *Cytogenet. Genome Res.* **140**, 75-78.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**, 1102-1104.

Tirosh, I., Weinberger, A., Bezalel, D., Kaganovich, M. and Barkai, N. (2008). On the relation between promoter divergence and gene expression evolution. *Mol. Syst. Biol.* **4**, 159.

Turner, B. J., Balsano, J. S., Paul, J. M. and Rasch, E. M. (1983). Clonal diversity and evolutionary dynamics in a diploid-triploid breeding complex of unisexual fishes (*Poecilia*). *Evolution* **37**, 798-809.

Vinogradov, A. E. (1998). Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry* **31**, 100-109

Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J. and Comai, L. (2009). Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr. Biol.* **19**, 1128-1132.

Wang, J., Tian, L., Lee, H.-S., Wei, N. E., Jiang, H., Watson, B., Madlung, A., Osborn, T. C., Doerge, R. W., Comai, L. et al. (2006). Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* **172**, 507-517.

Wang, H., Jiang, J., Chen, S., Qi, X., Fang, W., Guan, Z., Teng, N., Liao, Y. and Chen, F. (2014). Rapid genetic and epigenetic alterations under intergeneric genomic shock in newly synthesized *Chrysanthemum morifolium* x *Leucanthemum paludosum* hybrids (Asteraceae). *Genome Biol. Evol.* **6**, 247-259.

Xiao, J., Song, C., Liu, S., Tao, M., Hu, J., Wang, J., Liu, W., Zeng, M. and Liu, Y. (2013). DNA methylation analysis of allotetraploid hybrids of red crucian carp (*Carassius auratus* red var.) and common carp (*Cyprinus carpio* L.). *PLoS ONE* **8**, e56409.

Yoo, M.-J., Szadkowski, E. and Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**, 171-180.

Zhu, W.-G., Srinivasan, K., Dai, Z., Duan, W., Druhan, L. J., Ding, H., Yee, L., Villalona-Calero, M. A., Plass, C. and Otterson, G. A. (2003). Methylation of adjacent CpG sites affects Sp1/Sp3 binding and activity in the *p21^{Cip1}* promoter. *Mol. Cell. Biol.* **23**, 4056-4065.