Hybrid asexuality as a primary reproductive barrier: on the interconnection between asexuality and speciation

Karel Janko[1,2], Jan Pačes[1,3], Hilde Wilkinson-Herbots[4], Rui J Costa[4], Jan Roslein[1,5], Pavel Drozd[2], Nataliia Iakovenko[2], Jakub Rídl[3], Jan Kočí[1,2], Radka Reifová[6], Věra Šlechtová[1], Lukáš Choleva[1]

1, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Laboratory of fish genetics, Rumburska 89, 27721 Libechov,Czech Republic

2, Faculty of Science, University of Ostrava, Department of Biology and Ecology, Chitussiho 10, 71000 Ostrava, Czech Republic

3, Institute of Molecular Genetics, Czech Academy of Sciences, Laboratory of Genomics and Bioinformatics, Víděňská 1083, 14220 Prague 4, Czech Republic

4, University College London, Department of Statistical Science, WC1E 6BT, London, United Kingdom

5, Institute of Vertebrate Biology, Czech Academy of Sciences, Kvetna 8, 603 65 Brno, Czech Republic

6, Faculty of Science, Charles University in Prague, Department of Zoology, Vinicna 7, 12844 Prague 2, Czech Republic

Abstract

Speciation usually proceeds in a continuum from intensively hybridizing populations until the formation of irreversibly isolated species. Restriction of interspecific gene flow may often be achieved by gradual accumulation of postzygotic incompatibilities with hybrid infertility typically evolving more rapidly than inviability. A reconstructed history of speciation in European loaches (*Cobitis*) reveals that accumulation of postzygotic reproductive incompatibilities may take an alternative, in the literature largely neglected, pathway through initiation of hybrids' asexuality rather than through a decrease in hybrids' fitness. Combined evidence show that contemporary *Cobitis* species readily hybridize in hybrid zones, but their gene pools are isolated as hybridization produces infertile males and fertile but clonally reproducing females that cannot mediate introgressions. Nevertheless, coalescent analyses indicated intensive historical gene flow during earlier stages of *Cobitis* diversification, suggesting that non-clonal hybrids must have existed in the past. Revealed patterns imply that during the initial stages of speciation, hybridization between little diverged species produced recombinant hybrids mediating gene flow, but growing divergence among species caused disrupted meiosis in hybrids resulting in their clonality, which acts as a barrier to gene flow.

Comparative analysis of published data on other fish hybrids corroborated the generality of our findings; the species pairs producing asexual hybrids were more genetically diverged than those pairs producing fertile sexual hybrids but less diverged than species pairs producing infertile hybrids. Hybrid asexuality therefore appears to evolve at lower divergence than other types of postzygotic barriers and might thus represent a primary reproductive barrier in many taxa.

Introduction

Speciation may be accomplished by various mechanisms that restrict interspecific gene flow. These are typically categorized into pre- and post-zygotic reproductive isolation mechanisms (hereafter referred to as preRIMs and postRIMs) and could be driven by a few divergent loci of large effect as well as by many loci with putatively small and additive effects (Parchman et al. 2013), (rev. in (Seehausen et al. 2014)). Since species formation can rarely be studied in real time, speciation research has often taken a comparative approach by surveying a number of species pairs that putatively follow similar evolutionary trajectories from weakly isolated populations to irreversibly isolated species, but have progressed to different distances toward speciation. Although the rate at which nascent species proceed along the speciation continuum is probably non-linear (Orr and Turelli 2001; Bolnick and Near 2005; Matute et al. 2010) and taxon-dependent (Edmands 2002), a pervasive observation is that hybridization capability decreases as one moves from closely to distantly related pairs of taxa (Rykena 2002; Russell 2003; Sánchez-Guillén et al. 2014) – a pattern also referred to as the 'speciation clock' or 'incompatibility clock' (Bolnick and Near 2005). Comparative studies showed that hybrid infertility tends to evolve sooner than hybrid inviability (Price and Bouvier 2002; Russell 2003); species pairs producing viable and fertile hybrids generally exhibit lower genetic distances among themselves than those that produce infertile or partly infertile hybrids, and generally the highest divergences occur between species pairs producing inviable hybrids. The frequently observed asymmetry of hybrid incompatibilities has been put into the context of the Dobzhansky-Muller model, assuming the stochastic accumulation of many incompatibilities with small effects (Welch 2004).

Hybridization is not only integral to speciation. It has been known for many decades that asexuality also appears to be tightly associated with the loss of sexual reproduction in many hybrid forms (Ernst 1918; Choleva et al. 2012). However, the causal effects of hybridization on the production of clonal gametes are very poorly understood and the situation is further complicated by the variety of distinct cytological mechanisms in which asexual reproduction may be achieved (Stenberg and Saura 2009). Some explanations assume that the coexistence of distinct genomes within progenitor egg cells may deregulate the pathways underlying meiosis (Schultz 1973; Avise 2008) or cause heterochrony in the expression of genes involved in sexual reproduction (Carman 1997). Others assume the existence of some genetic peculiarities inherent to only certain sexual species or populations, which makes them predisposed to produce hybrid clones - the “phylogenetic constraint hypothesis”, (e.g. Hotz et al. 1985; Murphy et al. 2000). Yet, some theories even suggest that instead of directly triggering asexuality, hybridization and asexuality may arise independently but confer some evolutionary advantage when combined together (rev. in (Kearney et al. 2009)).

Dating back to (Ernst 1918), a set of theories invoked a generalizing framework for understanding the link between hybridization and asexuality, suggesting that the initiation and type of hybrid asexuality depend on the genetic distance between hybridizing species. Based on the observation that the proportion of unreduced gametes is higher in hybrids between distantly rather than closely related species, Moritz et al. (1989) formulated the “balance hypothesis” stating that parthenogenesis can arise only when the genomes of parental species are divergent enough to disrupt meiosis in hybrids, yet not divergent enough to seriously compromise hybrid viability or fertility. This theory is congruent with the observations from some taxa that parental species known to produce asexual hybrids appear to be somewhat genetically distant rather than being each other’s closest relatives at the species level (Moritz, Densmore, et al. 1992; Moritz, Uzzell, et al. 1992; Moritz, Wright, et al. 1992; Jančúchová-Lásková et al. 2015).

In this paper, we elaborate on the neglected fact that hybrid initiation of asexuality and evolution of RIMs operate on conceptually similar backgrounds and we show that both processes are intermingled. Indeed, both the balance hypothesis and the speciation clock link the establishment of clonal reproduction or hybrid incompatibilities, respectively, to the genetic distance between hybridizing species. From this perspective, the establishment of hybrid asexuality may be considered as a special case of the accumulation of Dobzhansky-Muller incompatibilities that disrupt critical processes - sexual reproduction in this case. In this respect, it is noteworthy that hybrid asexuality is commonly observed only in one direction of cross (Wirtz 1999; Janko et al. 2003), which is similar to that observed with hybrid infertility and inviability. As clonal hybrids are unlikely to mediate interspecific gene flow (e.g. Keller et al. (2007; but see Mikulíček et al. (2014) for the rare counter-example), hybrid asexuality can generally be considered as an effective barrier between hybridizing species even if other forms of preRIMs or postRIMs are absent.

We examine a) whether the formation of asexual hybrids does depend on genetic distance between parental species; b) whether the establishment of reproductive barriers can be primarily accomplished by the formation of hybrid asexuality and c) the extent to which the processes of speciation and asexuality are interconnected. We focus on the model example of European spined loaches (*Cobitis*). Several phylogenetic lineages of these bottom-dwelling fish colonized Europe during the Tertiary period. Mediterranean Europe has many isolated rivers that are inhabited by diverse *Cobitis* species clustering in deeply divergent phylogenetic lineages (*Cobitis* Lineage I – V *sensu* (Bohlen et al. 2006)). However, vast areas of non-Mediterranean Europe were colonized by a single lineage (*Cobitis* Lineage V) that mainly comprises three species (*C. elongatoides*, *C. tanaitica* and *C. taenia*). These species display a parapatric distribution, with overlapping zones located in Central Europe, the lower Danube, and Southern Ukraine (Fig 1). According to nuclear DNA markers, the Danubian *C. elongatoides* is a sister lineage to both other broadly distributed *C. taenia* and *C. tanaitica* (Choleva et al. 2014). Two other species (*C. taurica* and *C. pontica*) have recently been described from terminal stretches of isolated inflows of the Black Sea, which also appear to be very closely related to *C. taenia* and *C. tanaitica* in nuclear genetic markers (Janko, Flajšhans, et al. 2007). Interestingly, coalescent analysis of combined nuclear and mitochondrial DNA (mtDNA) loci (Choleva et al. 2014), showed that *C. tanaitica*, in spite of being closely related to *C. taenia* in the nuclear markers studied, was subjected to intensive gene flow with *C. elongatoides* in the mitochondrion, leading to the fixation of an introgressed *C. elongatoides*-like mitochondrion over the entire *C. tanaitica* distribution range. *C. tanaitica* thus represents one of the most massive mtDNA introgressions in the animal kingdom (Choleva et al. 2014). Such a massive mtDNA introgression is surprising as *Cobitis* hybrids reported in previous studies appeared either as infertile hybrid males or as clonally reproducing hybrid females (Janko, Bohlen, et al. 2007; Janko, Flajšhans, et al. 2007; Choleva et al. 2012). Although some of these hybrid lineages achieved considerable evolutionary ages (i.e. the so called Hybrid Clades I and II are as ancient as 0.25 and 0.35 Mya, respectively (Janko et al. 2005; 2012)), their clonality theoretically prevents any interspecific gene flow. Understanding the mechanisms of introgressive hybridization between *Cobitis* species and their change over time may therefore shed light on the interconnection between the formation of RIMs and the formation of asexual hybrids.

We addressed this question with three complementary approaches. First, we performed a population genetic analysis of the *C. elongatoides* – *C. tanaitica* hybrid zone to test if there is any ongoing introgressive hybridization. Second, we analyzed the reproductive modes in hybrids between several differently related *Cobitis* species to reveal the extent of the currently observed reproductive isolation. Third, we employed phylogenomic and coalescent analyses of transcriptomic data to estimate levels and timing of historical gene flow among species. The combined evidence of these three approaches implied that the capability of introgressive hybridization diminishes with growing divergence between species. It appeared, however, that the restriction in gene flow has not necessarily been accomplished by classical pre- or post-RIMs but rather by the production of asexual hybrids, which might constitute an important component of reproductive isolation. The generality of our findings was corroborated by amending Russell's (2003) comparative study (see above), including genetic distances among pairs of fish species producing clonal hybrids.

Results

We performed three independent tests to determine whether contemporary species currently hybridize or have been interconnected by historical gene flow.

Analysis of C. elongatoides – C. tanaitica hybrid zone and test of ongoing introgressive hybridization

1. Firstly, we tested for ongoing gene flow between species by performing population genetic analysis of *C. elongatoides* – *C. tanaitica* hybrid zone similar to *C. elongatoides* – *C. taenia* analysed in (Janko et al. 2012). 818 *Cobitis* specimens were captured and identified on 26 localities all over the lower Danubian River Basin. Allozyme analysis of diagnostic loci revealed the presence of *C. elongatoides* at 11 sites, *C. tanaitica* at seven sites and *C. strumicae* at 10 localities (Fig 1 and Table 1). Apart from sexual species, we found various hybrid biotypes, which were mostly polyploid but we also encountered 32 diploid hybrids heterozygotic for species-specific alleles at all diagnostic allozyme loci (Table 1).

Microsatellite analysis was successfully performed on 105 diploid individuals. Most diploids had unique combinations of alleles (Table S1), but 28 diploids clustered in eight groups of identical multilocus genotypes (MLG A – MLG H). Such groups of identical individuals were considered as clone mates because there was negligible probability of the same genotype being shared by two or more individuals arising from independent sexual events ($p < 10^{-5}$). Furthermore, the genetic distances among two groups of individuals or MLGs were significantly lower than would be expected from independent sexual events ($p < 0.01$). As in (Janko et al. 2012), both such groups were assigned as multilocus lineages (MLL) suggesting that altogether, 29 diploid individuals could be assumed to form eight clonal lineages. Allozymes indicated a hybrid state of all such clonal individuals. No mtDNA haplotypes were shared by *C. elongatoides* and *C. tanaitica* (Fig 2). One clone (MLG B) possessed a haplotype E1, which was shared with *C. elongatoides,* while the remaining individuals assigned as hybrids possessed haplotypes clustering in the old “Hybrid clade I” defined in (Janko et al. 2005).

Analysis of the combined microsatellite and allozyme data by the Structure provided the likelihood values that converged during the runs and results did not notably change between replicates. Regardless of the locality, the optimal number of clusters *K* = 2 was selected as the best-fit partitioning of the diploid dataset. Altogether, we found three types of diploid individuals. First, we identified those individuals, where the parameter *q* ranged between 0.993 and 0.999 (presumably *C. elongatoides* individuals). Second, we identified those individuals with *q* values between 0.002 and 0.011 (43 presumably *C. tanaitica* individuals). Finally, all 32 individuals identified as hybrids by allozymes had intermediate *q* values (0.554 - 0.783; Fig 3A).

Data analysis using the NewHybrids software was slightly sensitive to the type of prior applied to *Θ* but not to *π*. However, consistent with the Structure, all specimens presumed to be *C. elongatoides* and *C. tanaitica* were always assigned as the pure parental species (*p* > 0.95). Under Jeffrey's prior, the NewHybrids assigned all but one of the abovementioned hybrids into the F1 class with the probability exceeding 95 %; however, it could not decide in the case of MLL E between F1 (*p* = 0.53), F2 (*p* = 0.10), and B1 (*p* = 0.35) states (Fig 3B). Under the uniform prior, most hybrids were assigned into F1 class, but NewHybrids could not decide between F1 and B1 states of MLL H, MLL E and individual 09BG19K22 (Fig 3B and Table S1).

Analysis of reproductive modes of interspecific Cobitis hybrids

2. Second, we tested whether *Cobitis* hybrids can mediate interspecific gene flow by analyzing reproductive modes of natural and artificial hybrids between several differently related *Cobitis* species.

2.1. Reproductive modes of natural C. elongatoides-tanaitica hybrids: We successfully analyzed five backcrossed families stemming from four natural triploid EEN and one diploid EN hybrid females (letters E and N stand for haploid *elongatoides* and *tanaitica* genomes, respectively). Backcross progeny consistently expressed all maternal alleles, suggesting the production of unreduced gametes and lack of segregation. In one allele, single progeny differed from the maternal allele by a single repeat, indicating a mutation event. A part of the progeny also contained a haploid set of paternal alleles indicating that the sperm's genome is sometimes incorporated, leading to a ploidy increase (Table 2 and S2). We also compared allozyme profiles of eggs and somatic tissues of six EEN females, in order to test for hemiclonal reproduction (i.e., hybridogenesis). Hybridogenesis was supposed to lead to reduced, albeit non-segregating gametes (Cimino 1972; Uzzell et al. 1980; Carmona et al. 1997), which consistently express allozyme products of one parental species only. Nevertheless, we found that allozyme profiles of all eggs were identical to the somatic tissues of their maternal individuals, further suggesting that natural hybrids do not exclude any genome (Table S3).

2.2. Reproductive modes of artificial F1 hybrids: In addition to *C. elongatoides – C. taenia* crosses published in (Choleva et al. 2012), we newly obtained two *C. taenia – C. pontica* primary crosses. Family No. 1 (Tables 2 and S2) produced a number of F1 hybrids, of which one F1 x F1 couple successfully spawned and produced F2 progeny. Second reproducing family (No. 17; Tables 2 and S2) reached maturity in a number of four males and three females in a single aquarium. From this family we obtained two clutches of F2 progeny spawned by single F1 female, which mated with two different F1 males. Both hybrid sexes were viable and fertile in both families No. 1 and No. 17, as evidenced by successful production of F2 progeny. F2 progeny mostly possessed one allele from the mother and the other from the father. Such inheritance patterns suggest the sexual reproduction of *C. taenia-pontica* hybrids in both families. However, family No. 1 also contained different types of F2 progeny: 15 individuals contained the complete set of maternal alleles and 9 of them also possessed the haploid set of paternal alleles (Table 2). Such patterns indicate that *C. taenia-pontica* hybrid females produced not only recombinant sexual gametes but partly also unreduced gametes. The unreduced eggs either developed into a clonal progeny via gynogenesis or into triploid progeny after a true fertilization with a haploid sperm.

Estimation of levels and timing of historical gene flow among the species using transcriptome data

3. Thirdly, we estimated levels and timing of historical gene flow among the species by analyzing single nucleotide polymorphisms (SNP) variability of *C. elongatoides*, *C. taenia, C. tanaitica* and *C. pontica* transcriptomes. As an outgroup we used *C. strumicae* - a species belonging to subgenus *Bicanestrinia* that diverged from the ingroup between 12.4 – 17.6 Mya (Bohlen et al. 2006), which was sampled in isolated Black Sea tributaries outside of the distribution ranges of ingroup species in order to avoid any possible reproductive contact.

The assembly of mRNA from five *C. taenia* specimens comprised 20,385 contigs (potential mRNAs) and was used as a reference for mapping of reads from all species (two individuals of each ingroup species and one outgroup). In total, we identified 187,205 SNPs (Appendix S4), which were then analyzed by fitting eight coalescent models assuming different scenarios of species divergence and connectivity (Fig 4; (Wilkinson-Herbots 2008; Wilkinson-Herbots 2012; Wilkinson-Herbots 2015; Costa and Wilkinson-Herbots 2016), *related manuscript 1*). The results are described in Table 3. Because the available models only allow fitting pairs of taxa, we separately analyzed six datasets corresponding to all pairwise combinations of the four analyzed species. Most models were successfully fitted to all datasets but two models did not converge in the case of the *C. taenia* – *C. tanaitica* dataset.

Isolation-with-initial-migration model IIM7 (where the label '7' indicates the number of parameters of the model) assumes that gene flow occurred from the time t_0 when the ancestral species split until time t_1 when the descendant species became completely isolated from each other. This model fitted all pairwise species datasets significantly better than both other models of Wilkinson-Herbots (Wilkinson-Herbots 2008; Wilkinson-Herbots 2012), namely the isolation model I4 assuming no gene flow since the initial split of the species, and the isolation-with-migration model IM4 assuming ongoing hybridization from the initial split until the present (Likelihood Ratio Test; LRT $p << 10^{-10}$ for all comparisons involving *C. elongatoides*, and $p < 0.01$ for all other species comparisons). To test if the better fit of IIM7 is due to the correct assessment that interspecific gene flow ceased recently, or merely due to allowing an additional size change at time t_1, we employed five more models that allow an additional population size change and also relax the assumption of equal population sizes during the migration stage (Fig 4; (Costa and Wilkinson-Herbots 2016), *related manuscript 1*).

Pairwise datasets of the three closely related *C. taenia, C. tanaitica* and *C. pontica* were better fitted by the isolation-with-initial-migration models (IIM7, IIM8) compared to the I4 and I7 isolation models (LRT $p < 0.01$ for all comparisons). The generalized isolation-with-migration GIM9 model gave estimates of 0 for the current migration rate and hence reduced to the IIM8 model (where these models could be fitted). However, these models consistently suggested that the analyzed species were interconnected by a very intensive gene flow at a level close to one migrant sequence per generation or even more until a time t_1 of approximately 0.2 – 0.5 Mya. Such high levels of gene flow are considered close to panmixia (e.g. (Lowe and Allendorf 2010)) suggesting that those three taxa might have formed a single substructured species between times t_0 and t_1 and speciated only recently. To explore this possibility, we additionally fitted the isolation I6 model allowing one additional size change of the ancestral population before the isolation phase. The I6 model also estimated the speciation time at around 0.5 Mya and indeed provided a very good fit for all three datasets, being selected as the best model for two of these. Thus, coalescent analysis suggests that *C. taenia, C. tanaitica* and *C. pontica* are currently isolated and possibly speciated only recently.

The datasets of species pairs including *C. elongatoides* and any one of *C. taenia, C. tanaitica* or *C. pontica* were fitted better by the isolation-with-initial-migration model IIM8 than by isolation models I7 and I4 (LRT $p < 0.0002$ in all cases) and both isolation-with-migration models, IM4 and IM5 (LRT $p << 10^{-10}$ in all cases). The IIM8 model also fitted the data better that the IIM7 model (which assumes equal population sizes during the migration stage) in the case of *C. elongatoides* – *C. taenia* and *C. elongatoides* – *C.tanaitica* species pairs (lower AIC scores and LRT p < 0.023 and 0.001, respectively) but not for *C. elongatoides* – *C. pontica* species pairs where IIM7 had a better AIC score. The divergence time estimates were consistent across species comparisons and place the initial split of *C. elongatoides* from the other species at roughly 9 Mya. It is estimated that *C. elongatoides* exchanged genes with the other species at average rates of less than 0.1 migrants per generation until time t_1, for which ML estimates vary between 1.65 and 0.49 Mya depending on the data set and model but have generally quite large confidence intervals (Table 3). Given that the speciation times of the other three species are much more recent (see above) than their split from *C. elongatoides*, our results suggest that the detected gene flow occurred predominantly between *C. elongatoides* and the common ancestor of the other three species. The most complex GIM9 model did not improve the fit to the data compared to the IIM8 model (as indicated by the AIC scores of the models and by LRT*;* $p > 0.59$ for all cases). Moreover, the GIM9 model estimated that recent migration rates were much lower than historical ones, thus virtually converging to the IIM8 model and confirming that *C. elongatoides* has been historically exchanging genes with the other species but became isolated in recent times.

Comparative analysis of genetic divergence between hybridizing fish species and types of reproductive isolation including hybrid asexuality

4. Finally, we tested the generality of our findings, implying that the hybrids' asexuality may represent an intermediate stage of species diversification process. To do so, we investigated the general correlation between genetic divergence of hybridizing pairs of fish species and the dysfunction of their F1 fish hybrids. Dysfunction was measured in terms of infertility, inviability or asexual reproduction of hybrids (see Methods section for details how we amended the dataset from Russell's (2003) comparative study to incorporate asexuality). We found that the Kimura 2-parameter (K2P) corrected divergences of cytochrome *b* gene sequences from fish species pairs that produce asexual hybrids range from 0.051 to 0.166 (mean = 0.122; s.e. = 0.038). This appears as intermediate between those species pairs producing both sexes of hybrids fertile and viable (Russell's (2003) hybrid class 0; mean = 0.079; s.e. = 0.054) and those pairs that produce viable but infertile hybrids of both sexes (Russell's hybrid class 2; mean = 0.179; s.e. = 0.025) (Fig 5 and Appendix S5). Although some species of the *Hexagrammos* and *Poecilia* genera were involved in more types of hybridization producing asexuals, we treated each species only once and repeated the analysis several times to account for all combinations. The Shapiro-Wilk test did not reject normality in any of the hybrid classes tested ($p > 0.05$)and regardless of species pairs considered, parental divergences in the asexual hybrid class were always significantly lower than those of Russell's (2003) class 2 (Student's *t*-test, $p < 0.01$) and always significantly higher than that of Russell's (2003) hybrid class 0 (*t*-test, $p < 0.05$). Interestingly, the range of divergences among asexual hybrids is notably similar to the divergences of hybrids where the functionality of one sex is lower than that of the other (hybrid classes 0.5 - 1.5; mean = 0.118; s.e. = 0.040).

Discussion

Natural asexual organisms are considered as treasured evolutionary models because they could provide clues about the paradox of sex. Our study offers conceptually novel view on this phenomenon because we found that hybrid asexuality may form an inherent stage of the speciation process and can form an effective reproductive barrier, which arises earlier in speciation than hybrid sterility and inviability.

Accomplishment of speciation in spite of ongoing hybridization

Studied *Cobitis* species readily hybridise both in natural and experimental conditions (Janko, Flajšhans, et al. 2007; Choleva et al. 2012; Janko et al. 2012), suggesting the absence of strong preRIMs. However, two independent lines of evidence suggest that contemporary interspecific gene flow between *C. elongatoides* and the remaining *Cobitis* species is unlikely and that speciation has been completed. First, according to experimental crossings of *C. elongatoides-taenia* (Janko, Bohlen, et al. 2007; Choleva et al. 2012) and *C. elongatoides-tanaitica* (present study) females, hybrids do not produce reduced gametes through the ”standard” sexual process and do not produce recombinant progeny. Allozyme analyses of oocytes further argue against the production of reduced gametes through hybridogenesis; contrary to evidence of all known hybridogenetic vertebrates, which exclude one parental species' genome (Cimino 1972; Uzzell et al. 1980; Carmona et al. 1997), we found that oocytes of *Cobitis* hybrid females expressed allozyme alleles of both parental species. Unreduced gametes either develop clonally or occasionally incorporate the sperm's genome, leading to polyploidy, but are unlikely to enable the interspecific gene flow. Hybrid males also do not appear to mediate the gene flow since *C. elongatoides-taenia* hybrid males are infertile (Choleva et al. 2012) and *C. elongatoides-tanaitica* hybrid males have not been observed in nature.

Second, the completion of speciation is also supported by analyses of *C. elongatoides* – *C. taenia* and *C. elongatoides* – *C. tanaitica* hybrid zones. The range overlap between *C. elongatoides* and C. *tanaitica* covers most of the Danube watershed below the Iron Gates Gorge and water bodies in the Dobrudja region and is larger than the Central European *C. elongatoides* – *taenia* hybrid zone (Janko et al. 2012). The Lower Danubian hybrid zone also appears more complex, since some Danubian tributaries are inhabited by phylogenetically distant *C. strumicae* that according to Bohlen et al. (2006), probably invaded Danubian stretches during the recent postglacial expansion from the Mediterranean Basin. Nevertheless, both zones have similarities with respect to major evolutionary aspects. Polyploid asexuals dominate in both zones (in the Danube drainage, they also invaded rivers inhabited by *C. strumicae* (Choleva et al. 2008)), while diploids within both zones could consistently be assigned into two types. The first type was represented by one or the other pure sexual species (q parameter close to 0 or 1). The second type was represented by hybrids. However, unlike classical patterns where the distribution of the hybrids' q values follows a continuum, the q index for *Cobitis* was sharply unimodal around intermediate values and most hybrids were unambiguously assigned as F1 (NewHybrids software, $p > 0.95$), except three Danubian diploid lineages, where we could not reject the B1 state although the F1 state was still preferred. Interestingly, these three lineages belong to the old clonal Hybrid clade I and possess a number of private microsatellite alleles, suggesting that their assignment by both the Structure and NewHybrids software might have been affected by mutational divergence from contemporary sexual species. Problems in disentangling the backcrossed and ancient F1 hybrid origins were also debated in other asexual lineages (see e.g. (Barbiano et al. 2013)). Most diploid hybrids from both zones also clustered into lineages of (nearly) identical individuals that may be assigned as clones. In the Danube, the MLG B represents a recently arisen clone as it shares its mtDNA haplotype with *C. elongatoides* but all remaining diploid hybrids cluster in the so-called Hybrid clade I representing an ancient asexual lineage that arose from hybridization between *C. elongatoides* and *C. tanaitica* about 0.35 Mya (Janko et al. 2005). Contemporary hybrids of *C. elongatoides* thus do not appear able to produce recombinant progeny mediating interspecific gene flow.

In contrast, RNAseq data suggest that hybrids that were able to mediate gene flow must have existed in the past. Coalescent analyses clearly rejected isolation models suggesting that gene flow existed after the initial divergence of *C. elongatoides* from *C. taenia, C. pontica* and *C. tanaitica.* Interestingly, datasets including *C. elongatoides* were significantly better fitted by isolation-with-initial-migration models than by the simpler isolation-with-migration models, which implies that gene flow involving *C. elongatoides* occurred in the historical period after its initial divergence around 9 Mya but did not occur after t_1, suggesting the isolation of contemporary populations (Fig 4 and Table 3). The most complex GIM9 model gave a worse AIC score than the IIM8 model but congruently suggested a drastic drop in gene flow intensity after t_1 (e.g., the estimated migration rate between *C. elongatoides* and *C. tanaitica* changed from a historical value of 0.170 to a mere 0.012 after t_1; (Table 3)). As the IM,IIM and GIM models assumed gene flow since the initial split of *C. elongatoides* but compare only pairs of extant species, it is reasonable to conclude that the inferred gene flow occurred between *C. elongatoides* and the common *C. taenia – C. tanaitica – C. pontica* ancestor that diversified only recently around ca 0.5 Mya. Unfortunately, the large confidence intervals of the t_1 estimates prevent an unambiguous conclusion as to whether nuclear gene flow continued after the *C. taenia – C. tanaitica – C. pontica* speciation.

Conversely, we found no evidence of gene flow between the closely related *C. taenia, C. tanaitica* and *C. pontica*. These datasets were initially best fitted by the IIM7 model, which estimated a very intensive gene flow between t_0 and t_1, suggesting that those taxa in fact might have formed a single species possibly with some weak substructure, and truly speciated more recently at t_1 estimated at around 0.5 Mya for the split of *C. tanaitica* from the other species and around 0.23 Mya for the split of *C. taenia – C. pontica* (note that the t_1 estimates were in agreement among distinct models within the same dataset). Isolation-I6 model provided slightly older estimates of t_1 but improved the fit to the data for two of the species pairs, further suggesting that those species may have speciated only recently around t_1 without notable subsequent gene flow.

The present results are consistent with our previous analyses of nine nuclear and one mtDNA loci but two differences were noted. First, Choleva et al. (2014) found significant traces of mitochondrial but not nuclear *C. elongatoides – C. tanaitica* gene flow, which led to the somewhat paradoxical impression that the nucleus has not been affected by hybridization while *C. tanaitica*'s mitochondrion has suffered one of the most massive introgressions among animals (a complete mtDNA replacement). The current analysis of the extended dataset indicated gene flow in the nuclear compartment too, which is more plausible biologically. Second, the previous model-based inference indicated *C. taenia – C. tanaitica* gene flow in nucleus, which contrasted with the available field data (2014); although karyotype differences between both species make their hybrids easily recognizable (*C. taenia* has 2n = 48 chromosomes, while other species have 2n = 50), no *C. taenia –*

tanaitica hybrids have ever been observed in nature and the ranges of both species are closely adjacent but sympatry has not been documented (Janko, Flajšhans, et al. 2007). The present analysis is therefore more in line with current knowledge about *Cobitis* and the discrepancy with the previous analysis of few loci may potentially reflect the tendency of Bayesian IM algorithms to inflate the estimates of gene flow when the number of loci is low and splitting times are recent (Cruickshank and Hahn 2014; Hey et al. 2015).

Of course, any model-based inferences should be treated with caution. For example, having analysed only pairs of species, our models may be affected by intractable interactions with other species. IM models are robust to this type of violation, for low or moderate levels of gene flow (2010), and the inferred lack of hybridization between *C. taenia* – *C. tanaitica* – *C. pontica* implies that the pairwise comparison of any one of these species with *C. elongatoides* is unlikely to be affected by the existence of the other species. However, other problems may be important, such as uncertainty regarding the relative mutation rates of the different loci (which have been estimated by comparing them with the outgroup), additional population size changes, geographical structure or oscillating rather than constant intensity of gene flow between t_0 and t_1. Furthermore, coalescent models assume evolution under neutrality or negative selection, which effectively decreases the locus-specific mutation rate (Nadachowska and Babik 2009), but other types of selection might have affected our data.

It is therefore a great advantage that model inferences are supported by independent types of data; the existence of gene flow between *C. elongatoides* and other species is supported by fixation of an *C. elongatoides*-like mitochondrial lineage in *C. tanaitica* (Choleva et al. 2014). On the other hand, small but non-zero divergences between contemporary *C. elongatoides* and *C. tanaitica* mtDNA haplotypes, clonal reproduction of the studied hybrids, and the presumed lack of introgressive hybridization in hybrid zones imply that such gene flow ceased and does not occur at present.

Simultaneous evolution of asexuality and RIMs

Several complementary approaches showed that the initial divergence of *C. elongatoides* has been coupled with production of hybrids whose reproductive mode enabled more or less intensive gene flow with other species or their common ancestor. However, as these species diverged from each other, the introgressions became restricted since the major type of hybrids became asexual, which are considered effectively infertile from the point of view of gene flow. Such patterns conform to the "balance hypothesis" (Moritz et al. 1989), which predicts that hybridization between gradually diverging species would initially produce mostly sexual hybrids while successful asexuals would arise at intermediate stages when a hybrid's meiosis is disrupted but fertility is not yet significantly reduced. Simultaneously, this scenario is consistent with gradual decline in the species' capability of introgressive hybridization that is expected to evolve along the speciation continuum from weakly separated entities towards strongly isolated species (e.g., (Seehausen et al. 2014)). The establishment of asexual hybrids and the gradual accumulation of RIMs thus appear interconnected processes that correlate with divergence of hybridizing species.

Artificial crossing experiments of *Cobitis* taxa support this idea. Crossings of distantly related *C. elongatoides* and *C. taenia* produced only clonally reproducing F1 females and infertile F1 males (Choleva et al. 2012), while mating of the closely related *C. taenia* and *C. pontica* species (this study) resulted in fertile F1 progeny of both sexes that mostly produced recombinant progeny. Such data also conform to the empirical observation that one sex often acquires infertility earlier than the other one (Russell 2003; Bolnick and Near 2005). Unfortunately, lack of data on *Cobitis* sex chromosomes prevent speculations as to whether infertility of hybrid males could be related to Haldane's rule or other processes such as the increased sensitivity of spermatogenesis compared to oogenesis to perturbations in hybrids (Wu and Davis 1993). Interestingly, we noted that some *C. taenia-pontica* female hybrids reproduced clonally. This observation extends the list of asexual hybrids within the family Cobitidae to species pairs containing *taenia-pontica* (laboratory hybrids from this study), *elongatoides-taenia* and *elongatoides-tanaitica* genomes (Janko, Flajšhans, et al. 2007) as well as two cases from other Cobitidae genera (Zhang et al. 1998; Kim and Lee 2000). Such phylogenetically widespread and independent emergence of asexual hybrids upon combining various species pairs is not consistent with the "phylogenetic constraint hypothesis". Instead, such observations provide further support to the hypothesis that hybrid asexuality constitutes an intermediate step in the species diversification process.

Hybrid asexuality as an intermediate step in speciation continuum?

Our findings can be extrapolated to other species. We included known asexual hybrids into Russell's (2003) comparative study of fish hybrids and found that asexual hybrids appear at intermediate levels of parental divergences between species pairs producing fertile and viable hybrids (hybrid class 0) and those producing infertile hybrids of both sexes (hybrid classes 2) (Fig 5, Appendix S5). In fact, the divergence of asexual-producing species pairs (mean K2P distance 0.122 $^{\pm}$ 0.038) is similar to those producing hybrids with lowered fertility of hybrids of one sex (Russell's hybrid classes 0.5 - 1.5; K2P distance 0.118 $^{\pm}$ 0.040), which is in line with the observation that populations of asexual fish are often constituted by females only. While Russell (2003) corroborated that severity of a hybrid's dysfunction correlates with the divergence of the hybridizing fish, our amendment suggest that hybrids with a skewed reproductive mode towards asexuality tend to appear at an intermediate stage of the diversification process. The relatively wide range of divergence for the asexual hybrids also agrees with the expectation that accumulation of RIMs follows a variable-rate clock and that different types of incompatibilities accumulate in a noisy manner (Edmands 2002). Similarly in reptiles, divergences among hybridizing species that produce asexual hybrids are significantly higher than among those producing bisexual hybrids (Jančúchová-Lásková et al. 2015).

There also appears to be a continuum in the ability of species pairs to produce hybrid asexuals. At one end, there are dynamically hybridizing species pairs that produce diverse assemblages of asexual hybrids, e.g. *Cobitis* (Choleva et al. 2012), *Pelophylax* (Hotz et al. 1985), *Poeciliopsis* (Schultz 1973); whereas at the other end there are monoclonal hybrid lineages stemming from one or few ancient events while attempts to cross their contemporary sexual relatives do not produce clones, e.g., *Poecilia* (Stöck et al. 2010). *Phoxinus* represents an intermediate case, where historical hybridizations produced a highly diverse asexual assemblage but new clones can no longer be produced (Angers and Schlosser 2007). Similarly, some natural asexual complexes comprise only of asexual hybrids (e.g. *Cobitis*) while clonal and sexual hybrid females co-occur and fertile hybrid males may exist in others (e.g. *Fundulus, Rutilus rutilus x Abramis bramma*). Such a variety of patterns may reflect that different species proceeded different distances along the same continuum.

The interconnection of asexuality with speciation has been addressed by relatively few studies so far and present study showed that the interactions between asexuality and speciation are much deeper than previously believed. For example Dubois (2011) analyzed asexuals from the perspective of a species concept, Fontaneto et al. (2007) investigated their potential to establish species-like entities and some researchers suggested that asexual lineages may occasionally revert to sex and give rise to new sexual species, e.g. (Vrijenhoek 1998; Cunha et al. 2008). We found that the speciation continuum may contain an inherent, previously unnoticed, stage of asexuality. The production of asexual rather than sexual hybrids helps to establish an effective barrier to gene flow even in the absence of other typical forms of pre- and postRIMs. Asexuality is phylogenetically widespread but rare phenomenon. However, the generally short life span of asexual lineages (Butlin et al. 1999) and transiency of phase when diverging species produce asexual hybrids may imply that many currently reproductively incompatible species might have historically produced asexual hybrids that have gone extinct. In theory, the stage of hybrid asexuality might have been quite important in the speciation clocks of many groups capable of creating asexuals including arthropods, vertebrates or plants.

Materials and Methods

All experimental procedures involving fish were approved by the Institutional Animal Care and Use Committee of the Institute of Animal Physiology and Genetics AS CR (IAPG AS CR), according with directives from State Veterinary Administration of the Czech Republic, permit number 124/2009, and by the permit number CZ 00221 issued by Ministry of Agriculture of the Czech Republic. Crossing experiments were conducted under the supervision of L. Choleva, holder of the Certificate of competency according to §17 of the Czech Republic Act No. 246/1992 coll. on the Protection of Animals against Cruelty (Registration number CZ 02361), provided by Ministry of Agriculture of the Czech Republic, which authorizes animal experiments in the Czech Republic.

1. Material: Fig 1 and Table 1 indicate the origin of investigated fish specimens used in this study. A priori classification of captured specimens into taxonomic units (species and hybrid types) was based on previously verified diagnostic allozyme loci and sequencing of the first intron of the nuclear S7 gene following (Janko, Flajšhans, et al. 2007). Ploidy was routinely estimated from gene dose effect using allozymes as described in (Janko, Flajšhans, et al. 2007) and also by flow cytometry in cases where we disposed sufficient amount of fixed tissue.

2. Detection of current hybridization and introgression: To test whether contemporary species hybridize and are interconnected by ongoing gene flow, we performed extensive analyses of hybrid zones.

2.1. sample processing: 818 sampled specimens from *C. elongatoides* – *C. tanaitica* hybrid zone were routinely genotyped by allozyme or PCR-RFLP method described in (Janko, Flajšhans, et al. 2007), which allows fast and inexpensive a priori identification of *Cobitis* genomes as well as ploidy. A subset of individuals was also analyzed by flow cytometry in order to confirm the allozyme-based genotyping. Individuals identified as diploids were subsequently subject to microsatellite analysis following the protocols of (De Gelas et al. 2008; Choleva et al. 2012). In total, nine microsatellite loci in two multiplexes were analyzed to study the genetic admixture among diploid individuals from the Hybrid zone (multiplex 1: loci Cota_006, Cota_010, Cota_027, Cota_068, Cota_093, Cota_111; multiplex 2: loci Cota_032, Cota_033, Cota_041). We also amplified and sequenced 1190 bp fragment of the cytochrome *b* gene according to (Janko, Flajšhans, et al. 2007) in a subset of diploid individuals.

The GenAlEx 6.5 software (Peakall and Smouse 2006) was used to identify clusters of individuals sharing the same microsatellite multilocus genotype (MLG) and to calculate the probability, conditional on observed variability studied loci, that individuals sharing the same MLG arose by independent sexual events (see (Janko et al. 2012) for details). While some MLG may represent distinct clones, others may belong to the same clonal lineage (the so called 'multilocus lineage' (MLL) and differ only by scoring errors or post formational mutations. We used the approach described in (Janko et al. 2012) to identify groups of MLGs forming such mutilocus lineages. Phylogenetic relationships among reconstructed mtDNA haplotypes were estimated using a median-joining network (Bandelt et al. 1999), and drawn using NETWORK software (http://www.fluxus-engineering.com/netwinfo.htm). To put our samples into a phylogenetic context of previous knowledge, we included previously published haplotypes (Choleva et al. 2008; Janko et al. 2012).

2.2 Detection of interspecific gene flow: microsatellite and allozyme data were used to detect admixed individuals from the *C. elongatoides* – *C. tanaitica* hybrid zone by two Bayesian clustering methods that are based on different algorithms but have similar power for hybrid detection (Vähä and Primmer 2006). We first used admixture model implemented in Structure 2.3.3 (Pritchard et al. 2000) to compute the parameter *q*, i.e. the proportion of an individual's genome originating in one of the two inferred clusters, corresponding to the parental species *C. elongatoides* and C. *tanaitica*. The analysis was based on runs with 10^6 iterations, following a burn-in period of $5*10^4$ iterations. Ten independent runs for number of populations varying from *K*= 1 to *K*= 10. The best value of *K* was chosen following the method proposed by Evanno et al. (2005), with Structure Harvester (Earl and vonHoldt 2011), which takes into account the rate of change in the log likelihood between successive *K* values.

Second, the Bayesian clustering method implemented in NewHybrids 1.1 (Anderson and Thompson 2002) was used to compute the posterior probability that an individual in a sample belongs to one of the defined hybrid classes or parental species. The eight genotypic classes investigated were: *C. elongatoides*, *C. tanaitica*, F1 hybrid, F2 hybrid, and two types of backcross to either *C. elongatoides* or *C. tanaitica*. The two backcross types included those having 75 % of their genome originated from the backcrossing species (B1 generation) and those having 95 % of their genome originated from the backcrossing species (further-generation backcrosses). Posterior distributions were evaluated by running five independent analyses to confirm convergence. We started with different random seeds, performed 10^4 burn-in iterations and followed by 500,000 Monte Carlo Markov Chain iterations without using prior allele frequency information. Analyses were run for four combinations of prior distributions (uniform or Jeffreys for *Θ* and *π* parameters) to explore the robustness of the results (Anderson and Thompson 2002).

Clonal reproduction causes strong linkage among markers, which violates the assumptions made by both Structure and NewHybrids. To minimize the effect of clonal propagation of identical genomes analyses we followed (Janko et al. 2012) and used only one representative of each unique MLL. Since each MLL incorporates several slightly different MLGs, we repeated the analysis several times with randomly chosen representatives of each MLL. We also repeated those analyses with either Cota_006 or Cota_041 locus removed, since De Gelas et al. (2008) reported possible linkage between those loci. The microsatellite locus Cota_027 was removed since it does not amplify in *C. elongatoides*.

3. Experimental crossing and analysis of reproductive mode of hybrids: In order to understand the mode of reproduction of *Cobitis* hybrids, we performed crossing experiments of selected fish couples under conditions described in (Choleva et al. 2012). Obtained progeny was either bred for subsequent crossing experiment or reared till the resorption of yolk sac and sacrificed for subsequent microsatellite analysis. Microsatellite multiplex 1 was subsequently analyzed in parents and progeny to verify their mode or heredity. Two types of crossing were performed:

3.1 In the one type of experiment, we analyzed the reproductive mode of naturally occurring diploid and triploid hybrids. To do so, we individually crossed wild caught EN and EEN females (see Table 2) with males of the sexual species and genotyped parental individuals as well as their progeny. Several authors (Cimino 1972; Uzzell et al. 1980; Carmona et al. 1997) also reported more complex reproductive modes when one parental genome is excluded prior to meiosis while the other is clonally transmitted. This type of so-called hybridogenetic reproduction leads to reduced, albeit non-segregating gametes and may be detected by comparison of allozyme profiles of somatic and germinal tissues. In order to test for premeiotic genome exclusion in *Cobitis* hybrids, we have therefore compared allozyme profiles of somatic tissues of hybrid females with those from their ova using the standard allozyme methods adapted for *Cobitis* (Janko, Flajšhans, et al. 2007).

3.2 The other type of experiment represents part of long-term project aimed at *de novo* creation of F1 hybrids among different *Cobitis* species and analysis of their mode of inheritance. Part of the results concerning the *C. elongatoides* – *C. taenia* crosses were reported previously (Choleva et al. 2012). We individually crossed males and females of different species and reared their progeny until sexual maturity. The progeny were subsequently combined either to produce backcross or second filial generation. In 2000, we have successfully produced several clutches of *C. elongatoides-taenia* hybrids which were backcrossed between 2004 and 2009 and the results were published in (Choleva et al. 2012). In 2006, we have obtained successful spawning of several couples combining *C. pontica* and *C. taenia* individuals and reared their progeny until maturity. After reaching the maturity, several pairs of *C. pontica-taenia* hybrids were combined for F2 or backcross generation and we obtained three clutches of F2 progeny (families No. 1 and 17 in Table 2). The individuals used for crossing experiments are listed in Table 2 and S2.

4. Detection of historical admixture: In order to test for periods of historical or contemporary gene flow on a genome-wide scale, we applied coalescent models to SNP data from mRNA.

4.1. mRNA sequencing and assembly: the mRNA sequencing concerned liver and oocyte tissues of 5 individuals of C. taenia*,* 2 of each *C. elongatoides, C. tanaitica* and *C. pontica* species as well as an outgroup species *C. strumicae* belonging to the subgenus Bicanestrinia sampled from the isolated Sredecka River (Fig 1). RNA isolations were made with the Ambion® ToTALLY RNA™ Total RNA Isolation Kit. cDNA libraries were constructed by SMARTer PCR cDNA Synthesis Kit (Clontech) according to the manufacturer's instructions with the following exceptions: modified CDS-T22 primer (5'-AAGCAGTGGTATCAACGCAGAGTTTTTGTTTTTTTCTTTTTTTTTTVN-3') was used instead of 3' BD SMART CDS Primer II A; first strand synthesis time was prolonged to 2 hours; and PCR conditions during second strand synthesis were modified (denaturation: 95°C/ 2 min, amplification: 18 times 95°C/ 10 sec, 65°C/ 30 sec, 72°C/ 3 min, final extension: 72°C/ 5 min). cDNA was normalized by Trimmer cDNA normalization Kit (Evrogen, Moscow, Russia). 1 µg of each normalized cDNA was used for sequencing library preparations according to the Roche Rapid Library Preparation Protocol (Roche, Welwyn Garden City, UK). Libraries were tagged using MID adapters (Roche), pooled (usually 2 samples per one large sequencing region) and sequenced using GS FLX+ chemistry (454 Life Sciences, Roche).

Initial 1886536 reads (648620753 bp) obtained from *C. taenia* liver and oocytes were filtered to remove low quality part of reads, adaptors and primer sequences using Trimmomatic software (Bolger et al. 2014). Technical PCR multiplicates were removed from 454 data set using cdhit-454 software (Niu et al. 2010). Resulting 1707769 reads (568470258 base pairs) were used for cDNA assembly using Newbler (Software Release: 2.6 20110517_1502) from Roche with parameters: minimum read length: 40; minimum overlap length: 40; minimum match: 90 %; minimum contig length 300 bp. Assembled loci were blasted against *Cobitis takatsuensis* mitochondrion (NC_015306.1) using blastn (Camacho et al. 2009) and contigs matching to mitochondrion were removed from subsequent analysis in order to analyse the phylogeny and gene flow in nucleus only. Without prior genome knowledge, correct detection of SNPs may be compromised by undetected paralogy, copy number variation, and repetitive parts of genomes. In those cases, reads map to the contigs assembled from two (or more) different loci resulting in spurious heterozygotes calling based on between-paralogue variation (Gayral et al. 2013). We removed from analysis all contigs too similar to each other, where read mapping was not unique, and kept only those contigs, where mapping of all reads did not show any changes in different programs. We further excluded contigs indicative of paralogous mapping, where we observed excessive heterozygosity with identical heterozygotic states being observed simultaneously at same SNP positions of distant outgroup (*C. strumicae*) as well as all scored ingroup species. First assembly (28338 cDNA contigs, 31831582 bp with N50 1301 bp and 95.39 % bases with quality score Q40 and better) was cleaned for duplicates and paralogs, which left us with transcriptome of 20385 potential mRNAs (average length 1096.5 bp, total length 22355325 bp and N50 1246 bp).

On this reference transcriptome we separately mapped 454 reads from all species including outgroup and two individuals from each ingroup species (for the mapping we selected the two *C. taenia* individuals with the highest coverage), with the Newbler software. For each individual, all highly confident SNPs with sufficient coverage were inserted into the database using our own SQL scripts. To detect if a particular SNP was sequenced in any animal, we prepared new version of the reference transcriptome with the bases replaced by N in positions where the particular SNP was detected. This new transcriptome was then used for new mapping and detection of SNP presence/absence in all individuals. Given that the reference transcriptome was prepared from *C. taenia* but more distant species were subsequently mapped onto it, our approach introducing 'N' on variant positions minimized the unbalanced mapping of more distant species onto the model transcriptome. Finally, we sorted identified SNPs from each locus into individual-locus-specific matrices.

4.2 Detection of interspecific gene flow from SNP data: SNP data obtained from RNAseq were analyzed using a recently introduced coalescent-based maximum-likelihood method ((Wilkinson-Herbots 2008; Wilkinson-Herbots 2012; Wilkinson-Herbots 2015; Costa and Wilkinson-Herbots 2016); *related manuscripts 1*) which is computationally inexpensive and estimates simultaneously the population sizes, migration rates as well as population splitting times including some scenarios of time-variable migration rates.

Three models published in (Wilkinson-Herbots 2008; Wilkinson-Herbots 2012) were fitted to the data: a) a strict isolation model with four parameters (I4-model) assuming that an ancestral population with size parameter Θ_a split at time t_0 into two isolated descendant populations of sizes Θ_{c1} and Θ_{c2} (throughout the paper the "population size" parameter is defined as $\Theta_i = 4N_i\mu$, where N_i is the effective diploid size of species $_i$ and μ is the mutation rate per sequence per generation, averaged over the loci included in the analysis; wherever an index 'c' accompanies the parameter name, it will always indicate the values relevant for current populations, while an index 'a' indicates the states of ancestral populations before the split) b) an "isolation with migration" model with four parameters (IM4-model) where an ancestral population of size Θ_a split at time t_0 into two descendant populations of equal size Θ_c interconnected by gene flow at rate M_c ($M_c = 4N_c * m_c$, where m_c is the proportion of migrants per generation, and where the index 'c' refers to migration between current populations) and c) an "isolation with initial migration" model (IIM7-model with 7 parameters) assuming that an ancestral population of size $Θ_a$ split at time t_0 into two descendant populations of equal size $Θ$ interconnected by gene flow at rate M, which lasted until time t_1 since when two descendant populations of sizes $Θ_{c1}$ and $Θ_{c2}$ evolved in isolation until the present (Fig 4). Although both the I4- and IM4-models are nested within the IIM7-model, which facilitates their comparison with the latter model, it is problematic that the IIM7 model allows an additional population size change, which makes it difficult to evaluate whether its better fit is due to the correct assessment that gene flow was followed by isolation, or merely due to the inclusion of an additional change of population size.

We therefore explored the effect of differences or changes in population size by fitting four additional nested models that relax the assumption of equal sizes during the migration stage and some also allow additional population size changes. The mathematical details of these relaxed models are given in (Costa and Wilkinson-Herbots 2016) and in *related manuscript 1*. These include d) an extended isolation model with seven parameters (I7) assuming that an ancestral population of size $Θ_a$ split at time t_0 into two isolated descendant populations of sizes $Θ_1$ and $Θ_2$, and allowing an additional population size change at time t_1 resulting in their current sizes $Θ_{c1}$ and $Θ_{c2}$, respectively; e) a more general IM-model with five parameters (IM5) where an ancestral population of size $Θ_a$ split at time t_0 into two descendant populations of sizes $Θ_{c1}$ and $Θ_{c2}$ exchanging genes at constant rate M until the present; f) an extended IM-model with nine parameters (GIM9) where an ancestral population of size $Θ_a$ split at time t_0 into two descendant populations of sizes $Θ_1$ and $Θ_2$ interconnected by gene flow at rate M until t_1, from which time onwards both species are at their current sizes $Θ_{c1}$ and $Θ_{c2}$ and gene flow occurs at its current rate M_c; g) an updated IIM-model with eight parameters (IIM8) which assumes that an ancestral species of size $Θ_a$ split at time t_0 into two descendant species of sizes $Θ_1$ and $Θ_2$ interconnected by gene flow at rate M until time t_1, after which both species are at their current sizes $Θ_{c1}$ and $Θ_{c2}$ and evolve in complete isolation (Fig 4). Note that all models described above are nested within the GIM9-model. In the case of the three closely related species, an additional six-parameter isolation model (I6) was used which allows for one additional population size change of the ancestral species (Fig 4). This additional model was used because of the very high gene flow estimates obtained with models that allow migration after the split time t_0, suggesting that the nascent populations formed a single, effectively panmictic species. Therefore, the I6-model represents the scenario where an ancestral species of size Θ_a underwent a change of size at time t_0 and subsequently remained a single panmictic species of size Θ_{a1} until time t_1, when it split into two isolated populations of sizes Θ_{c1} and Θ_{c2}. Mathematically, the I6-model is merely a special case of the IIM7 model, in the limit as the migration rate *M* tends to infinity. All models were ranked according to their AIC score and we also calculated the evidence ratio for each model, providing a relative measure of how much less likely a given model is compared to the best-fitting model, given the set of candidate models considered and the data (Anderson and Thompson 2002); Table 3. In addition, Likelihood Ratio tests were performed to compare the fit of pairs of nested models, where we have assumed that the use of the χ2 distribution with the appropriate number of degrees of freedom is conservative ((Self and Liang 1987; Wilkinson-Herbots 2015; Costa and Wilkinson-Herbots 2016)).

The currently available implementation of the above models allows the analysis of only two species at once, and we therefore prepared separate datasets for each of the six pairwise species comparisons: *C. elongatoides – C. taenia, C. elongatoides – C. tanaitica, C. elongatoides – C. pontica, C. taenia – C. tanaitica, C. taenia – C. pontica and C. tanaitica – C. pontica*. The datasets for each pairwise comparison were represented by locus-specific alignments of SNP positions (we did not use invariant positions) with three rows – two rows corresponded to sites with successfully resolved allelic states in both compared species and the third row contained SNPs of the outgroup species *C. strumicae*. The coalescent models assume free recombination among loci but no recombination within loci and require three sets of data for each pairwise species analysis. These consist of the number of nucleotide differences between pairs of sequences sampled a) both from species 1, b) both from species 2, and c) one from each species. To estimate all model parameters simultaneously, data from all three types of sequence pairs must be included but each pair of sequences must come from a different, independent locus (Wilkinson-Herbots 2012). Therefore, we randomly divided the analyzed loci into three non-overlapping subsets, two of which were used to obtain numbers of differences between pairs of sequences from the same species (either species 1 or species 2), and the last set of loci was used to obtain data on the number of differences between one sequence drawn from the first species and one sequence from the second species. The intraspecific data were simply calculated as the number of heterozygous positions at each locus, which in fact represents the number of nucleotide differences between a pair of alleles brought together by segregation into a sampled individual. The preparation of the third dataset was more complex since it requires the comparison of two haploid sequences from different species, while our sequences originated from diploid individuals. When these possessed multiple heterozygotic positions in the same locus, it prevented unambiguous reconstruction of their phase. Therefore, we extracted the longest possible alignment of SNPs where both compared individuals (species 1 and species 2) had at most one heterozygous position by trimming the per-locus alignments of SNPs (similar to the trimming procedure done in (Lohse et al. 2011)). Subsequently we randomly phased the remaining heterozygous SNP markers into either one or the other base and compared the number of differences between both species (this procedure in fact represents the comparison of two alleles drawn from species 1 and 2 respectively while keeping the allelic states resolved in the outgroup on a shortened alignment).

The comparison with an outgroup sequence enabled us to estimate the relative mutation rates at all loci (Wang and Hey 2010). We set the divergence time between the *C. strumicae* and the ingroup to 17.7 Mya according to penalized likelihood time-calibrated tree of Majtánová et al. (2016) and computed 95% confidence intervals based on the profile likelihood as described in (Costa and Wilkinson-Herbots 2016). Because of the computational time required, such confidence intervals were provided only for the best-fitting model for each dataset. In some cases, the likelihood profiling broke down prematurely, since forcing one parameter too far from its maximum-likelihood estimate rendered the whole model impossible to fit. In such cases we indicate the nearest value that could be obtained, keeping in mind that the true 95% CI should be wider (Table 3).

In order to incorporate intrapopulation variability into our estimates, we combined the data from two individuals of each species. In doing so, we randomly sampled each locus from either one or the other individual (assuming free recombination among loci) but we always kept the sequence of SNPs within each locus from a single randomly selected individual to avoid the introduction of false recombination.

5. Testing the general applicability of our findings: We performed a comparative study of sexual and asexual fish hybrids in order to test the general validity of patterns revealed in *Cobitis*. Russell (2003) investigated the correlation between the genetic divergence between hybridizing fish species (distances in the cytochrome *b* gene corrected with the K2P) and the dysfunction of their F1 fish hybrids (infertility or inviability). Russell (2003) characterized each hybrid case by the postzygotic isolation index ranging from 0 (both hybrid sexes fertile) to 4 (both sexes inviable). The index value 2 assigned the stage when both hybrid sexes are viable but infertile, therefore preventing any effective interspecific gene flow. We introduced additional type of postzygotic isolation index (5) to those fish hybrids that have been documented to transmit their genomes clonally (gynogenesis, androgenesis) or hemiclonally (hybridogenesis). Altogether, performed literature search lead us to 13 cases of fish asexual hybrids, which were added to the database of Russell (2003). In single case we modified Russell's (2003) data since he assigned *R. rutilus x A. bramma* hybrids with postzygotic isolation index 0.5 but Слынько (2000) showed that such hybrids produce clonal gametes and can reproduce via androgenesis. Therefore, *R. rutilus x A. bramma* hybrids were assigned with index value 5. In accordance with Russell's data, the cytochrome *b* gene divergence was calculated from available sequences of parental taxa using the K2P correction using the Mega 5.0 software (Tamura et al. 2011) (Appendix S5). The genetic distances of the group 5 was compared with the other types of hybrids using the *t*-test after the normality of data was evaluated with the Shapiro-Wilk test.

To avoid phylogenetic dependence, Russell (2003) considered each species in one cross only. Published cases of asexual hybrids concern non-overlapping species pairs with three exceptions. In *Cobitis, C. elongatoides* has been involved in at least two crosses leading to naturally occurring asexuals (*C. elongatoides-taenia* and *C. elongatoides-tanaitica*) but since the original *C. tanaitica*-like mitochondrion has been lost (see above), we considered *C. elongatoides – C. taenia* cross only. On the other hand, *Hexagrammos octogramus* and *Poeciliopsis monacha* produce asexual hybrids by mating with two (*H. otaki* and *H. agrammus*) and three (*P. lucida, P. occidentalis, P. latipina*) congenerics, respectively. Therefore, we performed the *t*-tests several times with only one such cross per species.

Acknowledgements

Authors are profoundly obliged to Roger Butlin and Diego Fontaneto for very inspiring comments during the MS writing. We also thank Sarka Pelikanova, Jana Cechova (*in memoriam*), Koen DeGelas and Bart Hellemans for help with DNA and RNA in laboratory and to Jana Kopecka and Vera Slechtova for help with allozyme analysis. Jan Kotusz and Petya Ivanova helped during the field work and we are very obliged for it too.

Funding information

The study was supported by grant no. 206/09/1298, P506/10/1155, P506/12/P857 and 13-12580S provided by the Czech Science Foundation (www.gacr.cz), grant no. KJB600450902 provided by the Grant Agency of Academy of Sciences of the Czech Republic (www.cas.cz) and by the Academy of Sciences of the Czech Republic (ASCR) Grant No. RVO 67985904 and by TEWEP (CZ.1.05/2.1.00/03.0100 IET). RJC's research was supported by the Engineering and Physical Sciences Research Council [grant number EP/K502959/1].

Author contribution

KJ formulated the central hypothesis of the study and drafted the manuscript; KJ, HP and LC conceived and designed the experiments; KJ, NI, JK, JRi, RR and LC performed the experiments; KJ, HP, HWH, RDC, PD, JK, JRo and LC analysed the data; HWH and RJC developed the mathematical models; KJ, HWH and RR contributed to the MS writing,

None of the authors have any competing interests

References to related manuscript 1:

Mathematical details of amendments that were used to extend published coalescent models are described in manuscript attached as "related manuscript 1":

Related Manuscript 1: The Generalised Isolation-With-Migration Model: a Maximum-Likelihood Implementation for Multilocus Data Sets. Rui J. Costa and Hilde Wilkinson-Herbots

References

Anderson EC, Thompson EA. 2002. A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. Genetics 160:1217–1229.

Angers B, Schlosser IJ. 2007. The origin of Phoxinus eos-neogaeus unisexual hybrids. Mol. Ecol. 16:4562–4571.

Avise IJ. 2008. Clonality : The Genetics, Ecology, and Evolution of Sexual Abstinence in Vertebrate Animals: The Genetics, Ecology, and Evolution of Sexual Abstinence in Vertebrate Animals. University of California. Oxford University Press

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37–48.

Barbiano LA da, Gompert Z, Aspbury AS, Gabor CR, Nice CC. 2013. Population genomics reveals a possible history of backcrossing and recombination in the gynogenetic fish Poecilia formosa. Proc. Natl. Acad. Sci. 110:13797–13802.

Bohlen J, Perdices A, Doadrio I, Economidis PS. 2006. Vicariance, colonisation, and fast local speciation in Asia Minor and the Balkans as revealed from the phylogeny of spined loaches (Osteichthyes; Cobitidae). Mol. Phylogenet. Evol. 39:552–561.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30:2114–2120.

Bolnick DI, Near TJ. 2005. Tempo of Hybrid Inviability in Centrarchid Fishes (teleostei: Centrarchidae). Evolution 59:1754–1767.

Butlin RK, Schön I, Martens K. 1999. Origin, age and diversity of clones. J. Evol. Biol. 12:1020–1022.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Carman JG. 1997. Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol. J. Linn. Soc. 61:51–94.

Carmona JA, Sanjur OI, Doadrio I, Machordom A, Vrijenhoek RC. 1997. Hybridogenetic reproduction and maternal ancestry of polyploid Iberian fish: the Tropidophoxinellus alburnoides complex. Genetics 146:983–993.

Choleva L, Apostolou A, Ráb P, Janko K. 2008. Making it on their own: sperm-dependent hybrid fishes (Cobitis) switch the sexual hosts and expand beyond the ranges of their original sperm donors. Philos. Trans. R. Soc. B Biol. Sci. 363:2911.

Choleva L, Janko K, De Gelas K, Bohlen J, Šlechtová V, Rábová M, Ráb P. 2012. Synthesis of Clonality and Polyploidy in Vertebrate Animals by Hybridization Between Two Sexual Species. Evolution 66:2191–2203.

Choleva L, Musilova Z, Kohoutova-Sediva A, Paces J, Rab P, Janko K. 2014. Distinguishing between Incomplete Lineage Sorting and Genomic Introgressions: Complete Fixation of Allospecific Mitochondrial DNA in a Sexually Reproducing Fish (Cobitis; Teleostei), despite Clonal Reproduction of Hybrids. PLoS ONE 9:e80641.

Cimino MC. 1972. Egg-Production, Polyploidization and Evolution in a Diploid All-Female Fish of the Genus Poeciliopsis. Evolution 26:294–306.

Costa RJ, Wilkinson-Herbots H. 2016. Efficient Maximum-Likelihood Inference For The Isolation-With-Initial-Migration Model With Potentially Asymmetric Gene Flow. ArXiv160103684 Q-Bio Stat [Internet]. Available from: http://arxiv.org/abs/1601.03684

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.

Cunha C, Doadrio I, Coelho MM. 2008. Speciation towards tetraploidization after intermediate processes of non-sexual reproduction. Philos. Trans. R. Soc. B Biol. Sci. 363:2921–2929.

De Gelas K, Janko K, Volckaert FAM, De Charleroy D, Van Houdt JKJ. 2008. Development of nine polymorphic microsatellite loci in the spined loach, Cobitis taenia, and cross-species amplification in the related species C. elongatoides, C. taurica and C. tanaitica. Mol. Ecol. Resour. 8:1001–1003.

Dubois A. 2011. Species and “strange species” in zoology: Do we need a “unified concept of species”? Comptes Rendus Palevol 10:77–94.

Earl DA, vonHoldt BM. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Edmands S. 2002. Does parental divergence predict reproductive compatibility? Trends Ecol. Evol. 17:520–527.

Ernst A. 1918. Bastardierung als Ursache der Apogamie im Pflanzenreich. Eine Hypothese zur experimentellen Vererbungs- und Abstammungslehre. Nabu Press

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Fontaneto D, Herniou EA, Boschetti C, Caprioli M, Melone G, Ricci C, Barraclough TG. 2007. Independently Evolving Species in Asexual Bdelloid Rotifers. PLoS Biol 5:e87.

Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N, et al. 2013. Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. PLoS Genet. [Internet] 9. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3623758/

Hey J, Chung Y, Sethuraman A. 2015. On the occurrence of false positives in tests of migration under an isolation-with-migration model. Mol. Ecol. 24:5078–5083.

Hotz H, Mancino G, Bucciinnocenti S, Ragghianti M, Berger L, Uzzell T. 1985. Rana ridibunda varies geographically in inducing clonal gametogenesis in interspecies hybrids. J. Exp. Zool. 236:199–210.

Jančúchová-Lásková J, Landová E, Frynta D. 2015. Are genetically distinct lizard species able to hybridize? A review. Curr. Zool. 61:155–181.

Janko K, Bohlen J, Lamatsch D, Flajšhans M, Epplen JT, Ráb P, Kotlík P, Šlechtová V. 2007. The gynogenetic reproduction of diploid and triploid hybrid spined loaches (Cobitis: Teleostei), and their ability to establish successful clonal lineages—on the evolution of polyploidy in asexual vertebrates. Genetica 131:185–194.

Janko K, Culling MA, Rab P, Kotlik P. 2005. Ice age cloning-comparison of the Quaternary evolutionary histories of sexual and clonal forms of spiny loaches (Cobitis; Teleostei) using the analysis of mitochondrial DNA variation. Mol. Ecol. 14:2991–3004.

Janko K, Flajšhans M, Choleva L, Bohlen J, Šlechtová V, Rábová M, Lajbner Z, Šlechta V, Ivanova P, Dobrovolov I, et al. 2007. Diversity of European spined loaches (genus Cobitis L.): an update of the geographic distribution of the Cobitis taenia hybrid complex with a description of new molecular tools for species and hybrid determination. J. Fish Biol. 71:387–408.

Janko K, Kotlik P, Ráb P. 2003. Evolutionary history of asexual hybrid loaches (Cobitis: Teleostei) inferred from phylogenetic analysis of mitochondrial DNA variation. J. Evol. Biol. 16:1280–1287.

Janko K, Kotusz J, De Gelas K, Slechtová V, Opoldusová Z, Drozd P, Choleva L, Popiołek M, Baláž M. 2012. Dynamic formation of asexual diploid and polyploid lineages: multilocus analysis of Cobitis reveals the mechanisms maintaining the diversity of clones. PloS One 7:e45384.

Kearney M, Fujita MK, Ridenour J. 2009. Lost sex in the reptiles: constraints and correlations. In: Lost Sex. Dordrecht Netherlands: Springer. p. 447–474.

Keller B, Wolinska J, Tellenbach C, Spaak P. 2007. Reproductive isolation keeps hybridizing Daphnia species distinct. Limnol. Oceanogr. 52:984–991.

Kim I, Lee E. 2000. Hybridization experiment of diploid-triploid cobitid fishes, Cobitis sinensis-longicorpus complex (Pisces, Cobitidae. Folia Zool. Supplement A:17–22.

Lohse K, Harrison RJ, Barton NH. 2011. A General Method for Calculating Likelihoods Under the Coalescent Process. Genetics 189:977–987.

Lowe WH, Allendorf FW. 2010. What can genetics tell us about population connectivity? Mol. Ecol. 19:3038–3051.

Majtánová Z, Choleva L, Symonová R, Ráb P, Kotusz J, Pekárik L, Janko K. 2016. No Evidence for Increased Rate of Chromosomal Evolution in Asexuals: Karyotype Stability in Diploids. PLoS One in press.

Matute DR, Butler IA, Turissini DA, Coyne JA. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. Science 329:1518–1521.

Mikulíček P, Kautman M, Demovič B, Janko K. 2014. When a clonal genome finds its way back to a sexual species: evidence from ongoing but rare introgression in the hybridogenetic water frog complex. J. Evol. Biol. 27:628–642.

Moritz C, Brown WM, Densmore LD, Wright JW, Vyas D, Donnellan S, Adams M, Baverstock P. 1989. Genetic diversity and the dynamics of hybrid parthenogenesis in Cnemidophorus (Teiidae) and Heteronotia (Gekkonidae. In: Evolution and Ecology of Unisexual Vertebrates. New York State Museum, Albany, New York. p. 268–280.

Moritz C, Densmore L, Wright J, Brown W. 1992. Mitochondrial DNA analyses and the origin and relative age of parthenogenetic Cnemidophorus: phylogenetic constraints on hybrid origins. Evolution 46:184–192.

Moritz C, Uzzell T, Spolsky C, Hotz H, Darevsky I, Kupriyanova L, Danielyan F. 1992. The maternal ancestry and approximate age of parthenogenetic species of Caucasian rock lizards (Lacerta: Lacertidae). Genetica 87:53–62.

Moritz C, Wright JW, Brown WM. 1992. Mitochondrial DNA Analyses and the Origin and Relative Age of Parthenogenetic Cnemidophorus: Phylogenetic Constraints on Hybrid Origins. Evolution 46:184–192.

Murphy RW, Fu J, Macculloch RD, Darevsky IS, Kupriyanova LA. 2000. A fine line between sex and unisexuality: the phylogenetic constraints on parthenogenesis in lacertid lizards. Zool. J. Linn. Soc. 130:527–549.

Nadachowska K, Babik W. 2009. Divergence in the face of gene flow: the case of two newts (amphibia: salamandridae). Mol. Biol. Evol. 26:829–841.

Niu B, Fu L, Sun S, Li W. 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. BMC Bioinformatics 11:187.

Orr HA, Turelli M. 2001. The Evolution of Postzygotic Isolation: Accumulating Dobzhansky-Muller Incompatibilities. Evolution 55:1085–1094.

Parchman TL, Gompert Z, Braun MJ, Brumfield RT, McDonald DB, Uy J a. C, Zhang G, Jarvis ED, Schlinger BA, Buerkle CA. 2013. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Mol. Ecol. 22:3304–3317.

Peakall ROD, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288–295.

Price TD, Bouvier MM. 2002. The Evolution of F1 Postzygotic Incompatibilities in Birds. Evolution 56:2083–2089.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Russell S. 2003. Evolution of intrinsic post-zygotic reproductive isolation in fish. Ann. Zool. Fenn.:321–329.

Rykena S. 2002. Experimental hybridization in green lizards (Lacerta s. str.), a tool to study species boundaries. Mertensiella 13:78–88.

Sánchez-Guillén RA, Córdoba-Aguilar A, Cordero-Rivera A, Wellenreuther M. 2014. Genetic divergence predicts reproductive isolation in damselflies. J. Evol. Biol. 27:76–87.

Schultz RJ. 1973. Unisexual fish: laboratory synthesis of a “species.” Science 179:180–181.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström Å, et al. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176–192.

Self SG, Liang K-Y. 1987. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions. J. Am. Stat. Assoc. 82:605–610.

Stenberg P, Saura A. 2009. Cytology of Asexual Animals. In: Schön I, Martens K, Dijk P, editors. Lost Sex. Springer Netherlands. p. 63–74.

Stöck M, Lampert KP, Möller D, Schlupp I, Schartl M. 2010. Monophyletic origin of multiple clonal lineages in an asexual fish (Poecilia formosa). Mol. Ecol. 19:5204–5215.

Strasburg JL, Rieseberg LH. 2010. How Robust Are “Isolation with Migration” Analyses to Violations of the IM Model? A Simulation Study. Mol. Biol. Evol. 27:297–310.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

Uzzell T, Hotz H, Berger L. 1980. Genome exclusion in gametogenesis by an interspecific Rana hybrid: Evidence from electrophoresis of individual oocytes. J. Exp. Zool. 214:251–259.

Vähä J-P, Primmer CR. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Mol. Ecol. 15:63–72.

Vrijenhoek RC. 1998. Clonal Organisms and the Benefits of Sex. In: Carvalho GR, editor. Advances in molecular ecology. Amsterdam: IOS Press. p. 151–172.

Wang Y, Hey J. 2010. Estimating Divergence Parameters With Small Samples From a Large Number of Loci. Genetics 184:363–379.

Welch JJ. 2004. Accumulating Dobzhansky-Muller Incompatibilities: Reconciling Theory and Data. Evolution 58:1145–1156.

Wilkinson-Herbots H. 2015. A fast method to estimate speciation parameters in a model of isolation with an initial period of gene flow and to test alternative evolutionary scenarios. ArXiv151105478 Q-Bio Stat [Internet]. Available from: http://arxiv.org/abs/1511.05478

Wilkinson-Herbots HM. 2008. The distribution of the coalescence time and the number of pairwise nucleotide differences in the “isolation with migration” model. Theor. Popul. Biol. 73:277–288.

Wilkinson-Herbots HM. 2012. The distribution of the coalescence time and the number of pairwise nucleotide differences in a model of population divergence or speciation with an initial period of gene flow. Theor. Popul. Biol. 82:92–108.

Wirtz P. 1999. Mother species-father species: unidirectional hybridization in animals with female choice. Anim. Behav. 58:1–12.

Wu CI, Davis AW. 1993. Evolution of postmating reproductive isolation: the composite nature of Haldane’s rule and its genetic bases. Am. Nat. 142:187–212.

Zhang Q, Arai K, Yamashita M. 1998. Cytogenetic mechanisms for triploid and haploid egg formation in the triploid loach Misgurnus anguillicaudatus. J. Exp. Zool. 281:608–619.

Слынько ЮВ. 2000. Система размножения межродовых гибридов плотвы (Rutilus rutilus L.), леща (Abramis brama L.) и синца (Abramis ballerus L.)(Leuciscinae: Cyprinidae). Available from: http://www.dissercat.com/content/sistema-razmnozheniya-mezhrodovykh-gibridov-plotvy-rutilus-rutilus-l-leshcha-abramis-brama-l

Figures, Tables and Legends

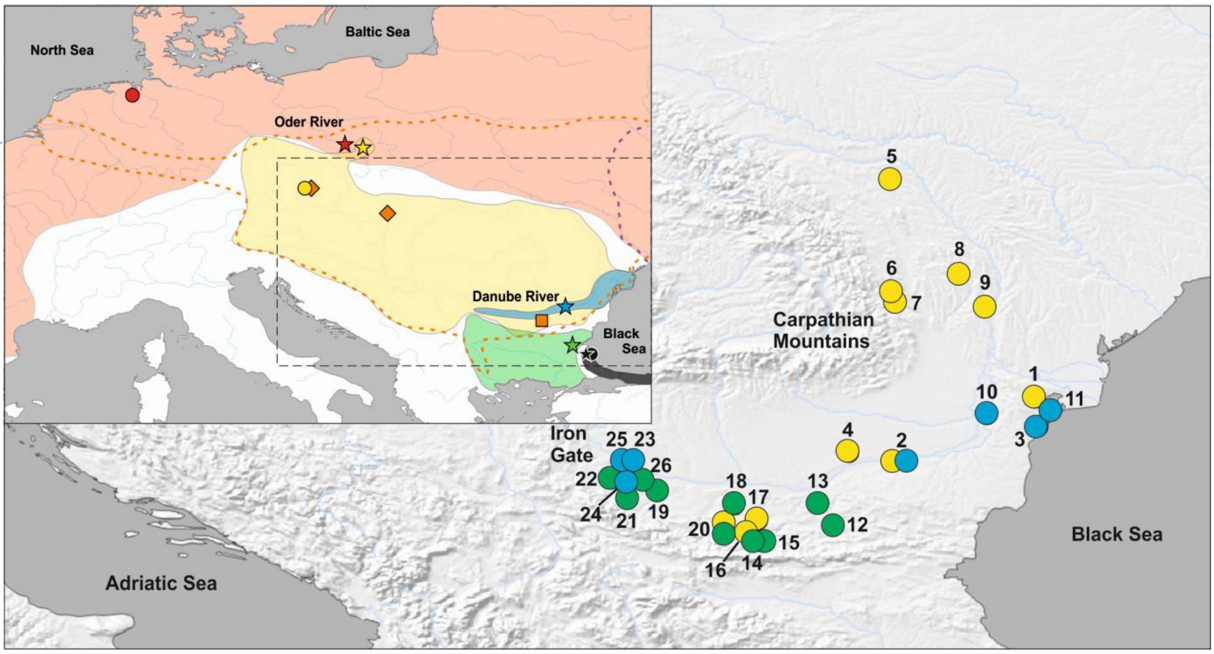

Fig 1. Map of sampling sites and distribution of species and hybrid biotypes.

Yellow circles indicate localities with *C. elongatoides* samples, blue represent *C. tanaitica* and green indicate *C. strumicae*. Locality numbers correspond with Table 1. The inset shows European distribution of studied sexual species. Red stands for *C. taenia* distribution range, yellow for *C. elongatoides*, blue for *C. tanaitica*, black for *C. pontica*, and green for *C. strumicae*. Stars indicate sampling sites of individuals used for transcriptome analyses, circles of those used in crossing experiments (the orange square stands for diploid and diamonds for triploid *C. elongatoides-tanaitica* hybrids used in crossings). Orange dotted line delimits the distribution of the ancient clonal lineage, the so called Hybrid Clade I, purple dotted line the distribution of the so called Hybrid Clade II.

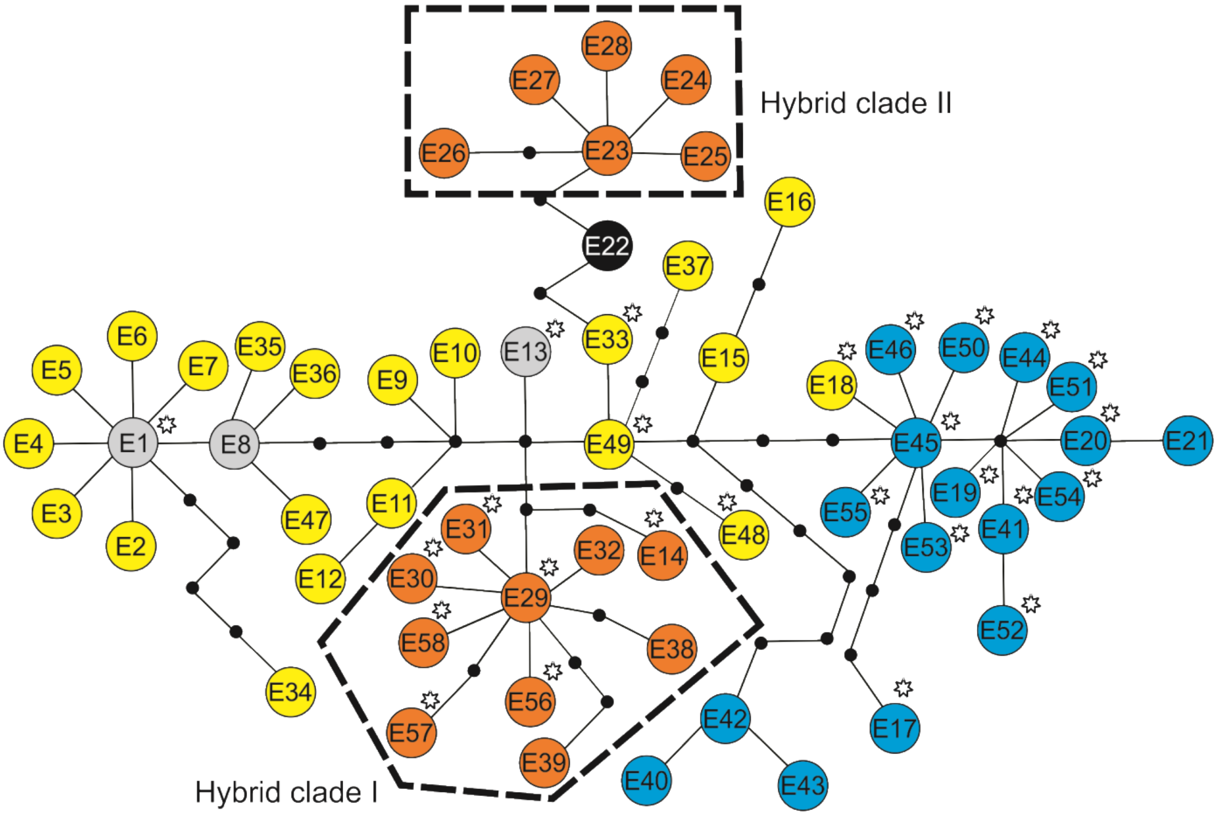

Fig 2. Median-joining haplotype network showing phylogenetic relationships among *C. elongatoides*-like haplotypes of the cytochrome *b* gene.

The network was constructed from previously published haplotypes and those from the current study (with asterisk). Yellow colour denotes haplotypes sampled in *C. elongatoides*; blue in *C. tanaitica*; black in *C. pontica*; orange in *C. elongatoides-tanaitica* hybrid (Hybrid clade I) and *C. elongatoides-taenia* hybrid (Hybrid clade II). Light grey circles denote haplotypes shared by both *C. elongatoides* and hybrids. Small black circles represent missing (unobserved) haplotypes.

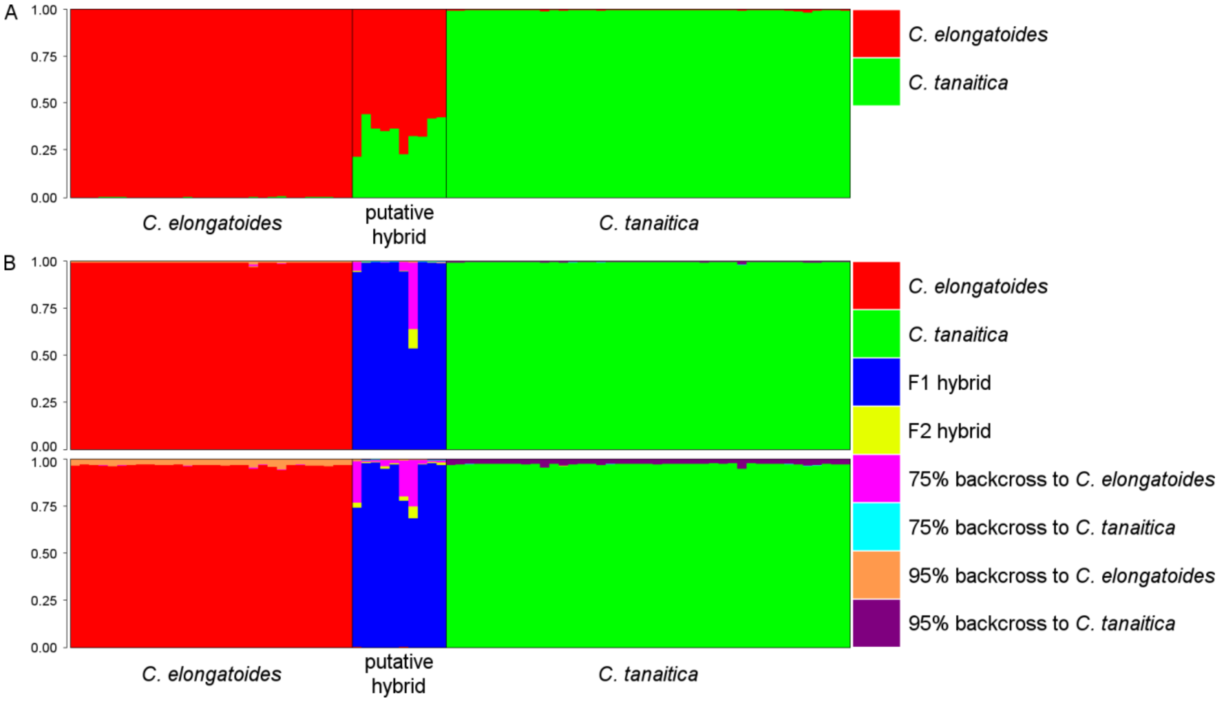

Fig 3. Population genetic analyses of the hybrid zone

(A) Individual proportion of membership to one of the two species-specific clusters according to Structure for *K* = 2. Each vertical bar represents one individual and colours show proportion of their assignment to respective clusters corresponding to sexual species. For the visual guidance, the individuals are grouped into a priori defined biotypes according to diagnostic allozyme markers (horizontal axis). (B) Classification of individual's genotype according to NewHybrids. Each vertical bar represents one individual. Each colour represents the posterior probability of an individual belonging to one of the eight different genotypic classes. Individuals are sorted as in (A). Upper pane represents the results with Jeffreys prior and lower pane with the uniform prior.

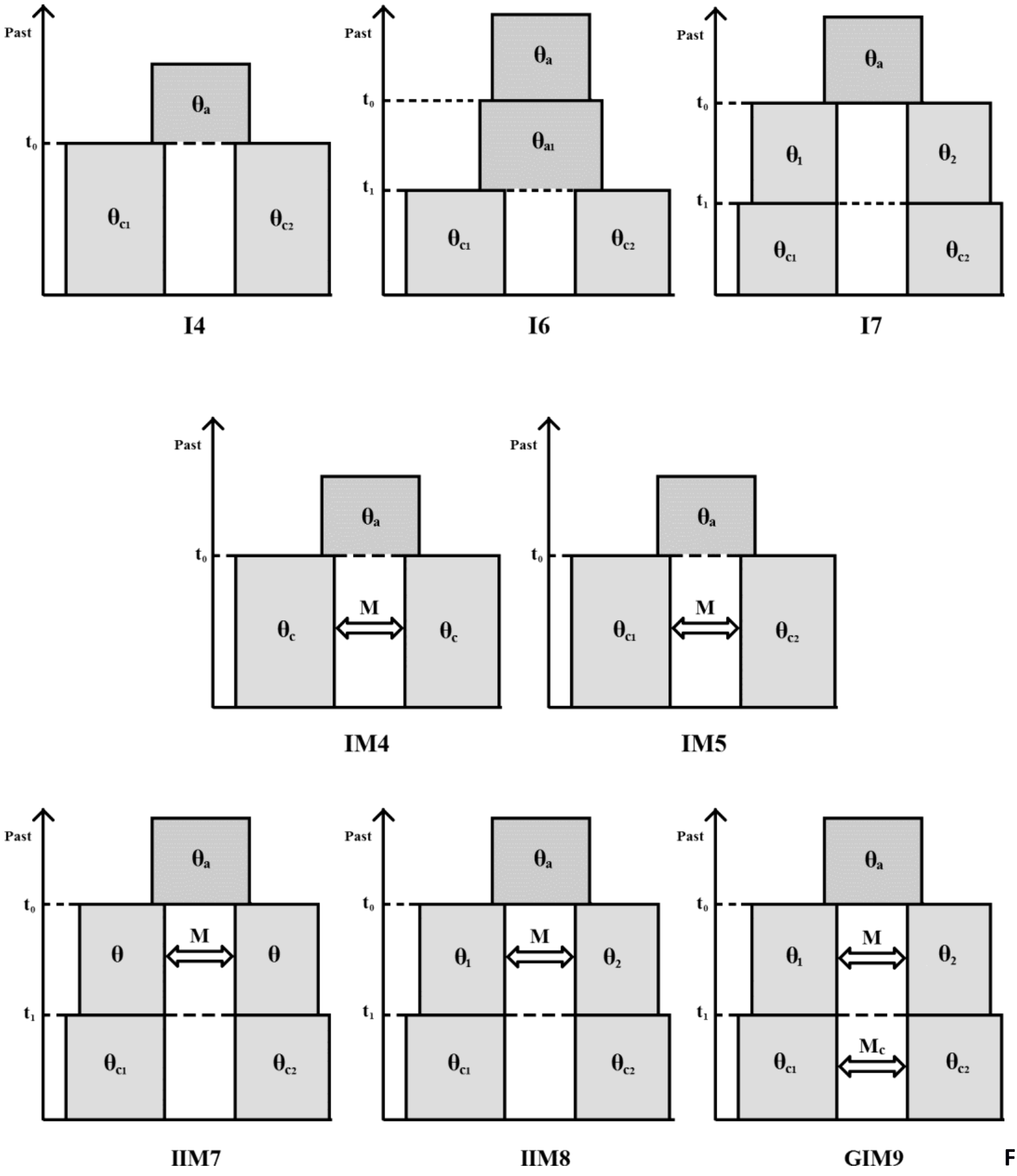

Fig 4.

Schematic view of the eight coalescent models.

Parameter names are the same as in Table 3. Arrows along the side of model diagrams indicate the respective time periods.

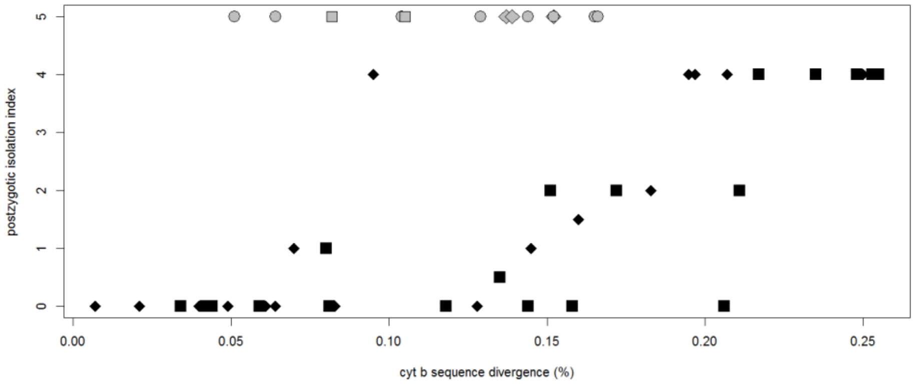

Fig 5. Correlation between intrinsic post-zygotic reproductive isolation and K2P corrected distances in cytochrome *b* gene between hybridizing species.

Reproductive isolation index is defined according to Russell's study as follows: 0, both hybrid sexes are fertile; 0.5, one sex fertile, the other sometimes infertile; 1, one sex fertile, the other infertile but viable; 1.5, one sex infertile but viable, the other sometimes still fertile; 2, both sexes viable but infertile; 2.5, one sex viable but infertile, the other sex only sometimes viable; 3, one sex viable, the other missing; 3.5, one sex sometimes viable, the other not; 4, both sexes inviable; 5, hybrids of at least one sex are known to form asexual lineages (highlighted in grey color). Species pairs where one species occurred more than once in the analysis are indicated by grey diamonds (*Poeciliopsis monacha*) and grey squares (*Hexagrammos octogramus*), respectively.

Table 1. Locality information from the Danube River hybrid zone.

site no.	site	country	coordinates		biotype	haplotype	biotype	haplotype	biotype	haplotype	biotype	biotype	biotype
					EE		NN		EN		3n EN	SS	3n ENS
1	B. Topraichioiului	Romania	44.931833	28.725650	2(1)						4		
2	Danube R.	Romania	44.080154	26.731695	1		24(18)	E19(1), E20(2), E45(5), E46(1), E50(2)	9(9/MLL H=5; MLG B=4)	E1(3), E14(1), E30(2)	118		
3	Lacul Sinoe	Romania	44.547919	28.776184			50(18)	E19(1), E20(3), E41(1), E44(1), E45(6), E51(1), E52(1), E54(1)			9		
4	Comana	Romania	44.181348	26.137802	4						21		
5	Ibăneasa R.	Romania	47.971325	26.716598	1						6		
6	Tazlău R.	Romania	46.405153	26.729092	1(1)	E13(1)					2		
7	Tazlău R.	Romania	46.279012	26.768735	1(1)						2		
8	Bârlad R.	Romania	46.673388	27.667935	2(2)	E13(1), E33(1)					10		
9	Elan Creek	Romania	46.201751	28.033022	4(4)	E18(2)					10		
10	Lacul Hazarlâc	Romania	44.707846	28.063956			5(2)	E45(1), E53(1)	6(6/MLL H=6)	E30(6)	11		
11	Lacul Razim	Romania	44.757561	28.942496			1		1(1)	E29(1)	12		
12	Stara reka R.	Bulgaria	43.163320	25.923570								7	3

13	Jantra R.	Bulgaria	43.469008	25.725494					12(12/MLL E=4; MLG C=3;MLG A=2;MLG G=1;MLL H=1)	E30(2), E31(6), E56(1), E57(1), E58(1)	87	28	110
14	Vidima R.	Bulgaria	42.958872	24.845437								2	
15	Vidima R.	Bulgaria	42.964136	24.975734								5	
16	Osam R.	Bulgaria	43.229727	24.873110	3(3)	E49(2)							
17	Osam R. trib.	Bulgaria	43.084336	24.746859	18(18)	E48(1), E49(7)					1		
18	Vit R.	Bulgaria	43.442467	24.546828					1(1)	E31(1)	57	11	12
19	Cibrica R.	Bulgaria	43.674934	23.451674								12	3
20	Katuneshka R.	Bulgaria	43.233101	24.412033	1						2	19	5
21	Tzibritza R. trib.	Bulgaria	43.601129	23.049294								11	5
22	Archar R.	Bulgaria	43.813449	22.840116					1			4	5
23	Danube R.	Bulgaria	44.079952	23.030558			4(3)	E17(1), E45(1)			2		
24	Danube R.	Bulgaria	43.794074	23.076914			1(1)	E55(1)			2		
25	Danube R.	Bulgaria	44.094977	22.987080			1(1)	E45(1)			9		
26	Lom R.	Bulgaria	43.813003	23.245475					3(3/MLL H=2; MLL G=1)	E31(2)	11	9	34

Notes: Biotypes are indicated as follows: EE, *C. elongatoides*; NN, *C. tanaitica*; EN, *C. elongatoides-tanaitica* diploid hybrid; 3n EN, *C. elongatoides-tanaitica* triploid hybrid; SS, *C. strumicae*; 3n ENS, *C. elongatoides-tanaitica-strumicae* triploid hybrid. Numbers of biotypes used for microsatellite analyses are indicated in parentheses. Haplotype distribution of the cytochrome *b* gene is shown with absolute sample frequencies in parentheses. Distribution of detected multilocus lineages (MLLs) and multilocus genotypes (MLGs) is also shown.

Table 2. Crossing experiments.

female parent	biotype	origin	male parent	biotype	origin	progeny			Family ID
						sexual	clonal	polyploid	
EENF1	EEN	Okna R., Slovakia	EEM1	EE	Okna R., Slovakia	0	6	1	No. 401
EENF24	EEN	Okna R., Slovakia	EEM8	EE	Okna R., Slovakia	0	16	0	No. 424
EENF25	EEN	Okna R., Slovakia	EEM8	EE	Okna R., Slovakia	0	9	0	No. 425
EENF10	EEN	Ipeľ R., Slovakia	EEM9	EE	Nová Říše, Czech R.	0	9	0	No. 413
097CENNF1	EN	Jantra R., Bulgaria	09EXPM7C	PP	Veleka R., Bulgaria	0	6	4	No. 8
10EXF1F9C08	TP	laboratory hybrid	10EXF1M9C08	TP	laboratory hybrid	11	6	9	No. 1
F1TFPM062	TP	laboratory hybrid	F1TFPM065	TP	laboratory hybrid	6	0	0	No. 17; clutch A
F1TFPM062	TP	laboratory hybrid	F1TFPM066	TP	laboratory hybrid	1	0	0	No. 17; clutch B

Notes: E, haploid *C. elongatoides* genome; N, haploid *C. tanaitica* genome; P, haploid *C. pontica* genome; T, haploid *C. taenia* genome; For each family, we indicate different types of progeny: “sexual” denotes a number of progeny obtained from segregating gametes; “clonal”, a number of progeny obtained from clonal gametes; “polyploid”, a number of progeny obtained from fertilized clonal gametes. Note that several clutches from different F1 individuals occurred in the experimental family No. 17.

Table 3. Summary of the coalescent models.

data	model	AIC	evid ratio	-ln$L$	$M=4Nm$	$M_c=4Nm_c$	$\theta_{c1}=4N_{c1}\mu$	$\theta_{c2}=4N_{c2}\mu$	$\theta_1=4N_1\mu$	$\theta_2=4N_2\mu$	$\theta_a=4N_a\mu$	$\theta_{a1}=4N_a\mu$	t_1 (Mya)	t_0 (Mya)
EE-TT dataset	IIM8	5171.922	1.000	2577.961	0.137 (0.050 , 0.308)		0.287 (0.190 , 0.374)	0.122 (0.080 , 0.161)	1.336 (0.920 , 1.910)	3.471 (1.512 , 35.370)	0.989 (0.530 , 1.378)		0.678 (0.376 , 0.997)	8.782 (7.212 , 10.151)
2258 loci	IM9	5173.636	0.424	2577.818	0.116	0.027	0.288	0.123	1.372	3.601	1.033		0.693	8.605
30557 SNP	IIM7	5175.056	0.209	2580.528	0.125		0.236	0.116	1.560		1.008		0.544	8.785
	I7	5186.492	0.001	2586.246			0.284	0.125	1.660	4.763	1.395		0.712	6.993
	IM5	5244.716	0.000	2617.358		0.045	0.556	0.210			0.759			10.019
	IM4	5305.478	0.000	2648.739		0.032	0.362				0.722			9.830
	I4	5364.354	0.000	2678.177			0.677	0.356			2.132			5.236
EE-NN dataset	IIM8	6547.768	1.000	3265.884	0.166 (0.042 , 0.576)		0.474 (0.326 , 0.584)	0.289 (0.193 , 0.367)	1.294 (0.898 , 2.015)	7.972 (1.847, >63.289)	0.827 (0.189 , 1.276)		1.490 (0.672 , 2.149)	9.151 (7.811 , 11.011)
2601 loci	IM9	6549.590	0.402	3265.795	0.170	0.012	0.486	0.300	1.344	17.206	0.860		1.650	8.990
36748	IIM7	6555.918	0.017	3270.959	0.083		0.308	0.216	1.409		0.927		0.660	8.840

data	model	AIC	evid ratio	-ln$L$	$M=4Nm$	$M_c=4Nm_c$	$\theta_{c1}=4N_{c1}\mu$	$\theta_{c2}=4N_{c2}\mu$	$\theta_1=4N_1\mu$	$\theta_2=4N_2\mu$	$\theta_a=4N_a\mu$	$\theta_{a1}=4N_a\mu$	t_1 (Mya)	t_0 (Mya)
SNP														
	I7	6560.946	0.001	3273.473			0.457	0.284	1.613	8.860	1.255		1.440	7.470
	IM5	6615.996	0.000	3302.998		0.043			0.578	0.388	0.538			9.463
	IM4	6628.952	0.000	3310.476		0.037	0.485				0.540			10.380
	I4	6715.904	0.000	3353.952			0.695	0.578			1.997			5.750
EE-PP dataset	IIM7	7516.946	1.000	3751.473	0.055 (0.022 , 0.117)		0.257 (0.061 , 0.410)	0.315 (0.081 , 0.481)	1.223 (0.968 , 1.671)		1.115 (0.754 , 1.533)		0.495 (0.0664 , 1.184)	8.811 (7.871 , 10.165)
2781 loci	IIM8	7518.776	0.401	3751.388	0.058		0.257	0.326	1.276	1.189	1.114		0.510	8.817
39923 SNP	IM9	7520.776	0.147	3751.388	0.058	0.000	0.257	0.326	1.276	1.189	1.114		0.510	8.817
	I7	7531.118	0.001	3758.559			0.272	0.338	1.450	1.363	1.484		0.587	7.371
	IM5	7571.364	0.000	3780.682		0.035	0.564	0.644			0.906			9.787
	IM4	7571.404	0.000	3781.702		0.037	0.615				0.922			9.820
	I4	7634.916	0.000	3813.458			0.701	0.784			1.854			6.542
NN-PP dataset	I6	6233.448	1.000	3110.724			0.251 (0.193 , 0.316)	0.370 (0.285 , 0.470)			2.898 (2.195 , 3.976)	1.227 (1.052 , 1.395)	0.669 (0.485 , 0.875)	5.878 (4.151 , 7.964)

data	model	AIC	evid ratio	-ln$L$	M=4Nm	M_c=4Nm$_c$	θ_{c1}=4$N_{c1}\mu$	θ_{c2}=4$N_{c2}\mu$	θ_1=4$N_1\mu$	θ_2=4$N_2\mu$	θ_a=4$N_a\mu$	θ_{a1}=4$N_a\mu$	t_1 (Mya)	t_0 (Mya)
2753 loci	IIM7	6234.280	0.660	3110.140	2.550		0.216	0.343	0.527		2.809		0.476	5.609
39307 SNP	IIM8	6235.592	0.342	3109.796	4.496		0.208	0.389	0.813	0.382	2.809		0.503	5.646
	IM9	6237.592	0.126	3109.796	4.495	0.000	0.209	0.390	0.816	0.383	2.820		0.505	5.669
	IM5	6248.156	0.001	3119.078		0.294	0.262	0.405			2.887			6.170
	I4	6259.512	0.000	3125.756			0.207	0.301			1.464			0.554
	IM4	6259.806	0.000	3125.903		0.368	0.328				2.885			6.130
	I7	6265.512	0.000	3125.756			0.207	0.301	1.780	1.194	1.146		0.554	0.554
TT-PP dataset	I6	4742.940	1.000	2365.470			0.080 (0.049 , 0.117)	0.254 (0.153 , 0.380)			2.627 (1.874 , 3.965)	1.149 (0.989 , 1.307)	0.294 (0.166 , 0.451)	5.518 (3.449 , 8.269)
2394 loci	IIM7	4744.286	0.510	2365.143	4.293		0.069	0.247	0.521		2.605		0.230	5.250
32490 SNP	IIM8	4746.158	0.200	2365.079	2.418		0.067	0.217	0.390	0.675	2.601		0.208	5.190
	IM9	4748.158	0.074	2365.079	2.418	0.000	0.068	0.217	0.391	0.677	2.612		0.209	5.211
	I4	4759.906	0.000	2375.953			0.069	0.207			1.350			0.257
	IM5	4761.082	0.000	2375.541		0.089	0.079	0.236			1.405			0.365

data	model	AIC	evid ratio	-ln*L*	M=4Nm	M_c=4Nm$_c$	θ_{c1}=4$N_{c1}\mu$	θ_{c2}=4$N_{c2}\mu$	θ_1=4$N_1\mu$	θ_2=4$N_2\mu$	θ_a=4$N_a\mu$	θ_{a1}=4$N_a\mu$	t_1 (Mya)	t_0 (Mya)
	I7	4765.906	0.000	2375.953			0.069	0.207	0.055	0.447	1.350		0.257	0.257
	IM4	4844.382	0.000	2418.191		0.207	0.147				1.446			0.450
TT-NN dataset	IIM7	4637.858	1.000	2311.929	2.139 (<2.139 , 18.504)		0.113 (0.071 , 0.151)	0.213 (0.142 , 0.278)	0.713 (0.554 , 0.853)		2.907 (1.990 , >3.450)		0.458 (<0.362 , 0.675)	7.898 (<7.447 , 11.891)
2298 loci	I6	4641.116	0.196	2314.558			0.133	0.234			2.813	1.666	0.620	8.162
30987 SNP	I7	4642.192	0.115	2314.096			0.090	0.218	89722.250	0.494	1.731		0.395	0.886
	I4	4643.410	0.062	2317.705			0.130	0.226			1.852			0.605
	IM5	4645.410	0.023	2317.705		0.000	0.130	0.226			1.852			0.605
	IM4	4675.530	0.000	2333.765		0.000	0.171				1.817			0.630
	IIM8				NO CONVERGENCE		NO CONVERGENCE				NO CONVERGENCE			
	IM9				NO CONVERGENCE		NO CONVERGENCE				NO CONVERGENCE			

Notes: For each dataset we indicate the species-pair included (EE, *C. elongatoides*; TT, *C. taenia*; NN, *C. tanaitica*; PP, *C. pontica*), the number of loci analysed, number of SNP used in each pairwise species comparison, and the results obtained for the seven and eight coalescent models, respectively. The models are ranked according to their AIC score, with the best-fitting model shown at the top and the worst model at the bottom. For each model we indicate the number of parameters and give the values of the AIC score, the evidence ratio, the negated loglikelihood (-ln*L*), and the ML estimates obtained, with 95% confidence intervals based on the profile likelihood shown in brackets for the best-fitting model only. When computing these confidence intervals, it was not always possible to obtain the precise upper or lower bound of the interval (in some cases the computational procedure broke off before the required confidence level was reached). In such cases the upper or lower bounds of the intervals are indicated as less than (<) or greater than (>) the nearest value that could be obtained.

Supporting Information

Table S1. Microsatellite markers and cytochrome *b* gene haplotypes. This Table lists the alleles found in each sampled individual (specified by a respective ID number) on the basis of eight microsatellite markers. The cytochrome b gene haplotype is also specified.

Table S2. Allelic profiles of parents and their progeny within crossing experiments on the basis of seven microsatellite markers.

Table S3. Oocyte analysis. For each female we indicate her biotype, origin, allozyme profile and the number of analyzed eggs.

Appendix S4. SNP dataset. Every contig is represented by ten lines, starting with its name (ref_acc) on the first line and followed by nine lines corresponding to nine individuals. Individuals' genotypes are indicated by letters as follows: s, *C. strumicae*; p, *C. pontica*; e, *C. elongatoides*; t, *C. taenia*; n, *C. tanaitica*. Variant positions are expressed using IUPAC notation and invariant positions are excluded.

Appendix S5. Summary of genetic divergences of fish species producing asexual hybrids.