bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Para-allopatry in hybridizing fire-bellied toads (*Bombina bombina and B. variegata*): inference from transcriptome-wide coalescence analyses

Beate Nürnberger[1], Konrad Lohse[1], Anna Fijarczyk[1,2], Jacek M. Szymura[2] and Mark L. Blaxter[1]

[1] Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK

[2] Department of Comparative Anatomy, Institute of Zoology, Jagiellonian University, Gronostajowa 9, 30-387 Kraków, Poland

Corresponding Author: Dr. Beate Nürnberger, e-mail: b.nurnberger@ed.ac.uk, Telephone: +44 131 229 9795

Running title: Para-allopatry in *Bombina*

Key words: ecological speciation, hybrid zone, RNA-seq, genome coalescence

Data archiving: NCBI Sequence Read Archive (RNA-seq data), afterParty (http:/....., transcriptome assemblies), Github (source code for bioinformatic analyses).

Words: 7481 - Tables: 4 - Figures: 5

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Ancient origins, profound ecological divergence and extensive hybridization make the fire-bellied toads *Bombina bombina* and *B. variegata* (Anura: Bombinatoridae) an intriguing test case of ecological speciation. Narrow *Bombina* hybrid zones erect barriers to neutral introgression whose strength has been estimated previously. We test this prediction by inferring the rate of gene exchange between pure populations on either side of the intensively studied Kraków transect. We developed a software pipeline to extract high confidence sets of orthologous genes from *de novo* transcriptome assemblies, fitted a range of divergence models to these data and assessed their relative support with analytic likelihoods calculations. There was clear evidence for post-divergence gene flow, but, as expected, no perceptible signal of recent introgression via the nearby hybrid zone. The analysis of two additional *Bombina* taxa (*B. v. scabra* and *B. orientalis)* validated our parameter estimates against a larger set of prior expectations. Despite substantial cumulative introgression over millions of years, adaptive divergence of the hybridizing taxa is essentially unaffected by their lack of reproductive isolation. Extended distribution ranges also buffer them against small-scale environmental perturbations that have been shown to reverse the speciation process in other, more recent ecotypes.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The central role of environmental heterogeneity in driving the evolution of novel ecotypes has been impressively illustrated in the rapidly accumulating literature on ecological speciation. In particular, numerous studies have demonstrated how divergent adaptation can lead to partial reproductive isolation by causing assortative mating within ecotypes, reduced performance of hybrids or both (*e.g.* Hatfield and Schluter 1999; Rundle et al. 2000; Via et al. 2000; Jiggins et al. 2001; Linn et al. 2003; Bearhop et al. 2005; Ramsey et al. 2003; Huber et al. 2007; see Nosil 2012 for a detailed treatment). Such barriers to gene flow can arise even on a historical time scale (Hendry et al. 2007). If this momentum can be maintained then, intriguingly, ecological speciation may run its course relatively swiftly.

The hypothesis of rapid ecological speciation appears to be contradicted by the millions of years of divergence that are typically required before substantial pre- and/or postzygotic isolation is observed in experimental settings (Coyne and Orr 1997; Sasa et al. 1998; Price and Bouvier 2002; Bolnick and Near 2005). interbreeding in nature may cease much earlier, if divergent adaptation includes novel habitat preferences and/or shifts in breeding/flowering time (Schemske 2010). But in the absence of irreversible barriers to reproduction, the new ecotypes are reliant on the particular environmental conditions under which they evolved (Boughman 2013). Subsequent change may increase gene flow levels so that phenotypic divergence is eroded again (Taylor et al. 2006; Seehausen et al. 2008; Vonlanthen et al. 2012; Grant and Grant 2014) or it may create periods of allopatry that facilitate the accumulation of habitat-independent gene flow barriers (Butlin et al. 2008; Bierne et al. 2013). In that case, speciation is initiated by ecological divergence but may be completed by a variety of processes. Much work

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

has been done on recently evolved ecotypes (Faria et al. 2014), many of which will not persist long enough to become reproductively isolated (Futuyma 1987; Dynesius and Jansson 2014). It is therefore of interest to determine the level of gene exchange between taxa with a long history of ecological divergence in allo- and parapatry.

Here, we report on one such taxon pair at a relatively advanced stage of the speciation process and estimate the rate of gene flow between two pure populations on a either side of a current hybrid zone. This long term average reflects introgression past genetic barriers that arise from ecological divergence and any habitat-independent incompatibilities. We discuss our results in the context of theory about genetic architectures that produce more or less efficient barriers to gene flow.

The European fire-bellied toads, *Bombina bombina* and *B. variegata,* are clearly defined taxa that differ profoundly in a large number of traits, including features of morphology, anatomy, life history and the mating system (see Szymura 1993 for an overview). Many of these differences reflect the distinct ecological niches that the toads occupy. *B. bombina* reproduces in semi-permanent lowland ponds, whereas *B. variegata* uses ephemeral breeding sites in more mountainous terrain. These aquatic habitats place opposing demands on the tadpoles: cryptic behavior to avoid invertebrate predators *versus* rapid growth and development as desiccation looms (Rafińska 1991; Werner and Anholt 1993). Morphological and behavioral differences between the tadpoles (Vorndran et al. 2002) match established patterns in specialized anuran species along the aquatic permanence gradient (*e.g.* Relyea 2001). Many differences between the adults are also likely adaptations, for example of *B. variegata* to an ever-shifting distribution of breeding sites in the landscape and of *B. bombina* to the large mating aggregations that form

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in ponds. Such simultaneous selection on multiple traits (multifarious selection) may generate particularly efficient barriers to gene flow (Rice and Hostert 1993; Nosil et al. 2009).

B. bombina and *B. variegata* diverged several million years ago (Pabijan et al. 2013) and have experienced repeated cycles of allo- and parapatry (Fijarczyk et al. 2011). Present day hybrid zones formed after the last glacial maximum (less than 10,000 years ago; Szymura 1993, 1996). They are typically 2 - 9 km wide and located along the edge of mountainous regions (Fig. 1). Recombinant marker genotypes (< 10 diagnostic loci) predominate in intensively sampled transects in Poland, Hungary, Romania and Ukraine (Szymura and Barton 1991; Gollmann 1987; Vines et al. 2003; Yanchukov et al. 2006) and make up about a quarter of individuals in the Pešćenica (Croatia) transect (MacCallum et al. 1998). We expect that all individuals in a hybrid zone are recombinant to some degree. The temporal stability of phenotypic clines at 50 and 70 year sampling intervals (Szymura and Barton 1991; Yanchukov et al. 2006) proves that natural selection counteracts the erosion of taxon differences. Direct evidence for habitat-independent natural selection comes from increased mortality of hybrid embryos under uniform conditions (Szymura and Barton 1986; Kruuk et al. 1999b).

Classic cline theory links the patterns seen in nature to the three main determinants of hybrid zones: gene flow, selection and recombination. It predicts that the observed significant linkage disequilibria among independently segregating markers (Nürnberger et al. 2003) result from the influx of pure genotypes into the hybrid zone center and increase the 'effective' selection pressure on each one of the selected loci (Szymura and Barton 1986; Barton and Gale 1993; Kruuk et al. 1999a). The associated allele frequency clines thus become steeper (Slatkin 1973) which further strengthens linkage disequilibria and effective selection. This positive feedback (Barton 1983) can explain the sharp step in allele frequency in the center of all clinal *Bombina*

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transects (Szymura and Barton 1991; Yanchukov et al. 2006). It helps to maintain adaptive taxon differences and also generates a barrier to gene flow for neutral variants (Nagylaki 1976; Barton 1986).

The strength of this barrier reflects the overall level of reproductive isolation and can be estimated from the shape of clines at marker loci (Szymura and Barton 1986; Barton and Gale 1993). The two Polish transects near Kraków and Przemyśl (width: ~ 6 km) provide a robust estimate of this barrier into the *B. variegata* gene pool (Szymura and Barton 1991). Under a diffusion model of dispersal and based on an estimated dispersal range of ~ 1 km per generation (Szymura and Barton 1991), this barrier is equivalent to 51 km of unimpeded habitat. Assuming a generation time of three years (Szymura 1988), it would delay the introgression of neutral *B. bombina* alleles by about 8000 years (Barton and Gale 1993). The estimated barrier strength also implies that the mean fitness of hybrid populations in the zone center is reduced by 42% relative to that of pure populations (Szymura and Barton 1991). Mitochondrial genomes might be expected to introgress more freely (Toews and Brelsford 2012), but this is not observed in *Bombina*: concordant mtDNA and nuclear marker clines together with very weak cytonuclear disequilibria suggest that mtDNA may itself be subject to selection in hybrids (Szymura et al. 1985; Yanchukov et al. 2006; Hofman and Szymura 2007; Fijarczyk et al. 2011; but see Vörös et al. 2006).

Thanks to the genome-wide linkage disequilibria in the hybrid zone, it was possible to draw all of these inferences from the small number of available diagnostic markers. Direct estimates of gene exchange between pure populations, however, require genome-wide genetic resources. Their development has so far been hampered by the 'big' genomes (Dufresne and Jeffery 2011) of *B. bombina* and *B. variegata*; for both taxa, C-value estimates just over the threshold of 10 pg

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(~ 10 Gbp) have been obtained (Gregory 2015). There is, however, no indication of polyploidy (Olmo et al. 1982). The two sequenced anuran genomes (*Silurana (Xenopus) tropicalis*, Hellsten et al. 2010; *Nanorana parkeri*, Sun et al. 2015) are too distant from *Bombina* (MRCA about 250 Mya) to be useful. We therefore chose *de novo* transcriptome assembly as a cost-effective method to generate reduced representation genomic data (Davey et al. 2011).

Evolutionary or population genomic analyses of next generation sequence data take a locus-centered view (Cahais et al. 2012) and require the definition of orthologous sequences across samples. With RNA-seq data, the task of ortholog definition is non-trivial, because variants of a single locus (splice isoforms, alleles) are recovered along with sequences from recent paralogs. The identification of orthologs across samples cannot succeed unless paralogous sequences from each sample are included in the search, but the result is ambiguous if splice isoforms are clustered as well. Similarly, the discovery and genotyping of single nucleotide polymorphisms (SNP) based on read mapping runs the risk of either conflating variation from several loci (false positives) or failing to detect allelic variants across different isoforms (false negatives, De Wit et al. 2015).

These issues have been variously addressed in recent *de novo* transcriptome assemblies. With the help of a reference genome, paralogs can be identified by phylogenetic methods (*e.g.* Heger and Ponting 2007; Osborne et al. 2013). Alternatively, variant call sets have been screened for suspect genotype distributions in order to remove SNPs that are likely derived from more than one locus (*e.g.* Gayral et al. 2013; Harris et al. 2015). In the absence of a reference genome and starting from either one or two samples per taxon, we devised criteria to identify putative paralogs within transcriptome assemblies. These were included in a clustering step across taxa to generate a high-confidence ortholog gene set for which SNPs were called. Each orthogroup

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contained sequences from four *Bombina* taxa with a fully resolved phylogeny based on complete mitochondrial genomes (Zheng et al. 2009; Pabijan et al. 2013): the hybridizing pair *B. bombina* and *B. v. variegata* as well as *B. v. scabra,* a subspecies from the Southern Balkans (Fig. 1), and *B. orientalis* (East Asia), the sister taxon of the European lineages.

To quantify nuclear divergence and introgression, we used a recently developed coalescent approach (Lohse et al. 2015) that fits models of divergence with or without gene flow and assesses their support via analytically computed likelihoods. This inference is extremely efficient compared to multilocus methods that require computationally intensive Markov chain Monte Carlo sampling (Hey and Nielsen 2004) or simulations (*e.g.* Shafer et al. 2015). Models were fit by two approaches: either using the site frequency spectrum (SFS) or multilocus data summarized by the blockwise SFS, *i.e.* the full configurations of mutations in blocks of sequence. With these coalescent analyses, we address the following questions: (i) Does the timing and order of divergence between the four *Bombina* lineages match that previously inferred from mitochondrial data*?,* (ii) Is there evidence for post-divergence gene flow between the hybridizing taxa away from the hybrid zone and, if so, how does it compare to the estimate between the *B. variegata* subspecies (genetic barrier to gene flow *vs*. greater spatial separation)? and (iii) Is there any evidence for a recent burst of introgression as a result of the current hybrid zone?.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Sample origins

Roche 454 sequence data were generated from two females: *B. bombina (Bb)*, Kozłów, Poland (50º29'N, 20º02'E); *B. v. variegata (Bvv)*, Mogielica, Poland (49º39'N, 20º16'E). Illumina data came from four individuals: *B. bombina*, male, Sędziejowice, Poland (50°34'N, 20°39'E); *B. v. variegata*, male, Biała Woda, near Szczawnica, Poland (49°24'N, 20°34'E); *B. v. scabra (Bvs)*, female, Rhodopes, Bulgaria; *B. orientalis (Bo)*, female, Northern clade, Korea. The European sampling locations are indicated in Figure 1. The *B. bombina* and *B. v. variegata* sample sites are on either side of the Kraków transect (Szymura and Barton 1986, 1991) and their distances to the cline center are approximately 61 km (*Bb*, Kozłów), 72 km (*Bb*, Sędziejowice), 36 km (*Bvv*, Mogielica) and 61 km (*Bvv*, Biała Woda).

RNA extraction, library preparation and sequencing

For Roche 454 sequencing, total RNA was extracted from liver, using RNAzol®RT, and mRNA was isolated using Dynabeads Oligo$(dT)_{25}$. The RevertAidTM Premium First Strand cDNA Synthesis Kit was used to synthesize the first cDNA strand with a primer complementary to the poly(A) tail with disrupted sequence: 5'-TTTTTCTTTTTTCTTTTTTV-3'. The RNA in the RNA:DNA hybrid was cleaved with RNAse H and the second strand synthesized using DNA polymerase I and T4 DNA polymerase. No additional PCR reactions were performed and the cDNA library was not normalized. Two cDNA libraries were sequenced independently (in half a FLX plate each) on a 454 GS FLX Titanium system (454 Life Sciences, Roche) at the Functional Genomics Center in Zürich (Switzerland).

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For Illumina RNA-seq, frozen samples were thawed and transferred into 1 ml Trizol, homogenized in a TissueLyser and passed five times through a 0.9 mm (20 gauge) needle attached to a syringe (Gayral et al. 2013), followed by 5 min incubation at RT and centrifugation for 1 min at 14,000 rpm. The samples were Chloroform extracted once and then mixed with 200 µl of 70% Ethanol. Each sample was divided into 2 - 3 aliquots, each of which was added to a Qiagen miRNA spin column and processed according the manufacturer's instructions. cDNA library preparation and paired-end sequencing on an Illumina HiSeq instrument were carried out by Edinburgh Genomics.

Transcriptome assemblies

Roche 454 data - Reads were cleaned by removing residues of poly(A) tails with SnoWhite v.1.1.4 (Barker et al. 2010). Reads with low-complexity regions (including stretches of di- and trinucleotide repeats which could inflict spurious assemblies) and exact read duplicates were also removed. Low quality ends were trimmed with PRINSEQ-lite 0.17 (Schmieder and Edwards 2011). Reads which showed significant sequence identity (using BLASTN) to rRNA sequences of *Silurana tropicalis* were discarded.

De novo transcriptome assemblies of *B. bombina* and *B. variegata* were conducted separately for each species. We took a combinatorial approach to optimize each assembly (Kumar and Blaxter 2010). Reads were first assembled using Newbler (v2.6, Roche) and Mira v. 3 (Chevreux et al. 2004) independently, after which both assemblies were co-assembled with CAP3 (Huang and Madan 1999). For Mira and Newbler we used settings recommended for *de novo* transcriptome assembly from reads produced by the 454 platform (Mira: no clipping of poly(A) stretches). For CAP3, we used an overlap length cutoff equal to 80, an overlap percent identity cutoff equal to 95, and switched off the clipping option. To validate the assembly, reads

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were mapped back to assembled contigs using BLAT (Kent 2002), retaining only contigs with reads that mapped over at least 95% of their length (discarded contigs: 11% of *B. bombina* and 14% of *B. variegata)*.

Illumina data - Adapter removal was carried out with Cutadapt (Martin 2011) as implemented in the wrapper Trim Galore! v. 0.3.7 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and quality trimming was performed with Trimmomatic (Bolger et al. 2014), clipping any sequence downstream of four consecutive bases with mean quality below 25. Only sequences with a minimum length of 50 nucleotides (nt) were retained (Table 1).

De novo assembly was carried out with Trinity (v. 2014-04-13p1, Grabherr et al. 2011) in paired-end mode with the addition of all retained single reads. Default settings were chosen except for the minimum K-mer count (2), the maximum expected length between fragment pairs (400) and the minimum percent identity required for the merging of paths during transcript reduction (98). Assemblies were searched for contaminants by BLAST analysis (megablast) against the NCBI nucleotide database and contigs with close to full length matches (ID > 0.90) to Bacteria and Archaea were removed (37 - 219 contigs per taxon). For each assembly, a BLASTX comparison to the *Silurana (Xenopus) tropicalis* proteome (ftp.xenbase.org/pub/Genomics/JGI/Xentr7.1/Xentr7_2_Stable_Protein.fa.gz, last accessed 16 Sept 2014) was carried out with expect-value cutoff = 0.01. Read support for each contig was quantified with RSEM (Li and Dewey 2011) using a Trinity script. All contigs with fpkm (fragments per kilobase per million reads) = 0 were removed from the assemblies (2514 - 4004 contigs per taxon). Open reading frame (ORF) predictions were computed with TransDecoder

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which is bundled with Trinity. Finally, assembly completeness was assessed with CEGMA (Parra et al. 2007).

Reference transcriptomes

For each taxon, we aimed to define a reference transcriptome that comprised all assembled gene copies (including paralogs) but excluded splice isoforms. Details of the analyses are presented in the Supporting Information and summarized here. Sequence identities (ID) were computed for pairwise nucleotide alignments within Trinity components. In the first instance, all contig pairs with overall sequence identity below 0.98 were designated as paralogs. In the case of markedly heterogeneous ID scores along the alignment, we distinguished poorly aligned tracts due to structural variation near alignment termini from those due to the alternate use of duplicated exons. The latter case represents paralogy below the level of the gene (*cf.* Gabaldón and Koonin 2013) and contigs were designated accordingly, even if the mean ID was greater or equal 0.98. In the case of *B. bombina* and *B. v. variegata*, additional pairwise comparisons were carried out, using BLASTN, of the Roche 454 contigs to the Trinity contigs. Paralogs were identified as before. For each taxon, the reference consisted of one contig per Trinity component (with maximum ORF length), all associated paralogs and, if available, all Roche 454 contigs (separated into paralogs and non-paralogs).

Orthologue identification and alignment

We used OrthoMCL (Li et al. 2003) with default parameters to cluster predicted peptide sequences from all four reference transcriptomes (Fig. 2). The complete proteome of *S. tropicalis* was added as a 'backbone' for clustering and to provide additional information on multi-gene families. From the initial clustering (inflation value = 1.5), we retained only those groups that comprised the following: (1) at least one contig from each *Bombina* taxon, (2) no paralog pairs from any *Bombina* taxon and (3) no more than three *S. tropicalis* contigs (see Fig.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2b for examples). From this subset of clusters we retained only those that were robust to changes in the inflation value (range: 0.5 - 5.0) in order to counter common errors of graph-based methods such as OrthoMCL, *i.e*. over-clustering and incorrect handling of domain recombination (Kristensen et al. 2011). Finally, pairwise peptide alignments between *Bombina* taxa were computed within robust orthogroups. Note that we apply stringent criteria to obtain a set of robust orthogroups rather than an accurate list of paralogs. False positives in that list are therefore of no concern.

Variant detection

For the coalescence analyses, only the Illumina datasets were considered (corresponding to one diploid individual per taxon). For each taxon, reads were mapped to the reference transcriptome (for *B.b*. and *B.v.v.*: Trinity contigs only) with Bowtie2 v.2.2.4 (Langmead and Salzberg 2012, mode = sensitive). Variant detection was carried out with GATK (McKenna et al. 2010). Duplicates were removed, reads from isoforms were split as required (Split'N'Trim) and base qualities were recalibrated once to convergence. Raw variants were called with the HaplotpeCaller (genotyping mode = Discovery, stand_emit_conf = 10, stand_call_conf = 20) and filtered following recommendations for RNAseq data (DePristo et al. 2011; Van der Auwera et al. 2013).

Peptide alignments between taxon pairs within orthogroups were back-translated into the nucleotide sequences of the reference contigs with PAL2NAL (Suyama et al. 2006). For a given orthogroup and taxon, an alternate haplotype was constructed from the variant call format (VCF) file and added to the between-taxon alignments. These were parsed for variation at fourfold degenerate sites.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Coalescent analyses

We used two complementary strategies to estimate models of divergence with and without gene flow analytically for each pairwise comparison of *B. v. variegata* against increasingly diverged taxa (*B. v. scabra, B. bombina, B. orientalis)*. For an unrooted alignment of a pair of diploid samples, there are four types of sites: i) heterozygote in taxon A (k_A), ii) heterozygote in taxon B (k_B), iii) double heterozygote (2 distinct alleles only, k_{AB}) and iv) alternate homozygotes (k_{AABB}). First, we based inference on the average frequencies of these site types $E[k_i]$, essentially the folded, joint site frequency spectrum (SFS) for two species. Second, we used the joint distribution of the four site types in the short blocks of fixed length (150 consecutive fourfold degenerate sites) to fit models of divergence and gene flow. The configuration of mutations in each block can be summarized as a vector of site counts $\underline{k} = \{k_A, k_B, k_{AB}, k_{AABB}\}$. Assuming blocks are unlinked, the overall model likelihood is given as the product of $p(\underline{k})$, the probability of each observed blockwise configuration of mutations across blocks. We will refer to this as the blockwise SFS (bSFS) throughout.

These two approaches trade off power against bias. The SFS ignores all linkage information and summarizes the data into four site frequencies (Tab. 2). In principle, this allows us to co-estimate four parameters. However, the near absence of shared heterozygote sites, $E[k_{AB}]$, restricted our analyses to three-parameter models. This limitation is offset by more robust inference. In particular, point estimates based on the SFS are not affected by recombination or heterogeneity in mutation rate. In contrast, the analyses based on the bSFS assume a constant mutation rate across blocks and no recombination within them. Violations of these assumptions produce biased estimates. This is of particular concern for transcriptome data, given the added potential for recombination in introns. But by considering the joint distribution of linked mutations, the bSFS allows us to fit more complex models.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lohse et *al.* (2015) have recently used a recursion for the generating function of genealogies (Lohse et al. 2011) to derive expressions for both $E[k_i]$ and $p(k)$ under a model of isolation with unidirectional migration (IM). The model assumes an ancestral population of constant effective size N_{anc} that splits into two populations A and B at time T (measured in $2N_{anc}$ generations) followed by unidirectional migration at a constant rate of $M=4N_{anc}\ m$ individuals per generation, where m is the proportion of immigrants per generation. We refer to the null model without migration as Div (Fig. 3)

Asymmetry in the frequency of unique heterozygous sites in each species (k_A and k_B) may be due to differences in effective population size, asymmetry in post-divergence gene flow or both. We therefore also considered a model in which N_e in one of the descendant species was allowed to change instantaneously after divergence, while N_e of the other species remained constant (Fig. 3). We contrasted the relative support for this model without gene flow (Div_2, M = 0) with that for its counterpart, IM, with gene flow but a single N_e for all three populations. These models with three parameters were fit using the SFS and the bSFS. Finally, we used the bSFS to fit models that allow for both post-divergence gene flow and asymmetries in N_e (IM_2). In each case, we tested for gene-flow and asymmetries in N_e in either direction (i.e. N_A = a N_{anc} or N_B = b N_{anc} and M_{AB} or M_{BA}).

We assume that orthoblocks are randomly distributed and, given the large size of the *Bombina* genome (2600 cM, Morescalchi 1965), that the effects of physical linkage between them can be ignored (a rough calculation gives a maximum density of one orthoblock per 1.8 cM). Each orthoblock can therefore be treated as an independent replicate of the coalescent. To obtain a set of unlinked SNPs, we randomly sampled a single SNP per orthoblock and bootstrapped the

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analysis across 1000 such sub-sampled data sets. 95% confidence intervals of parameter estimates were estimated as 1.96 SD of parameter estimates across bootstrap replicates. The dataset comprised more than 1000 blocks with 150 fourfold degenerate sites and an average of 2.7 - 6.0 variable sites per block (depending on the comparison, Table 2). Maximum likelihood estimates were obtained using the *Nmaximise* function in *Mathematica* (for details see Lohse et al. 2015).

Estimation of synonymous substitution rates

In order to convert coalescence time estimates, T, into absolute times, t, we derived an estimate of the substitution rate per generation, μ, at fourfold degenerate sites from a recent anuran phylogeny based on nine nuclear genes (*rag1, rag2, bdnf, pomc, cxcr4, slc8a1, scl8a3, rho, h3a,* Irisarri et al. 2012b). That study demonstrated considerable rate variation between Neobatrachians and Archaeobatrachians (including *Bombina*). We therefore used data from two taxa in the latter group, *Alytes* and *Discoglossus* (MRCA 153 Mya, Roelants et al. 2011), which form a sister lineage to *Bombina.* From the concatenated nuclear dataset (Irisarri et al. 2012a), 896 fourfold degenerate sites were extracted and the substitution rate was estimated with PAML v. 4.7. (Yang 2007). Given the three year generation time of *Bombina*, we obtained $\mu = 9.15 \times 10^{-9}$.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Assembly overview

As is typical for Trinity assemblies, a large fraction (~ 60%) of relatively short contigs with low coverage had no similarity (BLASTX) with *S. tropicalis* and no ORF prediction. For example in the *B. v. variegata* assembly, contigs without ORF (n = 56,517) were shorter (median: 411 *vs.* 1214 *nt*) and had lower coverage (fpkm median: 1.14 *vs.* 2.19, mean: 3.76 *vs.* 27.3) than those with ORF prediction (n = 24,329). Similarly, in the Roche 454 assemblies, for *B. v. variegata*, contigs without ORF (n = 19,088) had a median length of 458 *nt* compared to 700 *nt* in contigs with ORF (n = 9739). The number of annotated Trinity components was close to the total gene count in other anurans (*S. tropicalis*: 26,800, *N. parkeri*: 23,400; Sun et al. 2015) and over 90% of 248 ultra-conserved core eukaryote genes (CEGs) were either partially or completely (> 70% of total protein length) assembled for each taxon.

Paralogs

We focused our search for paralogs on those subsets of the assemblies in which ORFs had been identified, because all downstream analyses were based on coding sequence. In the Trinity assemblies, there were between 2791 (*Bvs*) and 3715 (*Bvv*) components per taxon with ORFs and more than one contig. Paralogs were identified in about 25% of these components in each taxon. In *B. v. variegata* for example, 10,753 contigs were considered in the paralog search and partitioned as follows: 3,715 contigs (with maximum ORF length) representing the components, 1,767 paralogs and a remainder of 5,271 contigs, that was removed from further analysis and presumably contained alleles and isoforms. For the other taxa, the paralog counts were 1,375 (*Bvs*), 1,737 (*Bb*) and 1,422 (*Bo*). A BLASTN search of Roche 454 contigs with

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ORFs against the Trinity database generated matches for 9,172 and 7,866 contigs in *B. v. variegata* and *B. bombina*, respectively. Of these Roche 454 contigs, 3,375 (*Bvv*) and 2,626 (*Bb*) were designated as paralogs. This relatively greater proportion reflects the overall lower identities of these BLAST-derived sequence pairs compared to those within Trinity components. Note that only the most similar Trinity/Roche 454 paralog pairs are expected to affect the definition of orthogroups (see the Supporting Information for additional information).

Orthologs

Of the 15,775 OrthoMCL orthogroups with more than one *Bombina* taxon, 1,719 were excluded because they comprised at least one pair of paralogous contigs. There were 6,772 groups from which at least one of the four *Bombina* taxa was missing. An excess (> 3) of *S. tropicalis* contigs excluded 3,845 groups. *Bombina* paralogs were the sole reason for exclusion in 716 cases (Figure 4a). After these filtering steps, 4,978 orthogroups remained. Among these, 4,040 groups were robust to changes in the inflation value and used to generate pairwise alignments among taxa. The inclusion of presumptive paralogs in the clustering analysis not only led to the removal of groups with in-paralogs but also allowed for a sequence-based decision which transcript of a paralog pair to include in a given group. Figure 4b gives an example where paralog pairs were split into two orthogroups that most likely reflect alternate exon use in a gene with homology to acid phosphatase 1 in *S. tropicalis*. Note that in this case an alternative clustering based on just one contig per component resulted in an orthogroup that mixes isoforms and would lead to erroneous estimates of sequence divergence (see Supporting Information B for a comparison of both clustering approaches).

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Variant detection

In each taxon, genotype qualities (a Phred-scaled measure of the confidence in the reported genotype) reached the maximum reported value (99) for more than 2/3 of SNPs and were 40 or over for over 95% of SNPs. We observed a small number of mostly low quality (< 20) genotype calls (0.5 - 3.4% per taxon) that did not include the reference allele (*i.e.* alternate homozygotes). With only one diploid individual per taxon, these cases point to assembly errors and were removed from further analysis.

Average heterozygosity (k_A, k_B) at fourfold degenerate sites ranged from 0.1 % in *B. bombina* to 0.7 % in *B. orientalis* (Table 2). Per site synonymous divergence (k_{AABB}) was lowest between the subspecies *B. v. scabra* and *B. v. variegata*, intermediate between *B. variegata* and *B. bombina* and greatest between *B. bombina* and *B. orientalis* (Table 2). This is expected given the order of taxon divergence previously inferred from protein and mitochondrial data (Pabijan et al. 2013).

Coalescence analysis

In both the *B. v. variegata/B. v. scabra* and the *B. variegata/B. bombina* comparison, an IM history fit the data significantly better than a history of isolation without gene flow. This was true regardless of whether we allowed for differences in N_e or not (Table 3) or whether the inferences was based on the SFS or the bSFS. Estimates of gene flow (*M*) were low (< 0.1) in both comparisons, but *M* was significantly greater between *B. v. variegata* and *B. v. scabra* than between *B. v. variegata* and *B. bombina*, both in the SFS and the bSFS (Table 4).

Our estimates of the divergence time also agreed well between SFS and bSFS analyses (Table 4). We converted estimates of *T* into absolute time using $t = T * 2N * g$, where is *g* is generation time, and $N = \theta / (4\mu)$, using an estimate of μ from a related pair of Archaeobatrachians (*cf.*

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods). Based on the bSFS analysis, this resulted in ancestral population sizes of 0.33 - 2 x 10^5. The divergence times for *B. v. variegata* and *B. v. scabra* were 3.6 and 2.5 mya for the SFS and the bSFS analysis, respectively. The analogous figures for the other comparisons were 3.9 and 3.3 mya (*Bvv/Bb*) and 4.6 mya (*Bvv/Bo*).

The SFS and bSFS analyses differed in several respects. First, the blockwise analyses showed no support for gene flow in the allopatric comparison of *B. variegata* and *B. orientalis.* In other words, allowing for gene flow as well as differing population sizes (IM_2) did not improve the fit compared to a scenario of strict divergence with varying N_e (Div_2, Table 3). In contrast, the SFS approach favored a model of strong gene flow (M = 0.136) from *B. v. variegata* to *B. orientalis* (IM) over Div_2. Second, using the SFS the best fitting model involved gene flow and always into the gene pool with relatively greater heterozygosity (*e.g*. from *B. bombina* to *B. v. variegata*). In contrast, the bSFS analyses could separately account for varying N_e and, in the case of the two *B. variegata* subspecies, gave significantly better support for gene flow in the opposite direction (*i.e.* from *B. v. scabra* to *B. v. variegata*). Finally, estimates of *M* were lower in the bSFS than the SFS analyses (Table 4).

To assess the relative fit of the best supported models, we computed the expected SFS under the models inferred from the bSFS (Table 2). We also tested how much of the gene flow signal is due to the presence of shared heterozygous sites. For *B. variegata* and *B. bombina,* only one block contained shared heterozygous sites and removing this block had no effect on the results. Instead, most of the signal for gene flow came from an excess of blocks with no fixed differences (k_{AABB}= 0, Fig. 5). For *B. variegata* and *B. bombina* there are 134 such blocks which is highly unlikely under the best fitting divergence scenario (p= 0.002 given a total of 1,479 blocks and assuming binomial sampling), but not under the IM_2 model (p = 0.89). We can use

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analogous coalescent calculations to ask whether there is an excess of invariant blocks (k = {0,0,0,0}), as would perhaps be expected if introgression between *B. variegata* and *B. bombina* occurred in a recent burst following the formation of the hybrid zone, or was partially adaptive. There were 95 invariant blocks, which is not unexpected under the IM_2 model (p = 0.235).

Discussion

The coalescent analyses provide evidence that hybridization between *B. bombina* and *B. v. variegata* results in continued gene exchange between these anciently diverged taxa, but the estimated long term average of gene flow was very low (M = 0.020 or 0.027, depending on the analysis). Thus despite the abundance of fertile hybrids in contact zones, the effective rate of gene exchange between the two pure populations was no greater than one immigrant per 100 - 150 years. These values of M are at least 40 times smaller than an estimate by the same method for two sympatric *Heliconius* species that hybridize infrequently (Lohse et al. 2015). However when integrated over time, the *Bombina* figures imply that a substantial portion of genes has been affected by past gene flow: based on the results from the bSFS analysis, this proportion for *B. v. variegata* is $1 - e^{-0.020 * 16.8} = 0.285$ (Table 4). The analogous figure for *B. bombina* is 0.24.

Before we explore our results in more detail, we comment on aspects of our methodology. We devised an analysis pipeline that identified putative paralogs based on nucleotide sequence identity and excluded all cross-taxon gene clusters with paralog pairs from further analysis. This filtering step will also have eliminated the more serious cases of incorrectly phased variants (Haas et al. 2013). By using the same tissue (liver) for all samples, we avoided problems due to tissue-specific expression. However, some errors will remain. For any recent paralog pairs

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

above our identity threshold (0.98), only one contig was used in the clustering step and so orthology may have been mis-assigned. Finally, differential gene loss (Rasmussen and Kellis 2012) may have resulted in some paralogous clusters.

Interpretation of the coalescent results requires weighing the strengths and weaknesses of both inference schemes. Recombination within blocks reduces the variance in coalescence times in the bSFS scheme and so leads to underestimates of M and N_e (Wall 2003). Our blocks of 150 fourfold degenerate sites were on average 1500 nt long, which is longer than the average CDS length in *S. tropicalis* (1323 nt) and *N. parkeri* (1382 nt, Sun et al. 2015). In these fully sequenced genomes, the average gene length is 18.0 kb and 24.4 kb, respectively (Sun et al. 2015). This suggests an average span of 20 kb or more for the genomic complements of our blocks and a likely influence of recombination on the parameter estimates (recall the relatively lower values for M, Table 4). Overall, though, the biases appear small, because the estimates for a given model agreed well between bSFS and SFS analyses. Selective sweeps at linked sites in the ancestral population may have the opposite effect of inducing additional variation in coalescence times (Coop and Ralph 2012). The excess of blocks with no or few divergent sites in the European comparisons may in principle be explained in this way. But no such excess is seen in the allopatric comparison (*Bvv/Bo*) which suggests that it is instead a genuine signal of past migration.

While the SFS analysis makes fewer assumptions, it confounds gene flow with differences in effective population size since the lineage split. Thus, when assuming a single N_e, we always inferred gene flow into the taxon with higher heterozygosity and obtained unrealistic estimates for the allopatric pair (Table 4). In the bSFS analyses with two N_e parameters, M estimates were reduced to zero as expected (*Bvv/Bo*) or the direction of gene flow was reversed (*Bvv/Bvs*). The

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bSFS contains additional information to distinguish gene flow from asymmetries in N_e: a larger N_e in one descendant population increases heterozygosity uniformly whereas gene flow generates a small number of blocks with large numbers of heterozygous sites in the population receiving migrants.

The observed heterozygosities per taxon (Table 2) fully match our expectations from previous studies. *B. bombina,* with the lowest estimate (k_B = 0.00106), has recently expanded its range from glacial refugia near the Black Sea and has displayed minimal genetic diversity in all previous *Bombina* comparisons (Szymura 1988, 1993; Fijarczyk et al. 2011). The difference in heterozygosity between the *B. variegata* subspecies agrees with previously reported differences in N_e (Fijarczyk et al. 2011). Finally, the *B. orientalis* specimen (max. estimate in this study, k_B = 0.00748) comes from Korea, an important Pleistocene refugium in Asia (Zheng et al. 2009).

While the blockwise analyses allow a more detailed reconstruction of the divergence history, they still fit a very simplistic model. In particular, the assumption of unidirectional gene flow is unlikely to apply to *Bombina*. In fact for the hybridizing taxa, both the support for the two versions of the IM_2 model (A→B *vs*. B→A, Table 3) and the M estimates under these models are very similar (M in the less favored direction (*Bvv*→*Bb*) is 0.0127). Note also that the support for asymmetric N_e is very weak and M for the IM model with a single N_e (*Bb* →*Bvv*) is also 0.02. For the two *B. variegata* subspecies, the best supported model ($IM_{2,B\rightarrow A}$) matches previous findings of relatively more gene flow from the Balkans into the Carpathians (Fijarczyk et al. 2011). Given these limitations and considerations of bias above, *M* estimates based on the bSFS should be interpreted as a minimum, long-term average of gene flow following the lineage split.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The species divergence times we have inferred here are by necessity more recent than the divergence at any particular gene (which includes an additional coalescence time in the common ancestral population). However, the difference between our estimates and previous calibrations for mitochondrial genes is surprising. For example for the split between *B. v. variegata* and *B. bombina*, we obtain 3.3 Mya compared to the 6.4 or 8.9 Mya divergence at mtDNA genes (two different fossil calibrations, Pabijan et al. 2013; see also Heled and Drummond 2015). While some of this difference may reflect coalescent stochasticity (*i.e*. by chance, the coalescence of mitochondrial genomes may considerably predate the taxon split), it likely reflects differences in calibration.

The new divergence time estimate is consistent with the appearance of the first *Bombina bombina* fossils in the late Pliocene (Rage and Roček 2003) and with analyses of allozymes (Nei's D, Szymura 1993). But it too is clearly beset by uncertainties. We therefore conclude cautiously that *B. bombina* and *B. v. variegata* diverged in the mid-Pliocene, at a time when much of their present-day distribution range in Central Europe was still covered by remnants of the Paratethys Sea. Subsequent climatic oscillations during the Quarternary will have resulted in repeated range shifts (*cf.* Futuyma 1987), presumably with recurring opportunities for gene exchange that have left a clear signal in present-day transcriptomes.

However, our dataset contains no signal of increased introgression due to ongoing hybridization. While the proportion of invariant blocks is significantly larger than expected under strict divergence, it is consistent with a constant rate of gene flow. This observation agrees with the prediction from cline theory (cf. Introduction): considering the barrier effect (51 km) and the distance of the *B. v. variegata* sample from the hybrid zone (61 km), it should take roughly 18,500 years for genetic variants to cross the Kraków transect and arrive at Biała Woda.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Assuming similar barrier strengths, the *B. bombina* sample (71 km) should be insulated even longer from introgression *via* that route. Selectively favored alleles, however, will introgress much more quickly (Barton 1979). For example, an allele with a 1% selective advantage is expected to traverse the Kraków transect in a mere 245 years (*via* expression in Barton and Gale, 1993, p. 30 top). Our data set may well contain a few blocks of this kind, but larger samples of individuals and long range linkage information would be needed to detect them.

The strength of the genetic barrier to gene flow varies among different *Bombina* contact zones. Near Apahida, Romania, the two taxa meet in an extended habitat mosaic at a distance of 20 km from the nearest pure species range (Vines et al. 2003). Plausible total selection strengths can be found that could balance the high estimated gene flow rates and maintain adaptive taxon differences indefinitely. But with, say, 20 selected loci, neutral marker clines would dissipate within ~ 160 years. A very dense spacing of the total selection pressure would be required (~ one selected locus every 4 cM, *i.e.* about 650 loci) to maintain such clines over 10,000 years. In fact, there are reasons to believe that this particular contact zone is much more recent (Vines et al. 2003). Spatial and temporal heterogeneity in rates of gene exchange is a common theme in the hybrid zone literature (Harrison and Larson 2014) and often linked to environmental conditions, as in this example. Episodic spates of gene flow appear to be just as much a product of environmental heterogeneity as any initial adaptive divergence. But given the slow rate of diffusion, neutral introgression should attenuate quickly with increasing distance from the hybrid zone, even in Apahida.

A post-zygotic barrier to neutral gene flow depends on physical linkage between neutral and selected loci and thus becomes stronger as a given total selection stength is more finely distributed across the genome (Barton and Bengtsson 1986). In *Bombina*, the number of loci

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

under selection has been estimated to be at least 55 and is most likely considerably larger (Szymura and Barton 1991). Abundant trait differences also suggests that multifarious selection may promote reproductive isolation, not just because many loci are involved but also because of the pleiotropic interactions they may have (Nosil and Harmon 2009). Under Fisher's (1930) geometric model, such interactions lengthen the adaptive walk towards a new phenotypic optimum and, as a consequence, reduce the mean fitness of hybrids upon secondary contact (Chevin et al. 2014). The largest drop in mean hybrid fitness is expected after the first few mutations have fixed, but the total fitness reduction at the end of an adaptive walk is only moderate (< 50%, Barton 2001). This polygenic model of adaptive divergence thus predicts that hybrids would continue to be conduits of gene flow unless adaptive substitutions also trigger strongly deleterious epistatic interactions (i.e. Dobzhansky-Muller incompatibilities, Chevin et al. 2014). Such incompatibilities exist in *Bombina* (*cf.* Introduction). They may or may not result from ecological divergence, but their joint effect is clearly insufficient to cause reproductive isolation. We do not know at present why this is so but note that one important driver of speciation is missing: these toads do not have heteromorphic sex chromosomes and so Haldane's Rule (Orr 1997) does not apply.

As it stands, *B. bombina* and *B. variegata* play important roles in their respective ecological communities, just like ‘good’ biological species, despite their gene pools fraying at the edges. With an estimate of m on the order of 10^{-6} only a few dozen dispersal ranges away from the hybrid zone, adaptive divergence can proceed uninhibited by hybridization. Similarly low gene flow rates were obtained for hybridizing sunflower species (Sambatti et al. 2012). Under these para-allopatric conditions (Coyne and Orr 2004), there is no antagonism between selection and gene flow. Consequently, any speciation mechanism may increase reproductive isolation between well separated, pure population of either taxon without constraints on the underlying

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genetic architecture. In contrast, local adaptation under selection-gene flow balance between two panmictic populations favors a clustered genomic distribution of selected loci (Felsenstein 1981; Ortíz-Barrientos et al. 2002) which can gradually evolve in redundant genetic systems (Yeaman and Whitlock 2011). Especially when such traits under divergent natural selection also cause assortative mating (magic traits, Gavrilets 2004; Servedio et al. 2011), novel ecotypes may rapidly form. But as long as prezygotic barriers are incomplete or labile under environmental perturbations, hybrids will be generated and post-zygotic barriers to gene flow are likely to be inefficient.

The conundrum of incomplete speciation in *Bombina* illustrates that speciation tends to defy rules (Coyne and Orr 1989) and argues for an open-minded exploration of possible paths towards reproductive isolation. Rapid evolution of reproductive isolation between recent ecotypes is feasible whenever it can overcome the challenges of a limited gene pool from which to draw mutations that drastically reduce gene flow, the trend towards a clustered genomic distribution of selected traits and the need to build up irreversible barriers to gene flow before further environmental change threatens to reverse diversification. This suggests that many young ecotypes will be ephemeral, because they do not persist long enough to reach complete isolation (Futuyma 1987). Those ecotypes that adapt to abundant and widely distributed resource should stand a relatively much better chance to become reproductively isolated in their own time. This does not diminish the role of ecology in generating biological diversity. To the contrary, it underscores that relatively young ecotypes can play important roles in their respective communities and even prompt diversification in other taxa (Abrahamson and Blair 2007), irrespective of whether they are reproductively isolated. We hasten to add that this view has a very long history indeed (Darwin 1859; see Mallet 2008).

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data archiving

Raw Illumina reads were submitted to SRA with accession numbers PRJEB7348-PRJEB7351. *De novo* assemblies of Roche 454 and Illumina data including orthology information are available on the AfterParty platform (link). The source code for the paralog filter written in Perl can be found at github/.... (see note in Supporting Information A).

Acknowledgments

We thank N. Wrobel (Edinburgh Genomics) and M. Kuenzli (Functional Genomics Center, Zürich) for help with RNA extraction and Roche 454 sequencing, respectively. W. Babik, B. Elsworth, M. Jones, G. Koutsovoulos, S. Kumar and D. Laetsch generously shared their bioinformatics and computing expertise. We thank N. Barton for comments on a previous version of this manuscript. BN was supported by a Daphne Jackson Fellowship funded by the Natural Environment Research Council. AF received funds from the European Science Foundation (FroSpects) and Society-Environment-Technology (www.set.uj.edu.pl) for an exchange visit in Edinburgh. Collecting permits were issued by the Minister of the Environment, Republic of Poland, DLKOPiK-op/ogiz-4200/II-6/1441/07/msz.s. The study was approved by the First Local Ethical Committee on Animal Testing, Jagiellonian University, Kraków (10/OP/2006). We acknowledge financial support from the Polish Ministry of Science and Education (N303 11832/2735) and the National Science Centre (2013/09/B/NZ8/03349) to JMS.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

595 **References**

Abrahamson, W. G., and C. P. Blair. 2007. Sequential radiation through host-race formation: herbivore diversity leads to diversity in natural enemies: Pp. 188–202 *in* K. J. Tilmon, ed. Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects. University of Chicago Press, Berkeley, CA.

Barker, M. S., K. M. Dlugosch, L. Dinh, R. S. Challa, N. C. Dracopoli, M. G. King, and L. H. Rieseberg. 2010. EvoPipes.net: Bioinformatic Tools for Ecological and Evolutionary Genomics. Evol. Bioinforma. Online 6:143–149.

Barton, N. H. 1979. Gene flow past a cline. Heredity 43:333–339.

Barton, N. H. 1983. Multilocus clines. Evolution 37:454–471.

Barton, N. H. 1986. The effects of linkage and density-dependent regulation on gene flow. Heredity 57:415–426.

Barton, N. H., and B.-O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. Heredity 56:357–376.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University Press, Oxford.

Bearhop, S., W. Fiedler, R. W. Furness, S. C. Votier, S. Waldron, J. Newton, G. J. Bowen, P. Berthold, and K. Farnsworth. 2005. Assortative mating as a mechanism for rapid evolution of a migratory divide. Science 310:502–504.

Bierne, N., P.-A. Gagnaire, and P. David. 2013. The geography of introgression in a patchy environment and the thorn in the thigh of ecological speciation. Curr. Zool. 59:72–86.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics btu170.

Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution 59:1754–1767.

Boughman, J. W. 2013. Ecological speciation: selection and the origin of species. Curr. Zool. 59:1–6.

Butlin, R. K., J. Galindo, and J. W. Grahame. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:2997–3007.

Cahais, V., P. Gayral, G. Tsagkogeorga, J. Melo-Ferreira, M. Ballenghien, L. Weinert, Y. Chiari, K. Belkhir, V. Ranwez, and N. Galtier. 2012. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. Mol. Ecol. Resour. 12:834–845.

Chevin, L.-M., G. Decorzent, and T. Lenormand. 2014. Niche dimensionality and the genetics of ecological speciation. Evolution 68:1244–1256.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chevreux, B., T. Pfisterer, B. Drescher, A. J. Driesel, W. E. G. Müller, T. Wetter, and S. Suhai. 2004. Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. Genome Res. 14:1147–1159.

Coop, G., and P. Ralph. 2012. Patterns of neutral diversity under general models of selective sweeps. Genetics 192:205–224.

Coyne, J. A., and H. A. Orr. 1997. Patterns of speciation in *Drosophila* revisited. Evolution 51:295–303.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, Mass.

Coyne, J. A., and H. A. Orr. 1989. Two rules of speciation. Pp. 180–207 *in* D. Otte and J. A. Endler, eds. Speciation and its consequences. Sinauer Associates, Sunderland, Massachusetts.

Darwin, C. 1859. On the origin of species by mean of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499–510.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498.

De Wit, P., M. H. Pespeni, and S. R. Palumbi. 2015. SNP genotyping and population genomics from expressed sequences – current advances and future possibilities. Mol. Ecol. 24:2310–2323.

Dufresne, F., and N. Jeffery. 2011. A guided tour of large genome size in animals: what we know and where we are heading. Chromosome Res. 19:925–938.

Dynesius, M., and R. Jansson. 2014. Persistence of within-species lineages: a neglected control of speciation rates. Evolution 68:923–934.

Faria, R., S. Renaut, J. Galindo, C. Pinho, J. Melo-Ferreira, M. Melo, F. Jones, W. Salzburger, D. Schluter, and R. Butlin. 2014. Advances in Ecological Speciation: an integrative approach. Mol. Ecol. 23:513–521.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. Evolution 35:124–138.

Fijarczyk, A., K. Nadachowska, S. Hofman, S. N. Litvinchuk, W. Babik, M. Stuglik, G. Gollmann, L. Choleva, D. Cogalniceanu, T. Vukov, G. Dzukić, and J. M. Szymura. 2011. Nuclear and mitochondrial phylogeography of the European fire-bellied toads *Bombina bombina* and *Bombina variegata* supports their independent histories. Mol. Ecol. 20:3381–3398.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford University Press, Oxford.

Futuyma, D. J. 1987. On the role of species in anagensis. Am. Nat. 130:465–473.

Gabaldón, T., and E. V. Koonin. 2013. Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14:360–366.

Gavrilets, S. 2004. Fitness Landscapes and the Origin of Species. Princeton University Press, Princeton, New Jersey.

Gayral, P., J. Melo-Ferreira, S. Glémin, N. Bierne, M. Carneiro, B. Nabholz, J. M. Lourenco, P. C. Alves, M. Ballenghien, N. Faivre, K. Belkhir, V. Cahais, E. Loire, A. Bernard, and N. Galtier. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. PLoS Genet 9:e1003457.

Gollmann, G. 1987. *Bombina bombina* and *B. variegata* in the Mátra mountains (Hungary): new data on distribution and hybridization (Amhibia, Anura, Discoglossidae). Amphib.-Reptil. 8:213–224.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Grant, P. R., and B. R. Grant. 2014. Speciation undone. Nature 507:178–179.

Gregory, T. R. 2015. Animal Genome Size Database. http://www.genomesize.com

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. J. Hered. 105:795–809.

Harris, S. E., R. J. O'Neill, and J. Munshi-South. 2015. Transcriptome resources for the white-footed mouse (*Peromyscus leucopus*): new genomic tools for investigating ecologically divergent urban and rural populations. Mol. Ecol. Resour. 15:382–394.

Hatfield, T., and D. Schluter. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution 53:866–873.

Heger, A., and C. P. Ponting. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Res. 17:1837–1849.

Heled, J., and A. J. Drummond. 2015. Calibrated birth–death phylogenetic time-tree priors for Bayesian inference. Syst. Biol. 64:369–383.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, and D. S. Rokhsar. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. Science 328:633–636.

Hendry, A. P., P. Nosil, and L. H. Rieseberg. 2007. The speed of ecological speciation. Funct. Ecol. 21:455–464.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760.

Hofman, S., and J. M. Szymura. 2007. Limited mitochondrial DNA introgression in a *Bombina* hybrid zone. Biol. J. Linn. Soc. 91:295–306.

Huang, X., and A. Madan. 1999. CAP3: a DNA sequence assembly program. Genome Res. 12:868–877.

Huber, S. K., L. F. De León, A. P. Hendry, E. Bermingham, and J. Podos. 2007. Reproductive isolation of sympatric morphs in a population of Darwin's finches. Proc. R. Soc. Lond. B Biol. Sci. 274:1709–1714.

Irisarri, I., D. San Mauro, F. Abascal, A. Ohler, M. Vences, and R. Zardoya. 2012a. Data from: The origin of modern frogs (Neobatrachia) was accompanied by the acceleration in mitochondrial and nuclear substitution rates.

Irisarri, I., D. San Mauro, F. Abascal, A. Ohler, M. Vences, and R. Zardoya. 2012b. The origin of modern frogs (Neobatrachia) was accompanied by acceleration in mitochondrial and nuclear substitution rates. BMC Genomics 13:626.

Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302–305.

Kent, W. J. 2002. BLAT—The BLAST-Like Alignment Tool. Genome Res. 12:656–664.

Kristensen, D. M., Y. I. Wolf, A. R. Mushegian, and E. V. Koonin. 2011. Computational methods for gene orthology inference. Brief. Bioinform. 12:379–391.

Kruuk, L. E. B., S. J. E. Baird, K. S. Gale, and N. H. Barton. 1999a. A comparison of multilocus clines by environmental adaptation or by selection against hybrids. Genetics 153:1959–1971.

Kruuk, L. E. B., J. S. Gilchrist, and N. H. Barton. 1999b. Hybrid dysfunction in fire-bellied toads (*Bombina*). Evolution 53:1611–1616.

Kumar, S., and M. L. Blaxter. 2010. Comparing *de novo* assemblers for 454 transcriptome data. BMC Genomics 11:571.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. 13:2178–2189.

Linn, C., J. L. Feder, S. Nojima, H. R. Dambroski, S. H. Berlocher, and W. Roelofs. 2003. Fruit odor discrimination and sympatric host race formation in Rhagoletis. Proc. Natl. Acad. Sci. USA 100:11490–11493.

Lohse, K., M. Chmelik, S. H. Martin, and N. H. Barton. 2015. Strategies for calculating blockwise likelihoods under the coalescent.

Lohse, K., R. J. Harrison, and N. H. Barton. 2011. A general method for calculating likelihoods under the coalescent process. Genetics 189:977–987.

MacCallum, C. J., B. Nürnberger, N. H. Barton, and J. M. Szymura. 1998. Habitat preference in the *Bombina* hybrid zone in Croatia. Evolution 52:227–239.

Mallet, J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:2971–2986.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:pp. 10–12.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Morescalchi, A. 1965. Osservazioni sulla cariologia di *Bombina*. Boll.Zool. 32:207–219.

Nagylaki, T. 1976. Clines with variable migration. Genetics 83:867–886.

Nosil, P. 2012. Ecological Speciation. Oxford University Press, Oxford.

Nosil, P., and L. Harmon. 2009. Niche dimensionality and ecological speciation. Pp. 127–154 *in* R. K. Butlin, J. R. Bridle, and D. Schluter, eds. Speciation and patterns of diversity. Cambridge University Press, Cambridge.

Nosil, P., L. J. Harmon, and O. Seehausen. 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.

Nürnberger, B., S. Hofman, B. Förg-Brey, G. Praetzel, A. Maclean, J. M. Szymura, C. M. Abbott, and N. H. Barton. 2003. A linkage map for the hybridising toads *Bombina bombina* and *B. variegata* (Anura: Discoglossidae). Heredity 91:136–142.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Olmo, E., A. Morescalchi, V. Stingo, and G. Odierna. 1982. Genome characteristics and the systematics of Discoglossidae (Amphibia, Salientia). Monit. ZoolItalNS 16:283–299.

Orr, H. A. 1997. Haldane's Rule. Annu. Rev. Ecol. Syst. 28:195–218.

Ortíz-Barrientos, D., J. Reiland, J. Hey, and M. A. F. Noor. 2002. Recombination and the divergence of hybridizing species. Genetica 116:167–178.

Osborne, O. G., T. E. Batstone, S. J. Hiscock, and D. A. Filatov. 2013. Rapid speciation with gene flow following the formation of Mt. Etna. Genome Biol. Evol. 5:1704–1715.

Pabijan, M., A. Wandycz, S. Hofman, K. Węcek, M. Piwczyński, and J. M. Szymura. 2013. Complete mitochondrial genomes resolve phylogenetic relationships within *Bombina* (Anura: Bombinatoridae). Mol. Phylogenet. Evol. 69:63–74.

Parra, G., K. Bradnam, and I. Korf. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067.

Price, T. D., and M. M. Bouvier. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56:2083–2089.

Rafińska, A. 1991. Reproductive biology of the fire-bellied toads, *Bombina bombina*and *B. variegata* (Anura: Discoglossidae): egg size, clutch size and larval period length differences. Biol. J. Linn. Soc. 43:197–210.

Rage, J.-C., and Z. Roček. 2003. Evolution of anuran assemblages in the Tertiary and Quaternary of Europe, in the context of palaeoclimate and palaeogeography. Amphib.-Reptil. 24:133–167.

Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the Monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). Evolution 57:1520–1534.

Rasmussen, M. D., and M. Kellis. 2012. Unified modeling of gene duplication, loss, and coalescence using a locus tree. Genome Res. 22:755–765.

Relyea, R. A. 2001. The relationship between predation risk and antipredator responses in larval anurans. Ecology 82:541–554.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? Evolution 47:1637–1653.

Roelants, K., A. Haas, and F. Bossuyt. 2011. Anuran radiations and the evolution of tadpole morphospace. Proc. Natl. Acad. Sci. 108:8731–8736.

Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatic sticklebacks. Science 287:306–308.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sambatti, J. B. M., J. L. Anderson, D. Ortiz-Barrientos, E. J. Baack, and L. H. Rieseberg. 2012. Reconciling extremely strong barriers with high levels of gene exchange in annual sunflowers. Evolution 66:1459–1473.

Sasa, M. M., P. T. Chippindale, and N. A. Johnson. 1998. Patterns of postzygotic isolation in frogs. Evolution 52:1811–1820.

Schemske, D. W. 2010. Adaptation and *The Origin of Species*. Am. Nat. 176:S4–S25.

Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864.

Seehausen, O., G. Takimoto, D. Roy, and J. Jokela. 2008. Speciation reversal and biodiversity dynamics with hybridization in changing environments. Mol. Ecol. 17:30–44.

Servedio, M. R., G. S. V. Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: “magic” but not rare? Trends Ecol. Evol. 26:389–397.

Shafer, A. B. A., L. M. Gattepaille, R. E. A. Stewart, and J. B. W. Wolf. 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: in silico evaluation of power, biases and proof of concept in Atlantic walrus. Mol. Ecol. 24:328–345.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733–756.

Sun, Y.-B., Z.-J. Xiong, X.-Y. Xiang, S.-P. Liu, W.-W. Zhou, X.-L. Tu, L. Zhong, L. Wang, D.-D. Wu, B.-L. Zhang, C.-L. Zhu, M.-M. Yang, H.-M. Chen, F. Li, L. Zhou, S.-H. Feng, C. Huang, G.-J. Zhang, D. Irwin, D. M. Hillis, R. W. Murphy, H.-M. Yang, J. Che, J. Wang, and Y.-P. Zhang. 2015. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. Proc. Natl. Acad. Sci. 112:E1257–E1262.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.

Szymura, J. M. 1993. Analysis of hybrid zones with *Bombina*. Pp. 261–289 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University Press, New York.

Szymura, J. M. 1988. Regional differentiation between hybridizing species *Bombina bombina* (L.) and *B. variegata* (L.) (Salientia, Discoglossidae) in Europe. Rozprawy Habilitacyjne No. 147. Uniwersytet Jagielloński, Kraków.

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B variegata*, near Cracow in Southern Poland. Evolution 40:1141–1159.

Szymura, J. M., and N. H. Barton. 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparison between transects and between loci. Evolution 45:237–261.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Szymura, J. M., C. Spolsky, and T. Uzzell. 1985. Concordant change in mitochondrial and nuclear genes in a hybrid zone between two frog species (genus *Bombina*). Experientia 41:1469–1470.

Taylor, E. B., J. W. Boughman, M. Groenenboom, M. Sniatynski, D. Schluter, and J. L. Gow. 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. Mol. Ecol. 15:343–355.

Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21:3907–3930.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Pp. 11.10.1–11.10.13 *in* Current Protocols in Bioinformatics. John Wiley & Sons, Inc.

Via, S., A. C. Bouck, and S. Skillman. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. ii. selection against migrants and hybrids in the parental environments. Evolution 54:1626–1637.

Vines, T. H., S. C. Köhler, M. Thiel, I. V. Ghira, T. R. Sands, C. J. MacCallum, N. H. Barton, and B. Nürnberger. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. Evolution 57:1876–1888.

Vonlanthen, P., D. Bittner, A. G. Hudson, K. A. Young, R. Müller, B. Lundsgaard-Hansen, D. Roy, S. Di Piazza, C. R. Largiader, and O. Seehausen. 2012. Eutrophication causes speciation reversal in whitefish adaptive radiations. Nature 482:357–362.

Vorndran, I. C., E. Reichwaldt, and B. Nürnberger. 2002. Does differential susceptibility to predation in tadpoles stabilize the *Bombina* hybrid zone? Ecology 83:1648–1659.

Vörös, J., M. Alcobendas, I. Martínez-Solano, and M. García-Paris. 2006. Evolution of *Bombina bombina* and *Bombina variegata* (Anura: Discoglossidae) in the Carpathian Basin: a history of repeated mt-DNA introgression across species. Mol. Phylogenet. Evol. 38:705–718.

Wall, J. D. 2003. Estimating ancestral population sizes and divergence times. Genetics 163:395–404.

Werner, E. E., and B. R. Anholt. 1993. Ecological consequences of the trade-off between growth and mortality rates mediated by foraging activity. Am. Nat. 142:242–272.

Yanchukov, A., S. Hofman, J. M. Szymura, S. Mezhzherin, S. Y. Morozov-Leonov, N. H. Barton, and B. Nürnberger. 2006. Hybridization of *Bombina bombina* and *B. variegata* (Anura, Discoglossidae) at a sharp ecotone in Western Ukraine: comparisons across transects and over time. Evolution 60:583–600.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Yeaman, S., and M. C. Whitlock. 2011. The architecture of adaptation under migration-selection balance. Evolution 65:1897–1911.

Zheng, Y., J. Fu, and S. Li. 2009. Toward understanding the distribution of Laurasian frogs: a test of Savage's biogeographical hypothesis using the genus *Bombina*. Mol. Phylogenet. Evol. 52:70–83.Abrahamson, W. G., and C. P. Blair. 2007. Sequential radiation through host-race formation: herbivore diversity leads to diversity in natural enemies: Pp. 188–202 *in* K. J. Tilmon, ed. Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects. University of Chicago Press, Berkeley, CA.

Barker, M. S., K. M. Dlugosch, L. Dinh, R. S. Challa, N. C. Dracopoli, M. G. King, and L. H. Rieseberg. 2010. EvoPipes.net: Bioinformatic Tools for Ecological and Evolutionary Genomics. Evol. Bioinforma. Online 6:143–149.

Barton, N. H. 1979. Gene flow past a cline. Heredity 43:333–339.

Barton, N. H. 1983. Multilocus clines. Evolution 37:454–471.

Barton, N. H. 1986. The effects of linkage and density-dependent regulation on gene flow. Heredity 57:415–426.

Barton, N. H., and B.-O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. Heredity 56:357–376.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University Press, Oxford.

Bearhop, S., W. Fiedler, R. W. Furness, S. C. Votier, S. Waldron, J. Newton, G. J. Bowen, P. Berthold, and K. Farnsworth. 2005. Assortative mating as a mechanism for rapid evolution of a migratory divide. Science 310:502–504.

Bierne, N., P.-A. Gagnaire, and P. David. 2013. The geography of introgression in a patchy environment and the thorn in the thigh of ecological speciation. Curr. Zool. 59:72–86.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics btu170.

Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution 59:1754–1767.

Boughman, J. W. 2013. Ecological speciation: selection and the origin of species. Curr. Zool. 59:1–6.

Butlin, R. K., J. Galindo, and J. W. Grahame. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:2997–3007.

Cahais, V., P. Gayral, G. Tsagkogeorga, J. Melo-Ferreira, M. Ballenghien, L. Weinert, Y. Chiari, K. Belkhir, V. Ranwez, and N. Galtier. 2012. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. Mol. Ecol. Resour. 12:834–845.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chevin, L.-M., G. Decorzent, and T. Lenormand. 2014. Niche dimensionality and the genetics of ecological speciation. Evolution 68:1244–1256.

Chevreux, B., T. Pfisterer, B. Drescher, A. J. Driesel, W. E. G. Müller, T. Wetter, and S. Suhai. 2004. Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. Genome Res. 14:1147–1159.

Coop, G., and P. Ralph. 2012. Patterns of neutral diversity under general models of selective sweeps. Genetics 192:205–224.

Coyne, J. A., and H. A. Orr. 1997. Patterns of speciation in *Drosophila* revisited. Evolution 51:295–303.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, Mass.

Coyne, J. A., and H. A. Orr. 1989. Two rules of speciation. Pp. 180–207 *in* D. Otte and J. A. Endler, eds. Speciation and its consequences. Sinauer Associates, Sunderland, Massachusetts.

Darwin, C. 1859. On the origin of species by mean of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499–510.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498.

De Wit, P., M. H. Pespeni, and S. R. Palumbi. 2015. SNP genotyping and population genomics from expressed sequences – current advances and future possibilities. Mol. Ecol. 24:2310–2323.

Dufresne, F., and N. Jeffery. 2011. A guided tour of large genome size in animals: what we know and where we are heading. Chromosome Res. 19:925–938.

Dynesius, M., and R. Jansson. 2014. Persistence of within-species lineages: a neglected control of speciation rates. Evolution 68:923–934.

Faria, R., S. Renaut, J. Galindo, C. Pinho, J. Melo-Ferreira, M. Melo, F. Jones, W. Salzburger, D. Schluter, and R. Butlin. 2014. Advances in Ecological Speciation: an integrative approach. Mol. Ecol. 23:513–521.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. Evolution 35:124–138.

Fijarczyk, A., K. Nadachowska, S. Hofman, S. N. Litvinchuk, W. Babik, M. Stuglik, G. Gollmann, L. Choleva, D. Cogalniceanu, T. Vukov, G. Dzukić, and J. M. Szymura. 2011. Nuclear and

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mitochondrial phylogeography of the European fire-bellied toads *Bombina bombina* and *Bombina variegata* supports their independent histories. Mol. Ecol. 20:3381–3398.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford University Press, Oxford.

Futuyma, D. J. 1987. On the role of species in anagensis. Am. Nat. 130:465–473.

Gabaldón, T., and E. V. Koonin. 2013. Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14:360–366.

Gavrilets, S. 2004. Fitness Landscapes and the Origin of Species. Princeton University Press, Princeton, New Jersey.

Gayral, P., J. Melo-Ferreira, S. Glémin, N. Bierne, M. Carneiro, B. Nabholz, J. M. Lourenco, P. C. Alves, M. Ballenghien, N. Faivre, K. Belkhir, V. Cahais, E. Loire, A. Bernard, and N. Galtier. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. PLoS Genet 9:e1003457.

Gollmann, G. 1987. *Bombina bombina* and *B. variegata* in the Mátra mountains (Hungary): new data on distribution and hybridization (Amhibia, Anura, Discoglossidae). Amphib.-Reptil. 8:213–224.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Grant, P. R., and B. R. Grant. 2014. Speciation undone. Nature 507:178–179.

Gregory, T. R. 2015. Animal Genome Size Database.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. J. Hered. 105:795–809.

Harris, S. E., R. J. O'Neill, and J. Munshi-South. 2015. Transcriptome resources for the white-footed mouse (*Peromyscus leucopus*): new genomic tools for investigating ecologically divergent urban and rural populations. Mol. Ecol. Resour. 15:382–394.

Hatfield, T., and D. Schluter. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution 53:866–873.

Heger, A., and C. P. Ponting. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Res. 17:1837–1849.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Heled, J., and A. J. Drummond. 2015. Calibrated birth–death phylogenetic time-tree priors for Bayesian inference. Syst. Biol. 64:369–383.

Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, and D. S. Rokhsar. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. Science 328:633–636.

Hendry, A. P., P. Nosil, and L. H. Rieseberg. 2007. The speed of ecological speciation. Funct. Ecol. 21:455–464.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760.

Hofman, S., and J. M. Szymura. 2007. Limited mitochondrial DNA introgression in a *Bombina* hybrid zone. Biol. J. Linn. Soc. 91:295–306.

Huang, X., and A. Madan. 1999. CAP3: a DNA sequence assembly program. Genome Res. 12:868–877.

Huber, S. K., L. F. De León, A. P. Hendry, E. Bermingham, and J. Podos. 2007. Reproductive isolation of sympatric morphs in a population of Darwin's finches. Proc. R. Soc. Lond. B Biol. Sci. 274:1709–1714.

Irisarri, I., D. San Mauro, F. Abascal, A. Ohler, M. Vences, and R. Zardoya. 2012a. Data from: The origin of modern frogs (Neobatrachia) was accompanied by the acceleration in mitochondrial and nuclear substitution rates.

Irisarri, I., D. San Mauro, F. Abascal, A. Ohler, M. Vences, and R. Zardoya. 2012b. The origin of modern frogs (Neobatrachia) was accompanied by acceleration in mitochondrial and nuclear substitution rates. BMC Genomics 13:626.

Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302–305.

Kent, W. J. 2002. BLAT—The BLAST-Like Alignment Tool. Genome Res. 12:656–664.

Kristensen, D. M., Y. I. Wolf, A. R. Mushegian, and E. V. Koonin. 2011. Computational methods for gene orthology inference. Brief. Bioinform. 12:379–391.

Kruuk, L. E. B., S. J. E. Baird, K. S. Gale, and N. H. Barton. 1999a. A comparison of multilocus clines by environmental adaptation or by selection against hybrids. Genetics 153:1959–1971.

Kruuk, L. E. B., J. S. Gilchrist, and N. H. Barton. 1999b. Hybrid dysfunction in fire-bellied toads (*Bombina*). Evolution 53:1611–1616.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kumar, S., and M. L. Blaxter. 2010. Comparing *de novo* assemblers for 454 transcriptome data. BMC Genomics 11:571.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. 13:2178–2189.

Linn, C., J. L. Feder, S. Nojima, H. R. Dambroski, S. H. Berlocher, and W. Roelofs. 2003. Fruit odor discrimination and sympatric host race formation in Rhagoletis. Proc. Natl. Acad. Sci. USA 100:11490–11493.

Lohse, K., M. Chmelik, S. H. Martin, and N. H. Barton. 2015. Strategies for calculating blockwise likelihoods under the coalescent.

Lohse, K., R. J. Harrison, and N. H. Barton. 2011. A general method for calculating likelihoods under the coalescent process. Genetics 189:977–987.

MacCallum, C. J., B. Nürnberger, N. H. Barton, and J. M. Szymura. 1998. Habitat preference in the *Bombina* hybrid zone in Croatia. Evolution 52:227–239.

Mallet, J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:2971–2986.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:pp. 10–12.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Morescalchi, A. 1965. Osservazioni sulla cariologia di *Bombina*. Boll.Zool. 32:207–219.

Nagylaki, T. 1976. Clines with variable migration. Genetics 83:867–886.

Nosil, P. 2012. Ecological Speciation. Oxford University Press, Oxford.

Nosil, P., and L. Harmon. 2009. Niche dimensionality and ecological speciation. Pp. 127–154 *in* R. K. Butlin, J. R. Bridle, and D. Schluter, eds. Speciation and patterns of diversity. Cambridge University Press, Cambridge.

Nosil, P., L. J. Harmon, and O. Seehausen. 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nürnberger, B., S. Hofman, B. Förg-Brey, G. Praetzel, A. Maclean, J. M. Szymura, C. M. Abbott, and N. H. Barton. 2003. A linkage map for the hybridising toads *Bombina bombina* and *B. variegata* (Anura: Discoglossidae). Heredity 91:136–142.

Olmo, E., A. Morescalchi, V. Stingo, and G. Odierna. 1982. Genome characteristics and the systematics of Discoglossidae (Amphibia, Salientia). Monit. ZoolItalNS 16:283–299.

Orr, H. A. 1997. Haldane's Rule. Annu. Rev. Ecol. Syst. 28:195–218.

Ortíz-Barrientos, D., J. Reiland, J. Hey, and M. A. F. Noor. 2002. Recombination and the divergence of hybridizing species. Genetica 116:167–178.

Osborne, O. G., T. E. Batstone, S. J. Hiscock, and D. A. Filatov. 2013. Rapid speciation with gene flow following the formation of Mt. Etna. Genome Biol. Evol. 5:1704–1715.

Pabijan, M., A. Wandycz, S. Hofman, K. Węcek, M. Piwczyński, and J. M. Szymura. 2013. Complete mitochondrial genomes resolve phylogenetic relationships within *Bombina* (Anura: Bombinatoridae). Mol. Phylogenet. Evol. 69:63–74.

Parra, G., K. Bradnam, and I. Korf. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067.

Price, T. D., and M. M. Bouvier. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution 56:2083–2089.

Rafińska, A. 1991. Reproductive biology of the fire-bellied toads, *Bombina bombina*and *B. variegata* (Anura: Discoglossidae): egg size, clutch size and larval period length differences. Biol. J. Linn. Soc. 43:197–210.

Rage, J.-C., and Z. Roček. 2003. Evolution of anuran assemblages in the Tertiary and Quaternary of Europe, in the context of palaeoclimate and palaeogeography. Amphib.-Reptil. 24:133–167.

Ramsey, J., H. D. Bradshaw, and D. W. Schemske. 2003. Components of reproductive isolation between the Monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). Evolution 57:1520–1534.

Rasmussen, M. D., and M. Kellis. 2012. Unified modeling of gene duplication, loss, and coalescence using a locus tree. Genome Res. 22:755–765.

Relyea, R. A. 2001. The relationship between predation risk and antipredator responses in larval anurans. Ecology 82:541–554.

Rice, W. R., and E. E. Hostert. 1993. Laboratory experiments on speciation: what have we learned in 40 years? Evolution 47:1637–1653.

Roelants, K., A. Haas, and F. Bossuyt. 2011. Anuran radiations and the evolution of tadpole morphospace. Proc. Natl. Acad. Sci. 108:8731–8736.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatic sticklebacks. Science 287:306–308.

Sambatti, J. B. M., J. L. Anderson, D. Ortiz-Barrientos, E. J. Baack, and L. H. Rieseberg. 2012. Reconciling extremely strong barriers with high levels of gene exchange in annual sunflowers. Evolution 66:1459–1473.

Sasa, M. M., P. T. Chippindale, and N. A. Johnson. 1998. Patterns of postzygotic isolation in frogs. Evolution 52:1811–1820.

Schemske, D. W. 2010. Adaptation and *The Origin of Species*. Am. Nat. 176:S4–S25.

Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864.

Seehausen, O., G. Takimoto, D. Roy, and J. Jokela. 2008. Speciation reversal and biodiversity dynamics with hybridization in changing environments. Mol. Ecol. 17:30–44.

Servedio, M. R., G. S. V. Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: “magic” but not rare? Trends Ecol. Evol. 26:389–397.

Shafer, A. B. A., L. M. Gattepaille, R. E. A. Stewart, and J. B. W. Wolf. 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: in silico evaluation of power, biases and proof of concept in Atlantic walrus. Mol. Ecol. 24:328–345.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733–756.

Sun, Y.-B., Z.-J. Xiong, X.-Y. Xiang, S.-P. Liu, W.-W. Zhou, X.-L. Tu, L. Zhong, L. Wang, D.-D. Wu, B.-L. Zhang, C.-L. Zhu, M.-M. Yang, H.-M. Chen, F. Li, L. Zhou, S.-H. Feng, C. Huang, G.-J. Zhang, D. Irwin, D. M. Hillis, R. W. Murphy, H.-M. Yang, J. Che, J. Wang, and Y.-P. Zhang. 2015. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. Proc. Natl. Acad. Sci. 112:E1257–E1262.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.

Szymura, J. M. 1993. Analysis of hybrid zones with *Bombina*. Pp. 261–289 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford University Press, New York.

Szymura, J. M. 1988. Regional differentiation between hybridizing species *Bombina bombina* (L.) and *B. variegata* (L.) (Salientia, Discoglossidae) in Europe. Uniwersytet Jagielloński, Kraków.

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B variegata*, near Cracow in Southern Poland. Evolution 40:1141–1159.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Szymura, J. M., and N. H. Barton. 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparison between transects and between loci. Evolution 45:237–261.

Szymura, J. M., C. Spolsky, and T. Uzzell. 1985. Concordant change in mitochondrial and nuclear genes in a hybrid zone between two frog species (genus *Bombina*). Experientia 41:1469–1470.

Taylor, E. B., J. W. Boughman, M. Groenenboom, M. Sniatynski, D. Schluter, and J. L. Gow. 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. Mol. Ecol. 15:343–355.

Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21:3907–3930.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Pp. 11.10.1–11.10.13 *in* Current Protocols in Bioinformatics. John Wiley & Sons, Inc.

Via, S., A. C. Bouck, and S. Skillman. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. ii. selection against migrants and hybrids in the parental environments. Evolution 54:1626–1637.

Vines, T. H., S. C. Köhler, M. Thiel, I. V. Ghira, T. R. Sands, C. J. MacCallum, N. H. Barton, and B. Nürnberger. 2003. The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. Evolution 57:1876–1888.

Vonlanthen, P., D. Bittner, A. G. Hudson, K. A. Young, R. Müller, B. Lundsgaard-Hansen, D. Roy, S. Di Piazza, C. R. Largiader, and O. Seehausen. 2012. Eutrophication causes speciation reversal in whitefish adaptive radiations. Nature 482:357–362.

Vorndran, I. C., E. Reichwaldt, and B. Nürnberger. 2002. Does differential susceptibility to predation in tadpoles stabilize the *Bombina* hybrid zone? Ecology 83:1648–1659.

Vörös, J., M. Alcobendas, I. Martínez-Solano, and M. García-Paris. 2006. Evolution of *Bombina bombina* and *Bombina variegata* (Anura: Discoglossidae) in the Carpathian Basin: a history of repeated mt-DNA introgression across species. Mol. Phylogenet. Evol. 38:705–718.

Wall, J. D. 2003. Estimating ancestral population sizes and divergence times. Genetics 163:395–404.

Werner, E. E., and B. R. Anholt. 1993. Ecological consequences of the trade-off between growth and mortality rates mediated by foraging activity. Am. Nat. 142:242–272.

Yanchukov, A., S. Hofman, J. M. Szymura, S. Mezhzherin, S. Y. Morozov-Leonov, N. H. Barton, and B. Nürnberger. 2006. Hybridization of *Bombina bombina* and *B. variegata* (Anura, Discoglossidae) at a sharp ecotone in Western Ukraine: comparisons across transects and over time. Evolution 60:583–600.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Yeaman, S., and M. C. Whitlock. 2011. The architecture of adaptation under migration-selection balance. Evolution 65:1897–1911.

Zheng, Y., J. Fu, and S. Li. 2009. Toward understanding the distribution of Laurasian frogs: a test of Savage's biogeographical hypothesis using the genus *Bombina*. Mol. Phylogenet. Evol. 52:70–83.

Table 1. Data summary. Listed are the read counts before and after quality control (QC) and assembly statistics for the Illumina and Roche 454 datasets (assembly software in parentheses). Abbr. = taxon abbreviation. Total assembled bases (Trinity) are computed from the longest isoform per component. Only the contigs with read support are listed.

		Illumina (Trinity)						Roche 454 (Newbler/Mira/CAP3)			
Taxon	Abbr.	raw sequence pairs (x 10^6)	reads pairs after QC (x 10^6)	single reads after QC (x 10^6)	components ("genes")	contigs	total assembled bases (x 10^6)	raw reads	reads after QC	contigs	total assembled bases (x 10^6)
B. v. variegata	*Bvv*	55.3	42.3	7.7	67,300	80,878	47.4	636,551	611,417	22,009	15.9
B. v. scabra	*Bvs*	34.4	25.9	5.2	53,909	62,920	36.9				
B. bombina	*Bb*	47.1	35.9	6.9	61,724	73,131	44.5	691,857	667,125	28,827	19.8
B. orientalis	*Bo*	46.9	35.3	6.9	52,484	62,261	35.7				

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Average per site frequencies of the four mutation types for each pairwise comparison (A versus B). Taxon A is *B. v. variegata*. Observed data are listed separately for the site frequency spectrum (SFS) and the blockwise site frequency spectrum (bSFS). For the latter, the expected values from the best fitting model (Table 3, bold) are also given. The number of orthoblocks differs slightly between analyses, because all blocks that violated the four gamete test were excluded from the blockwise analysis.

Taxon B	Method	Source	k_A	k_B	k_{AB}	k_{AABB}	# ortho-blocks
B. v. scabra	SFS	obs	0.00178	0.00275	0.0000444	0.0136	1352
	bSFS	obs	0.00174	0.00274	0.0000198	0.0136	1347
		exp	0.00168	0.00266	0.0000188	0.0134	
B. bombina	SFS	obs	0.00181	0.00106	0.0000180	0.0187	1481
	bSFS	obs	0.00177	0.00106	0.0000090	0.0187	1479
		exp	0.00155	0.00119	0.0000039	0.0185	
B. orientalis	SFS	obs	0.00172	0.00748	0.0000616	0.0308	1190
	bSFS	obs	0.00162	0.00744	0	0.0308	1179
		exp	0.00154	0.00752	0	0.0312	

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Support (ΔL) relative to the best model (*). Taxon comparisons are listed as in Table 2. SFS = site frequency spectrum, bSFS = blockwise analysis. Div = strict divergence; fixed N_e, IM = divergence with migration, fixed N_e; Div_2 = strict divergence, asymmetric N_e; IM_2 = divergence with migration, asymmetric N_e. Migration is unidirectional (A→B or B→A). See text for details.

* The IM_2 models collapsed to a strict divergence model (Div_2)

Taxon B	Method	Div	$IM_{A\rightarrow B}$	$IM_{B\rightarrow A}$	Div_2	$IM_{2,A\rightarrow B}$	$IM_{2,B\rightarrow A}$
B. v. scabra	SFS	-30.2	**0***	-13.2	-22.4		
	bSFS	-39.3	-26.1	-31.1	-39.3	-25.6	**0***
B. bombina	SFS	-14.4	-4.24	**0***	-11.7		
	bSFS	-58.0	-26.3	-0.3	-43.6	-3.0	**0***
B. orientalis	SFS	-64.7	**0***	-55.8	-14.7		
	bSFS	-223	-223	-223	**0**	Div_2^*	Div_2^*

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4. Maximum likelihood estimates of parameters under the best supported model (Table 3). The divergence time (*T*) is given in units of 2 N_{anc} generations (95% CI in brackets). *a* is a scalar for the effective population size of one of the descendant populations (in parentheses, cf. Figure 3). n/a = not applicable.

Taxon B	Method	θ	*M*	*T*	*a*
B. v. scabra	SFS	0.00179	0.078 (0.055, 0.099)	12.5 (9.2, 15.7)	n/a
	bSFS	0.00129	0.032	12.0	2.07 (B)
B. bombina	SFS	0.00107	0.027 (0.013, 0.041)	22.4 (16.6, 28.1)	n/a
	bSFS	0.00121	0.020	16.8	0.99 (B)
B. orientalis	SFS	0.00277	0.136 (0.120, 0.150)	∞	n/a
	bSFS	0.00752	0	3.75	0.2 (A)

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1. Distribution of *B. bombina* and *B. variegata* in Central and Eastern Europe. Included are the sampling locations (triangles) and the approximate locations of intensively studied hybrid zones. Dashed lines represent national boundaries.

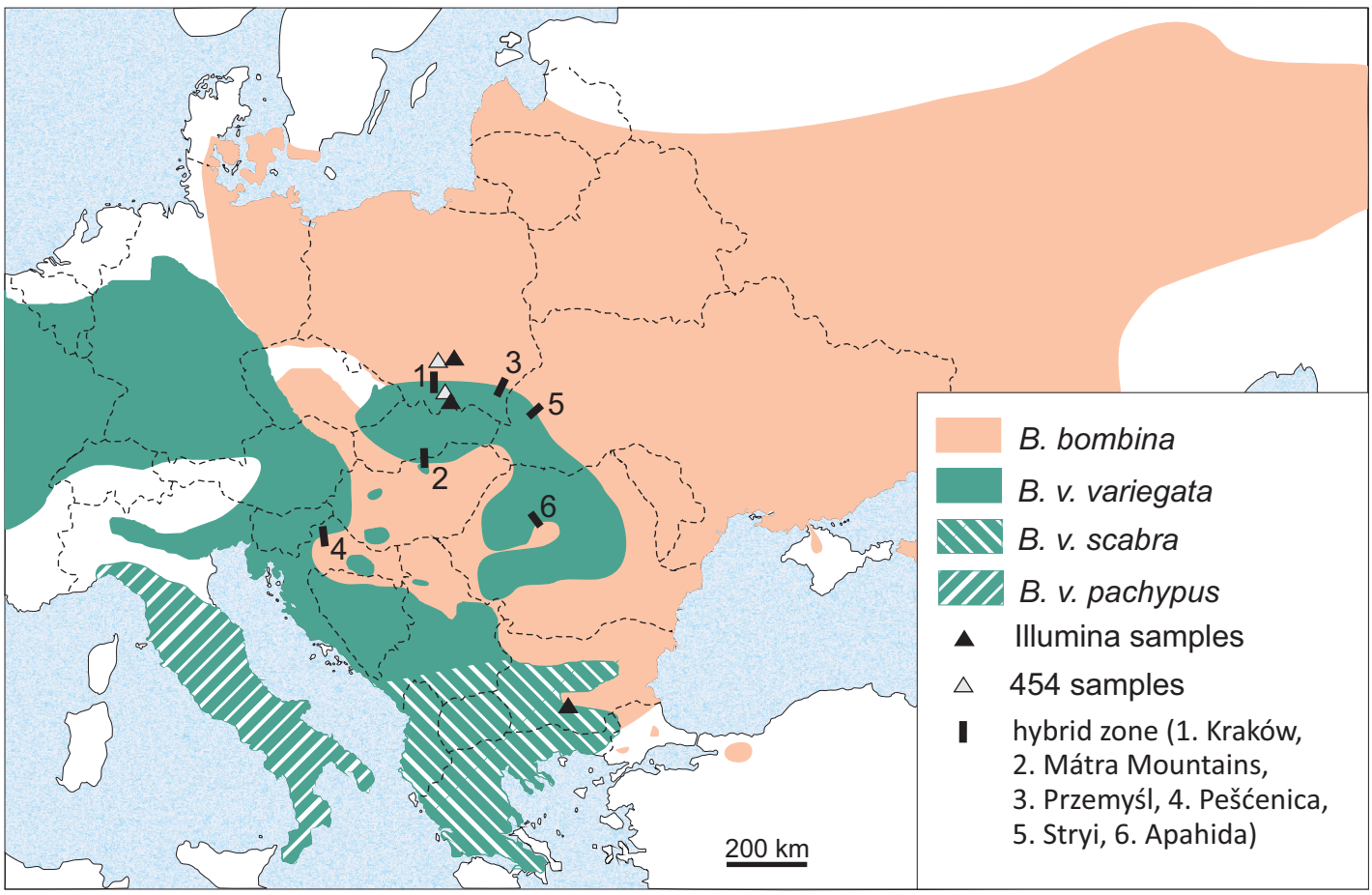
Figure 2. OrthoMCL clustering and filtering of raw groups. a. workflow. (taxon abbreviations as in Table 1, *St* = *Silurana tropicalis*) and b. examples of OrthoMCL clusters. Two clusters (indicated by an asterisk) pass the filtering criteria 1-3 (see text). Paralogue pairs (P) are linked by heavy lines.

Figure 3. Sketches of the four coalescence models. The block widths indicate relative effective population sizes in the ancestral lineage and its two descendants A and B. Arrows mark the direction of gene flow.

Figure 4. Filtering of OrthoMCL clusters. a. Counts of excluded clusters according to criteria 1 - 3 (see text for details). b. Examples of alternative exon use generating two orthogroups (rows of boxes), each with a member of a paralogue pair from each taxon. Shown are positions 41 - 76 (total: 158) of the aligned peptide sequence with differences between orthogroups (top *versus* bottom variant). A alternative set of reference transcriptomes comprising only the longest contig per component results in one cluster which mixes isoforms (heavy bars under the boxes).

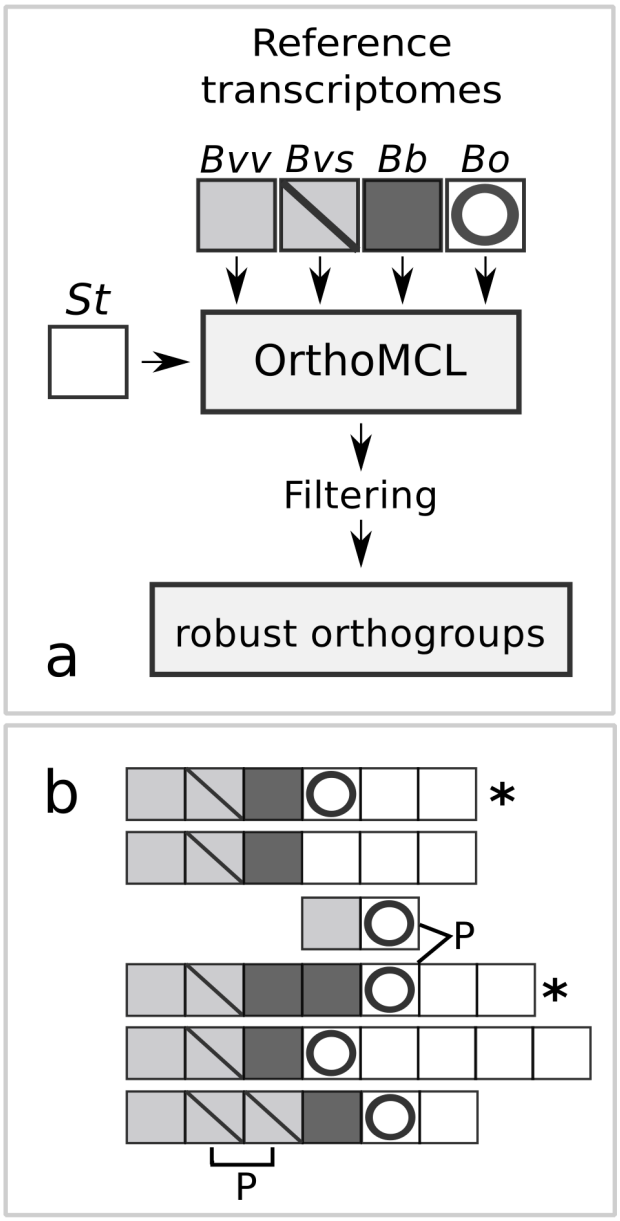
Figure 5. The distribution of fixed differences between *B. v. variegata* and *B. bombina* (k_{AABB}) in blocks of 150 fourfold degenerate sites. The divergence model predicts a narrower distribution (gray dotted) than observed (black), *i.e.* there is an excess both of blocks with no fixed differences and of blocks with a large number of fixed differences ($k_{AABB} > 6$). An IM_2 model with migration and asymmetry in N_e (gray, dashed) gives a good fit to the data.

bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



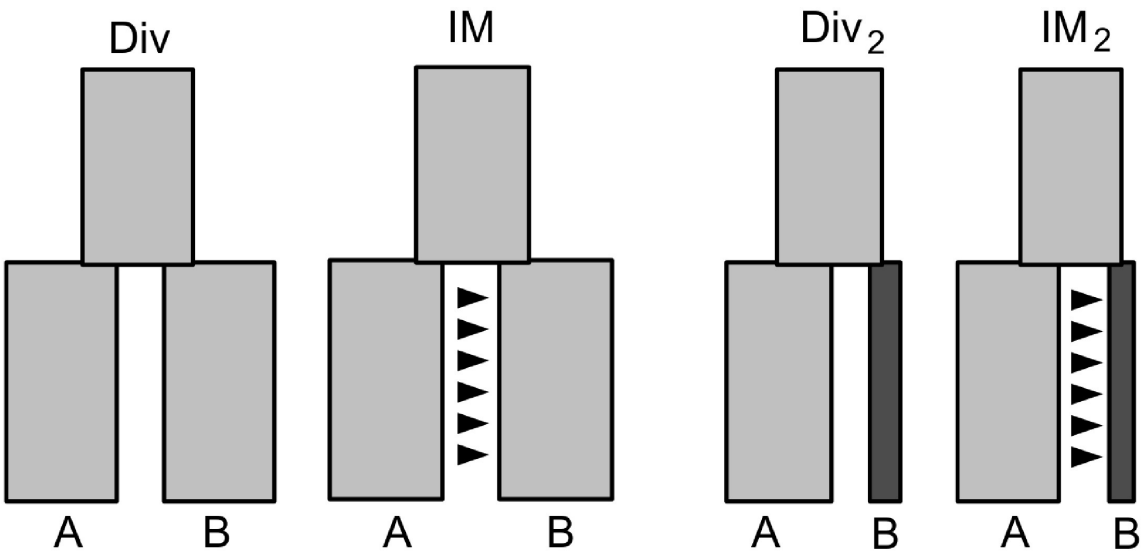
bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2



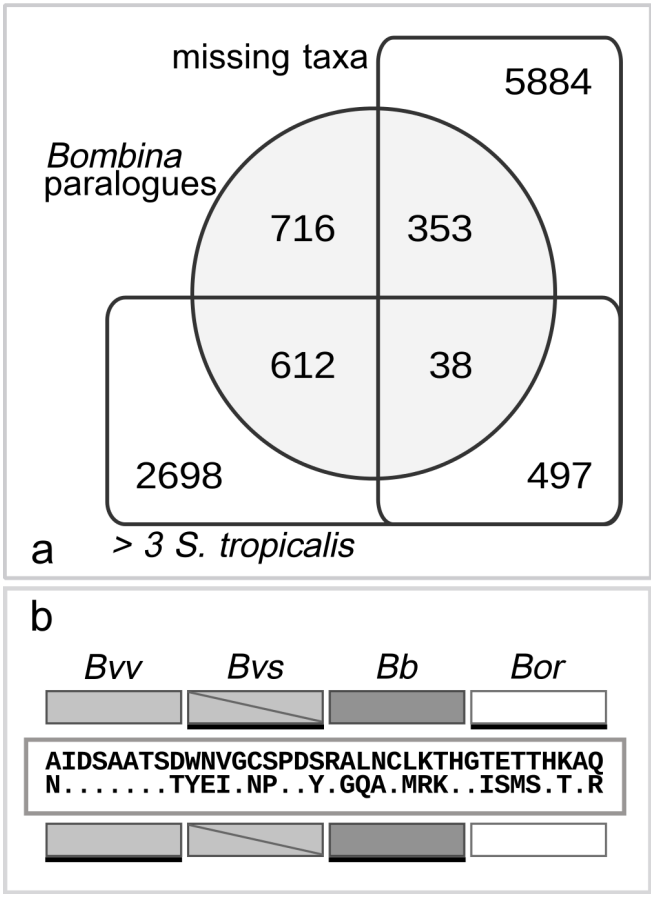
bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/030056; this version posted October 28, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

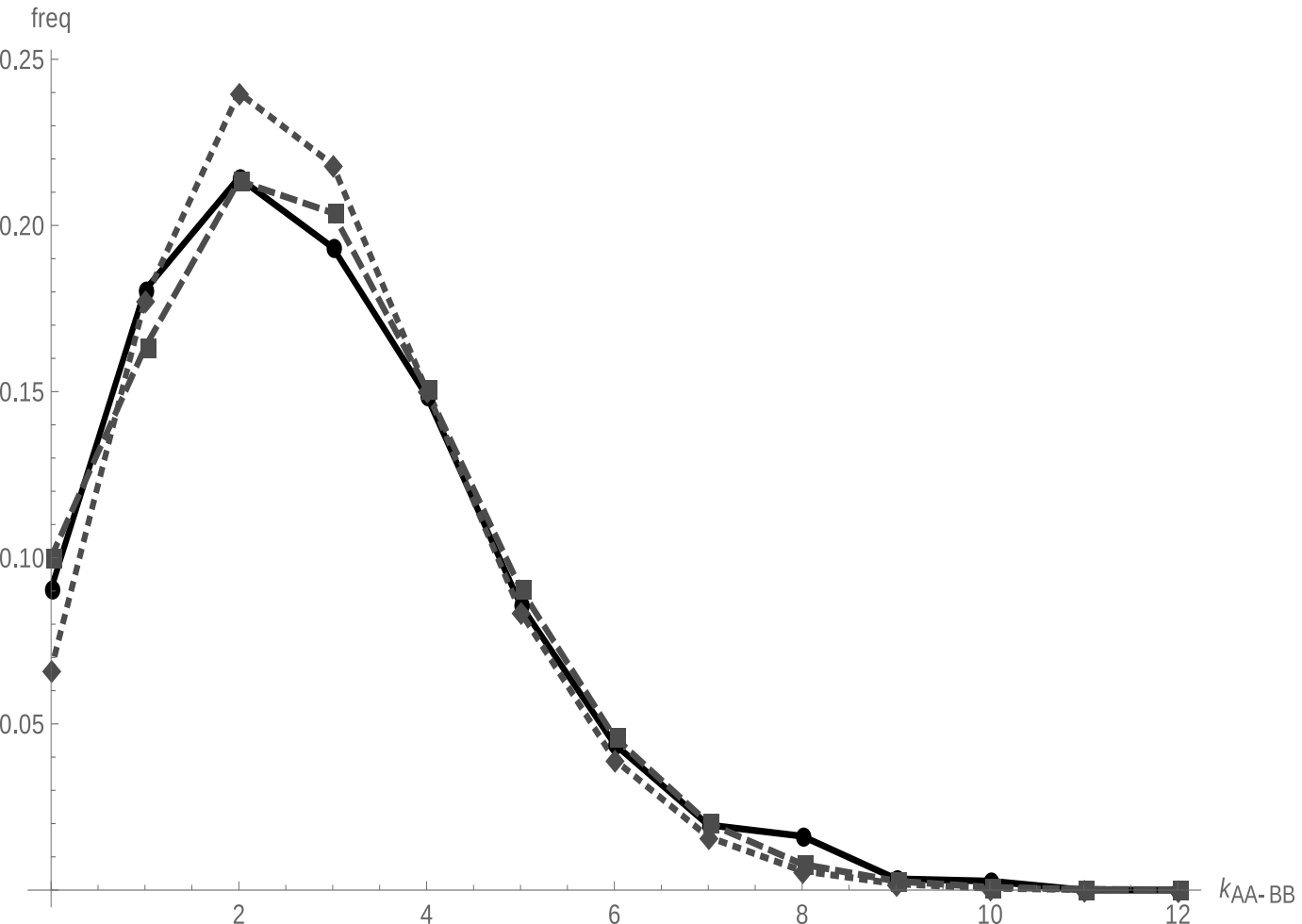



Figure 5