Gene expression regulation and lineage evolution: the North and South tale of the hybrid polyploid Squalius alburnoides complex

Irene Pala1,2,*, Manfred Schartl3, Miguel Brito4, Joana Malta Vacas4 and Maria Manuela Coelho1

1Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências da
Universidade de Lisboa, Lisboa, Portugal
2Department of Biology, Lund University, Lund, Sweden
3Physiologische Chemie I, Biozentrum, University of Würzburg, Würzburg, Germany
4Escola Superior de Tecnologias da Saúde de Lisboa, Lisboa, Portugal

The evolution of hybrid polyploid vertebrates, their viability and their perpetuation over evolutionary time have always been questions of great interest. However, little is known about the impact of hybridization and polyploidization on the regulatory networks that guarantee the appropriate quantitative and qualitative gene expression programme. The Squalius alburnoides complex of hybrid fish is an attractive system to address these questions, as it includes a wide variety of diploid and polyploid forms, and intricate systems of genetic exchange. Through the study of genome-specific allele expression of seven housekeeping and tissue-specific genes, we found that a gene copy silencing mechanism of dosage compensation exists throughout the distribution range of the complex. Here we show that the allele-specific patterns of silencing vary within the complex, according to the geographical origin and the type of genome involved in the hybridization process. In southern populations, triploids of S. alburnoides show an overall tendency for silencing the allele from the minority genome, while northern population polyploids exhibit preferential biallelic gene expression patterns, irrespective of genomic composition. The present findings further suggest that gene copy silencing and variable expression of specific allele combinations may be important processes in vertebrate polyploid evolution.

Keywords: allopolyplloid; gene expression; allele silencing; hybrid lineage evolution

1. INTRODUCTION

Polyplody has been proposed as an important driving force of evolution, and the success of polyploid lineages confirms it as a successful evolutionary transition and a potentially relevant factor in evolutionary diversification (Otto & Whitton 2000). Ploidy rise brought upon by a hybridization event creates an epigenetic instability state that can only be overcome if regulation mechanisms, which contribute to gene copy perpetuation and heterosis, outrun disadvantages and potentiate species adaptation and viability. Therefore, it is important to investigate the functional basis of how polyploids overcome an initial period of instability, and establish processes that allow evolutionary flexibility and efficient competition with their diploid counterparts.

Hybrid polyploidy success has been documented in numerous plant species (Otto & Whitton 2000) and also in animals (Dowling & Secor 1997), but the regulatory changes that contribute to genome stabilization and regulation in the presence of distinct chromosome sets are still elusive. Successful polyploid vertebrates including Rana esculenta (Hotz et al. 1999) and the Bufo viridis complex (Stöck et al. 2002), and asexual hybrid lineages such as the Amazon molly, Poecilia formosa (reviewed in Lampert & Schartl 2008), have been extensively studied regarding different molecular and reproductive aspects that contribute to their viability (Lamatsch & Stöck 2009). In all these systems, several aspects (such as the relevance of heterosis and the role of gene duplicates in lineage-specific evolution) have been comprehensively addressed (Comai 2005; Pignatta & Comai 2009), but the impact of polyploidization on the mechanisms regulating gene expression has not yet been clarified. Several deviations in the expected gene expression patterns have been reported in polyploid plants (Auger et al. 2005), including organ-specific silencing (Adams et al. 2003), epigenetic regulation of duplicates (Comai et al. 2000; Shaked et al. 2001), parent-of-origin-specific control of gene expression (Alleman & Doctor 2000) and a still unclear effect designated as ‘odd ploidy response’ (Guo et al. 1996). In vertebrates, little has been done to understand the effects of ploidy rise on gene regulation and their impact on the evolutionary potential of populations. The Squalius alburnoides complex was the first system in which dosage compensation by gene copy silencing was reported in a polyploid vertebrate context (Pala et al. 2008).
This complex of hybrid fish presents several unique features that make it an ideal system to address the question of gene expression regulation and its impact on the persistence of hybrid lineages over evolutionary time (reviewed in Alves et al. 2001). It is an allopolyploid complex, resulting from interspecific hybridization between two Iberian cyprinid species: *Squalius pyrenaicus* as the maternal ancestor (contributing with the so-called P genome), and a still undetermined species (closely related to *Anaecypris hispanica*) as the paternal ancestor (A genome; reviewed in Alves et al. 2001). Presently, the complex is widely distributed in the Iberian Peninsula, where it occurs in sympathy with two species of the genus *Squalius*: *S. pyrenaicus* (P genome) in southern basins and *S. carolitertii* (C genome) in northern basins (figure 1a). These species form a reproductive complex with the hybrids, leading to diploid and polyploid offspring (figure S1, electronic supplementary material). The two bisexual species act as sources of new genetic material, and contribute to the maintenance of gene flow between the different genome-specific gene copies and processes of hybridogenesis (Carmona et al. 1997; Alves et al. 2001; Pala & Coelho 2005). The escape from strictly asexual reproduction and the complexity of the genetic exchange routes definitively contribute to the success of *S. albunoides*. In southern populations, an additional form designated as ‘nuclear non-hybrid’ (AA; carrying mtDNA of *S. pyrenaicus*) occurs and actively contributes to the maintenance of the genetic diversity of the complex (Alves et al. 1999).

We used the *S. albunoides* system to approach the question of gene expression regulation and evolution of polyploid taxa. As an allopolyploid, the *S. albunoides* complex offers the advantage of allowing the distinction between the different genome-specific gene copies and how they contribute to overall expression. The presence of lineages within the complex, established differentially in time, including different forms and showing distinct dynamics of genetic exchange and different evolutionary potential, further allows us to explore the question of whether the mechanisms of gene silencing observed in triploids of the southern populations of *S. albunoides* (Pala et al. 2008) would also be present in other independent populations and in different genomic compositions.

We have studied the expression pattern of a total of seven widely expressed and tissue-restricted genes in different organs of diploid (2N), triploid (3N) and tetraploid (4N) forms in Douro, Mondego and Tejo River basins.
the hypothesis that the type of genomes brought together upon the hybridization process might be an important factor in the establishment of specific mechanisms of gene expression regulation in vertebrate allopolyploids.

2. MATERIAL AND METHODS

(a) Samples, DNA and RNA extraction, and genotyping

Samples of *S. alburnoides* and of the sympatric *Squalus* species were collected from three locations corresponding to the northern and southern distribution ranges of the complex (figure 1b). Individuals were sacrificed with an overdose of the anaesthetic MS222. Organs and fin clips were collected for RNA and DNA extraction, which were performed as described in Pala et al. (2008). Blood samples were drawn from the caudal vein, stabilized in buffer (40 mM citric acid trisodium salt, 0.25 M sucrose and 5% dimethyl sulphoxide) and immediately frozen at −80°C. The determination of genotype identity of individual samples was performed by a conjoined approach of flow-cytometry measurements (Dawley & Goddard 1988) and the analysis of microsatellite variation.

Genotypes were determined by cross-species amplification of three microsatellite loci (LC03, LC04 and LC05) using the Multiplex PCR Kit (Qiagen) and identification of alleles specific to the respective genomes (P, C and A), as previously described (Pala & Coelho 2005; Crespo-López et al. 2007; Cunha et al. 2008). In the cases for which it was not possible to identify the three or four distinct alleles in triploids and tetraploids, an additional locus (LC01) was used (table S1, electronic supplementary material). Amplification products were analysed with an automated sequencer (ABI 310 Genetic Analyser).

(b) Sequence determination and genome expression

Sequences of known teleost orthologues of a total of seven genes—ubiquitously expressed (β-actin, rp88, gapdh, gdf1) and organ-specific (β-actin, dmrt1 and eye-specific (rhodopsin)—were used as templates for the design of gene-specific primers (table S2, electronic supplementary material). Primers were initially tested on cDNA samples of *S. pyrenaicus* (PP), *S. carolitertii* (CC) and *S. alburnoides* (AA). Amplifications were performed according to the following PCR conditions: pre-heating at 94°C for 2 min 30 s, 35 cycles at 94°C for 45 s, 52°C (amh, dmrt1)/55°C (rhodopsin, β-actin, rp88, ef1a and gapdh) for 40 s and 72°C for 1 min 15 s, and a final extension at 72°C for 10 min. Polymorphic sites for the three genomes (P, C and A) were identified for the seven genes by sequence alignment using SEQUENCHER v. 4.0 (Gene Codes Corporation). In hybrid samples, the presence of cDNAs derived from single genome copies or from both genomes was determined through sequence comparison and based on the identified polymorphic sites between genomes (P, C and A). Genome control sequences, representing the P, C and A genomes and obtained from *S. pyrenaicus, S. carolitertii* and nuclear non-hybrid *S. alburnoides*, were also analysed. Two to four forward and reverse sequences of each gene were obtained per individual per organ, using independently synthesized cDNA samples and PCR amplifications. PCR amplifications for replicate reactions were also performed independently, following the procedures described in Pala et al. (2008).

(c) Quantitative real-time PCR analysis

Primers and specific TaqMan probes for β-actin, rp88, gapdh (Pala et al. 2008) and ef1a (EFreal-F5’-CCGTC TGCCACTTCAGGATG-3’; EFreal-R5’-CATACACGGCT TT GAGGACACC-3; EF/6VIC 5’TCCACACGACC ACGGGACAGT-3’) were used for amplification of muscle, liver and gonad samples of representative specimens of the CAA and CCA triploid (3–5) and CCAA tetraploid (1–2) forms from the Douro and Mondego River basins. Reactions and analysis were conducted as described in Pala et al. (2008).

3. RESULTS

(a) Gene expression patterns differ according to organ and geographical location

The sequence analysis of genome-specific polymorphisms for the different genes in the various organs revealed contrasting results between southern and northern populations of *S. alburnoides* (table 1).

In the sample from the southern population (Tejo Basin), the four housekeeping genes (β-actin, rp88, ef1a and gapdh) showed differential expression patterns according to organ. In muscle and liver, diploid PA individuals showed the expected simultaneous expression of A and P alleles (referred from now on as biallelic expression). Triploids, on the other hand, exhibited a variety of expression profiles: for β-actin and rp88 most muscle and liver samples exhibited expression exclusively from the A genome, although biallelic expression was also observed. For ef1a, two triploid samples in muscle and one in liver showed biallelic expression while in the remaining one only A genome transcripts were identified. For gapdh, biallelic expression of P and A alleles was identified in all triploid muscle samples, while in liver samples all genes exhibited a predominantly monoallelic expression of A genome transcripts, with a small number of individuals exhibiting biallelic expression. The most paradigmatic example of this was the β-actin gene, of which transcripts in triploids corresponded exclusively to the A genome.

In individuals from northern populations (Mondego and Douro Basins), in liver and muscle samples, biallelic expression was observed with no exceptions for all gene/organ combinations irrespective of ploidy level and genomic constitution, including CA, CCA, CAA and CCAA *S. alburnoides*. The same was true for β-actin and rp88 transcripts in eye samples of triploid (CCA and CAA) and tetraploid individuals: both C and A alleles were expressed (figure S2, electronic supplementary material). A tendency towards a preferential biallelic expression was also observed in eye samples of southern triploid (PAA) specimens: four out of seven samples showed expression of P and A alleles of the β-actin gene, and rhodopsin transcripts resulted from biallelic expression in all samples analysed.

Brain samples from triploid and tetraploid individuals from Mondego and Douro showed an exception to the overall tendency of biallelic expression of C and A genomes in these basins: although C and A β-actin expression was observed in all samples, the gapdh gene showed monoallelic A-genome-exclusive expression in CAA individuals (figure S2, electronic supplementary material).
In adult gonads (table 1) of southern triploid individuals, all four housekeeping genes (\(\beta\)-actin, \(rpl8\), \(ef1a\) and \(gapdh\)) showed exclusive expression of A genome alleles. The \(amh\) gene showed biallelic expression in triploid PAA, while \(dmrt1\) exhibited both monoallelic A expression (nine samples) and biallelic (three samples) genome expression. Diploid PA exhibited a biallelic expression pattern in all genes analysed (\(amh\), \(dmrt1\) and \(\beta\)-actin). In northern samples, patterns were divergent. The \(amh\) gene was the only gene for which expression of both C and A alleles was observed in all diploid, triploid and tetraploid forms of \(S. alburnoides\). The remaining five genes exhibited a distinctive expression profile between the two forms of triploids: while in CCA samples the observed expression pattern was biallelic (as in diploid and tetraploid samples from the same locations), in CAA individuals \(dmrt1\), \(\beta\)-actin, \(rpl8\), \(ef1a\) and \(gapdh\) expression was exclusively from the A genome. Exclusive expression of alleles from the minority genome was never observed in triploids.

4. DISCUSSION
In the present work, we have addressed the question of whether the mechanisms of gene silencing and regulation of transcripts to the diploid level, described in triploids of the southern populations of \(S. alburnoides\) (Pala et al. 2008), would also be present in other populations and in different genomic contexts. We have studied the expression patterns of housekeeping and tissue-specific genes in individuals of different genomic composition and ploidy degrees, from different locations within the distribution range of the \(S. alburnoides\) complex. Based on a representative sample of the forms that constitute the complex in different locations, we have attempted to understand the global effects of hybridization and ploidy rise on gene expression regulation in this system and evaluate their evolutionary implications.

The method was further validated by an independent quantitative real-time PCR analysis comparing definite cell numbers (I. Matos & M. M. Coelho 2009, unpublished data).

Table 1. Relative frequency of allele-specific transcripts of P, C and A genomes in different organs of individuals from southern (Tejo) and northern (Mondego and Douro) populations of the complex.

<table>
<thead>
<tr>
<th>Ploidy</th>
<th>Species</th>
<th>Genotype</th>
<th>Muscle</th>
<th>Liver</th>
<th>Gonad</th>
</tr>
</thead>
<tbody>
<tr>
<td>2n</td>
<td>(S. pyrenaicus)</td>
<td>PP</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>(S. alburnoides)</td>
<td>AA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(S. alburnoides)</td>
<td>PA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3n</td>
<td>(S. alburnoides)</td>
<td>PAA</td>
<td>(\beta)-actin</td>
<td>0.67</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(rpl8)</td>
<td>0.67</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ef1a)</td>
<td>0.33</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(gapdh)</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(amh)</td>
<td>0.33</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(dmrt1)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(S. alburnoides)</td>
<td>CAA</td>
<td>(\beta)-actin</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(rpl8)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ef1a)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(gapdh)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(amh)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(dmrt1)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Only \(amh\), \(dmrt1\) and \(\beta\)-actin were analysed.*

(b) Quantitative expression
The relative expression ratios obtained from the comparison between average ct values of triploid and tetraploid samples and the diploid controls were always approximately 1 (figure 2), implying dosage compensation by regulation of gene expression to the diploid level in polyploids. Despite qualitative differences in allele representation, a similar effect of dosage compensation was observed in these populations compared with the one described in the southern Tejo basin (Pala et al. 2008).
Gene expression and dosage compensation

In southern populations of the Tejo River basin, the problem of keeping balance of expression regulatory networks in an uneven-numbered genome is apparently overcome by allele silencing (Pala et al. 2008), similar to what has been observed in plant species following polyploidization (Adams et al. 2003, 2004). In plants, it has been proposed that adapted polyploids would avoid extinction by reducing gene redundancy and thus undergoing an evolutionary pathway leading to functional diploidization (Paterson et al. 2004; Wang et al. 2005). The need for correct gene balance in a polyploid context as a way to tolerate the presence of extra genome copies and to maintain stable aneuploids has been further demonstrated in plants (Birchler & Veitia 2007) and discussed also as a possibility in vertebrates (Mable 2007). Recently, dosage effects on transcription levels have been revealed in salmon (Ching et al. 2010), and similar levels of gene expression have been reported in triploid and diploid individuals. In the S. alburnoides complex, the necessity for balanced gene expression is apparent in both southern (Pala et al. 2008) and northern populations, with higher ploidy forms exhibiting gene expression regulated to the diploid levels. The similar quantitative outcome, apparently ensured in both geographical locations, is nevertheless accompanied by a distinct usage of specific alleles. In plant allopolyploids, further support for the possibility of different genome-specific allele usage producing similar transcript levels comes from the study of genome-wide expression dominance. The expression of one or the other genome was found to be dependent on the specific genomic combinations that were brought together in the hybrid, irrespective of the level of gene expression (Rapp et al. 2009).

Thus, the preferential monoallelic expression observed in the southern triploids of S. alburnoides, and the usage of both C and A allele copies in northern populations, are both in line with gene expression patterns observed in other polyploid taxa. Variable patterns of expression according to organ, silencing of genome specific alleles and non-additive gene expression have all been observed in plant polyploids (Adams et al. 2004; Auger et al. 2005; Paterson 2005; Salmon et al. 2005) and pointed out as consequences of the integration of divergent regulatory hierarchies.

Population differences in gene expression patterns of polyploids of the same species and ploidy level have also been reported in allopolyploid plants of the genus Tragopodon (Tate et al. 2006), and independent polyploidy events have been pointed out as a likely factor contributing to the observed variation. A similar effect apparently occurs in the S. alburnoides fish complex: patterns of gene expression diverge even within the same hybrid system, as shown by inequality of allele usage according to geographical origin and genome combination. Distinctive patterns of gene expression were also observed in gonads of the two triploid forms (CAA, CCA) in northern populations, contrasting with the overall identity in expression patterns in somatic organs. The patterns of
gene expression in CAA gonads were similar to the ones observed in southern PAA samples. This may be related to the type of reproduction and mechanism of gamete production of the different forms, namely the exclusion of the minority genome (Carmona et al. 1997; Alves et al. 1998; Pala & Coelho 2005). Conversely, the overall biallelic expression observed in CCA females is not so easy to integrate. Owing to their low abundance in most populations, no extensive information about the reproductive modes of CCA females has been gathered, and the present results might indicate that they do not follow the same process of hybridogenesis as other triploid females (Crespo-López et al. 2006).

(b) Hybridization events and genome usage

The variation in allele expression dynamics observed here could be related to the difference in the origin and the timing of constitution of the three lineages of the complex (ranging from less than 0.7 to 0.01 Myr, according to Sousa-Santos et al. 2007). Regulation of gene expression can be variable in ancient and newly formed polyploids (Wang et al. 2004; Adams & Wendel 2005), and the disparity in the timing of the onset of hybridization events could be a factor to be taken into account.

An alternative possibility could be the difference in the composition of the genomes involved in the maintenance of the complex in each of the three basins. The C genome, introduced through the bisexual species S. caroliterti in northern populations, was a later addition to the complex, after initial hybridization events involving only S. pyrenaicus (Alves et al. 1997; Cunha et al. 2004). Genomic stress induced by two distinct genomes being brought together in the same cell nucleus often leads to different epigenetic modifications of homologous (Riddle & Birchler 2003) and differential capacity of regulatory interactions (Comai 2000; Adams & Wendel 2004; Rapp et al. 2009). Both the P and C genome apparently have a good functional ‘affinity’ with the A genome, but it is possible that regulatory elements that have to interact with both heterologous genomes in hybrids respond differentially or have distinct interaction capacity.

Another important difference between the northern and southern populations of the complex, which has been pointed out as a key factor in the maintenance of its diversity and contributing actively to the evolutionary success of S. alburnoides, is the presence of nuclear non-hybrid males of AA genotype (Alves et al. 2002; Crespo-López et al. 2006, 2007). The preferential expression of A genome alleles in southern populations could be related to the establishment of this nuclear ‘non-hybrid’ lineage. A different preference for expressed alleles might have been established in northern populations in which the process of A allele recombination and shifting, mediated by the nuclear non-hybrid form, is not present.

The heterogeneity of the genomes brought together in plant allopolyploids has been proposed as a major factor underlying patterns of non-additive gene expression, as suggested by studies in different ploidy levels of cotton (Flagel et al. 2008) and Senecio (Hegarty et al. 2006). In these species, the effects of genome merger apparently supplant the impact of genome doubling. The differential patterns of gene expression according to genomic composition, reported here for the S. alburnoides complex, also point towards a strong influence of the type of genomes involved in the hybridization events occurring in each geographical location. However, this might not be the sole explanation for the observed patterns. Dosage restrictions and ploidy level might be additional key players in the regulation of gene expression in this system, as indicated by the similar behaviour of diploids of different geographical origins in terms of qualitative and quantitative gene expression, and the switch to differential patterns of genome specific allele usage in triploids of the northern and southern distribution of S. alburnoides.

The elucidation of the mechanism underlying dosage compensation and preferential allele usage, as well as the characterization of the factors that modulate differential gene expression, will hopefully contribute to a better understanding of how vertebrate polyploid genomes are regulated and how the different adopted strategies can influence their odds on the race for adaptability and evolutionary success.

The authors thank Carina Cunha for providing samples of the Douro River basin and Judite Alves for providing the means for RNA sample quantification and Maria Ana Aboim for help in the field.

This work was supported by the Project PTDC/BIA-BDE/65856/2006 to M.M.C. and by a grant (SFRH/BD/19286/2004) to I.P., both from the Fundação para a Ciência e a Tecnologia (FCT; part of the Ministry of Science of Portugal); M.S. is supported by the Deutsche Forschungsgemeinschaft (SF6 567).

REFERENCES


