JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12308

Comparative analysis reveals that polyploidy does not decelerate diversification in fish

S. H. ZHAN*, L. GLICK†, C. S. TSIGENOPOULOS‡, S. P. OTTO§ & I. MAYROSE†

*University of British Columbia, Vancouver, BC, Canada
†Department of Molecular Biology and Ecology of Plants, Tel Aviv University, Tel Aviv, Israel
‡Institute of Marine Biology and Genetics, Hellenic Centre for Marine Research, Crete, Greece
§Department of Zoology, University of British Columbia, Vancouver, BC, Canada

Keywords:

Acipenseridae;
binary state speciation and extinction;
Botiidae;
chromosome number;
Cyprinidae;
Salmoniformes;
teleost fish.

Abstract

While the proliferation of the species-rich teleost fish has been ascribed to an ancient genome duplication event at the base of this group, the broader impact of polyploidy on fish evolution and diversification remains poorly understood. Here, we investigate the association between polyploidy and diversification in several fish lineages: the sturgeons (Acipenseridae: Acipenseriformes), the botiid loaches (Botiidae: Cypriniformes), Cyprininae fishes (Cyprinidae: Cypriniformes) and the salmonids (Salmonidae: Salmoniformes). Using likelihood-based evolutionary methodologies, we co-estimate speciation and extinction rates associated with polyploid vs. diploid fish lineages. Family-level analysis of Acipenseridae and Botiidae revealed no significant difference in diversification rates between polyploid and diploid relatives, while analysis of the subfamily Cyprininae revealed higher polyploid diversification. Additionally, order-level analysis of the polyploid Salmoniformes and its diploid sister clade, the Esociformes, did not support a significantly different net diversification rate between the two groups. Taken together, our results suggest that polyploidy is generally not associated with decreased diversification in fish – a pattern that stands in contrast to that previously observed in plants. While there are notable differences in the time frame examined in the two studies, our results suggest that polyploidy is associated with different diversification patterns in these two major branches of the eukaryote tree of life.

Introduction

From vertebrates to fungi, polyploidy (or whole genome duplication) is widely recognized as a key feature of eukaryotic genomes (Taylor *et al.*, 2003; Jaillon *et al.*, 2004; Kellis *et al.*, 2004; Dehal & Boore, 2005). Polyploidy reaches its zenith in plants, with all seed plants thought to have experienced one or more genome duplications in their evolutionary past (Bowers *et al.*, 2003; Cui *et al.*, 2006; Soltis *et al.*, 2009; Van de Peer *et al.*, 2009; Jiao *et al.*, 2011). While polyploidy is widespread in plants, it is more sparsely documented in animals (instances summarized in Otto & Whitton, 2000; Mable *et al.*, 2011). Nevertheless, genomewide analyses have revealed several ancient genome duplications in animals: two episodes early in vertebrate evolution (Dehal & Boore, 2005) and one specific to teleost fish (Taylor *et al.*, 2003). Evolutionarily recent cases are reported in amphibians and reptiles, and most notably in fish where entire polyploid lineages have been described (reviewed in Otto & Whitton, 2000; Mable *et al.*, 2011). Polyploids often differ markedly from their diploid progenitors in morphological, physiological and life-history characteristics (Levin, 1983; Ramsey & Schemske, 2002), and these differences may contribute to the establishment and success of polyploid species in novel ecological settings. It is thus hypothesized that polyploidy may serve as an important mechanism for niche differentiation and ecological diversification,

Correspondence: Itay Mayrose, Department of Molecular Biology and Ecology of Plants, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Tel.: +972 3 640 7212; fax: +972 3 640 9380; e-mail: itaymay@post.tau.ac.il

especially in harsh environments (reviewed in Otto, 2007; Fawcett & Van de Peer, 2010).

A long-standing debate concerning polyploidy is whether it influences a lineage's evolutionary success. Historically, researchers have focused on plants because of the rich documentation of polyploidy in this group. Polyploids were traditionally regarded as evolutionary 'dead ends' because of the hypothesized deleterious effects associated with ploidy level increase, such as gene dosage imbalance of the sex chromosomes (Orr, 1990), reduced fertility in heteroploid hybrids (Ramsey & Schemske, 2002), and inefficiency of selection when genes are masked by multiple copies (Haldane, 1933; Fisher, 1935; Wright, 1969). It was further argued that if polyploids were more successful than their diploid relatives, polyploidy should have replaced diploidy as the predominant genetic system in extant eukaryotes (Stebbins, 1971). Supporting these views, a statistical analysis showed that the high prevalence of polyploidy in plants can be explained by frequent polyploid formation and slow reversal to diploidy rather than elevated lineage diversification following polyploidy (Meyers & Levin, 2006). Recent comparative analyses of plant genomes, however, revealed signatures of ancient polyploidization events (i.e. palaeopolyploidy) that occurred multiple times during flowering plant evolution (Van de Peer *et al.*, 2009; Jiao *et al.*, 2011), indicating that all extant flowering plants have experienced at least one round of polyploidy in their evolutionary past. This suggests that rather than being evolutionary 'dead ends', polyploids can indeed persist and even blossom into diverse and successful clades. Greater genetic degrees of freedom, increased heterosis, different niche tolerances and altered colonizing abilities, as well as molecular mechanisms such as functional divergence of duplicated genes and buffering of crucial functions, are a few of the hypotheses proposed to explain the success of polyploid lineages (Werth & Windham, 1991; Soltis & Soltis, 2000; Taylor *et al.*, 2001; Comai, 2005; Chapman *et al.*, 2006; Otto, 2007; Semon & Wolfe, 2007).

Although genomic evidence has rekindled the 'polyploid-success' view, large-scale phylogenetic investigations have suggested otherwise, at least for relatively short evolutionary time scales. Using a comprehensive phylogenetic and cytological data set of vascular plants, Wood *et al.* (2009) reported that polyploidy accompanied 15% and 31% of speciation events in angiosperms and in ferns, respectively. However, they discovered no significant association between polyploid incidence and elevated diversification in plants. Using likelihood-based methodologies, Mayrose *et al.* (2011) further found that recently formed polyploid plant lineages experience lower diversification rates compared to their diploid congeners as a consequence of both lower speciation and higher extinction rates. While the current picture in plants illustrates that neopolyploids generally diversify less rapidly, this hypothesis has not been rigorously investigated in animals.

Among animals, fish exhibit the most appreciable degree of polyploid incidence (reviewed in Leggatt & Iwama, 2003; Le Comber & Smith, 2004; Mable *et al.*, 2011). Polyploid assemblages have been well documented in Acipenseridae (Birstein *et al.*, 1997; Ludwig *et al.*, 2001), Ostariophysi [e.g. Botiidae [Šlecthtová *et al.*, 2006)], and most notably in Cyprinidae (Machordom & Doadrio, 2001a; Tsigenopoulos *et al.*, 2002, 2010; Levin *et al.*, 2012). In addition, the families Salmonidae (Allendorf & Thorgaard, 1984; Johnson *et al.*, 1987) and Catostomidae (Uyeno & Smith, 1972; Ferris, 1984) are thought to have undergone genome duplication in their ancestry. Genomic analyses established that the exceptionally species-rich ray-finned fish descended from a polyploid ancestor, highlighting the potentially profound impact of polyploidy on fish evolution (Taylor *et al.*, 2003). A later study by Hoegg *et al.* (2004) narrowed the phylogenetic window of the ancient polyploidization event, pinpointing it to the branch leading to the radiation of the teleost fish – the main constituent of the ray-finned fish clade. Hoegg and colleagues suggested that the teleost-specific ancient polyploidy was linked to the evolutionary success and phenotypic diversification of teleost fish. More recently, Santini *et al.* (2009) tested the same association using a model-based method that incorporates both phylogenetic and diversity information. Their analysis detected a significant rise in diversification rate around the timing of the ancient polyploidization event. However, the authors cautioned that the ancient polyploidization event (or any other transition along the same branch in the tree) may explain merely ~10% of extant teleost diversity, as much of the remaining diversity may be ascribed to two subsequent radiations that are not associated with known genome duplication events. Importantly, these studies focused on a single ancient polyploidy event, whose link to higher diversification may be coincidental. To improve our understanding of the contribution of polyploidy to fish evolution, more events must be considered and the generality of the association robustly investigated.

In the current study, we assess the link between polyploidy and diversification in a few fish lineages where polyploid species have been extensively documented. By applying likelihood-based phylogenetic methodologies, we estimate the diversification rates of polyploids and their diploid kin, and explore the relative contribution of speciation and extinction to the evolutionary fate of polyploid fish lineages. As our methods are comparable to those used by Mayrose *et al.* (2011), our study enables the first quantitative comparison of polyploid diversification patterns between plants and fish, thus providing important insights into the long-term consequences of polyploidy in eukaryotes.

Materials and methods

Sequence data sets and ploidy level assignments

We gathered phylogenetic and ploidy level data from the literature for four fish groups that display notable variation in ploidy levels: (1) the sturgeons (Acipenseridae: Acipenseriformes), (2) the botiid loaches (Botiidae: Cypriniformes), (3) the Cyprininae subfamily (Cyprinidae: Cypriniformes) and (4) the salmonids and their relatives, the pikes and minnows (Salmonidae: Salmoniformes and Esocidae/Umbridae: Esociformes). Table 1 summarizes information regarding the fish groups examined in this study.

Sturgeon diversity consists of 25 species belonging to 4 genera (Ludwig, 2008). To reconstruct a sturgeon phylogeny, we assembled a multilocus data set using cytochrome *b* (*cytb*), 12S ribosomal RNA (12S rRNA), 16S ribosomal RNA (16S rRNA), cytochrome oxidase *c* subunit II (*COII*), NADH dehydrogenase subunit 5 (*ND5*), tRNA-Asp and tRNA-Phe. Sequence data were gathered from Birstein *et al.* (2002), Krieger *et al.* (2008), references cited in those studies, as well as additional sequences retrieved from NCBI GenBank; for *Acipenser dabryanus*, sequence data for all loci except *cytb* were extracted from the mitochondrial genome (GenBank accession: AY510085.1) published by Peng *et al.* (2007). Two paddlefish species, *Polyodon spathula* and *Psephurus gladius*, were included as outgroup taxa. The combined sequence data set encompasses the entire sturgeon diversity except *Pseudoscaphirhynchus fedtschenkoi*, which has been considered critically endangered and possibly extinct (Birstein, 1993). Ploidy level estimates for 21 species were taken from Table 2 in Peng *et al.* (2007). These estimates were based on microsatellite locus analysis (Ludwig *et al.*, 2001) and genome size data (Zhang *et al.*, 1999 and references therein), and agree well with the chromosome number distribution (available for 19 species; Table 3 in Ludwig *et al.*, 2001). The GenBank accession numbers and ploidy level estimates used to assemble the Acipenseridae data set are provided in Table S1.

The family Botiidae includes 47 species belonging to seven genera (Kottelat, 2004). The *cytb* sequence data from Šlecthtová *et al.* (2006) represent ~74% of extant botiid loach diversity. Three species (*Cobitis bilineata*, *Sabanejewia balcanica* and *S. larvata*) were used as outgroup taxa. Šlecthtová *et al.* (2006) inferred a single polyploidy event in the Botiidae that occurred along the ancestral lineage leading to the well-supported monophyletic group Botiinae. Thus, we treated all taxa in subfamily Botiinae as polyploid and all taxa in its sister subfamily Leptobotiinae as diploid.

Cyprininae, a remarkably diverse subfamily within Cyprinidae, is estimated to encompass over 1300 species belonging to roughly 110 genera (Yang *et al.*, 2010). We pooled together *cytb* sequence data from several sources to create a data set that is as taxonomically and cytologi-

Table 1 Fish groups examined in the current study.

Group	Taxa sampled*	Taxa overlap with Rabosky *et al.* subtrees†	Total no. of species in clade	Percentage of polyploids (%)	Diploid-to-polyploid transitions
Acipenseridae	24 [24]	0/24/0	25	57‡	3
Botiidae	35 [34]	5/30/4	47	69	1
Cyprininae	329 [420]	39/290/130	1300§	71¶	9**
Capoeta	18 [11]	8/10/1	22††	100¶	0
Pseudobarbus	7 [7]	0/7/0	7††	100¶	0
Schizothorax	33 [34]	2/31/3	59††	100¶	0
Sinocyclocheilus	33 [34]	2/31/3	57††	100¶	1
Salmoniformes	60 [59]	8/52/7	215‡‡	100	1
Esociformes	9 [11]	0/9/2	13‡‡	0	0

*Bracketed is the number of taxa in the subtrees extracted from the time-calibrated megaphylogeny published by Rabosky *et al.* (2013a).

†Taxa only in tree built for the current study/taxa in both trees/taxa only in Rabosky *et al.* subtree.

‡Calculated out of 21 species because ploidy level estimates for three species are not available.

§According to Yang *et al.* (2010).

¶Estimated using the best-fitting ChromEvol model, *excluding* 81 species whose ploidy level could not be reliably inferred according to a ChromEvol power analysis (see Materials and Methods).

**Inferred using the best-fitting ChromEvol model.

††Estimated by counting the number of valid species entries in FishBase, while excluding subspecies entries.

‡‡According to the California Academy of Sciences Catalog of Fishes database (Eschmeyer & Fong, 2012).

Table 2 Best-fitting BiSSE model (with the constraint $q_{PD} = 0$) according to AIC. The percent of trees (of 100) where each model received the best (lowest) AIC score is presented. The distribution of difference in AIC score (ΔAIC) relative to the best-fitting model is presented by the 5th and 95th percentile and the median.

	Group			
	Acipenseridae	Botiidae	Cyprininae	Salmoniformes/ Esociformes
Best-fitting model	M0	M0	Mse	Mse
% of trees supporting:				
Mse	0	0	100	100
Me	0	0	0	0
Ms	0	0	0	0
M0	100	100	0	0

Model	5/50/95 percentiles of ΔAIC compared to best-fitting model			
Mse	2.27/2.73/3.06	2.81/3.47/3.75	0/0/0	0/0/0
Me	0.96/1.16/1.41	2.00/2.00/2.00	63.44/68.67/75.15	13.80/16.08/19.22
Ms	0.29/0.76/1.18	1.19/1.56/1.79	27.30/32.94/38.40	12.99/15.19/18.83
M0	0/0/0	0/0/0	88.98/94.86/99.74	15.82/17.88/20.82

M0, diploid and polyploid lineages have equal rates of both speciation and extinction; Ms, speciation rates of diploid and polyploid lineages are allowed to vary, while extinction rates are equal; Me, extinction rates of diploid and polyploid lineages are allowed to vary, while speciation rates are equal; Mse, speciation rates as well as extinction rates of diploid and polyploid lineages are allowed to vary; AIC, Akaike Information Criterion; BiSSE, binary state speciation and extinction.

Table 3 MCMC-based estimates of evolutionary rates using the BiSSE model (with the constraint $q_{PD} = 0$). The rates of speciation ($\overline{\lambda}$) and extinction ($\overline{\mu}$) for diploids (D) and polyploids (P) were estimated from the median of the posterior distributions generated by MCMC sampling over all 100 trees. All rates are scaled relative to a total tree depth of 1, before pruning off the outgroup taxa. For each MCMC step, the diversification rate was derived as the speciation rate minus the extinction rate for diploids and polyploids (r_D and r_P, respectively). The posterior probability (PP) that diploids exhibit higher rates than polyploids is represented by the percentage of MCMC steps where the inequality was satisfied.

Group	$\overline{\lambda}_D$	$\overline{\lambda}_P$	$\overline{\mu}_D$	$\overline{\mu}_P$	$\overline{q}_{DP}$	PP($\lambda_D > \lambda_P$)	PP($\mu_D > \mu_P$)	PP($r_D > r_P$)
Acipenseridae	69.45	99.46	32.05	45.44	13.76	0.25	0.40	0.37
Botiidae	27.21	18.89	11.91	3.46	2.26	0.81	0.79	0.46
Cyprininae	21.80	121.45	1.95	91.69	1.55	**0.00**	**0.00**	*0.03*
Salmoniformes/ Esociformes	14.21	157.30	8.77	143.81	2.50	**0.00**	**0.00**	0.11

BiSSE, binary state speciation and extinction; MCMC, Markov chain Monte Carlo.
PP ≤ 0.025 is considered as support for higher rates in polyploids than diploids (highlighted in bold; marginal significance is in italics).

cally comprehensive as possible. For species belonging to the *Barbus sensu lato* group (including *Labeobarbus*, *Luciobarbus*, *Pseudobarbus* and *Barbus sensu stricto*), *cytb* sequence data were compiled from Machordom & Doadrio (2001a,b), Tsigenopoulos *et al.* (2002, 2003), Marková *et al.* (2010) and Tsigenopoulos *et al.* (2010); for species in the genera *Capoeta*, *Schizothorax* and *Sinocyclocheilus*, sequence data were primarily taken from Xiao *et al.* (2005), He & Chen (2006) and Levin *et al.* (2012), respectively; for the tribe Cyprinini *sensu stricto* (including *Carassioides*, *Carassius*, *Cyprinus* and *Procypris*; Yang *et al.*, 2010), most sequence data used for the current study were also used in Yang *et al.* (2010); additionally, for species belonging to various other genera within Cyprininae (as described in Yang *et al.*, 2010), sequence data were retrieved from NCBI GenBank. Only species considered valid according to FishBase (Froese & Pauly, 2012; http://www.fishbase.org/) were used. Sequence entries having ambiguous species designation (e.g. '*Barbus* sp.' and '*Capoeta* cf. *Banarescui*') as well as multiple subspecies entries were excluded (e.g. among *Carassius auratus* subspecies only *C. auratus langdorfii* was retained). Two non-Cyprininae species (*Tinca tinca* and *Gobio gobio*) were used as outgroup taxa. Chromosome numbers, used for the inference of ploidy levels (see below), were retrieved primarily from FishBase, Tsigenopoulos *et al.* (2002), references listed in Yang *et al.* (2010) and miscellaneous sources provided in the Supporting Information. Table S2 provides the GenBank accession numbers and chromosome numbers used for the Cyprininae data set.

In addition to these three fish groups, in which variation in ploidy level exists within the family, we

constructed a data set to examine the between-order consequences of polyploidy in Salmoniformes following its divergence from its sister clade Esociformes (the relationship between these two orders was demonstrated by Ishiguro *et al.*, 2003; López *et al.*, 2004; and Li *et al.*, 2008). Salmoniformes is thought to have evolved from a tetraploid ancestor (Allendorf & Thorgaard, 1984; Johnson *et al.*, 1987). Esociformes, on the other hand, did not undergo genome duplication after its divergence from Salmoniformes – a hypothesis supported by karyotypic data (Phillips & Ráb, 2001; Mank & Avise, 2006) and genome size data (C values of ~1.87–4.90 pg in salmonid species and ~0.85–2.70 pg in esocid and umbrid species according to the Animal Genome Size Database; Gregory, 2012). Thus, we regarded all salmonid species as polyploid and all esociform species as diploid. We assembled a *cytb* sequence data set that included 60 salmonid species from nine genera and nine esociform species from four genera. Two cyprinid species, *Barbus bocagei* and *Gobiobotia abbreviata*, were incorporated as outgroup taxa. Table S3 lists the GenBank accession numbers of the *cytb* sequences used for the Salmoniformes/Esociformes data set.

Phylogenetic reconstruction

Multiple sequence alignments were constructed using MUSCLE, version 3.8 (Edgar, 2004). The best-fitting nucleotide substitution model was selected using jModelTest, version 0.1.1 (Posada, 2008). Using Akaike Information Criterion (AIC) (Akaike, 1974), the GTR+G model was chosen as the most appropriate model for all single-locus *cytb* data sets (i.e. Botiidae, Cyprininae and Salmoniformes/Esociformes). The best-fitting model in the Acipenseridae data set was identified independently for each locus. The GTR+G model provided the best fit for 12S rRNA and *ND5*; the HKY+G model for 16S rRNA, *COII* and *cytb*; the SYM+G model for tRNA-Phe; and the K80+G model for tRNA-Asp. Next, a set of ultrametric Bayesian trees was sampled for each data set using MrBayes, version 3.2.1 (Huelsenbeck & Ronquist, 2001), under a relaxed molecular clock according to a Brownian motion model (Thorne *et al.*, 1998), running for 2 000 000 steps with a sampling frequency of once per 2000 steps; the initial 50% of the steps were discarded as burn-in. For the Acipenseridae data set, we conducted a partitioned MrBayes analysis allowing each defined locus to evolve at its best-fitting model identified by jModelTest. For each data set, tree topology was constrained so that the ingroup taxa formed a monophyletic group separate from the outgroup taxa. Outgroup taxa were pruned from the resulting MrBayes trees prior to the inference of ploidy levels and the diversification analysis. To summarize the distribution of MrBayes trees, 50% majority-rule consensus trees were built using *phyutility*, version 2.1.1 (Smith & Dunn, 2008; http://code.google.com/p/phyutility/) (Fig. S1). We assessed the quality of our phylogenetic reconstruction by comparing the MrBayes consensus trees to the phylogenies presented in the previous studies when available.

Inference of polyploidy

The Cyprininae data set contained species with available chromosome number information but for which ploidy level had not previously been determined. Thus, given the reconstructed phylogeny (a 50% majority-rule consensus tree built from the MrBayes trees was initially used; see below for a procedure that accounted for phylogenetic uncertainties), we next aimed to infer extant taxa as diploid or polyploid relative to the base chromosome number of the group examined. By doing so, we implicitly treated the root of the phylogeny as diploid. Thus, polyploids are defined here as those lineages that underwent a polyploidization event since the divergence from the common ancestor of the group examined. Specifically, the ChromEvol methodology (Mayrose *et al.*, 2010) was used to assign ploidy levels. This likelihood-based method assesses the fit of several models that allow for various types of chromosome number change along the phylogeny, infers the expected number of polyploid and dysploid (chromosome number changes by one, due, for example, to processes such as chromosome fission or fusion) transitions along each branch of the phylogeny, and reconstruct chromosome numbers at ancestral nodes of the tree. The models available in ChromEvol include six parameters for various types of chromosome number transition; ascending and descending dysploidy, polyploidy (i.e. doubling of the number of chromosomes), and 'demi-polyploidy' (i.e. multiplication of the chromosome number by 1.5, leading to, e.g. triplication events). Additional two-rate parameters allow the ascending and descending dysploidy rates to depend on the current number of chromosomes. We ran all eight available ChromEvol models (each including a different combination of the six rate parameters) and used AIC to select the best model. The expected number of ploidy transitions along each branch of the phylogeny was recorded based on the best-fitting model. In Mayrose *et al.* (2011), an extant taxon was categorized as a polyploid if the estimated expected number of ploidy transitions from the root to the tip exceeded a certain predefined threshold and as diploid otherwise. However, by arbitrarily setting a strict (or lenient) threshold for assigning polyploidy, the number of polyploid taxa may be underestimated (or overestimated). This misclassification may be particularly pronounced for groups with sparse chromosome number data. Thus, to prevent misestimating polyploid diversity, a simulation-based approach was developed (see below) and was applied to the Cyprininae data set in order to determine the optimal threshold that should be used (i.e. 0.48). Diversification analyses obtained using the 0.90 threshold (as in May-

rose *et al.*, 2011) resulted in similar conclusions regarding the relative diversification rates of diploids and polyploids (Table S10).

The ChromEvol methodology allowed us to categorize an extant species as polyploid or diploid regardless of whether chromosome number data were available for that specific taxon. However, because sampling of chromosome number data in certain clades may be comparatively sparse, ploidy levels may not always be estimated reliably. Thus, simulations were used to assign a confidence measure to the ploidy assignment of each extant taxon. Specifically, using the best-supported ChromEvol model, chromosome numbers were evolved from the root of the tree to the tips starting with the maximum *a posteriori* ancestral chromosome number estimated by the original ChromEvol analysis. In these simulations, we recorded for each tip taxon the evolutionary path leading to it (i.e. the number of polyploidization events from the root of the tree), and thus, the 'true' (simulated) ploidy level is known for all taxa. The resulting chromosome numbers at the tips of the tree were used as the data input to ChromEvol. To make the inference step as realistic as possible, simulated chromosome numbers at the tips were retained only for those species with available chromosome number data in the original data set and were converted to 'unknown' for species with missing chromosome number information prior to the inference step. An extant taxon was then inferred to be a polyploid if the estimated number of ploidy transitions from the root to the tip exceeded a certain threshold and as diploid otherwise. The assigned ploidy levels (either diploid or polyploid) were then compared with the ploidy levels that were actually simulated, and the number of correctly assigned taxa (i.e. true positives) was determined. This procedure was repeated for 100 randomly selected MrBayes trees and the threshold that resulted in the highest number of true positives was considered the optimal threshold for that data set (for Cyprininae, the optimal threshold was determined to be 0.48; varying this threshold in the range 0.44–0.65 resulted in nearly identical ploidy assessments, with true positive rates in the range 95.9–96.0%). Finally, using the optimal threshold, taxa with correctly inferred ploidy levels in at least 95% of the simulated runs were considered reliable, while the ploidy levels inferred for species that did not meet this cut-off were treated as unreliable.

A second complementary procedure was taken to account for phylogenetic uncertainties in the ChromEvol inferences – that is, to identify taxa whose inferred ploidy levels were sensitive to the underlying phylogeny. Specifically, ChromEvol was run on the set of 100 MrBayes trees using the best-fitting model determined from the initial ChromEvol analysis (which was based on the consensus tree), resulting in ploidy level estimates per tree as detailed above. Taxa whose ploidy level assignment was different than their consensus assignment (the assignment that was most commonly inferred) in more than 5% of the trees were deemed unreliable.

The ploidy status for all taxa whose ploidy assignment was flagged as unreliable according to either one of the two approaches described above was converted to missing data in subsequent binary state speciation and extinction (BiSSE) analyses (assigned as 'NA' when using the *diversitree* package, see below). The ploidy levels estimated by ChromEvol for Cyprininae are provided in Table S2.

BiSSE diversification analysis

To estimate diversification rates for diploids and polyploids, we applied the BiSSE model (Maddison *et al.*, 2007). BiSSE co-estimates six parameters: speciation rates of lineages in state P (polyploid) and D (diploid) (λ_P and λ_D, respectively); extinction rates of lineages in state P and D (μ_P and μ_D, respectively); and transition rates from P to D (q_{PD}) and D to P (q_{DP}). Using these estimates, the net diversification rate in each state (r_D and r_P) was calculated as, for example, $r_D = \lambda_D - \mu_D$. Because we defined polyploids as those species that had undergone a polyploidization event sometime since divergence from the base of the group examined, we forced the root state to the diploid state and fixed q_{PD} to zero (see Supporting Information for results allowing for polyploid-to-diploid reversals). This constraint is also compatible with the common assumption that polyploidy is largely an irreversible process (Meyers & Levin, 2006). Our analyses were performed using the 'skeletal' tree approach (FitzJohn *et al.*, 2009) implemented in the R package *diversitree*, version 0.9.3 (FitzJohn, 2012; www.zoology.ubc.ca/prog/diversitree/), which accounts for the sampling fraction of species in the given phylogeny out of the total number of species in the clade. Diversity estimates for the various groups analysed here were drawn from the literature (summarized in Table 1). Moreover, uneven sampling of polyploids and diploids in different clades of a phylogeny may influence estimates of diversification. Therefore, we accounted for uneven sampling using the 'split' extension of the BiSSE model (as implemented in *diversitree*) for the analysis of Salmoniformes and Esociformes as well as for Cyprininae. In the Salmoniformes and Esociformes phylogeny, we assumed that all members of Salmoniformes are polyploid and all members of Esociformes are diploid. Thus, we corrected for uneven sampling by specifying that 28% of polyploids and 100% of diploids in Salmoniformes were represented in the phylogeny, while 100% of polyploids and 69% of diploids in Esociformes were represented (Table 1). Similarly, we adjusted for the biased sampling in Cyprininae. We were able to confidently assign diversity estimates to four genera (*Capoeta*, *Pseudobarbus*, *Schizothorax* and *Sinocyclocheilus*), which were well corroborated to be monophyletic (with at least 95% posterior

probability support; Fig. S1d). Using genus diversity estimates from FishBase (Table 1), the sampling fractions for *Capoeta*, *Pseudobarbus*, *Schizothorax* and *Sinocyclocheilus* were estimated as 82%, 100%, 56% and 58%, respectively. Assuming that Cyprininae consists of 1300 species (Yang *et al.*, 2010), the 'background' clade (the rest of the phylogeny, excluding the above four genera) has a sampling fraction of ~21%. Unlike in the case of Salmoniformes and Esociformes, we assumed that within each specified clade, polyploids and diploids were being sampled at the same rates (e.g. in *Capoeta* 82% of polyploids and 82% of diploids were represented in the dataset). Results obtained using the complete sampling assumption were nearly identical (but with broader confidence intervals for the model parameters compared to those obtained while accounting for incomplete sampling; results not shown).

First, we conducted a maximum likelihood (ML) analysis to test whether (i) diploids and polyploids speciate at different rates, (ii) diploids and polyploids go extinct at different rates, or (iii) diploids and polyploids have both different speciation and different extinction rates. These three hypotheses can be tested by comparing the following BiSSE models, starting with the null model, M0, where $\lambda_D = \lambda_P$ and $\mu_D = \mu_P$. (i) Ms, where only speciation rates differ and $\mu_D = \mu_P$; (ii) Me, where only extinction rates differ and $\lambda_D = \lambda_P$; and (iii) Mse, in which both the speciation rates and the extinction rates are allowed to differ between diploids and polyploids. To account for uncertainty in the trees, we fitted the four models to 100 randomly selected post-burn-in MrBayes trees. The AIC model selection criterion was used to identify the best-fitting model given each individual tree. The best-fitting model across all trees was then chosen as the model that was best-supported most frequently (we note that in all four data sets examined, the best-supported model was also chosen across all 100 trees when tested individually). In addition, most model comparisons were nested (except Ms vs. Me), and all of the conclusions drawn from those comparisons were also supported by likelihood-ratio tests.

In addition to the ML-based analysis, the Markov chain Monte Carlo (MCMC) approach described in FitzJohn *et al.* (2009) was applied to obtain posterior probability distributions for each of the five parameters (λ_D, λ_P, μ_D, μ_P and q_{DP}), accounting for uncertainty in parameter estimation and incomplete sampling. Specifically, exponential priors (mean set to (log (number of tip taxa)/tree height) × 2 [or 0.5 for q_{DP}]) were placed on the five parameters. The BiSSE analysis was again conducted across the set of 100 MrBayes trees. For the first tree in the sample, the initial starting point was determined based on a heuristic estimated by *diversitree* according to the state-independent birth–death model. The subsequent 99 trees were started from the last point sampled in the previous tree. For each of the MrBayes trees, MCMC analysis was run for 1500 steps (except the initial tree, which was run for 2000 steps) and sampled every 10th step. The first 500 steps of the chain for each tree were regarded as burn-in and discarded from the analysis (first 1000 steps for the initial tree). The 100 chains (each corresponding to one tree sampled by MrBayes) were then concatenated to form a single sample. We note that individual MCMC chains converged rather quickly (graphical analysis suggests that the MCMC chains stabilized within several hundred steps) and that results were indistinguishable whether we pooled MCMC samples from 10, 50 or 100 trees (we nonetheless used the larger sample set).

To test whether estimated extinction and speciation rates differ between polyploids and diploids, we calculated the percentage of BiSSE MCMC steps in which the diploid rate was higher than that of polyploids (the posterior probability, PP, of diploids having a higher rate than polyploids). For example, to test whether extinction rates differ, we calculated the percentage of post-burn-in steps in which $\mu_D > \mu_P$; $PP(\mu_D > \mu_P) \geq 0.975$ is interpreted as significant support for the conclusion that diploids go extinct at a higher rate than polyploids while supports higher $PP(\mu_D > \mu_P) \leq 0.025$ polyploid extinction.

BiSSE analysis on time-calibrated phylogenies

The BiSSE analyses described above were repeated using time-calibrated phylogenies obtained through a recently assembled mega-phylogeny that encompasses 7822 extant fish species and spans the entire actinopterygiian diversity (Rabosky *et al.*, 2013a). This mega-phylogeny was reconstructed based on a 13-gene matrix and was time-calibrated using 60 fossil dates. The mega-phylogeny was downloaded from Dryad Digital Repository (Robosky *et al.*, 2013b), and the subtrees corresponding to the four fish groups investigated in our study were extracted. Table 1 provides information regarding the overlap between the taxa in the time-calibrated trees and the taxa in the trees reconstructed in the current study. For taxa found in the time-calibrated tree only, additional chromosome numbers were taken from FishBase and the literature (Table S2). Ploidy levels were assigned as detailed above except for Cyprininae, for which the simulation procedure detailed above was conducted along the single time-calibrated tree rather than a set of trees. The procedures for the diversification analyses followed those described above except that the BiSSE MCMC chain was run using a single tree for 20 000 generations (with the first 10 000 discarded as burn-in) instead of 1500.

Results

Phylogenetic distribution of polyploidy

The 50% majority-rule consensus tree for each group is presented in Fig. S1. The reconstructed Acipenseridae

phylogeny is highly similar to recently published Acipenseridae phylogenies (Peng *et al.*, 2007; Krieger *et al.*, 2008). All partitions in our phylogeny that received ≥ 95% posterior probability supported the same descendent species as the corresponding partitions in the phylogeny shown in Fig. 1 of Peng *et al.* (2007), while the three partitions that were not identical received low support (< 85 posterior probability) and involved species with the same ploidy levels. Our phylogeny confirms the basal relationships among the major lineages – the monophyly of the genus *Scaphirhynchus*, the basal clade containing *Acipenser oxyrinchus* and *A. sturio*, and the basal split to the monophyletic Atlantic clade and Pacific clade (Fig. S1a). Our phylogeny also recovered the three monophyletic lineages (one in the Atlantic clade and two in the Pacific clade) that experienced a common polyploidization event (as shown in Peng *et al.*, 2007).

The *cytb* Botiidae phylogeny reconstructed in this study (Fig. S1b) is in strong agreement with the *cytb* phylogeny published by Šlecthtová *et al.* (2006). Our phylogenetic analysis recovered the well-supported subfamilies Botiinae and Leptobotiinae, the monophyly of *Botia*, *Leptobotia*, *Parabotia*, *Sinibotia* and *Syncrossus*, and the paraphyly of *Yasuhikotakia*; the basal relationships among these lineages were supported, as demonstrated previously by Šlecthtová *et al.* (2006), except for *Chromobotia macranthus*, which is clustered with the *Sinibotia/Syncrossus/Yasuhikotakia* lineage with relatively poor support (Fig. S1b). Botiinae was previously inferred to have experienced a whole genome duplication event (Šlecthtová *et al.*, 2006), a conclusion supported by a ChromEvol analysis using the chromosome number data referred to in Šlecthtová *et al.* (2006) and using the consensus phylogeny of Fig. S1b (results not shown).

The evolutionary relationships among the major clades within Cyprininae (*B. sensu stricto*, *Capoeta*, *Labeobarbus*, *Schizothorax* and *Sinocyclocheilus*) are represented in our phylogenetic reconstruction of the subfamily (Fig. S1c). The monophyly of each of the genera *Capoeta*, *Pseudobarbus*, *Schizothorax* and *Sinocyclocheilus* was supported, as well as the monophyly of several clades consisting primarily of *Barbus* species. Given the consensus Bayesian phylogeny, we next aimed to use chromosome number data to infer shifts in ploidy levels. Because chromosome number data are incomplete and not uniformly distributed across the Cyprininae (available for 103 species of the 329 species in the reconstructed phylogeny) and the phylogeny of this group is still debated, we performed two tests – one based on parametric simulations and one to account for phylogenetic uncertainties – to determine the species for which ploidy levels can be reliably inferred (see Materials and Methods). This procedure resulted in 175 species inferred as polyploids, 73 as diploids and 81 as 'unreliable'. ChromEvol performs ML reconstruction of chromosome numbers at each internal node. Ploidy transitions along the Cyprininae phylogeny were then inferred using the ML ancestral reconstruction of chromosome numbers estimated using ChromEvol. The reconstructed haploid chromosome numbers consists of three ploidal levels (corresponding to haploid chromosome numbers of ~25, ~50, and ~75) with the root being at 25; chromosome number estimates at inferred ploidy shifts are indicated in Fig. S1c. We estimated ploidy transitions along 10 internal branches and three terminal branches of the Cyprininae phylogeny; 9 of these shifts involve transitions from diploidy to tetraploidy (seven along internal branches and two along terminal branches), while the other four shifts involve transitions to higher ploidy levels (three internal and one terminal). These ploidy shifts lead to or occur within Cyprininae clades that contain a notable number of reported polyploid taxa (*Capoeta*, *Carassius/Cyprinus*, *Luciobarbus*, *Sinocyclocheilus*, a *Barbus/Pseudobarbus* group described by Tsigenopoulos *et al.*, 2002, and a large assemblage of miscellaneous genera that include *Tor*, *Labeobarbus* and *Varicorhinus*).

Previous molecular phylogenetic studies investigated the evolutionary relationships among the major genera of Salmonidae (Crespi & Fulton, 2004) and among the subfamilies of Salmonidae (Coregoninae, Salmoninae, and Thymallinae; Yasuike *et al.*, 2010). To the best of our knowledge, however, no comprehensive species-level phylogeny exists for Salmonidae (the only family within Salmoniformes). In brief, our phylogeny recovers the monophyly of the major salmonid genera (*Salmo*, *Salvelinus* and *Oncorhynchus*) and the monophyly of each of the three subfamilies (Fig. S1d). Within Esociformes, all the genera (*Esox*, *Dallia*, *Novumbra* and *Umbra*) are each monophyletic, and the phylogenetic relationships among them that are supported by our phylogeny were also shown by López *et al.* (2004).

Comparing the diversification rates of diploids and polyploids

In the data sets examined in this study, the vast majority of diploid-to-polyploid transitions were inferred along internal branches of the phylogeny (Fig. S1). Additional ploidy level increases (e.g. tetraploidy to octaploidy), however, were inferred to occur along terminal branches in Acipenseridae (Fig. S1a) and in Cyprininae (Fig. S1c). This phylogenetic distribution of ploidy level shift contrasts with that observed in plants, where diploid-to-polyploid shifts mainly occur along terminal branches, not internal ones (Mayrose *et al.*, 2011). The difference in the phylogenetic distribution of ploidy shifts between fish and plants suggests that polyploidy may have a different impact on the persistence and diversification in the groups examined.

For Acipenseridae, the BiSSE ML analysis indicated that the best-supported model according to AIC was

M0, which assumes equal speciation and extinction rates of diploids and polyploids (Table 2). Thus, there is no significant support for the more complex models that allow for unequal diversification rates. The BiSSE MCMC results were consistent with this conclusion ($PP(\lambda_D > \lambda_P) = 0.25$ and $PP(\mu_D > \mu_P) = 0.40$). The median of the MCMC posterior distribution suggested that, if anything, polyploids tend to exhibit higher speciation rates and higher extinction rates than diploids (both by nearly 1.4-fold), but these differences were not statistically significant. Together, the MCMC analysis did not support the hypothesis that polyploidy is associated with different net diversification rates among sturgeons ($PP(r_D > r_P) = 0.37$; Table 3). We note that *Huso dauricus*, which was estimated to be diploid by microsatellite locus analysis (Ludwig *et al.*, 2001), was recently shown to be polyploid by karyotyping (Vasil'ev *et al.*, 2009). Our results were qualitatively the same whether we treated *H. dauricus* as diploid (presented here) or as polyploid (not shown).

According to the ML analysis, the best-fitting model for Botiidae was again M0, suggesting that diploid and polyploid lineages in this group do not differ substantially in their speciation or extinction rates. In agreement, the MCMC results did not support significantly different speciation rates ($PP(\lambda_D > \lambda_P) = 0.81$) or extinction rates ($PP(\mu_D > \mu_P) = 0.79$). Based on the MCMC posterior distributions, the median speciation and extinction rates were higher for diploids (roughly by 1.4-fold and 3.4-fold, respectively), but there was virtually no difference in the overall diversification rates between diploids and polyploids ($PP(r_D > r_P) = 0.46$; Table 3).

For Cyprininae, the ML analysis supported the hypothesis that polyploids possess both higher speciation rates and higher extinction rates than diploids, with Mse being the best-supported model. In agreement, the MCMC analysis showed that speciation rates are about 5.5-fold higher in polyploids vs. diploids ($PP(\lambda_D > \lambda_P) = 0$), and that extinction rates are about 47.1-fold higher in polyploids vs. diploids ($PP(\mu_D > \mu_P) = 0$). Given the lower absolute extinction rates, however, the net result was a marginally significant higher diversification rate of polyploid lineages ($PP(r_D > r_P) = 0.03$; Table 3).

Our BiSSE ML analysis also revealed that polyploid salmonids experience higher speciation rates and higher extinction rates than the diploid Esociformes, with Mse being the best-supported model. In agreement, the MCMC analysis indicated that the polyploid salmonids have significantly higher speciation rates (by about 11.1-fold; $PP(\lambda_D > \lambda_P) = 0$) and higher extinction rates (by about 16.4-fold; $PP(\mu_D > \mu_P) = 0$) than diploid Esociformes. We caution, however, that the association between polyploidy and higher speciation and extinction rates is based on the single polyploid transition leading to the salmonid clade, and so the evidence in this case is not replicated. Furthermore, the inflated speciation rates and extinction rates did not result in a significantly increased net diversification rate in the salmonids ($PP(r_D > r_P) = 0.11$; Table 3).

The above diversification analyses were repeated but this time using, for each group, a single time-calibrated tree extracted from a mega-phylogeny reconstructed for the ray-finned fishes (Rabosky *et al.*, 2013a). Results obtained for Acipenseridae, Botiidae and Cyprininae were very similar to those obtained using the sample of trees reconstructed here (see Table S9). For the Salmoniformes/Esociformes comparison, this analysis also supported significantly higher polyploid speciation ($PP(\lambda_D > \lambda_P) = 0$; Table S9) and extinction rates ($PP(\mu_D > \mu_P) = 0$; Table S9). However, this analysis supported a significantly higher net diversification rate for polyploids ($PP(r_D > r_P) = 0.01$).

All of the above BiSSE analyses were conducted under the assumption that polyploid-to-diploid reversals do not occur (i.e. $q_{PD} = 0$). Results obtained while allowing for polyploidy reversals generally mirror those obtained under the irreversibility assumption. In the ML analyses, the best models were the same, except for Acipenseridae for which Ms instead of M0 had a marginally superior, but not significant, AIC score (Table S6). ML parameter estimates with and without the irreversibility constraint are given in Table S5 and Table S7, respectively, while results obtained using MCMC reached the same conclusions and are given in Table S8.

Discussion

Recent literature surveys indicate that polyploidization is generally a rare occurrence among vertebrates, but it is particularly prominent in fish where entire polyploid assemblages have arisen (Mable *et al.*, 2011). Genome-wide analyses have unearthed ancient polyploidization events across the eukaryotic tree of life, promoting the view that polyploidy has played an important role in eukaryotic evolution. However, at least over a relatively short time scale, large-scale phylogenetic studies have suggested that recently formed polyploid plant species generally experience lower diversification rates compared to their diploid congeners (Mayrose *et al.*, 2011). To the best of our knowledge, the question whether polyploidy is associated with a shift in diversification patterns has not been rigorously explored in fish.

Our results suggest that polyploid fish lineages do not exhibit lower diversification rates compared to closely related diploids. In all four groups examined, the diversification rate for polyploids was higher than that of diploids in the majority of MCMC steps. Using the collection of Bayesian trees, this difference was marginally significant in the Cyprininae and not significant in the three other groups (Table 3). Results obtained using a single time-calibrated phylogeny were similar except that the higher net diversification of the polyploids Salmoniformes was significantly higher than their Esociformes diploid relatives. In addition, both the speciation

rates and the extinction rates were higher in polyploids than in diploids in the Acipenseridae, Cyprininae and Salmoniformes/Esociformes comparisons, but the differences were only significant in the latter two groups, while Botiidae exhibited the opposite trend (Tables 2 and 3). These general conclusions also held when we relaxed the assumption that polyploids could not revert to diploid (allowing $q_{PD} \neq 0$; see Tables S6 and S8).

The subfamily Cyprininae encompasses the largest number of known polyploidization events in fish, with multiple transitions leading to the tetraploid and hexaploid *Barbus* groups (Tsigenopoulos *et al.*, 2002). Because polyploidy has been frequently reported in Cyprininae and because this clade is so species-rich, it is tempting to hypothesize that the evolutionary success of Cyprininae is attributed to polyploidy. Our study indicates that polyploid Cyprininae lineages indeed diversify more rapidly than diploid lineages. These results nonetheless should be interpreted with caution owing to taxonomic biases as discussed below.

In a recent diversification analysis comparing diploid and polyploid plant lineages, Mayrose *et al.* (2011) demonstrated that polyploid plant species tend to undergo lower speciation rates and higher extinction rates than their diploid congeners, thereby leading to markedly lower polyploid net diversification rates. In stark contrast to plants, in none of the data sets analysed here did we find support for lower polyploid diversification than the reverse (Table 3). This may seem counterintuitive considering that polyploidy is widespread in plants and relatively uncommon in fish. Below, we discuss potential explanations for and caveats about this difference.

First, taxonomic biases may influence our results. While polyploidy has been investigated for decades in plants, this phenomenon has generally been much less appreciated by fish taxonomists. In particular, ploidy level is a central character used in the plant systematic literature, but it has only recently been incorporated in fish systematics (Mable *et al.*, 2011), perhaps due to the difficulty in identifying cryptic polyploids. Consequently, the frequency of polyploid fish may well be underestimated with nontrivial implications to downstream diversification analyses. In particular, polyploidization events are more likely to be recognized when they affect many animal species (i.e. at internal nodes leading to diverse groups), as events along terminal branches would go unnoticed until the ploidy level of the resulting species were measured. Such a bias would inflate the estimated diversification rate associated with polyploidy. Furthermore, in groups where polyploidy is known to occur, the frequency of polyploid fish may be overestimated, if phylogenetic and cytological studies tend to focus on those subclades exhibiting polyploidy, which may be the case with the Cyprininae.

Second, low or sparse taxonomic sampling may limit the power of our analyses, thus rendering our results tentative until sufficient data are collected. In the cases of Acipenseridae and Botiidae, while the clades are well sampled (96% and 74%, respectively), the overall species richness in these clades is rather small. Therefore, it is possible that there is insufficient power to draw robust conclusions from the BiSSE analyses for Acipenseridae and Botiidae even as more data accumulate. For the large Cyprininae group, however, the current number of species with sequence data is rather low (below 25%, assuming that the group contains over 1300 species). It is also important to note that while alternative Bayesian phylogenies were considered in the diversification analyses presented here, the taxonomy of the Cyprininae clade is unsettled and its phylogeny largely unresolved, adding considerate noise to the current data set. Thus, while our current results demonstrate higher diversification in polyploid Cyprininae fish, they should be interpreted cautiously pending future reanalyses with a more complete data set (in terms of both the number of species with available sequence data as well as the number of loci sampled for each species).

Third, polyploidy may be associated with different suites of characters in plants and animals, and it might be these associated characters that drive differences in diversification rates. In particular, polyploidy is associated with self-fertilization in plants (Barringer, 2007; Robertson *et al.*, 2010), perhaps as a preadaptation or an evolved response to reduce minority cytotype disadvantage (Levin, 1975). In fish, however, self-fertilization is rare (Alves *et al.*, 2001). While theory predicts that self-compatibility would increase the establishment success of polyploids (Rausch & Morgan, 2005), on a longer time scale having reduced levels of genetic mixing may make such lineages prone to extinction (Goldberg *et al.*, 2010). Interestingly, in animals, newly arisen polyploid taxa may avoid minority cytotype disadvantage through phenotypic shifts that cause polyploids to assortatively mate with other polyploids. For example, mating calls are altered in polyploid anurans (Keller & Gerhardt, 2001), and a similar mechanism has been hypothesized to operate in fish (reviewed in Mable *et al.*, 2011), thereby increasing the probability that polyploids establish without the long-term detrimental effects of self-fertilization.

Fourth, in plants, the high frequency of heteroploid speciation (i.e. speciation events involving a shift in ploidy) from the diploid state (estimated to be as high as 32% in plants; Mayrose *et al.*, 2011) may be a major component of the elevated speciation rates of diploids compared to neopolyploids, which may have a lower capacity to further speciate via polyploidy. In fish, heteroploid speciation appears to be less frequent – from 5% in Cyprininae to 21% in Acipenseridae (estimated using an extension of the BiSSE model described in Magnuson-Ford & Otto, 2012; see Data S1 and Table S4), thereby contributing less to the diploid speciation rate. Thus, the high polyploid abundance in plants may, in fact, be driven by the elevated speciation rates

of diploids generating polyploids. In addition, it seems that in fish, the rate of heteroploid speciation is not strikingly different between diploids and polyploids (e.g. in Acipenseridae, we inferred four heteroploid speciation events, with these occurring at 3 of the 14 nodes where the ancestral lineage was estimated to be diploid and one of the nine nodes where the ancestral lineage was estimated to be polyploid; Fig. S1a).

Fifth, it may be the case that polyploidization is particularly advantageous in lineages that have not undergone previous rounds of polyploidization. It is established that angiosperms have an extensive history of ancient polyploidy (Soltis *et al.*, 2009; Van de Peer *et al.*, 2009; Jiao *et al.*, 2011) and that polyploid formation in plants is a common and ongoing phenomenon. Fish, on the other hand, have undergone rather few ancient polyploidization events [two events early in vertebrate evolution (Dehal & Boore, 2005), and another one preceding the radiation of the ray-finned fish (Taylor *et al.*, 2003)], and polyploids are rarely reported except in certain lineages, such as Cyprininae. It is thus possible that the advantageous effects of polyploidy, such as increased genetic degrees of freedom, diminish with the number of prior rounds of polyploidization and might even be absent in many groups of plants where polyploidy has been particularly rampant.

Finally, it is important to consider the time frame of the analyses. The analyses in Mayrose *et al.* (2011) focused on a set of 63 genus-level plant groups. However, the currently available resolution of fish systematics has only allowed us to investigate a heterogeneous set of taxonomic ranks, all above the genus level. As stated by Levin (1983), 'chromosome doubling may propel a population into a new adaptive sphere, and render it capable of occupying habitats beyond the limits of its diploid progenitor. For this to occur, however, polyploids must survive long enough for chromosome doubling to influence subsequent evolution'. It is possible that the extra degree of genetic freedom of polyploids, provided by the additional paralogous gene set, is advantageous only over longer periods of evolutionary time. Similarly, the costs of polyploidy may be most acute over shorter time scales, such as reduced genetic variation when only one or a few individuals undergo polyploidization, a lack of prior adaptation to the phenotypic and ecological shifts induced by polyploidization, as well as minority cytotype disadvantage. It is thus possible that polyploidy does not affect diversification rates differently between animals and plants, but rather that older polyploid lineages enjoy an evolutionary success that the younger polyploid lineages do not. As we obtain richer data sets, with phylogenies spanning multiple taxonomic levels and more complete ploidy information, such possibilities can be explicitly examined. Nonetheless, our current investigation raises the possibility that polyploidy has had different evolutionary repercussions in fish and plants.

Acknowledgments

This study was supported by an Israel Science Foundation grant 1265/12 to IM.

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control,* **19**: 716–723.

Allendorf, F.W. & Thorgaard, G.H. 1984. Tetraploidy and the evolution of salmonid fishes. In: *Evolutionary Genetics of Fishes* (B.J. Turner, ed.), pp. 1–53. Plenum Press, New York, NY.

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

Barringer, B.C. 2007. Polyploidy and self-fertilization in flowering plants. *Am. J. Bot.* **94**: 1527–1533.

Birstein, V.J. 1993. Sturgeons and paddlefishes: threatened fishes in need of conservation. *Conserv. Biol.* **7**: 773–787.

Birstein, V.J., Hanner, R. & DeSalle, R. 1997. Phylogeny of the Acipenseriformes: cytogenetic and molecular approaches. *Environ. Biol. Fishes* **48**: 127–155.

Birstein, V.J., Doukakis, P., DeSalle, R. & McEachran, J.D. 2002. Molecular phylogeny of Acipenseridae: nonmonophyly of Scaphirhynchinae. *Copeia* **2002**: 287–301.

Bowers, J.E., Chapman, B.A., Rong, J. & Paterson, A.H. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438.

Chapman, B.A., Bowers, J.E., Feltus, F.A. & Paterson, A.H. 2006. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. *Proc. Natl. Acad. Sci. USA* **103**: 2730–2735.

Comai, L. 2005. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**: 836–846.

Crespi, B.J. & Fulton, N.J. 2004. Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. *Mol. Phylogenet. Evol.* **31**: 658–679.

Cui, L., Wall, P.K., Leebens-Mack, J., Lindsay, B.G., Soltis, D.E., Doyle, J.J. *et al.* 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* **16**: 738–749.

Dehal, P. & Boore, J.L. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**: e314.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.

Eschmeyer, W.N. & Fong, J.D. 2012. Species of Fishes by family/subfamily version (03/2012). http://research.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp.

Fawcett, J. & Van de Peer, Y. 2010. Angiosperm polyploids and their road to evolutionary success. *Trends Evol. Biol.* **2**: 16–21.

Ferris, S.D. 1984. Tetraploidy and the evolution of the catostomid fishes. In: *Evolutionary Genetics of Fishes* (B.J. Turner, ed.), pp. 55–93. Plenum Press, New York, NY.

Fisher, R.A. 1935. The sheltering of lethals. *Am. Nat.* **69**: 446–455.

FitzJohn, R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3**: 1084–1092.

FitzJohn, R.G., Maddison, W.P. & Otto, S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* **58**: 595–611.

Froese, R. & Pauly, D. 2012. FishBase version (02/2012). www.fishbase.org.

Goldberg, E.E., Kohn, J.R., Lande, R., Robertson, K.A., Smith, S.A. & Igic, B. 2010. Species selection maintains self-incompatibility. *Science* **330**: 493–495.

Gregory, T.R. 2012. Animal Genome Size Database (release 2.0). http://www.genomesize.com.

Haldane, J.B.S. 1933. *The Causes of Evolution*. Longwood Green, London.

He, D. & Chen, Y. 2006. Biogeography and molecular phylogeny of the genus *Schizothorax* (Teleostei: Cyprinidae) in China inferred from cytochrome *b* sequences. *J. Biogeogr.* **33**: 1448–1460.

Hoegg, S., Brinkmann, H., Taylor, J.S. & Meyer, A. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* **59**: 190–203.

Huelsenbeck, J.P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Ishiguro, N.B., Miya, M. & Nishida, M. 2003. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". *Mol. Phylogenet. Evol.* **27**: 476–488.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E. *et al.* 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**: 946–957.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E. *et al.* 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97–100.

Johnson, K.R., Wright, J.E.J. & May, B. 1987. Linkage relationships reflecting ancestral tetraploidy in salmonid fish. *Genetics* **116**: 579–591.

Keller, M.J. & Gerhardt, H.C. 2001. Polyploidy alters advertisement call structure in gray treefrogs. *Proc. Biol. Sci.* **268**: 341–345.

Kellis, M., Birren, B.W. & Lander, E.S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617–624.

Kottelat, M. 2004. *Botia kubotai*, a new species of loach (Teleostei: Cobitidae from the Ataran River basin (Myanmar), with comments on botiine nomenclature and diagnoses of two new genera. *Zootaxa* **401**: 1–18.

Krieger, J., Hett, A.K., Fuerst, P.A., Artyukhin, E. & Ludwig, A. 2008. The molecular phylogeny of the order Acipenseriformes revisited. *J. Appl. Ichthyol.* **24**: 36–45.

Le Comber, S.C. & Smith, C. 2004. Polyploidy in fishes: patterns and processes. *Biol. J. Linn. Soc.* **82**: 431–442.

Leggatt, R.A. & Iwama, G.K. 2003. Occurrence of polyploidy in the fishes. *Rev. Fish. Biol. Fish.* **13**: 237–246.

Levin, D.A. 1975. Minority cytotype exclusion in local plant populations. *Taxon* **24**: 35–43.

Levin, D.A. 1983. Polyploidy and novelty in flowering plants. *Am. Nat.* **122**: 1–25.

Levin, B.A., Freyhof, J., Lajbner, Z., Perea, S., Abdoli, A., Gaffaroğlu, M. *et al.* 2012. Phylogenetic relationships of the algae scraping cyprinid genus *Capoeta* (Teleostei: Cyprinidae). *Mol. Phylogenet. Evol.* **62**: 542–549.

Li, C., Lu, G. & Ortí, G. 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Syst. Biol.* **57**: 519–539.

López, J.A., Chen, W.-J. & Ortí, G. 2004. Esociform phylogeny. *Copeia* **2004**: 449–464.

Ludwig, A. 2008. Identification of Acipenseriformes species in trade. *J. Appl. Ichthyol.* **24**: 2–19.

Ludwig, A., Belfiore, N.M. & Pitra, C. 2001. Genome duplication events and functional reduction of ploidy levels in sturgeons (*Acipenser*, *Huso*, and *Scaphirhynchus*). *Genetics* **158**: 1203–1215.

Mable, B.K., Alexandrou, M.A. & Taylor, M.I. 2011. Genome duplication in amphibians and fish: an extended synthesis. *J. Zool.* **284**: 151–182.

Machordom, A. & Doadrio, I. 2001a. Evolutionary history and speciation modes in the cyprinid genus *Barbus*. *Proc. Roy. Soc. Lond. B Biol. Sci.* **268**: 1297–1306.

Machordom, A. & Doadrio, I. 2001b. Evidence of a Cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (*Luciobarbus*, Cyprinidae). *Mol. Phylogenet. Evol.* **18**: 252–263.

Maddison, W.P., Midford, P.E. & Otto, S.P. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* **56**: 701–710.

Magnuson-Ford, K.S. & Otto, S.P. 2012. Linking the investigations of character evolution and species diversification. *Am. Nat.* **180**: 225–245.

Mank, J.E. & Avise, J.C. 2006. Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. *Genetica* **127**: 321–327.

Marková, S., Šanda, R., Crivelli, A., Shumka, S., Wilson, I.F., Vukić, J. *et al.* 2010. Nuclear and mitochondrial DNA sequence data reveal the evolutionary history of *Barbus* (Cyprinidae) in the ancient lake systems of the Balkans. *Mol. Phylogenet. Evol.* **55**: 488–500.

Mayrose, I., Barker, M.S. & Otto, S.P. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol.* **59**: 132–144.

Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H. *et al.* 2011. Recently-formed polyploid plants diversify at lower rates. *Science* **333**: 1257.

Meyers, L.A. & Levin, D.A. 2006. On the abundance of polyploids in flowering plants. *Evolution* **60**: 1198–1206.

Orr, H.A. 1990. 'Why polyploidy is rarer in animals than in plants' revisited. *Am. Nat.* **6**: 759–770.

Otto, S.P. 2007. The evolutionary consequences of polyploidy. *Cell* **131**: 452–462.

Otto, S.P. & Whitton, J. 2000. Polyploidy: incidence and evolution. *Annu. Rev. Genet.* **34**: 401–437.

Peng, Z., Ludwig, A., Wang, D., Diogo, R., Wei, Q. & He, S. 2007. Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes). *Mol. Phylogenet. Evol.* **42**: 854–862.

Phillips, R.B. & Ráb, P. 2001. Chromosome evolution in the Salmonidae (Pisces): an update. *Biol. Rev.* **76**: 1–15.

Posada, D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**: 1253–1256.

Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J. & Alfaro, M.E. 2013a. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* **4**: 1958.

Rabosky, D.L., Santini, F., Eastman, J., Smith, S.A., Sidlauskas, B., Chang, J. & Alfaro, M.E. 2013b. Data from: Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Dryad Digital Repository*. doi:10.5061/dryad.j4802

Ramsey, J. & Schemske, D.W. 2002. Neopolyploidy in flowering plants. *Annu. Rev. Ecol. Syst.* **33**: 589–639.

Rausch, J.H. & Morgan, M.T. 2005. The effect of self-fertilization, inbreeding depression, and population size on autopolyploid establishment. *Evolution* **59**: 1867–1875.
Robertson, K., Goldberg, E.E. & Igic, B. 2010. Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae. *Evolution* **94**: 1527–1533.
Santini, F.S., Harmon, L.J., Carnevale, G. & Alfaro, M.E. 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evol. Biol.* **9**: 194.
Semon, M. & Wolfe, K.H. 2007. Consequences of genome duplication. *Curr. Opin. Genet. Dev.* **17**: 505–512.
Šlecthtová, V., Bohlen, J., Freyhof, J. & Ráb, P. 2006. Molecular phylogeny of the Southeast Asian freshwater fish family Botiidae (Teleostei: Cobitoidea) and the origin of polyploidy in their evolution. *Mol. Phylogenet. Evol.* **39**: 529–541.
Smith, S.A. & Dunn, C.W. 2008. Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. *Bioinformatics* **24**: 715–716.
Soltis, P.S. & Soltis, D.E. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl. Acad. Sci. USA* **97**: 7050–7057.
Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C. *et al.* 2009. Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**: 336–348.
Stebbins, G.L. 1971. *Chromosomal Evolution in Higher Plants*. Edward Arnold, London.
Taylor, J.S., Van de Peer, Y. & Meyer, A. 2001. Genome duplication, divergent resolution and speciation. *Trends Genet.* **17**: 299–301.
Taylor, J.S., Braasch, I., Frickey, T., Meyer, A. & Van de Peer, Y. 2003. Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* **13**: 382–390.
Thorne, J.L., Kishino, H. & Painter, I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**: 1647–1657.
Tsigenopoulos, C.S., Ráb, P., Naran, D. & Berrebi, P. 2002. Multiple origins of polyploidy in the phylogeny of southern African barbs (Cyprinidae) as inferred from mtDNA markers. *Heredity* **88**: 466–473.
Tsigenopoulos, C.S., Durand, J.D., Unlü, E. & Berrebi, P. 2003. Rapid radiation of the Mediterranean *Luciobarbus* species (Cyprinidae) after the Messinian salinity crisis of the Mediterranean Sea, inferred from mitochondrial phylogenetic analysis. *Biol. J. Linn. Soc.* **80**: 207–222.
Tsigenopoulos, C.S., Kasapidis, P. & Berrebi, P. 2010. Phylogenetic relationships of hexaploid large-sized barbs (genus *Labeobarbus*, Cyprinidae) based on mtDNA data. *Mol. Phylogenet. Evol.* **56**: 851–856.
Uyeno, T. & Smith, G.R. 1972. Tetraploid origin of the karyotype of catostomid fishes. *Science* **175**: 644–646.
Van de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L. & Vandepoele, K. 2009. The flowering world: a tale of duplications. *Trends Plant Sci.* **14**: 680–688.
Vasil'ev, V.P., Vasil'eva, E.D., Shedko, S.V. & Novomodny, G.V. 2009. Ploidy levels in the Kaluga, *Huso dauricus* and Sakhalin sturgeon *Acipenser mikadoi* (Acipenseridae, Pisces). *Dokl. Akad. Nauk SSSR* **426**: 275–278.
Werth, C.R. & Windham, M.D. 1991. A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Nat.* **137**: 515–526.
Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B. & Rieseberg, L.H. 2009. The frequency of polyploid speciation in vascular plants. *Proc. Natl Acad. Sci. USA* **106**: 13875–13879.
Wright, S. 1969. *Evolution and Genetics of Populations. II. The Theory of Gene Frequencies*. University of Chicago Press, Chicago, IL.
Xiao, H., Chen, S.Y., Liu, Z.M., Zhang, R.D., Li, W.X., Zan, R.G. *et al.* 2005. Molecular phylogeny of *Sinocyclocheilus* (Cypriniformes: Cyprinidae) inferred from mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **36**: 67–77.
Yang, L., Mayden, R.L., Sado, T., He, S., Saitoh, K. & Miya, M. 2010. Molecular phylogeny of the fishes traditionally referred to *Cyprinini sensu stricto* (Teleostei: Cypriniformes). *Zool. Scr.* **39**: 527–550.
Yasuike, M., Jantzen, S., Cooper, G.A., Leder, E., Davidson, W.S. & Koop, B.F. 2010. Grayling (Thymallinae) phylogeny within salmonids: complete mitochondrial DNA sequences of *Thymallus arcticus* and *Thymallus thymallus*. *J. Fish Biol.* **76**: 395–400.
Zhang, S.M., Yan, Y., Deng, H., Wang, D.Q., Wei, Q.W. & Wu, Q.J. 1999. Genome size, ploidy characters of several species of sturgeons and paddlefishes with comment on cellular evolution of Acipenseriformes. *Acta Zool. Sin.* **45**: 200–206.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Fifty-percent majority-rule consensus Bayesian phylogenies for the fish groups examined in this study: (a) Acipenseridae; (b) Botiidae; (c) Cyprininae; and (d) Salmoniformes and Esociformes.

Table S1 Ploidy level estimates for the Acipenseridae dataset.

Table S2 GenBank accession numbers, chromosome numbers, and ChromEvol results for the Cyprininae dataset.

Table S3 GenBank accession numbers for the Salmoniformes/Esociformes dataset.

Table S4 MCMC-based estimates of evolutionary rates under the BiSSE-ness model.

Table S5 ML parameter estimates of the best supported BiSSE model.

Table S6 Best supported BiSSE model according to AIC (allowing for polyploid-to-diploid reversals).

Table S7 ML parameter estimates of the best supported BiSSE model (allowing for polyploidto-diploid reversals).

Table S8 MCMC BiSSE parameter estimates (allowing for polyploid-to-diploid reversals).

Table S9 MCMC BiSSE analysis conducted on time-calibrated phylogenies.

Table S10 MCMC BiSSE analysis for Cyprininae using a threshold of 0.9 to assign ploidy levels.

Data S1 Methods.

Data deposited at Dryad: doi:10.5061/dryad.6h5v1

Received 12 September 2013; revised 18 November 2013; accepted 24 November 2013