Genome fractionation and loss of heterozygosity in hybrids and polyploids: mechanisms, consequences for selection and link to gene function

Karel Janko¹,², Oldřich Bartoš³,⁴, Jan Kočí¹,², Jan Roslein¹,², Edita Janková Drdová⁴, Jan Kotusz⁵, Jan Eisner⁶, Eva Štefková-Kašparová¹,⁷

¹, Laboratory of Fish Genetics, Institute of Animal Physiology and Genetics of the CAS, Rumburská 89, 277 21 Liběchov, Czech Republic
², Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic
³, Department of Zoology, Faculty of Science, Charles University, Viničná 7, 128 44 Prague, Czech Republic
⁴, Institute of Experimental Botany, Academy of Sciences of the Czech Republic, Prague 6, Czech Republic
⁵, Uniwersytet Przyrodniczy we Wroclawiu, Museum of Natural History, Wroclaw, Poland
⁶, Department of Mathematics and Biomathematics, Faculty of Science, University of South Bohemia, Branišovská 1760, 37005, České Budějovice, Czech Republic.
⁷, Czech Life Science university, Prague, Czech Republic
Abstract:

Hybridization and genome duplication may cause serious damages but may also open unique opportunities to invade new ecological niches or adapt to novel environments better than their parents. Following the initial merging or multiplications, the subgenomes of hybrids and polyploids undergo considerable changes, often eliminating segments of one parental genome, phenomena known as loss of heterozygosity (LOH) and genome fractionation. Mechanisms causing such changes are not well understood, and remain enigmatic particularly when hybridization is linked with asexual (clonal) reproduction that may enforce diverse array of genome evolutionary pathways ranging from long-term stasis to dynamic reformation.

Analysis of genome evolution in diploid and polyploid clonal hybrids between fish *Cobitis elongatoides* and either *C. taenia* or *C. tanaitica* species revealed that clonal genomes remain generally static on chromosome-scale level but undergo small-scale restructurations resulting in genome fractionation and LOH events. These events have complex molecular background in two distinct processes, the hemizygous deletions and conversions between orthologous subgenomes. The impact of both processes on clonal evolution is ploidy-dependent; while deletions frequently accumulated in polyploids, they appeared to be selected against in diploid asexuels where gene conversions prevailed. The incidence of genomic restructuration was not random with respect to individual genes, but it preferentially affected loci with unusually high transcription levels, genes under relatively strong purifying selection and also genes with particular functions, such as those related to endoplasmatic reticulum. Likelihood that given ortholog would be retained or lost correlated significantly with its parental origin, GC content (preferential loss of low-GC alleles) and expression (less expressed alleles tended to be replaced by more expressed ones). Contrary to expectations, however, we observed that the preferentially retained subgenome (the one derived from *C. elongatoides*) was not dominant at the transcription level as all hybrids were phenotypically more similar to the other parent whose genes were preferentially lost.

Our data show that the fate of subgenomes in asexual hybrids and polyploids depends on complex interplay of molecular mechanisms and selection that are affected by sequence composition, expression as well as parental ancestry.
Introduction:

The genome of a typical Metazoan possesses two alleles of each gene brought together by merging reduced gametes of two individuals belonging to the same species. Yet, these rules have often been alleviated as traces of whole genome duplications (WGD) and introgressive hybridizations have been documented in many taxa, vertebrates and humans included (Dehal and Boore 2005; Gittelman et al. 2016). Hybridization and polyploidization may cause serious problems, e.g. in transcription regulation leading to deregulation of transposable elements ((Dion-Côté et al. 2014; Lien et al. 2016)), but may also lead to creation of novel traits and acquisition of gene functions via sub-/neofunctionalization of duplicated genes (Yoo et al. 2014; Fridman 2015), potentially facilitating specialization to new niches (Madlung 2013).

Realizing their evolutionary significance and huge practical value to mankind, e.g. (Mason and Batley 2015), the research focus on hybridization and polyploidy intensified and lead to revelation of some prominent patterns. For instance, hybrid's phenotypes may range from intermediate forms to transgressive expression of novel traits (Bell and Travis 2005; Yoo et al. 2014; Bartoš et al. 2019) but often, one parental subgenome is more expressed than the other one (i.e. expression dominance; (Yoo et al. 2014; Alexander-Webber et al. 2016; Cheng et al. 2018). Hybrid's and polyploid's genomes evolve dynamically and often lose orthologous genes from one or the other parental species in processes referred to as loss of heterozygosity (LOH), genome fractionation or rediploidization in polyploids (Yoo et al. 2014; Lien et al. 2016; Du et al. 2020). These are often asymmetrical, e.g. (Alexander-Webber et al. 2016) and it has been proposed that loss of alleles from the less expressed subgenome may cause less severe effects and may therefore be preferred by selection (Yoo et al. 2014). However, the situation is likely more complex as orthologue loss/retention of may also be under the selection for proper dosage of molecular interactors and relative copy number of gene products (i.e. selection for stoichiometry (Birchler and Veitia 2012)) and hybrid populations may selectively filter orthologous genes according to their adaptive value in a given environment, e.g. (Gittelman et al. 2016; Lancaster et al. 2019; Smukowski Heil et al. 2019).

Thus, despite application of modern technologies, the question of why some genes tend to be retained in heterozygous or duplicated states, while others are subject to fractionation, still represents major evolutionary puzzle. It remains particularly unclear whether aforementioned patterns are driven by case-specific mechanisms or whether independent lineages follow similar evolutionary trajectories (Soltis et al. 2010; Deans et al. 2015).
Such a gap in current knowledge partly results from taxonomic bias in knowledge, especially towards plant species, since the incidence of hybridization and polyploidy have traditionally been underrated by zoologists. Moreover, direct tests of adaptive values of genomic rearrangements could be performed only under laboratory conditions in rapidly reproducing organisms (Smukowski Heil et al. 2017; Lancaster et al. 2019) since events like LOH are rather rare (Dukić et al. 2019). For practical reasons, most available data come from natural hybrids and polyploids, making it difficult to discern which patterns are direct consequences of genome merging and which evolved subsequently. In addition, many polyploids are of hybrid origin, which makes it challenging discerning what effects are inherent to polyploidy and which to hybridization itself. Finally, in many investigated hybrid and/or allopolyploid taxa, it is not even clear how they could establish themselves in natural environments since any new form is rare at the time of its emergence and therefore is threatened by frequency-dependent mating disadvantage and backcrossing with dominating ancestral populations (i.e. the minority cytotype exclusion principle: e.g. (Husband 2000)).

As one of the potential explanations for the establishment of such new strains, it has been hypothesized that they may alleviate initial caveats by taking advantage of asexual reproduction via unreduced gametes, which offers immediate reproductive isolation and clonal multiplication of novel genotypes unencumbered by maintaining two sexes (Cunha et al. 2008; Choleva and Janko 2013; Hojsgaard and Hörandl 2015; Dubey et al. 2019). The perception of asexual organisms indeed changed among biologists from being initially considered marginal evolutionary accident (Howell 1933) to current appreciation that they occur in all major Eukaryotic clades (Schön et al. 2009) and form dominant components of some ecosystems (Kearney 2005; Hojsgaard and Hörandl 2015). Emergence of asexual reproduction is tightly linked to hybridization and polyploidy (rev. in (Choleva and Janko 2013)), and may represent an inherent stage of the speciation process, representing a special type of Bateson-Dobzhansky-Muller (Janko et al. 2018). This paradigm shift coincides with increasing interest about the role recombination modification in evolution (Thompson and Jiggins 2014; Ortiz-Barrientos et al. 2016) suggesting that understanding evolutionary processes in asexual organisms may provide important insights into the mechanisms of genome evolution in general.

Unfortunately, there is no consensus about genomic consequences of asexuality. Clonal inheritance has been originally assumed to ensure stasis of genome with gradual accumulation of deleterious mutations (Muller 1964; Keightley and Otto 2006) and heterozygosity levels (Birky 1996; Mark Welch and Meselson 2000; Baloux et al. 2003) or modified dynamics of transposable elements (Hickey 1982). This view is currently challenged by indications of horizontal gene transfers in some asexual lineages (Gladysev et al. 2008;
Danchin et al. 2010) as well as by accumulating evidence that genomes may extremely quickly acquire aneuploidy or structural changes once the sex is lost, owing to relaxed constraints on the pairing of homologous chromosomes (Triantaphyllou 1981; Sunnucks et al. 1996; Normark 1999; Spence and Blackman 2000; Tucker et al. 2013). Heterozygosity may degrade quickly by hemizygous deletions and specially by gene conversions (Tucker et al. 2013), whose inherent GC bias may lead to increased guanine & cytosine contents in asexual genomes (Bast et al. 2018). The dynamics of asexual's genome may also be determined by its mode of origin since hybrid asexuals generally express high levels of heterozygosity, probably indicating efficient clonal transmission of parental genomes, while nonhybrid asexuals appeared to have lost most of their heterozygosity in consequence of pervasive gene conversions (Jaron et al. 2018).

With such a variety of patterns it is difficult to discern which mechanisms are taxon-specific and which relate to asexual reproduction per se. A major complication is that the so-called asexual organisms form very heterogeneous group employing a wide specter of gametogenetic mechanisms, ranging from ameiotic processes (apomixis) to those involving more or less distorted meiotic divisions (automixis) (Stenberg and Saura 2009; Stenberg and Saura 2013). Some types of automixis involve intragenomic recombinations or exclusions of large genomic parts, thereby decreasing heterozygosity among progeny (Bi and Bogart 2006) while other types are theoretically equivalent to apomixis. For example, many hybrid asexuals employ premeiotic endoduplication, which is a mechanism when normal meiosis is preceded by WGD in oogonial cells. In consequence, segregation and recombination presumably takes place between sister copies of chromosomes rather than between orthologs, resulting in clonal progeny (Lutes et al. 2010; Arai and Fujimoto 2013) Fig. 1a.

In this study we analyze causes and consequences of allelic recombination, conversion and loss of heterozygosity in a clonally reproducing vertebrate of hybrid origin, Cobitis. Actinopterygii. We focus on the so-called Cobitis taenia hybrid complex, which arose by hybridization of the species C. elongatoides (we denote its haploid genome ‘E’) with either of its two distant relatives, C. taenia (‘T’) or C. tanaitica (‘N’)(Choleva et al. 2012). Phylogenomic analysis (Janko et al. 2018) revealed that initial E-(TN) divergence occurred ca 9 Mya and was initially followed by intensive gene exchange. However, with ongoing divergence these species lost the capacity to produce sexual hybrids since crossings of current species lead to sterile hybrid males and fertile, but clonally reproducing, hybrid females. Hybrid females form unreduced gametes by premeiotic endoduplication but are gynogenetic, i.e. they require a sperm from sexual species to trigger development of their gametes (Janko et al. 2007; Choleva et al. 2012; Dedukh et al. 2020). Usually, sperm’s genome is degraded after fertilization but certain proportion of oocytes fuses with the sperm.
cells and consequently, diploid ET or EN females produce certain portion of triploid progeny that may have EET, ETT, EEN or ENN genomic constitution, depending on the sperm donor. Triploids also reproduce clonally. The hybridization between parental species is reciprocal but C. taenia is maternal ancestor of most elongatoides-taenia hybrids, while C. elongatoides is maternal ancestor of elongatoides-tanaitica hybrids (Janko et al. 2003). Because parental species lack obvious prezygotic reproductive barriers and glacial-interglacial cycles repeatedly brought their ranges into contact (Fig. 1b), new clones have been dynamically arising throughout much of Pleistocene and subsequently colonized Europe (Janko et al. 2005; Janko et al. 2012). Current asexual populations consist of recent clones with postglacial origin in Central European hybrid zone between C. elongatoides and C. taenia (these involve diploid ET and triploid EET and ETT forms) but also of ancient elongatoides-tanaitica hybrids (the so-called hybrid clade I, consisting of EN and EEN biotypes) which originated in Balkan hybrid zone ca 300 ky ago (Janko et al. 2005).

Cobitis hybrids thus offer excellent opportunity to investigate how genomes of natural asexual organisms evolve through time and across different ploidy levels. (Majtánová et al. 2016) demonstrated by karyotypic analysis that clonal hybrids maintain remarkable integrity of parental chromosomes without traces of large-scale recombinations and restructurations despite more than 300 ky of evolution since the initial hybridization event. This study investigates the dynamics of diploid and polyploid clonal genomes on the fine scale. To achieve this aim, we performed exome sequencing of sexual parental species and their clonal hybrids, and we subsequently compared such data to recently published gene expression profiles (Bartoš et al. 2019). This allowed us to identify mechanisms underlying the fractionation and heterozygosity loss in clonal genomes and to test how they relate to expression and function of affected genes.
3. Results

3.1. SNP detection and identification of species-specific variants

Exome-capture data were acquired from 46 specimens, including three parental species and their asexual hybrids, which were selected to cover all major phylogroups, ploidy types and hybrid genome compositions (see Fig. 1b and Table S1 for details) and we also included whole genome sequencing of one ET hybrid for control. Reads were mapped against published C. taenia reference transcriptome containing 20,600 contigs (Janko et al. 2018) and for simplicity, we restricted our analysis to 189,927 quality filtered bi-allelic SNPs (i.e. occurring in no more than two states across the entire dataset). The greatest genetic divergence was between C. elongatoides and the remaining two sister species, C. taenia and C. tanaitica, while hybrids appeared intermediate (multidimensional scaling (MDS), Fig. 2a).

Examination of pairwise genetic distances between hybrid individuals revealed 11 clusters of individuals that were nonrandomly similar to each other, therefore likely representing clonal lineages. We thus refer to them as independent Multilocus Lineages (MLL) sensu (Arnaud-Haond et al. 2007), see Fig. 2a and Table S2.

To characterize hybrid's SNP variation relative to their parental species, we followed (Ament-Velásquez et al. 2016) and divided all hybrids' SNPs into 10 categories (Table S2), of which five were particularly important for this paper. First, we detected 16,372 unique positions with variants occurring only among asexuals but not among sexuals (SNP categories "private-asexual 1a and 1b"). These SNPs presumably represent mutations acquired after clonal origin (Ament-Velásquez et al. 2016; Kočí et al. 2020). Notably, the proportions of such SNPs in genome of each hybrid were tightly correlated with its mtDNA distance from the nearest sexual counterpart (Pearson's r=0.971, df=23, p-value=8.66e-16), so that the ancient elongatoides-tanaitica hybrids had the highest amount of private asexual SNPs, while experimental F1 hybrids had the least amount of such SNPs, probably representing only rare sequencing errors. Second, we focused on SNP variants that were diagnostic between pairs of parental species, so that their hybrids should possess one or both parental variants, thereby allowing detection of LOH events. Throughout the entire dataset we therefore identified sites diagnosing C. elongatoides from C. taenia (referred to as E-T diagnostic sites; total of 37,988), C. elongatoides from C. tanaitica (E-N diagnostic sites; 30,281) and we also found SNPs differentiating C. elongatoides from the joint dataset of C. taenia and C. tanaitica (E-TN diagnostic sites; 27,311). According to the way how hybrids' SNPs were shared with these parental variants, we categorized them as “shared SNPs” of type sh3a (heterozygous for both parental variants), sh3b11 (homozygous for one parent's allele) and sh3b12 (homozygous for the other parent's allele) in Table S2.
3.2. Clonal lineages accumulate Loss of heterozygosity events in their evolution.

Hybrids were considerably more heterozygous than parental species (Fig. 2b; Wilcoxon rank sum test: $W = 520$, p-value $< 1e-9$) with no less than 98.5% of private asexual SNPs and the vast majority of diagnostic sites occurring in heterozygous states. Nevertheless, LOH was observed in some portion of diagnostic SNPs of every hybrid (categories sh3b11 and sh3b12 in Table S2). We verified the quality of base-calling and LOH detection by two approaches. We first compared SNP calling from exome-capture technology and whole-genome sequencing of the same ET hybrid (csc067) and found differences in only ~0.17% E-T diagnostic positions. We also compared two F1 hybrids against their parents and found homozygote states only in ~0.16% positions, where both parental individuals differed from each other. This indicates high reliability of LOH detection based on exome capture.

Two patterns were noted in the distribution of LOH SNPs. First, individual LOH sites were significantly more likely to be shared by individuals belonging to the same clonal lineage (MLL) than by individuals from different clones (Wilcoxon rank sum test, $W = 4540$, p-value $< 1e-9$; Fig. 2c). Second, the proportion of LOH sites in each individual significantly correlated with its proportion of private asexual mutations (Pearson’s $r = 0.955$, 95% c.i. = 0.902-0.980, p-value = 3.286e-14; Fig. 3a, Table S2). Hence, although we may not rule out existence of somatic mutations (López and Palumbi 2019), our data indicate that erosion of heterozygosity is heritable within clonal lineages, accumulates over clone’s evolutionary history, and therefore affects the germline.

We also performed the same analysis on E-TN diagnostic sites, in order to minimize potential effect of ancestral polymorphism when same allele might have been inherited from both parents at time of clonal origin but was subsequently lost in one parental species. Since C. taenia – C. tanaitica divergence predates origin of oldest Cobitis clones by hundreds of thousands of years (Janko et al. 2018), it may be posited that most such E-TN sites became diagnostic long before the origin of studied clones. Yet, the proportions of LOH at E-TN diagnostic sites also correlated significantly with private asexual SNPs ($r = 0.948$, 95% c.i. = 0.885-0.977, p-value = 2.145e-13) albeit with slightly less steep slope than in ET- and E-N sites, respectively (Fig. 3a). This suggests that some false-positives might have affected our dataset, but altogether the retention of ancestral polymorphism is an unlikely explanation of most observed LOH events.

3.3. Heterozygosity deteriorates by gene conversions and hemizygous deletions in asexuals.

We next investigated topological context of LOH sites to test if they might have been generated by point mutations. Since we used cDNA reference that excludes introns, we could not simply analyze the physical distance between SNPs in studies loci. Instead, we
tested if LOH sites within individual genes tend to occur in contiguous stretches. We thus compared observed distribution of LOH sites to permuted datasets with LOH sites randomly distributed across all genes (see the Methods for details). We observed that numbers of LOH occurring in contiguous stretches significantly exceeded simulated values, suggesting that LOH sites tend to occur in clusters. This is inconsistent with point mutations being the primary mechanism generating LOH events but instead suggests that most LOH events are created by processes affecting contiguous stretches of DNA, like gene conversions and gene deletions.

To distinguish between both candidate processes, we analyzed the sequencing coverage following (Tucker et al. 2013), who showed that conversions conserve the amount of allelic copies, while allelic deletions would result in coverage drop. To do so, we first confirmed that exome capture data are suitable for coverage comparisons of sequencing by analyzing among-individual variance in sequencing coverage across loci (Bragg et al. 2016). We found that capture efficiency considerably varied among loci but was highly correlates among individuals, thereby allowing inter-individual coverage comparisons to be made (see Appendix S1).

We then used normalized per-SNP coverages to calculate relative values of coverage for each hybrids’ LOH by comparing it to the coverages of the same site in parental species (see Methods). Following (Tucker et al. 2013) we predicted that conversions result in relative coverage ~1, while allelic deletions would result in coverage drop to values ~0.5 in diploids, or ~0.66 (single deletion) and ~0.33 (double deletion) in triploids. To investigate roles of both processes in LOH creation, we constructed for each hybrid biotype the histograms of relative coverages and tested their modality at aforementioned biologically relevant values. Ancient clones (EN and EEN) possessed relatively smooth distributions of relative normalized coverages with peaks close to 1, suggesting gene conversions as the main mechanism causing their LOH events. In contrast, recent clones (ET, EET and ETT) had additional peaks located at lower values (Fig. 4a-e), indicating simultaneous operation of both processes (Tucker et al. 2013).

To formally test whether observed data may be explained by single process or several simultaneously operating processes, we compared the fits of each histogram by single gamma distribution as a proxy for operation of only one process, or mixtures of more distributions with means fixed at aforementioned biologically relevant values. Nonlinear least square method was used to estimate the parameters controlling relative proportions of distributions in combined models (A and B parameters control the rate of conversions / hemizygous deletions in two-gamma distribution model while A, B & C control the rates of conversions & hemizygous & double deletions in three-distribution model). The most complex model assuming the occurrence of conversions and both single and double
deletions (three gamma distributions) did not significantly improve the fit to triploids’ data. However, mix of two gamma distributions assuming gene conversions and hemizygous deletions significantly outperformed any single-distribution in all datasets (F test in diploids: EN: df = {40,38}, F=2.583; ET: df = {41,39}, F=7.87; triploids: EEN: df = {44,42}, F=11.01; EET: df = {44,42}, F=112.6; ETT: df = {31,29}, F=57.49), suggesting that joint operation of both processes.

To further validate if deletions indeed occur in hybrids, we used Kolmogorov-Smirnov test to compare distributions of relative coverages at hybrids’ LOH sites with the distribution of coverages at the same sites in parental species, where no deletions are expected. This comparison indicated all hybrid biotypes have significant excess of low-coverage LOH sites, again suggesting that deletions cause coverage decrease in hybrids (see Fig. 4e and Fig. S1 for details).

Both tests thus documented the existence of LOH sites with decreased DNA content suggesting that **LOH events are caused by simultaneous operation of conversions and hemizygous deletions in all biotypes. However, double deletions in triploids are very rare or absent.**

### 3.4. Accumulation of LOH is biased with respect to parental subgenome, ploidy and hybrid type.

We noted that LOHs were non-randomly distributed among hybrids, and there were several trends behind such unevenness. First, there was a clear bias in retention of parental subgenomes (Fig. 3b). Virtually all LOH sites detected in triploids possessed allele of that parent which contributed two chromosomal sets. This type of bias in triploids is likely methodological, since sites where one allele of the diploid subgenome is lost would still appear heterozygous and hence escape our attention. However, significant bias was observed in diploid hybrids with **preferential retention of C. elongatoides allele** at ~80% LOH sites in ET and ~87% in EN hybrids.

Second, the **hemizygous deletions were significantly more common in triploid hybrids than in their diploid counterparts**. Specifically, the ratio of A/B parameters of combined Gama distributions suggest that deletions accounted for only ~21% LOH events in diploid hybrids between *C. elongatoides* and *C. taenia*, while their contribution rose to ~50% in triploid hybrid forms (EET and ETT); Fig. 4a-e. Similarly, in *elongatoides-tanaitica* hybrids, deletions explained accounted for less than 0.1% LOH events in diploid EN hybrids, while triploid EEN possessed ~18% deletions at LOH sites. These differences appeared highly significant after comparing data fitting by ‘free’ mixed gamma model using optimized A / B ratios and by ‘forced’ model with A / B ratio fixed to values estimated from triploids (EN with...
Finally, hemizygous deletions were more common among recent asexuals than in ancient clones, as evident from comparisons of A / B ratios between recent (ET) and ancient (EN) diploid clones (~21% vs. ~0.1%) as well as of recent (EET or ETT) and ancient (EEN) triploid clones (~50% vs ~18%); Fig. 4a-e.

3.5. Occurrence of LOH is related to sequence composition, allelic expression and gene function.

3.5.1. LOH depends on GC content but patterns are complex:

To investigate potential GC bias, we first inspected transcriptome-wide GC contents of sexual and asexual forms, measured either across all positions or only at the relatively neutral third codon positions, but found no significant differences between any biotypes. Next, we performed a more fine-scale analysis on E-TN diagnostic positions and separated all detected LOH sites into E-like or TN-like groups, depending on parental allele retained. Comparing parental sequences with each hybrid we inferred how many LOH sites underwent A/T -> G/C substitutions, G/C -> A/T substitutions or no change in GC content (i.e. A <-> T or G <-> C substitutions) and used contingency tables compared these counts with overall A/T - G/C differences between respective parental species across all E-TN diagnostic positions.

We found that E-like LOH events were significantly biased in favor of A/T -> G/C substitution in triploid (EET, EEN) and diploid (ET, EN) biotypes. This bias was ~25% on average and proved significant in every individual after FDR correction. However, we observed no such GC bias in TN-like LOH sites of any biotype (Fig. 5a). Our data thus indicate that GC-dependence has complex background and occurs only during loss of taenia/tanaitica allele, but not in the opposite direction.

3.5.2. LOH is affected by gene’s transcription:

We evaluated the effects of allele expression on LOH occurrence using recently published transcription profiles of livers and oocytes in sexual (C. elongatoides, C. taenia) and asexual (ET, EET, and ETT) females (Bartoš et al. 2019). While Bartoš et al. analyzed different individuals, they used the same reference transcriptome and thus we could investigate the transcription profiles of those genes which carried E-T diagnostic SNPs and were LOH-positive (either E-like or T-like LOH bearing), or LOH-negative, according to present exome-capture results. We applied two approaches to test for differences in gene and allele expression between these categories.

First, we found that LOH events generally tend to occur in genes with abnormally high expression levels. Specifically, we compared the TPM-normalized expression levels of LOH-positive and LOH-negative genes and found that the genes where LOH occurred had
significantly higher expression levels in liver tissue of all biotypes (FDR corrected WMW test
p values < 0.02 for all biotypes). Same trends, albeit insignificant, were observed in oocytes.
Second, published DeSeq2-normalized expression levels of both parental species were
compared to test if less expressed alleles tend to be preferentially lost during LOH events.
We adopted such an indirect approach of comparing both parental species instead of directly
inspecting allelic expressions in hybrids because (Bartoš et al. 2019)’s data in hybrids might
have been affected by undetected allelic deletions. Moreover, the interparental expression
divergence is a good predictor of relative allelic expression given pervasive cis-regulation
between subgenomes (Bartoš et al. 2019). We found that distributions of C. elongatoides / C.
taenia log2 fold change (log2FC) were very similar for LOH-negative genes and for genes
bearing E-like LOH (oocytes: mean log2FC -0.22 vs -0.21; livers: mean log2FC -0.14 vs -
0.16; note that prevalence of negative values is in agreement with slight but pervasive under
expression of C. elongatoides documented by (Bartoš et al. 2019). In contrast, log2FC
values were more negative (i.e. deviated towards C. taenia) in genes bearing T-like LOH
than in LOH-negative genes (oocytes: mean log2FC -0.34 vs -0.22; livers: mean log2FC -
0.22 vs -0.14) and Kolmogorov-Smirnov test proved that such a difference was significant in
oocytes (p value = 0.022). This suggests that the fate of hybrid’s alleles is affected by
expression levels inherited from parental species. However, the effect is again not
symmetrical with respect to subgenome ancestry, since less expressed alleles are
preferentially removed during E-like LOH events while no bias was observed in the reciprocal
direction.

3.5.3. LOH accumulate in genes with specific functions
Finally, we investigated whether LOH events accumulate in genes with specific functions. For
this purpose, we selected only loci with diagnostic SNPs that were successfully annotated
and performed two tests. First, as a proxy for selection regime of particular genes, we
investigated \( d_{\text{u}}/d_{\text{s}} \) values among orthologous sequences of C. elongatoides, C. taenia and C.
tanaitica (note that \( d_{\text{u}}/d_{\text{s}} \) values for all genes were published by (Kočí et al. 2020)). We
found in both elongatoides-taenia and elongatoides-tanaitica hybrid types that LOH-positive
genes were characterized by significantly lower \( d_{\text{u}}/d_{\text{s}} \) values than LOH-negative
genes (Linear mixed effect model with pairs of individuals as random factor; LRT p value for
elongatoides-taenia = 1.604095e-135; p value for elongatoides-tanaitica = 1.518951e-05).
Next, we searched if LOH positive genes are associated with particular Gene
Ontology terms using the GO::TermFinder (Boyle et al. 2004). Results are shown in Table 1,
which lists top 20 enriched GO terms of each category in each biotype. It shows that GO
terms associated with membrane coats and endoplasmatic reticulum were enriched among
LOH-positive genes in EET, EN and EEN biotypes with p-values corrected for multiple tests.
below alpha level 0.1. It also shows that some GO terms, whose corrected p value exceeded
the threshold level, were repeatedly encountered among top enriched GO terms in several
hybrid biotypes including independent *elongatoides-taenia* and *elongatoides-tanaitica* hybrid
types. These namely contained Cellular compartment type GO terms associated with cell-cell
junction and cell-substrate adherence junction and Biologic processes type GO terms of
*cellular biogenic amine metabolic process* and *Protein glycosylation*. 
4. Discussion

Genomes of asexual organisms may evolve in various ways ranging from fast restructuration to long-term conservation of heterozygosity. Such a diversity of patterns may reflect the variety of gametogenetic pathways used by such organisms so that hybrids employing premeiotic endoreplication, like *Cobitis*, can maintain integrity of parental subgenomes without any chromosomal-scale restructurations (Majtánová et al. 2016). However, present study showed that despite apparent stasis on large scale, the heterozygosity gained during original hybridization may gradually deteriorate by small-scale restructurations that affects genomic regions in relation to allelic origin, sequence composition and gene expression.

4.1. Large-scale stasis vs small-scale dynamics of asexual genomes.

After initial merging and duplication, allopolyploid organisms appear to deduplicate their genomes prominently via fractionation and deletions of orthologues (Yoo et al. 2014; Cheng et al. 2018; Du et al. 2020). Yet, we found that in polyploid loaches, or hybrids in general, deletions accounted for rather minor fraction of restructuration events. The majority of LOH sites had relative coverage ~1, thereby indicating higher incidences of recombinations between orthologues. Recombination may be followed by crossover (CO), which is expected to cause long stretches of LOH spanning till another recombination site or until the telomeric ends of paired chromosomes (Fig. 1c). But as (Majtánová et al. 2016)’s cytogenetic study ruled out any large-scale exchanges of chromosomal arms between subgenomes, it appears that most LOH events detected in this study were caused by gene conversions without crossovers.

Interestingly, organisms employing premeiotic endoreplication should form bivalents between sister copies of duplicated homologs, rather than between orthologous chromosomes (Lutes et al. 2010; Arai and Fujimoto 2013; Dedukh et al. 2020) and it is therefore unclear how such conversions may arise (Fig. 1c). Theoretically, they may result from errors in homology search during early meiosis if ExT bivalents form, but this explanation is unlikely for two reasons. First, *C. elongatoides* and *C. taenia* karyotypes are so divergent that most orthologous chromosomes may not form proper bivalents leading to sterility of those hybrid forms, which lack endoreplication, typically males (Dedukh et al. 2020). Hence, even if ectopic ExT pairings occur in asexual females, it would be unlikely to form proper bivalent. Second, (Dedukh et al. 2020) documented the occurrence of true crossovers in ExE and TxT bivalents in female hybrids (Fig. 1c), and hence hypothetical ExT bivalents would result in the exchange of large pieces of chromosomal arms, which have not been observed (Majtánová et al. 2016). An alternative explanation would therefore assume the role of mitotic conversions, which are important in DNA damage repair (Helleday 2003).
Indeed, mitotic conversions have been hypothesized to impact the evolution of asexuals (Omilian et al. 2006; Mandegar and Otto 2007) although, to our knowledge, they haven’t been directly observed in any multicellular asexual organism.

Whatever the underlying mechanism, the fact that LOH sites are heritable and shared among clone-mates suggests that LOH events occur in the germline. The genes affected by LOH events also do possess some characteristics typical of loci undergoing conversions. Namely, LOH positive genes have above-average expression levels, which is consistent with the hypothesis that DNA of transcriptionally active loci is more relaxed and hence prone to double strand breaks (DSB) followed by repair cascade including the recombination machinery (González-Barrera et al. 2002; Cummings et al. 2007). *Cobitis* hybrids also tend to replace the less expressed parental allele by the more expressed one, which is in line with growing evidence that more transcribed homoeologs are preferably utilized as templates during double strand break (DSB)-induced gene conversion (Schildkraut et al. 2006). Finally, the prevalence of AT->GC substitutions on some LOH sites conforms to expected GC bias in template preference (Duret and Galtier 2009; Williams et al. 2015).

Interestingly, however, our study revealed that processes affecting asexual genomes strongly depend on ancestry of allele acting as a template. Namely, the preferential retention of more transcribed allele was apparent only during *elongatoides* -> *taenia* allele replacement (T-like LOHs), while the GC bias was detected only in the opposite direction (E-like LOHs). On the other hand, overall GC contents were not notably affected by these processes as we found no differences between sexual and asexual forms at transcriptome-wide scale. This suggests that predicted increase in GC content (Bast et al. 2018) cannot be generally applied to all types of asexuals and that the ancestry of subgenomes should be taken into account.

### 4.2. Impact of LOH on evolution of hybrids and polyploids.

Genome rearrangements may bring both, the benefits and their accumulation may be facilitated by the lack of requirement of proper homology for chromosomal pairing in asexuals (Sunnucks et al. 1996). Consequently, conversions & deletions of genes or even chromosomal arms may potentially proceed at faster rates than mutation accumulation in some asexuals (Triantaphyllou 1981; Sunnucks et al. 1996; Normark 1999; Spence and Blackman 2000; Tucker et al. 2013), and hence slow the mutational deterioration (e.g. Muller’s ratchet process) by erasing deleterious mutations or increasing the fixation rate of beneficial mutations (Khakhlova and Bock 2006; Mandegar and Otto 2007). On the other hand, recombination may have mutagenic effects on its own (Arbeithuber et al. 2015).

In any case, recent analysis of mutation accumulation and fitness deterioration proposed several reasons why LOH events do not play important role in slowing the Muller’s
ratchet in asexual loaches (Kočí et al. 2020). In brief, only less than \(~1.5\%\) of private asexual SNPs occurred in homozygous states indicating quite efficient mechanism of clonal reproduction when the majority of newly acquired mutations occur at heterozygous state on one chromosome with little possibility of recombination or conversion. Consequently, observed rate of LOH accumulation was low occurring in only \(~10\%\) genes after \(~300\)kya of evolution in the oldest clone. This is orders of magnitude less than in aforementioned asexuals, where such processes have been hypothesized to interfere with the accumulation of deleterious mutations. Finally, efficiency of LOH in mutations erasing should increase with clonal age since in recent clones, the rare LOH events would likely happen on genes, where mutations haven't yet accumulated, while in older clones the emerging LOH event have higher chance to 'heal' previously mutated genes. However, such a process should lead to exponential correlation between the proportions of LOH and the private asexual SNPs, which was not found by (Kočí et al. 2020). This is not to say that LOH events may not counteract the ratchet in some asexuals, however, their role in removal of deleterious mutation is supposedly smaller in organisms with relatively efficient clonal reproduction, like *Cobitis*.

Nevertheless, we found evidence that LOH events considerably impact the evolution of studied organisms by other mechanisms. For example, we noticed that proportions of deletions were higher among young clones and polyploid hybrids than in ancient or diploid clones.

### 4.2.1. The effects of deletions are less severe in polyploids.

Hemizygous deletions cause aneuploidies on sub-chromosomal levels, and may therefore modify stoichiometry between interacting components of molecular complexes (Birchler and Veitia 2012) or between transcription factors and their binding sites, thereby affecting the gene regulation (Veitia et al. 2013). The magnitude of their effect probably scales with the length of deleted genomic region; long deletions affecting many genes have stronger effect than short-range deletions (Veitia et al. 2013). This may explain why *Cobitis* hybrids retained stable karyotypes with no chromosomal-scale deletions (Majtánová et al. 2016), but small-scale deletions of individual genes do occur and haven't been removed by selection. (Veitia et al. 2013) further postulated that the impact of aneuploidy should depend on allelic dosage. Consequently, hemizygous deletions would have weaker effects in triploids (changing allelic copy numbers from 1 to 2/3 relative to the rest of genome) than in diploids (change from 1 to 1/2) and double deletions in triploids (change from 1 to 1/3) would have the most severe effects. This may explain why triploids possessed higher proportion of hemizygous deletions than their diploid counterparts, but in the same time they had almost no double deletions.

The hypothesis that allelic deletions have mostly negative impact may also explain why young clones possess relatively higher proportion of deletions than old ones. Indeed,
selection-based removal of deleterious mutations requires some time proportional to the
collection coefficient, population size and genetic background, and hence, although recent
clones acquired lower absolute numbers of LOH events, they would have higher fraction of
deletions due to a time-lag necessary to remove these deleterious mutations (Johnson and
Howard 2007). In fact, we observed similar differences between young and old clones in
loads of nonsynonymous mutations (Kočí et al. 2020).

4.2.2. Biased genome fractionation and template-preference.

Another prominent pattern was the strong preference of elongatoides subgenome
retention at LOH sites (Fig. 3b). Biased genome fractionation is commonly observed among
hybrids and allopolyploids and may have various explanations, ranging from mechanistic
reasons when one ortholog induces the other’s loss, to natural selection preferring the
fixation of one allelic type in hybrid populations. For instance, it has been put in context to
expression dominance of one subgenome in hybrids, e.g. (Yoo et al. 2014; Alexander-
Webber et al. 2016; Cheng et al. 2018). Mechanisms causing such expression dominance
are unclear and may relate to various processes like cis-/trans divergence, unequal content
of transposable elements or levels of heterochromatinisation among parental species
(Woodhouse et al. 2014; Bottani et al. 2018). In any case, it has been proposed that once
expression dominance occurs, the loss of homoeologs from lower-expressed subgenome
would be preferred by selection due to less severe consequences (Yoo et al. 2014).

Interestingly, our data contrast this prediction, as the preferentially retained
subgenome – elongatoides – is clearly not dominant in hybrids. By contrast, recent study of
(Bartoš et al. 2019) documented significant bias towards taenia-like expression in ecologic
and phenotypic traits and an overall expression level dominance of the taenia subgenome in
hybrids’ transcriptomes. In fact, Bartoš et al.’ data indicates slight prevalence of taenia
subgenome transcripts in somatic tissue (~1,5%) and in the germline (~4%) of diploid
hybrids, suggesting that the expression dominance is not the causal explanation for
biased genome fractionation in Cobitis hybrids.

Selective elimination of one parental subgenome may be particularly adaptive in
gynogens by increasing their similarity to that parental species which provides them with
sperm, thereby increasing the chance to be fertilized (Beukeboom and Vrijenhoek 1998).
Interestingly, though, all investigated ET hybrids coexist with C. taenia, making it unlikely
that preferential loss of C. taenia alleles provides such type of sex-mimicry.

Our data thus suggest that causal link between transcriptome-wide expression
dominance and biased genome fractionation is more complex than predicted by
aforementioned hypotheses. For instance (Bartoš et al. 2019) documented that magnitude of
C. taenia expression dominance differs between somatic traits and the germline, suggesting
that **biased genome fractionation may also reflect tissue specific expression characteristics** and other traits which often escape researcher's attention.

We may also speculate that biased retention of *C. elongatoides* subgenome may reflect its special 'mechanistic' properties. Several mechanisms for biased template preference have been proposed, including different expression levels of orthologous alleles (Schildkraut et al. 2006), different GC contents (Duret and Galtier 2009; Williams et al. 2015) or the effect of maternal ancestry when maternal endonuclease systems may preferentially induce double strand breaks (DSB) on paternal chromosomes thereby causing biased DSB repair and unequal gene conversion (Wang et al. 2010). However, these hypotheses are unlikely to explain the prevalence of E-like LOH as *C. elongatoides* subgenome possesses neither higher expression levels nor different GC content and the maternal ancestor of all studied *elongatoides-taenia* hybrids was *C. taenia*, ruling out major role of maternal bias. But, it is possible that observed bias may reflect specific distribution of epigenetic markings, like methylation, which are known to affect recombination landscape (Mirouze et al. 2012) and may take unexpected and non-additive patterns in hybrids as compared to their parents (Hegarty et al. 2011).

Although we did not identify proximate reason for biased subgenome retention, our data do indicate that both subgenomes differ in their ability to induce LOH events in hybrids. For instance, we observed higher proportion of LOH events in EET triploids than in ETT triploids (Fig. 3). Given that detection of LOH events is limited in triploids by the presence of two conspecific allelic copies, higher fraction of allelic loss/replacements would escape our attention in ETT than in EET triploids if E-like LOH occur more frequently than T-like LOH. Moreover, we already mentioned that E-like and T(N)-like LOH events differ with respect to allelic expression or GC bias, altogether **suggesting some fundamental differences between hybrid’s subgenomes in the ability to induce allelic loss or replacement**.

**4.3. Loss of heterozygosity preferentially accumulates in particular gene pathways.**

Benefits of heterozygosity loss in hybrids has been directly tested in only a few studies (Smukowski Heil et al. 2017; Smukowski Heil et al. 2019). Even in these cases, the benefits of LOH appeared quite complex and specific for given allele, subgenome, and also for specific environmental conditions (Lancaster et al. 2019). Although we could not directly examine the effects of LOH, our study **revealed some parallel trends in heterozygosity loss among independent hybrid strains**. This suggests that LOH tend to accumulate in genes characterized by some common expression and functional properties, such as higher-than-average expression levels and lower-than-average $d_{el}/d_{es}$ ratios indicating strong purifying selection maintaining their functionality.
We also found that some gene pathways appear more affected by accumulation of LOH events than others, as apparent from significant enrichment of GO terms associated with endomembrane systems, endoplasmic reticulum and coated vesicles in some biotypes (Table 1). We have no explanation for such observation, but it is worth mentioning that genes with these GO terms often participate in multimeric protein complexes ensuring vesicle tethering, coating and transport to membranes. Since subunits of such complexes are coevolving to maintain proper functionality, it is tempting to speculate that hybrids would profit from removal of heterozygosity because mix of protein interactors from diverged orthologs may negatively impact composition of the entire complex.

Unfortunately, the power of GO analysis was weakened by relatively low number of annotated genes with diagnostic SNPs, so that some GOs appeared insignificant after p-value correction albeit all their genes carried LOH in some biotypes (Table 1). Interestingly, though, same or nested GO terms were sometimes recorded among top-enriched GO terms of different hybrid biotypes, thereby hypothetically indicating LOH accumulation also in other gene pathways which would therefore be worth of further study. For example, this concerned GO terms associated with cell-cell junction and cell-substrate adherence junction, containing genes expressed since the early zygotic embryogenesis that take part in cell migration, and cell-cell communication (Siddiqui et al. 2010; Goonesinghe et al. 2012; Matsui et al. 2015).

Conclusions

There is an increasing evidence that genomes of asexual, hybrid and polyploid taxa evolve dynamically with selective filtering of parts of parental subgenomes. Some common trends have been identified in genome evolution of unrelated hybrid/polyploid taxa, for example, in relation to gene/allelic expression. Although patterns revealed in the present study were generally consistent with several previously proposed mechanisms, some widely cited processes, such as preferential retention of transcriptionally dominant subgenome, were not supported. The lesson from asexual hybrid loaches thus shows that genome fractionation is a very complex process involving simultaneously operating mechanisms that range from a priori bias in template selection to selective fixation of adaptive LOH variants. Relative impact of involved mechanisms thus likely depends on reproductive mode, origin of particular subgenome, allelic sequence composition and transcription activity, properties of involved genes and environmental conditions. In combination with recent advances in understanding the effect of aneuploidies (Birchler and Veitia 2012; Veitia et al. 2013), the data acquired on taxa like asexual hybrid loaches can provide invaluable insight into the role of gene dosage in genome evolution in hybrids and neopolyploids. Investigation of genome
evolution in hybrid and polyploid taxa may also provide important information about basic biological processes related to interaction among genes, meiosis and mitosis.
2. Material & Methods

2.1. Studied specimens

The study is based on exome-capture data from 46 specimens including *C. paludica* as outgroup, the parental species *Cobitis elongatoides*, *C. tanaitica*, *C. taenia* and their asexual hybrids of various genomic compositions (see Fig. 1 and Table S1 for details). The specimens were a priori categorized into taxonomical units using flow cytometry and published PCR-RFLP markers (Janko et al. 2007). As in (Janko et al. 2018), we also included two laboratory *elongatoides-taenia* F1 hybrids with their parental individuals as a control of quality of base calling and LOH site detection.

2.2. DNA sequencing, SNP calling and identification of clonal lineages

Isolated gDNA was sheared with Bioruptor™ to proper fragment distribution, tagged by indices, pooled, hybridized to custom-designed exome-capture probes (Janko et al. 2018) and captured fragments were sequenced on Illumina NextSeq in 75 bp paired-end (PE) mode. To verify the robustness of exome-capture data and subsequent interpretation, we also performed whole-genome shotgun sequencing of single ET hybrid using the HiSeq X Ten sequencing platform in pair-end mode (average fragment length 200 bp, library preparation and sequencing performed by Macrogen®). Obtained reads were quality-trimmed by fqtrim tool (Pertea 2015); minimum read length 20 bp; 3’ end trimming if quality drops below 15 and aligned to *C. taenia* reference transcriptome that was published and cleaned from potentially paralogous contigs by (Janko et al. 2018). To identify mitochondrial variants, we also mapped the reads to published *C. elongatoides* mitochondrion (accession no. NC_023947.1). Mapping was performed with BWA MEM algorithm (Li and Durbin 2009) and resulting files were processed with Picard tools (Broad Institute 2015). Individuals’ variants were called with GATK v3.4 HaplotypeCaller tool and all individuals were jointly genotyped using the GenotypeGVCFs tool (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). Variant recalibration was based on available database of species-diagnostic positions (Janko et al. 2018) representing learning set for variant quality score recalibration tool VariantRecalibrator. Recalibrated variants were then filtered with ApplyRecalibration tool using 90 % tranche to filter all variants. All resulting highly confident
SNPs with coverage $\geq 10$ and genotype quality $\geq 20$ were transferred into the relational database using our own Python3/SQL scripts. SNP data of each specimen were subjected to clustering analysis by Plink v1.90b4 (Chang et al. 2015). To simplify the analysis, we focused solely on bi-allelic SNPs with at most two variants throughout the entire dataset. This resulted in removal of $\sim 1\%$ of positions. We also removed 103 positions where the two laboratory F1 hybrids differed from their parents, because such variants were suspiciously present in most of the other specimens and suggested potential sequencing or demultiplexing errors rather than real variants.

To identify which hybrid individuals putatively belong to the same clone, we followed (Arnaud-Haond et al. 2007). Specifically, we created a pairwise matrix of distances between all hybrid individuals calculated from SNP mismatches. We then investigated a histogram of such pairwise distances and found a saddle point, which putatively defines a threshold distance between pairs of individuals belonging to the same or different clones.

### 2.3. Selection of species-specific SNPs, Detection of Loss of Heterozygosity (LOH) events and evaluation of their topological clustering

All SNPs that passed through aforementioned filters were attributed into one of the 10 categories according to their distribution among biotypes, following (Ament-Velásquez et al. 2016). The most important category of SNPs for this study are the species-specific variants (categories shared sh3a-b), where parental species are fixed (monomorphic) for different alleles, thereby allowing for detection of so-called Loss of Heterozygosity (LOH) events, where hybrids appear homozygous contrary to the expectation. Having detected the LOH variants in hybrids, we test whether observed LOH events tend to be randomly distributed across individual's genes or rather tend to cluster, in which case the SNPs with LOH in given gene occur in tight proximity to each other with no discontinuation by heterozygous diagnostic sites. To do so, we identified within each gene the uninterrupted stretches of diagnostic sites with LOH and assigned each such cluster with a score ($S$) so that if the length of the LOH cluster = $n$, then $S = n^2$. Overall clustering score per animal is then simply represented by $\sum S$. Finally, to test whether observed clustering is nonrandom, we permuted for each hybrid individual its LOH sites across all diagnostic sites and genes, calculated $S$ and compared simulated values to empirical scores.
2.4. Analysis of sequencing coverage

To evaluate whether sequence capture data are suited to compare coverages among loci and across individuals, we investigated the variance in sequencing depth by counting in each sample the number of mapped reads per contig using Bedtools multicov v2.25.0, with default settings (Quinlan and Hall 2010). We then calculated the normalized coverage at each LOH site of every individual using the "total read count" approach (Dillies et al. 2013) and estimated the so-called ‘relative coverage’ by dividing hybrid’s normalized coverage at given site by normalized coverages of the same site in parental species. Since deletions are unlikely in sexual species, this allows detection of hemizygous deletions in hybrids. The expected values then depend on the ploidy of given hybrid: ~1 would indicate the same number of allelic copies indicating a conversion event, while hemizygous deletions would generate relative values ~0.5 and ~0.66 in diploid and triploid hybrids, respectively, while double allelic deletion in triploids would generate values ~0.33.

To reveal whether observed LOH events in hybrids are generated by gene conversion or hemizygous deletions, or their combination, we performed two tests. First, we simulated how the distribution of normalized coverages should look like, were it generated by conversions only. To do so, we generated simulated dataset for each hybrid individual, where coverage values at its LOH sites were sampled from exactly the same number of sites/genes in parental individuals. As such, we obtained realistic null expectation of relative coverages while taking into account used methodology of DNA sequencing and bioinformatic treatment. Such a null distribution simulated for each hybrid biotype has been compared with the empirical one by Kolmogorov-Smirnov test. The simulations have been repeated several times for each biotype, randomizing both the loci considered as well as the specimens used per each locus.

We then constructed histograms of relative coverages of all detected LOH sites for each hybrid biotype and tested their modality at values biologically relevant for gene conversion (~1) or deletion (~0.5 in diploids, or 0.66 and 0.33 in triploids) (Tucker et al. 2013). To test whether observed distributions deviate from unimodality and to evaluate the relative contribution of gene conversion and deletion processes, we applied the nonlinear least square method implemented in Gnuplot software to consecutively fit each histogram by gamma distributions with the shape parameter $k$ optimized by the fitting algorithm and mean ($\mu = \alpha/\beta$) fixed at aforementioned relevant values. In case of diploids, we fitted two distributions (or their mix), centered at 1 and 0.5, while in triploids we fitted three distributions (or their mixes) centered at 1, 0.66 and 0.33. Before fittings, we followed the Freedman-Diaconis rule to select the width of the bins in each histogram (Freedman and Diaconis 1981).
in order to take into account the properties of particular datasets of each biotype. In case of mix of distributions, we let the fitting algorithm estimate the optimal values of A, B and C parameters describing relative contributions of individual distributions.

2.5. Testing the effect of gene/allele expressions on LOH occurrence:

To evaluate potential effects of gene/allelic expression on the occurrence of LOH, we compared the present data with those published by (Bartoš et al. 2019).

Using present gDNA data we categorized loci based on presence or absence of LOH (LOH positive or negative) and its direction (E-like or T-like). We then analyzed the RNA expression of corresponding loci in data of Bartoš et al., (2019) using two tests:

- **Differences in overall gene expression**: To compare the expression levels of LOH-positive and LOH-negative genes in each hybrid biotype, we normalized original read counts by TPM method (Transcripts Per Kilobase Million) instead of DeSeq2 method used by Bartoš et al. (2019), since the TPM allowed for comparisons of multiple loci within each individual. The TPM normalization was performed according to (Mortazavi et al. 2008).

- **Allelic expression**: Because data of hybrids from Bartoš et al. (2019) could be affected by allelic deletions, we investigated expression divergence between parental species as a proxy for relative allelic expression in each locus. Specifically, we extracted the DeSeq2 normalized estimates of *C. elongatoides* and *C. taenia* expression levels and tested whether the direction of LOH event (either E-like or T-like) is related to log2 fold change (FC) between parental species. The test was performed by comparing the distributions of log2FC values in LOH-negative genes with those of either E-like LOH or T-like LOH-positive genes using Kolmogorov-Smirnov test.

2.6. Analysis of Gene Ontology Term enrichment.

The reference transcriptome was annotated with BLAST2GO tool v1.4.4 using GO database as of July 2019). From 20600 sequences, a subset of 13557 received BLASTx hit (e-value < 0.0001), from which 11314 was associated with significant GO Term annotation (default BLAST2GO settings). To identify GO terms potentially associated with LOH-positive genes, we performed GO enrichment analysis restricted to those genes, which possessed
diagnostic sites, thereby technically allowing detection of LOH. P-values were calculated from hypergeometric distribution implemented in GO::TermFinder (Boyle et al. 2004) using the list of LOH-positive genes as a testing dataset for each biotype. Since gene ontologies terms are a part of acyclic directed graphs (parent and child terms are not independent), we also corrected obtained p values with permutation-based correction provided in GO::TermFinder.
Acknowledgments

We would like to thank EMBL GeneCore crew for sequencing and advice on data processing. The research was supported by grants GAČR 17-09807S and 19-21552S, SGS19/PřF/2015 and SGS19/PřF/2017, and the Ministry of Education, Youth and Sports of the Czech Republic grant EXCELLENCE CZ.02.1.01/0.0/0.0/15_003/0000460 OP RDE.
References


Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. 2004. GO::TermFinder--open source software for accessing Gene Ontology information and
finding significantly enriched Gene Ontology terms associated with a list of genes.

Bioinformatics 20:3710–3715.


rediploidization. *Nature* [Internet] advance online publication. Available from:
http://www.nature.com/nature/journal/vaop/ncurrent/full/nature17164.html

López EH, Palumbi SR. 2019. Somatic Mutations and Genome Stability Maintenance in


Madlung A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited

Reproduction Does Not Apparently Increase the Rate of Chromosomal Evolution:
Karyotype Stability in Diploid and Triploid Clonal Hybrid Fish (Cobitis, Cypriniformes,

Mandegar MA, Otto SP. 2007. Mitotic recombination counteracts the benefits of genetic
1307.


Biotechnology* 33:436–441.


McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a
MapReduce framework for analyzing next-generation DNA sequencing data. *Genome

Mirouze M, Lieberman-Lazarovich M, Aversano R, Bucher E, Nicolet J, Reinders J,
Paszkowski J. 2012. Loss of DNA methylation affects the recombination landscape in


Muller HJ. 1964. The relation of recombination to mutational advance. *Mutation
Research/Fundamental and Molecular Mechanisms of Mutagenesis* 1:2–9.

Normark BB. 1999. EVOLUTION IN A PUTATTVELY ANCIENT ASEXUAL APHID
LINEAGE: RECOMBINATION AND RAPID KARYOTYPE CHANGE. *Evolution*
53:1458–1469.

Omlanian AR, Cristescu MEA, Dudycha JL, Lynch M. 2006. Ameiotic recombination in asexual
lineages of Daphnia. *Proceedings of the National Academy of Sciences* 103:18638–
18643.

Ortiz-Barrientos D, Engelstädter J, Rieseberg LH. 2016. Recombination Rate Evolution and


Fig. 1: The *Cobitis taenia* hybrid complex. a) Map of samples’ origin; sites’ numeric code corresponds to Table S1. b) Reproduction scheme of *Cobitis*; letters correspond to haploid genomes: E = *C. elongatoides*, T = *C. taenia*. c) example of gametogenetic pathways involving endoreplication and followed by meiotic pairing of either sister chromosomes or orthologous chromosomes. Note that the latter case may cause loss of heterozygosity among progeny via crossovers or gene conversions. Insert on the right demonstrates empirical evidence for the presence of bivalents and univalents in hybrid’s oocytes as from (Dedukh et al. 2020).
Fig. 2: a) MDS (multidimensional scaling; SVD algorithm) of samples based on filtered recalibrated SNPs; clustering visualization of first two coordinates (Plink v1.9b) to validate sample genetic origin; letters correspond to haplotype genomes as follows: E = C. elongatoides, N = C. tanaitica, T = C. taenia. Hybrid individuals are denoted by black color and numerals, which indicate the ID of determined clonal lineages, as in Table S2 (note that clone-mates tend to be clustered in the MDS plot). b) Heterozygosity of asexual and sexual samples across all filtered variant sites; missing sites are not included. c) Boxplots indicate the proportions of shared genotypes at LOH positions among all pairs of individuals belonging to the same (right) and different (left) clonal lineages. Note that we used the set of E-TN diagnostic sites to maintain compatibility of elongatoides-taenia and elongatoides-tanaitica hybrids.
**Fig. 3:** **a)** Correlation between proportion of private hybrid SNPs as the proxy for asexual's age, and proportion of LOH loci in each individual. For plotting points and their linear regression (solid line) we used E-T diagnostic sites for *elongatoides-taenia* hybrids and E-N diagnostic sites for *elongatoides-tanaitica* hybrids. Dashed line represents the linear regression fit for data calculated on E-TN diagnostic sites used for all individuals (points not shown). Letters correspond to haplotype genomes as follows: E = *C. elongatoides*, N = *C. tanaitica*, T = *C. taenia*. **b)** Barplot showing proportions of LOH events according to their genomic origin. Height of each bar represents number of unique SNPs that appeared as LOH in a given biotype.
Fig. 4: a-e) Histograms of relative coverage of LOH loci in ET, EET, ETT, EN and EEN biotypes, respectively. Arrows depict biologically meaningful values (for given ploidy); red lines represent mixture of two Gamma distribution; blue lines represent single Gamma distribution with mean centered at value 1. Coefficients A and B indicate proportions of both Gamma distributions in the combined model, A relates to the distribution assuming the mean relative coverage ~1, B to the distribution with the mean ~0.5 or 0.66 (for diploid or triploid biotype, respectively); f) orange represents empirical cumulative distribution function (ECDF) of relative coverages at LOH sites of ET biotype, black represents ECDF of relative coverages at the same sites, but taken from parental species, where no deletions are expected. ECDF curves for other biotypes are provided in Fig. S1.
Figure 5:

A: G/C bias at asexuals’ LOH sites as demonstrated for one representative of every biotype. Each individual is represented by two columns with E-like (left) and T-like (right) LOH events. Bar widths scale with absolute numbers of observed LOH events in each individual, while heights of color fields demonstrate proportions of LOH causing weak to strong, strong to weak and no GC change. The last bar represents differences between C. elongatoides and both other parental species at all E-TN diagnostic sites. Note that in E-like LOH events, all hybrids show consistent and significant increase in weak to strong substitution rates as compared against interparental divergence (no shift has been detected at TN-like LOH sites).

B: TPM normalized expression characteristics of LOH positive (blue) and LOH negative (red) genes in all Cobitis biotypes analyzed by Bartoš et al. 2019. For better orientation, the oocyte
data (left part) are depicted in lighter color tones, while liver data (right part) are in darker tones.
## Tables

### ET (292 LOH positive genes, 4582 annotated with diagnostic SNP)

| GO:0005576 | extracellular region pV 0.0028351 (22 vs 184) |
| GO:0005923 | bicellular tight junction pV 0.0161340 (3 vs 9) |
| GO:0070160 | tight junction pV 0.0219788 (3 vs 10) |
| GO:0000347 | THO complex pV 0.0222870 (2 vs 4) |
| GO:0000445 | THO complex part of transcription export complex pV 0.0222870 (2 vs 4) |
| GO:0016021 | integral component of membrane pV 0.0338391 (76 vs 987) |
| GO:0031224 | intrinsic component of membrane pV 0.0353113 (76 vs 989) |
| GO:0031680 | G-protein beta/gamma-subunit complex pV 0.0355939 (2 vs 5) |
| GO:0097431 | mitotic spindle pole pV 0.0355939 (2 vs 5) |
| GO:0005834 | heterotrimeric G-protein complex pV 0.0366592 (3 vs 12) |
| GO:0043296 | apical junction complex pV 0.0366592 (3 vs 12) |
| GO:1905360 | GTPase complex pV 0.0366592 (3 vs 12) |
| GO:0030687 | preribosome, large subunit precursor pV 0.0552198 (3 vs 14) |
| GO:0000346 | transcription export complex pV 0.0686844 (2 vs 7) |
| GO:0000152 | nuclear ubiquitin ligase complex pV 0.0878187 (2 vs 8) |
| GO:0031234 | extrinsic component of cytoplasmic side of plasma membrane pV 0.0897273 (3 vs 17) |
| GO:0043233 | cell-cell junction pV 0.0897273 (4 vs 27) |
| GO:000922 | spindle pole pV 0.0897273 (3 vs 17) |
| GO:0031234 | extrinsic component of cytoplasmic side of plasma membrane pV 0.0897273 (3 vs 17) |
| GO:0043230 | extracellular organelle pV 0.1083001 (2 vs 9) |
| GO:0070062 | extracellular exosome pV 0.1083001 (2 vs 9) |

### ETT (226 LOH positive genes, 4582 annotated with diagnostic SNP)

| GO:0005844 | polysome pV 0.0287870 (3 vs 14) |
| GO:0031514 | motile cilium pV 0.0318521 (2 vs 6) |
| GO:0005730 | nucleolus pV 0.0390230 (11 vs 123) |
| GO:0005732 | small nucleolar ribonucleoprotein complex pV 0.0431628 (2 vs 7) |
| GO:0071007 | U2-type catalytic step 2 spliceosome pV 0.0431628 (2 vs 7) |
| GO:0030684 | preribosome pV 0.0446155 (6 vs 53) |
| GO:004798 | nuclear transcription factor complex pV 0.0577484 (4 vs 30) |
| GO:0090575 | RNA polymerase II transcription factor complex pV 0.0577484 (4 vs 30) |
| GO:0031974 | membrane-enclosed lumen pV 0.0808439 (31 vs 488) |
| GO:0043233 | organelle lumen pV 0.0808439 (31 vs 488) |
| GO:0070013 | intracellular organelle lumen pV 0.0808439 (31 vs 488) |
| GO:0005657 | replication fork pV 0.0839463 (2 vs 10) |
| GO:0034708 | methyltransferase complex pV 0.0990699 (4 vs 36) |
GO:0071013 catalytic step 2 spliceosome pV 0.1015022 (3 vs 23)
GO:0005793 endoplasmic reticulum-Golgi intermediate compartment pV 0.1228584 (3 vs 25)
GO:0031461 cullin-RING ubiquitin ligase complex pV 0.1456830 (3 vs 27)
GO:0032040 small-subunit processome pV 0.1456830 (3 vs 27)
GO:0071944 cell periphery pV 0.1530158 (23 vs 373)
GO:0031461 cullin-RING ubiquitin ligase complex pV 0.1456830 (3 vs 27)
GO:0032040 small-subunit processome pV 0.1456830 (3 vs 27)
GO:0035097 histone methyltransferase complex pV 0.1697646 (3 vs 29)
GO:0031981 nuclear lumen pV 0.1815560 (23 vs 382)

EET (662 LOH positive genes, 4582 annotated with diagnostic SNP)
GO:0030117 membrane coat pV 0.0007544 (12 vs 31)
GO:0048475 coated membrane pV 0.0007544 (12 vs 31)
GO:0030120 vesicle coat pV 0.0014970 (9 vs 21)
GO:0030662 coated vesicle membrane pV 0.0137591 (9 vs 28)
GO:0031224 intrinsic component of membrane pV 0.0145280 (165 vs 989)
GO:0016020 membrane pV 0.0149709 (234 vs 1456)
GO:0030663 COPI-coated vesicle membrane pV 0.0297850 (4 vs 9)
GO:0030597 cell-cell junction pV 0.0322119 (11 vs 42)
GO:0012506 endomembrane system pV 0.0071884 (116 vs 405)
GO:0005783 endoplasmic reticulum pV 0.0075373 (116 vs 405)

EN (755 LOH positive genes, 3197 annotated with diagnostic SNP)
GO:005789 endoplasmic reticulum membrane pV 0.0006197 (46 vs 126)
GO:0072546 ER membrane protein complex pV 0.0007271 (5 vs 5)
GO:0042175 nuclear outer membrane-endoplasmic reticulum membrane network pV 0.0009193 (46 vs 128)
GO:0031026 endomembrane system pV 0.0041040 (107 vs 364)
GO:0012506 endomembrane system pV 0.0041040 (107 vs 364)
GO:0005783 endoplasmic reticulum pV 0.0041040 (107 vs 364)
GO:0005924 cell-substrate adherens junction pV 0.0098916 (5 vs 7)
GO:0005925 focal adhesion pV0.0098916 (5 vs 7)
<table>
<thead>
<tr>
<th>GO:0030055</th>
<th>cell-substrate junction pV 0.0098916 (5 vs 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0099081</td>
<td>supramolecular polymer pV 0.0121988 (19 vs 49)</td>
</tr>
<tr>
<td>GO:0099512</td>
<td>supramolecular fiber pV 0.0121988 (19 vs 49)</td>
</tr>
<tr>
<td>GO:0016528</td>
<td>sarcoplasm pV 0.0131308 (3 vs 3)</td>
</tr>
<tr>
<td>GO:0016529</td>
<td>sarcoplasmic reticulum pV 0.0131308 (3 vs 3)</td>
</tr>
<tr>
<td>GO:0017087</td>
<td>mitochondrial processing peptidase complex pV 0.0131308 (3 vs 3)</td>
</tr>
<tr>
<td>GO:0042383</td>
<td>sarcolemma pV 0.0131308 (3 vs 3)</td>
</tr>
<tr>
<td>GO:0030176</td>
<td>integral component of endoplasmic reticulum membrane pV 0.0134881 (12 vs 27)</td>
</tr>
<tr>
<td>GO:0031227</td>
<td>intrinsic component of endoplasmic reticulum membrane pV 0.0134881 (12 vs 27)</td>
</tr>
<tr>
<td>GO:0030016</td>
<td>myofibril pV 0.0147223 (6 vs 10)</td>
</tr>
<tr>
<td>GO:0030017</td>
<td>sarcomere pV 0.0147223 (6 vs 10)</td>
</tr>
<tr>
<td>GO:0005813</td>
<td>centrosome pV 0.0251815 (12 vs 29)</td>
</tr>
<tr>
<td>GO:0012505</td>
<td>endomembrane system pV 0.00017207 (190 vs 405)</td>
</tr>
<tr>
<td>GO:0005783</td>
<td>endoplasmic reticulum pV 0.0008575 (84 vs 166)</td>
</tr>
<tr>
<td>GO:0042175</td>
<td>nuclear outer membrane-endoplasmic reticulum membrane network pV 0.0009040 (67 vs 128)</td>
</tr>
<tr>
<td>GO:0005789</td>
<td>endoplasmic reticulum membrane pV 0.0009619 (66 vs 126)</td>
</tr>
<tr>
<td>GO:0072546</td>
<td>ER membrane protein complex pV 0.0007271 (5 vs 5)</td>
</tr>
<tr>
<td>GO:0030176</td>
<td>integral component of endoplasmic reticulum membrane pV 0.0087823 (17 vs 27)</td>
</tr>
<tr>
<td>GO:0031227</td>
<td>intrinsic component of endoplasmic reticulum membrane pV 0.0087823 (17 vs 27)</td>
</tr>
<tr>
<td>GO:0030496</td>
<td>midbody pV 0.0095585 (8 vs 10)</td>
</tr>
<tr>
<td>GO:005794</td>
<td>Golgi apparatus pV 0.0153109 (69 vs 145)</td>
</tr>
<tr>
<td>GO:005924</td>
<td>cell-substrate adherens junction pV 0.0154665 (6 vs 7)</td>
</tr>
<tr>
<td>GO:005925</td>
<td>focal adhesion pV 0.0154665 (6 vs 7)</td>
</tr>
<tr>
<td>GO:0030055</td>
<td>cell-substrate junction pV 0.0154665 (6 vs 7)</td>
</tr>
<tr>
<td>GO:0045121</td>
<td>membrane raft pV 0.0201926 (7 vs 9)</td>
</tr>
<tr>
<td>GO:0098589</td>
<td>membrane region pV 0.0201926 (7 vs 9)</td>
</tr>
<tr>
<td>GO:0098857</td>
<td>membrane microdomain pV 0.0201926 (7 vs 9)</td>
</tr>
<tr>
<td>GO:0097525</td>
<td>spliceosomal snRNP complex pV 0.0259453 (10 vs 15)</td>
</tr>
<tr>
<td>GO:0031090</td>
<td>organelle membrane pV 0.0351521 (157 vs 364)</td>
</tr>
<tr>
<td>GO:0030532</td>
<td>small nuclear ribonucleoprotein complex pV 0.0456027 (10 vs 16)</td>
</tr>
<tr>
<td>GO:0031300</td>
<td>intrinsic component of organelle membrane pV 0.0534485 (28 vs 56)</td>
</tr>
<tr>
<td>GO:0031301</td>
<td>integral component of organelle membrane pV 0.0534485 (28 vs 56)</td>
</tr>
</tbody>
</table>

**Table 1:** Top 20 enriched GO terms in cellular compartment GO category ranked by p value. For each hybrid biotype we indicate number of genes affected by LOH event and total number of annotated genes with diagnostic SNP relevant for given combination of parental
species. For each GO term, we indicate its ID, description and uncorrected p value, as well as numbers of LOH-positive genes and total number of genes with given GO in parentheses. Underlined GO terms are significant after correction for multiple test at alpha level=0.1. Colors are used to highlight GO terms shared between distinct biotypes so that the same color across biotypes indicates GO terms that are identical or nested.
## Supplementary files

<table>
<thead>
<tr>
<th>sample</th>
<th>biotype</th>
<th>sex</th>
<th>no. of reads</th>
<th>no. of bases</th>
<th>mapped reads</th>
<th>locality</th>
<th>site ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>csc008</td>
<td>TT</td>
<td>m</td>
<td>9328500</td>
<td>1890943172</td>
<td>7872341</td>
<td>Tnya river, Ukraine</td>
<td>22</td>
</tr>
<tr>
<td>csc009</td>
<td>EE</td>
<td>NA</td>
<td>10674754</td>
<td>2200211351</td>
<td>9160050</td>
<td>Kozačevski river, Bulgaria</td>
<td>3</td>
</tr>
<tr>
<td>csc011</td>
<td>EET</td>
<td>f</td>
<td>46169330</td>
<td>4546217825</td>
<td>24661920</td>
<td>Odra R. sites RQ, Poland</td>
<td>18</td>
</tr>
<tr>
<td>csc012</td>
<td>EEN</td>
<td>f</td>
<td>36325574</td>
<td>385572296</td>
<td>18876965</td>
<td>Odra R. sites 509, Poland</td>
<td>16</td>
</tr>
<tr>
<td>csc013</td>
<td>EEN</td>
<td>f</td>
<td>44514804</td>
<td>4386326838</td>
<td>25987623</td>
<td>Čierna voda, Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>csc014</td>
<td>EEN</td>
<td>f</td>
<td>47084588</td>
<td>4630637949</td>
<td>27469787</td>
<td>Čierna voda, Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>csc015</td>
<td>EEN</td>
<td>f</td>
<td>50128474</td>
<td>4925538969</td>
<td>25581023</td>
<td>Čierna voda, Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>csc016</td>
<td>EEN</td>
<td>f</td>
<td>47466730</td>
<td>4675165684</td>
<td>28108072</td>
<td>Čierna voda, Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>csc017</td>
<td>EEN</td>
<td>f</td>
<td>62562316</td>
<td>6142494088</td>
<td>32394947</td>
<td>Čierna voda, Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>csc018</td>
<td>EEN</td>
<td>f</td>
<td>55335802</td>
<td>5446458549</td>
<td>30131168</td>
<td>Čierna voda, Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>csc019</td>
<td>EE</td>
<td>NA</td>
<td>44761144</td>
<td>4406019008</td>
<td>23043480</td>
<td>Čierna voda, Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>csc020</td>
<td>EE</td>
<td>NA</td>
<td>55395494</td>
<td>5451569580</td>
<td>28769818</td>
<td>Čierna voda, Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>csc021</td>
<td>TT</td>
<td>m</td>
<td>42119812</td>
<td>2985717253</td>
<td>20350974</td>
<td>South Bug, Ukraine</td>
<td>19</td>
</tr>
<tr>
<td>csc022</td>
<td>TT</td>
<td>f</td>
<td>34584512</td>
<td>2448248483</td>
<td>16574306</td>
<td>Kodyma river, Ukraine</td>
<td>4</td>
</tr>
<tr>
<td>csc024</td>
<td>EE</td>
<td>NA</td>
<td>38482140</td>
<td>2734317798</td>
<td>18243498</td>
<td>Odra R. sites 801, Poland</td>
<td>11</td>
</tr>
<tr>
<td>csc025</td>
<td>EE</td>
<td>NA</td>
<td>37629664</td>
<td>266930656</td>
<td>18231557</td>
<td>Odra R. sites 801, Poland</td>
<td>11</td>
</tr>
<tr>
<td>csc027</td>
<td>EE</td>
<td>m</td>
<td>54542982</td>
<td>3908825666</td>
<td>28305773</td>
<td>Kamcia r., Bulgaria</td>
<td>6</td>
</tr>
<tr>
<td>csc028</td>
<td>TT</td>
<td>f</td>
<td>32664314</td>
<td>2308852524</td>
<td>16036910</td>
<td>South Bug, Ukraine</td>
<td>2</td>
</tr>
<tr>
<td>csc029</td>
<td>EE</td>
<td>NA</td>
<td>38364748</td>
<td>2723397165</td>
<td>18384671</td>
<td>Kozačevski river, Bulgaria</td>
<td>3</td>
</tr>
<tr>
<td>csc033</td>
<td>TT</td>
<td>f</td>
<td>28098246</td>
<td>2010306058</td>
<td>14112070</td>
<td>Krym Alma river, Ukraine</td>
<td>5</td>
</tr>
<tr>
<td>csc034</td>
<td>TT</td>
<td>m</td>
<td>27835994</td>
<td>1951741228</td>
<td>13518951</td>
<td>Odra R. sites 705, Poland</td>
<td>12</td>
</tr>
<tr>
<td>csc035</td>
<td>TT</td>
<td>m</td>
<td>41334443</td>
<td>2940164538</td>
<td>20820669</td>
<td>Odra R. sites 701, Poland</td>
<td>10</td>
</tr>
<tr>
<td>csc036</td>
<td>NN</td>
<td>m</td>
<td>31546236</td>
<td>22559091</td>
<td>16153019</td>
<td>Sucholužja, Ukraine</td>
<td>8</td>
</tr>
<tr>
<td>csc039</td>
<td>NN</td>
<td>NA</td>
<td>22567596</td>
<td>1603680625</td>
<td>10392416</td>
<td>Kuban river, Russia</td>
<td>14</td>
</tr>
<tr>
<td>csc046</td>
<td>EEN</td>
<td>f</td>
<td>35381744</td>
<td>2486858020</td>
<td>16471358</td>
<td>Odra R. sites RQ, Poland</td>
<td>18</td>
</tr>
<tr>
<td>csc052</td>
<td>ET</td>
<td>f</td>
<td>32020816</td>
<td>2249205626</td>
<td>14937219</td>
<td>Odra R. sites 701, Poland</td>
<td>10</td>
</tr>
<tr>
<td>csc055</td>
<td>EET</td>
<td>f</td>
<td>34678947</td>
<td>2432666387</td>
<td>15941455</td>
<td>Odra R. sites 510, Poland</td>
<td>15</td>
</tr>
<tr>
<td>csc056</td>
<td>EEN</td>
<td>f</td>
<td>32226860</td>
<td>2270486184</td>
<td>14849216</td>
<td>Sluč Vachovcy, Ukraine</td>
<td>20</td>
</tr>
<tr>
<td>csc057</td>
<td>ET</td>
<td>f</td>
<td>31545522</td>
<td>2190478704</td>
<td>14343725</td>
<td>Odra R. sites 701, Poland</td>
<td>10</td>
</tr>
<tr>
<td>csc067</td>
<td>ET</td>
<td>f</td>
<td>48437308</td>
<td>3434265969</td>
<td>23877555</td>
<td>Odra R. sites 701, Poland</td>
<td>10</td>
</tr>
<tr>
<td>csc069</td>
<td>ET</td>
<td>f</td>
<td>16831044</td>
<td>1632439963</td>
<td>10815080</td>
<td>laboratory hybrid</td>
<td>NA</td>
</tr>
<tr>
<td>csc070</td>
<td>E</td>
<td>m</td>
<td>20691516</td>
<td>2009443636</td>
<td>13009582</td>
<td>Olszyn, Poland</td>
<td>17</td>
</tr>
<tr>
<td>csc071</td>
<td>ET</td>
<td>f</td>
<td>14455828</td>
<td>1399281976</td>
<td>8718131</td>
<td>laboratory hybrid</td>
<td>NA</td>
</tr>
<tr>
<td>csc072</td>
<td>TT</td>
<td>f</td>
<td>12540166</td>
<td>1281069246</td>
<td>8556642</td>
<td>Olszyn, Poland</td>
<td>21</td>
</tr>
<tr>
<td>csc073</td>
<td>EN</td>
<td>f</td>
<td>24308246</td>
<td>2377911505</td>
<td>15929335</td>
<td>Ghiol lake, Romania</td>
<td>7</td>
</tr>
<tr>
<td>csc074</td>
<td>EN</td>
<td>f</td>
<td>19470508</td>
<td>1898963852</td>
<td>12481576</td>
<td>Ghiol lake, Romania</td>
<td>7</td>
</tr>
<tr>
<td>csc076</td>
<td>EET</td>
<td>f</td>
<td>36597710</td>
<td>3585778583</td>
<td>24666067</td>
<td>Odra R. sites 510, Poland</td>
<td>15</td>
</tr>
<tr>
<td>csc077</td>
<td>TT</td>
<td>NA</td>
<td>18349390</td>
<td>1778582689</td>
<td>11064568</td>
<td>Malaren lake, Sweden</td>
<td>13</td>
</tr>
<tr>
<td>csc078</td>
<td>EET</td>
<td>f</td>
<td>17368108</td>
<td>1693830781</td>
<td>10771162</td>
<td>Odra R. sites 510, Poland</td>
<td>15</td>
</tr>
<tr>
<td>csc079</td>
<td>EEN</td>
<td>f</td>
<td>22917494</td>
<td>2231070909</td>
<td>14125361</td>
<td>Odra R. sites 510, Poland</td>
<td>15</td>
</tr>
<tr>
<td>csc080</td>
<td>TT</td>
<td>NA</td>
<td>12337499</td>
<td>1200073223</td>
<td>7505927</td>
<td>Crimea, Ukraine</td>
<td>5</td>
</tr>
<tr>
<td>csc082</td>
<td>ETT</td>
<td>f</td>
<td>21529004</td>
<td>2091846439</td>
<td>13242462</td>
<td>Odra R. sites 704, Poland</td>
<td>9</td>
</tr>
<tr>
<td>csc086</td>
<td>ETT</td>
<td>f</td>
<td>20824472</td>
<td>2036604862</td>
<td>13308483</td>
<td>Odra R. sites 704, Poland</td>
<td>9</td>
</tr>
<tr>
<td>csc087</td>
<td>ETT</td>
<td>f</td>
<td>22230674</td>
<td>2174839517</td>
<td>14061887</td>
<td>Odra R. sites 701, Poland</td>
<td>10</td>
</tr>
<tr>
<td>csc089</td>
<td>ETT</td>
<td>f</td>
<td>50135182</td>
<td>4898480713</td>
<td>27586948</td>
<td>Odra R. sites 701, Poland</td>
<td>10</td>
</tr>
<tr>
<td>csc090</td>
<td>ETT</td>
<td>f</td>
<td>34147738</td>
<td>3331600457</td>
<td>19960304</td>
<td>Odra R. sites 701, Poland</td>
<td>10</td>
</tr>
</tbody>
</table>
Table S1: List of samples including sample name, biotype, sex, sequencing statistics and geographical origin.
<table>
<thead>
<tr>
<th></th>
<th>animal</th>
<th>biotype</th>
<th>MLL</th>
<th>pr1a (xy x y)</th>
<th>pr1b (y x y)</th>
<th>prh2a (xy x y)</th>
<th>prh2b (x y x y)</th>
<th>prh2b1 (x y y)</th>
<th>sh3a (x y x y)</th>
<th>sh3b11 (x y x)</th>
<th>sh3b12 (x y x)</th>
<th>sh4a (x y y x y)</th>
<th>sh4b (x y x y)</th>
<th>total SNPs</th>
<th>LOH proportion</th>
<th>diagnostic SNPs</th>
<th>private SNPs proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>elongatoides-tanaisca</td>
<td>csc012</td>
<td>EEN</td>
<td>5</td>
<td>6907</td>
<td>7</td>
<td>20054</td>
<td>23850</td>
<td>254</td>
<td>26346</td>
<td>2527</td>
<td>59</td>
<td>1195</td>
<td>645</td>
<td>8144</td>
<td>0.0894</td>
<td>28932</td>
<td>0.0845</td>
</tr>
<tr>
<td></td>
<td>csc013</td>
<td>EEN</td>
<td>9</td>
<td>6541</td>
<td>10</td>
<td>20704</td>
<td>24035</td>
<td>202</td>
<td>26851</td>
<td>2462</td>
<td>60</td>
<td>1214</td>
<td>639</td>
<td>82718</td>
<td>0.0859</td>
<td>29373</td>
<td>0.0792</td>
</tr>
<tr>
<td></td>
<td>csc014</td>
<td>EEN</td>
<td>12</td>
<td>7193</td>
<td>6</td>
<td>21336</td>
<td>23703</td>
<td>242</td>
<td>27108</td>
<td>2450</td>
<td>48</td>
<td>1231</td>
<td>637</td>
<td>83960</td>
<td>0.0844</td>
<td>29666</td>
<td>0.0858</td>
</tr>
<tr>
<td></td>
<td>csc015</td>
<td>EEN</td>
<td>16</td>
<td>7084</td>
<td>10</td>
<td>21142</td>
<td>23460</td>
<td>247</td>
<td>26870</td>
<td>2460</td>
<td>48</td>
<td>1221</td>
<td>633</td>
<td>83175</td>
<td>0.0854</td>
<td>29378</td>
<td>0.0853</td>
</tr>
<tr>
<td></td>
<td>csc016</td>
<td>EEN</td>
<td>20</td>
<td>7183</td>
<td>11</td>
<td>21291</td>
<td>23712</td>
<td>253</td>
<td>27096</td>
<td>2444</td>
<td>52</td>
<td>1229</td>
<td>636</td>
<td>83907</td>
<td>0.0843</td>
<td>29592</td>
<td>0.0857</td>
</tr>
<tr>
<td></td>
<td>csc017</td>
<td>EEN</td>
<td>24</td>
<td>7134</td>
<td>10</td>
<td>21174</td>
<td>23652</td>
<td>252</td>
<td>27063</td>
<td>2454</td>
<td>60</td>
<td>1237</td>
<td>641</td>
<td>83677</td>
<td>0.0850</td>
<td>29577</td>
<td>0.0854</td>
</tr>
<tr>
<td></td>
<td>csc018</td>
<td>EEN</td>
<td>28</td>
<td>7015</td>
<td>5</td>
<td>20385</td>
<td>24488</td>
<td>255</td>
<td>26926</td>
<td>2563</td>
<td>67</td>
<td>1214</td>
<td>657</td>
<td>83575</td>
<td>0.0890</td>
<td>29556</td>
<td>0.0840</td>
</tr>
<tr>
<td></td>
<td>csc019</td>
<td>EEN</td>
<td>32</td>
<td>8382</td>
<td>36</td>
<td>16798</td>
<td>28192</td>
<td>338</td>
<td>26797</td>
<td>2654</td>
<td>375</td>
<td>1069</td>
<td>774</td>
<td>83415</td>
<td>0.1016</td>
<td>29826</td>
<td>0.0769</td>
</tr>
<tr>
<td></td>
<td>csc020</td>
<td>EEN</td>
<td>36</td>
<td>6149</td>
<td>3</td>
<td>19369</td>
<td>24942</td>
<td>286</td>
<td>26146</td>
<td>2609</td>
<td>109</td>
<td>1108</td>
<td>657</td>
<td>81377</td>
<td>0.0942</td>
<td>28863</td>
<td>0.0756</td>
</tr>
<tr>
<td></td>
<td>csc021</td>
<td>EEN</td>
<td>40</td>
<td>7015</td>
<td>5</td>
<td>20385</td>
<td>24488</td>
<td>255</td>
<td>26926</td>
<td>2563</td>
<td>67</td>
<td>1214</td>
<td>657</td>
<td>83575</td>
<td>0.0890</td>
<td>29556</td>
<td>0.0840</td>
</tr>
<tr>
<td></td>
<td>csc022</td>
<td>EEN</td>
<td>44</td>
<td>8382</td>
<td>36</td>
<td>16798</td>
<td>28192</td>
<td>338</td>
<td>26797</td>
<td>2654</td>
<td>375</td>
<td>1069</td>
<td>774</td>
<td>83415</td>
<td>0.1016</td>
<td>29826</td>
<td>0.0769</td>
</tr>
<tr>
<td>elongatoides-taenia</td>
<td>csc030</td>
<td>EET</td>
<td>4</td>
<td>1673</td>
<td>1</td>
<td>28139</td>
<td>25039</td>
<td>109</td>
<td>35107</td>
<td>593</td>
<td>65</td>
<td>2237</td>
<td>798</td>
<td>93761</td>
<td>0.0184</td>
<td>35765</td>
<td>0.0179</td>
</tr>
<tr>
<td></td>
<td>csc031</td>
<td>EET</td>
<td>6</td>
<td>912</td>
<td>2</td>
<td>24942</td>
<td>28271</td>
<td>27</td>
<td>35290</td>
<td>6</td>
<td>231</td>
<td>1970</td>
<td>912</td>
<td>92563</td>
<td>0.0067</td>
<td>35527</td>
<td>0.0099</td>
</tr>
<tr>
<td></td>
<td>csc032</td>
<td>EET</td>
<td>7</td>
<td>12388</td>
<td>3</td>
<td>25388</td>
<td>28038</td>
<td>30</td>
<td>35565</td>
<td>12</td>
<td>228</td>
<td>1971</td>
<td>935</td>
<td>93408</td>
<td>0.0067</td>
<td>35805</td>
<td>0.0133</td>
</tr>
<tr>
<td></td>
<td>csc033</td>
<td>EET</td>
<td>10</td>
<td>1278</td>
<td>3</td>
<td>26137</td>
<td>28426</td>
<td>29</td>
<td>36548</td>
<td>13</td>
<td>232</td>
<td>2153</td>
<td>947</td>
<td>95766</td>
<td>0.0067</td>
<td>36793</td>
<td>0.0134</td>
</tr>
<tr>
<td></td>
<td>csc034</td>
<td>EET</td>
<td>13</td>
<td>6378</td>
<td>0</td>
<td>8145</td>
<td>5001</td>
<td>1</td>
<td>41845</td>
<td>35</td>
<td>37</td>
<td>340</td>
<td>38</td>
<td>53713</td>
<td>0.0017</td>
<td>41917</td>
<td>0.0007</td>
</tr>
</tbody>
</table>
Table S2: Counts of SNPs detected per each sample and category. In parentheses, we define each category by demonstrating states at SNP in asexual hybrid (H) and both parental species (P₁, P₂), separated by space (i.e. H P₁, P₂). Note that EN and EEN hybrids were classified against elongatoides—tanaitica comparisons, ET, EET and ETT hybrids against elongatoides—taenia and the two F1 hybrids against their direct parents.
Figure S1: Empirical cumulative distribution functions (ECDF) for individual biotypes. Orange represents empirical cumulative distribution function (ECDF) of relative coverages at LOH sites of given hybrid biotype, black represents ECDF of relative coverages at the same sites, but taken from parental species, where no deletions are expected. Significance of differences between distributions were tested by Kolmogorov-Smirnov test. Hybrid biotypes: a) ET biotype, b) EET biotype, c) ETT biotype, d) EN biotype, e) EEN biotype.