NewtCap: an efficient target capture approach to boost genomic studies in Salamandridae (True Salamanders and Newts)

Manon Chantal de Visser[1,2*], James France[1,2], Evan McCartney-Melstad[3,4,5], Gary M. Bucciarelli[4,5,6], Anagnostis Theodoropoulos[1,2], Howard Bradley Shaffer[4,5], Ben Wielstra[1,2]

1) Institute of Biology Leiden, Faculty of Science, Leiden University, Leiden, The Netherlands

2) Understanding Evolution, Naturalis Biodiversity Center, Leiden, The Netherlands

3) Nutcracker Therapeutics, Emeryville, California, USA

4) Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, California

5) La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, UCLA, Los Angeles, California

6) Department of Wildlife, Fish, and Conservation Biology, University of California, Davis, California, USA

* Corresponding author

Abstract

Salamanders have large and complex genomes, hampering whole genome sequencing, but reduced representation sequencing provides a feasible alternative. We present NewtCap: a sequence capture bait set that targets c.7k coding regions across the genomes of all True Salamanders and Newts (the family Salamandridae, also known as 'salamandrids'). We test the efficacy of NewtCap, originally designed for the Eurasian *Triturus* newts, in 30 species, belonging to 17 different genera, that cover all main Salamandridae lineages. We also test NewtCap in two other salamander families. We discover that NewtCap performs well across all Salamandridae lineages (but not in the salamander families Ambystomatidae and Hynobiidae). As expected, the amount of genetic divergence from the genus *Triturus* correlates negatively to capture efficacy and mapping success. However, this does not impede our downstream analyses. We showcase the potential of NewtCap in the contexts

of; 1) phylogenomics, by reconstructing the phylogeny of Salamandridae, 2) phylogeography, by sequencing the four closely related species comprising the genus *Taricha*, 3) hybrid zone analysis, by genotyping two *Lissotriton* species and different classes of interspecific hybrids, and 4) conservation genetics, by comparing *Triturus ivanbureschi* samples from several wild populations and one captive-bred population. Overall, NewtCap has the potential to boost straightforward, reproducible, and affordable genomic studies, tackling both fundamental and applied research questions across salamandrids.

Keywords

Hyb-seq, exon capture, Urodela, Caudata, High Throughput Sequencing, target enrichment

Introduction

One of the most challenging groups of animals to study genomically are the salamanders. These organisms have complex and large genomes that contain many repetitive elements compared to most other animals (e.g. they can be in the range of 10 to 40 times the size of a human genome; Gregory, 2002, Litvinchuk et al., 2007, Sessions, 2008, Sun et al., 2012, Gregory, 2024, Smith et al., 2019). This is the primary reason that conducting whole genome sequencing and *de novo* genome assembly for salamanders is extremely costly in terms of money, time, and computational resources (Lou et al., 2021, Calboli et al., 2011).

Fortunately, reduced representation sequencing strategies are paving the way toward more straightforward, reproducible, and affordable genomic studies, especially in organisms with large and complex genomes (Good, 2011, Zaharias et al., 2020). By focusing sequencing efforts on subsets of the genome rather than the entire genome, valuable time and resources can be conserved, allowing for a greater number of samples to be processed (Andermann et al., 2019). Different types of genome-subsampling techniques exist, broadly categorized as 'non-targeted' versus 'targeted', and their suitability varies depending on the study species and research objectives (da Fonseca et al., 2016, Lemmon and Lemmon, 2013, Jones and Good, 2016).

Non-targeted approaches, such as Restriction site-Associated DNA sequencing (RAD-seq) and related techniques, are widely used in genetic mapping and population studies, including salamander studies (e.g. Hu et al., 2019, Rancilhac et al., 2019, Hubbs et al., 2022, Andrews et al., 2016, France et al., 2024, Babik et al., 2024, Rodríguez et al., 2017, Burgon et al., 2021, Scott et al., 2024). While simple, scalable (Andermann et al., 2019), and having the key advantage of not

needing to know the sequences of any loci beforehand, such non-targeted approaches are known to yield missing data, underestimate genetic diversity, and call incorrect allele frequencies (Arnold et al., 2013, Rubin et al., 2012, Davey et al., 2013). This is due to restriction site polymorphisms that limit the phylogenetic signal for resolving deep, evolutionary relationships (Dodsworth et al., 2019, Rubin et al., 2012, Lowry et al., 2017).

On the other hand, targeted methods such as target capture sequencing offer a strategy to achieve higher resolution and specificity, as they focus on pre-selected (orthologous) loci (Mamanova et al., 2010, Harvey et al., 2016). With target capture sequencing, biotinylated RNA probes – or 'baits' – that are complementary to the loci of interest are used, causing them to bind to the 'target' regions, before streptavidin-coated magnetic beads are used to 'capture' them (Gnirke et al., 2009, Albert et al., 2007). The main advantage of using the target capture method over untargeted methods is that there will be higher efficiency because enrichment of specific (usually coding) genomic regions of interest is more effectively achieved across samples (Grover et al., 2012, Andermann et al., 2019, Heyduk et al., 2016, Fitzpatrick et al., 2024). Furthermore, the flanking regions of targets can also provide information on more variable genomic regions such as introns (Jones and Good, 2016, Zhou and Holliday, 2012), and off-target 'bycatch' reads can provide additional data as well (Featherstone and McGaughran, 2024, Guo et al., 2012).

Often custom target capture baits are designed, which can be based on a draft genome or transcriptome reference, however pre-designed baits can be ordered as well (Bi et al., 2012, Andermann et al., 2019, Jimenez-Mena et al., 2022). Over the last decade, many target capture bait sets have been tested and made publicly available for different types of organisms, ranging from micro-organisms to macro-organisms (e.g.; Andermann et al., 2019, Grover et al., 2012, Heyduk et al., 2016, Khan et al., 2024, Quek and Ng, 2024, Yu et al., 2023). For animals in particular, bait sets exist for certain groups of insects (e.g. Hymenoptera; Faircloth et al., 2015), snails (e.g. Eupulmonata; Teasdale et al., 2016), reptiles (e.g. Squamata; Schott et al., 2017, Singhal et al., 2017), fish (e.g. Acanthomorpha; Alfaro et al., 2018), and amphibians (e.g. Anura; Hutter et al., 2022). Salamander bait sets have also already been designed for particular genera (*Ambystoma* and *Triturus*; McCartney-Melstad et al., 2016, Wielstra et al., 2019). As the target capture approach generally handles a certain degree of sequence divergence well, a bait set designed for one genus has great potential to work in other genera too (Bragg et al., 2016, Portik et al., 2016). Thus, it is worth exploring the potential transferability of such existing bait sets to related taxa.

We introduce 'NewtCap': a target capture bait set and protocol for salamanders of the family Salamandridae (which includes the 'True Salamanders' and the 'Newts'). The bait set, originally designed for *Triturus* newts (Wielstra et al., 2019), targets c. 7k putative, orthologous, coding

regions. Here, we provide an updated version of the lab protocol that cuts down on costs, time and DNA input. Furthermore, we investigate the efficacy of NewtCap across 30 different species and 17 genera of the Salamandridae family and test its efficiency in two distantly related species of other salamander families. Besides checking general performance, we assess the usefulness of NewtCap in the contexts of; 1) phylogenomics, by inferring a Salamandridae phylogeny, 2) phylogeography, by investigating the relationships among multiple individuals of four closely related *Taricha* species, 3) hybrid zone analysis, by calculating the hybrid index and heterozygosity for two *Lissotriton* species and interspecific hybrids of different cross types, and 4) conservation genetics, by determining the genetic relatedness of a captive-bred *Triturus ivanbureschi* population to wild populations. Overall, we demonstrate that NewtCap is an important tool for molecular studies on salamandrids.

Materials and Methods

Sampling and DNA Extraction

We studied 73 individual salamanders in total, covering 30 species and 17 genera of the Salamandridae family (including nine interspecific hybrids of the genus *Lissotriton*), as well as two more distantly related samples (Table S1, see Zenodo) from the families Ambystomatidae and Hynobiidae (Marjanović and Laurin, 2013, Frost, 1985). We obtained DNA extractions or genetic data from previous studies (Wielstra et al., 2019, De Visser et al., in prep, Kazilas et al., 2024, Mars et al., 2024, Kalaentzis et al., 2024, Robbemont et al., 2023, Pasmans et al., 2006, Valbuena-Urena et al., 2013, Sequeira et al., 2022, France et al., in prep), as well as new samples for this study (provided by collaborators, see acknowledgments and Table S1 via Zenodo).

We used the Promega Wizard™ Genomic DNA Purification Kit (Promega, Madison, WI, USA), which is a salt-based DNA extraction protocol (Sambrook and Russell, 2001). The DNA from each sample was re-suspended in 100 µL 1 × TE buffer before concentrations and purity were assessed via spectrophotometry, using the DropSense96™ (Trinean, Gentbrugge, Belgium). As we used a minimum concentration of 150 ng/µL for library preparation, any samples found to be below this were concentrated with vacuum centrifugation.

NewtCap: Laboratory procedures

The target capture procedures are fully documented and included as Supplementary Materials (see under 'Data Accessibility'). These protocols are rigorously optimized versions of those previously

described (see Wielstra et al., 2019). Sonication of genomic DNA has been replaced with enzymatic fragmentation, resulting in a tenfold increase in library yield per ng of input DNA, in addition to time and cost savings. The volumes of all reagents in the library preparation were reduced by 75% compared to the manufacturers protocol to conserve reagents. Target capture protocols have been adapted from Mybaits V3.0 to the V4.0 kit and are fully compatible with Mybaits V5.0.

Library preparation

DNA libraries were constructed using the NEBNext Ultra™ II FS DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) following the manufacturers protocol, with quarter volumes of all reagents and with the following modifications: 1,000 ng of extracted genomic DNA was used as input (6.5 µL at 154 ng/µL). The enzymatic shearing time at 37 °C was adjusted to 6.5 minutes (as the minimum time of 15 minutes suggested by the manufacturer resulted in over-digestion). After NEB adapter ligation and cleavage with the NEB USER enzyme, NucleoMag™ magnetic separation beads (Macherey-Nagel, Düren, Germany) were used for double-ended size selection targeting an insert size of 300 bp. Libraries were indexed with 8 cycles of PCR amplification, using unique combinations of custom i5 and i7 index primers (Integrated DNA Technologies, Leuven, Belgium). NucleoMag™ beads were used again for a final cleanup before the libraries were resuspended in 22 µL of 0.1 × TE buffer. Library size distribution and concentration was measured using the Agilent 4150 TapeStation or 5200 Fragment analyzer system (Agilent Technologies, Santa Clara, CA, USA), using the D5000 ScreenTape or DNF-910 dsDNA Reagent Kit. We aimed for obtaining a final library concentration of at least 12 ng/µL.

Target Capture, Enrichment, and Sequencing

Libraries were equimolarly pooled in batches of 16, aiming for a total DNA mass of 4,000 ng (250 ng per sample). Vacuum centrifugation was then used to reduce the volume of each pool to 7.2 µL (556 ng/µL). We performed target capture with the MyBaits v4.0 kit (Arbor Biosciences, Ann Arbor, MI, USA) previously designed for *Triturus* newts (Wielstra et al., 2019), which targets 7,139 unique exonic regions (product Ref: # 170210-32). The manufacturers protocol was employed with the following deviations: Blocks C and O were replaced with 5 µL of *Triturus* derived C0t-1 DNA at 6,000 ng/µL (30,000 ng per pool). C0t-1 DNA is enriched in repetitive sequences and acts as a blocking buffer to non-specific targets in capture assays by hybridizing with repetitive sequences in the libraries (McCartney-Melstad et al., 2016). Tissue to produce C0t-1 DNA was derived from an invasive population of *T. carnifex* (Meilink et al., 2015).

The pooled libraries were incubated with the blocking buffer for 30 minutes, followed by hybridization for 30 hours at 62 °C. After capture of the hybridized baits with streptavidin coated beads and four cycles of washing, each pool was divided into equal-volume halves. The first half was subject to 14 cycles of PCR amplification, followed by bead cleanup, and resuspension in 22 µL of 0.1 × TE buffer. The concentration and fragment size distribution of the enriched pool were then measured with the TapeStation system, using the HS D5000 ScreenTape kit. If the final concentration was between 5 and 20 nM then the same protocol was employed for the second half of the pool. If not, the number of post-enrichment PCR cycles was altered to compensate. For each pool 16 GB (1 GB per sample) of 150 bp paired-end sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA) by BaseClear B.V. (Leiden, the Netherlands).

Bioinformatics for data pre-processing

We followed a standard pipeline for cleaning up the raw reads and mapping the trimmed sequence data against reference sequences in a Linux environment (adapted from Wielstra et al., 2019). We describe the main steps and provide the scripts we used publicly through GitHub: https://github.com/Wielstra-Lab/NewtCap_bioinformatics.

Quality Control and Read Clean-up

First, we clipped all reads to a maximum length of 150 bp using the BBDuk script by BBTools (BBMap – Bushnell B. – https://sourceforge.net/projects/bbmap/). Then we removed any leftover adapter contamination and low-quality bases (or reads) using Trimmomatic v.0.38 (Bolger et al., 2014). Adapter sequences for the TruSeq2 multiplexed libraries were identified and removed. Leading and trailing bases were trimmed if the Phred score was <5. We also removed reads in case the average Phred score in a sliding window (5' to 3') of a size of five base pairs dropped below 20, and we discarded all reads shorter than 50bp. We monitored the quality of our sequences before and after trimming with FastQC (Andrews, 2010) by using the summarizing 'quality_check' function of SECAPR (Andermann et al., 2018).

Read Mapping and Variant Calling

Cleaned reads were mapped to the set of 7,139 *T. dobrogicus* reference sequences with a maximum length of 450bp (based on *T. dobrogicus* transcripts, see Wielstra et al., 2019) that were initially

used for probe development. The reference FASTA file is provided as Supplementary Material. Mapping was performed using the MEM algorithm from Burrows-Wheeler Aligner v.0.7.17 (Li, 2013), and we stored results in BAM format using SAMtools v.1.18 (Li et al., 2009, Danecek et al., 2021). We added read group information using the AddOrReplaceReadGroups function of Picard v.2.25.1 and PCR duplicates were flagged with Picard's MarkDuplicates (http://broadinstitute.github.io/picard/). After doing so, the output was again saved in BAM format.

Next, we called variants using the HaplotypeCaller function of GATK v.4.1.3.0 (Auwera and O'Connor, 2020) including the -ERC GVCF option, and we used GATK's CombineGVCFs and GenotypeGVCFs functions to perform joint genotype calling to create multi-sample (ms) gVCF files as input for downstream analyses. In case samples still needed to be added or removed from msgVCF files after joint genotyping, we did so using the view function of BCFtools v.1.15.1 (Danecek et al., 2021, Li, 2011).

Downstream Analyses

Bait performance statistics

To evaluate the overall performance of the NewtCap baits, we calculated several statistics (Table S1). First, we counted the number of GB and reads contained in the raw FASTQ files for each sample. We also used the SAMtools' flagstat function to determine the total number of reads present in the (deduplicated) BAM files that passed quality control, as well as the percentage of these reads that successfully mapped to a reference sequence. We used SAMtools' coverage function to extract information on the mean depth of coverage, as well as the mean percentage of coverage, for each target, and then averaged these per sample to provide an estimate of the "success rate". Besides analyzing the contents of the BAM files, we checked how many sites were outputted in total in each separate gVCF file by counting lines (as each non-header line is one site). Then, we counted how many of those were considered SNPs, and how many were considered INDELs, by using the stats function of BCFtools (note that this is done before merging any files or applying any SNP filtering).

To estimate how the performance of NewtCap correlates with the level of genetic divergence from *T. dobrogicus*, the species that was used for bait design and as a reference for read mapping, we performed several statistical analyses. For these, we used a rough estimate of divergence times for lineages of Salamandridae (from De Visser et al., in prep). We explored the relationship between this estimated genetic divergence from *T. dobrogicus* and the following performance variables; 1) the amount of GB of data obtained, 2) the number of obtained reads, 3) the percentage of mapped reads, 4) the percentage of reads marked as PCR duplicates, 5) the mean read depth number after

deduplication, 6) the mean coverage of the sequence bases after deduplication, 7) the number of SNPs found in the gVCFs files, and 8) the number of INDELs found in the gVCF files. We calculated the correlation coefficients (r) between either of these variables and the estimated genetic divergence from *T. dobrogicus*. Depending on whether our data met the assumptions for parametric or non-parametric testing, we either employed the Pearson correlation method or the Spearman's rank correlation method. We determined the level of significance using a two-tailed test, with the p-value threshold of $p<0.00625$ to indicate statistical significance (implementing a Bonferroni correction on the usual threshold $p<0.05$ for the eight tests performed, as $0.05/8 = 0.00625$). These analyses, including assumption checks such as testing for normality, were performed in Microsoft Excel 2024 (https://office.microsoft.com/excel).

Concatenated phylogenetic analyses in RAxML

To investigate the usefulness of NewtCap in the context of phylogenomics, we built phylogenetic trees with a Maximum Likelihood method using RAxML. We used NewtCap to reconstruct an existing Salamandridae tree that was originally built based on an independent set of 5,455 nuclear genes (derived from transcriptome data, see Rancilhac et al., 2021), by including 23 samples that cover at least one representative genus for each of the main clades *in sensu* (as described in Rancilhac et al., 2021). To further check the performance of NewtCap across True Salamanders as well as for non-salamandrid species, and to explore the position of the root, we added three additional samples in a second, extended analysis, namely the Salamandridae species *Mertensiella caucasica* (family Salamandridae) and two non-salamandrid species *Ambystoma mexicanum* (family Ambystomatidae) and *Paradactylodon gorganensis* (family Hynobiidae). The raw msgVCF files for these two analyses respectively comprised 812,574 sites and 812,603 sites from across 7,135 targets. Note that not all sites in the raw and intermediate (ms)gVCF files were necessarily polymorphic (i.e. invariant - or monomorphic - sites may still have been included, unless we specifically stated that these have been removed).

We applied quality filtering on the msgVCFs (which contained only the samples chosen for these phylogenetic analyses). First, we removed sites that showed heterozygote excess ($p<0.05$) using BCFtools v.1.15.1 , in order to rid paralogous targets (see Wielstra et al., 2019). Then, we used VCFtools v.0.1.16 to enforce the following strict filtering options; discarding INDELs ("--remove-indels"), filtering out sites with for instance poor genotype and mapping quality scores ("QD<2", "MQ<40", "FS>60", "MQRankSum <−12.5", "ReadPosRankSum<−8", and "QUAL<30"; Poplin et al., 2017), and discarding sites with over 50% missing data ("--max-missing

= 0.5”). At this stage, the intermediate msgVCF files contained, in total, 625,182 SNPs from across 7,103 targets, and 621,754 SNPs from across 7,095 targets, for the sets of 23 and 26 samples.

We converted the files into PHYLIP format with the ‘vcf2phylip.py’ script (Ortiz, 2019), which, by default, requires a minimum of four samples representing each SNP variant. Next, we performed an ascertainment bias correction by using the ‘ascbias.py’ script (https://github.com/btmartin721/raxml_ascbias) to remove sites considered invariable by RAxML. The final PHYLIP file was used as input for RAxML v.8.2.12 (Stamatakis, 2014), which we ran with 100 rapid bootstrap replicates, under the ASC_GTRGAMMA model, and with the Lewis ascertainment correction (Lewis, 2001). We obtained the best-scoring Maximum Likelihood tree out of the concatenation analyses.

To gain insights into the potential of NewtCap in a phylogeographical context, we also built a phylogenetic tree for the New World newt genus *Taricha* using RAxML. As input, we used a msgVCF file containing nine samples: eight *Taricha* samples covering four distinct species within the genus (*T. granulosa*, *T. rivularis*, *T. sierrae* and *T. torosa*, with two samples per species) and one sample of the sister-genus *Notophthalmus* as outgroup (Table S1). The raw msgVCF file comprised 209,072 sites from 7,059 targets. Subsequently, we performed the filtering steps as described above, which left the intermediate msgVCF file with 180,385 SNPs from across 7,048 targets. We visualized all phylogenies using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Hybrid analyses and conservation genetics

To assess the usefulness of NewtCap in the context of hybridization studies, we used the R packages ‘*triangulaR’* (Wiens and Colella, 2024), ‘*ggplot2’* (Wickham, 2011) and ‘*vcfR’* (Knaus and Grunwald, 2017) to build triangle plots. We investigated the target capture data of in total 15 *Lissotriton* individuals; three of each parental species (*L. vulgaris* and *L. montandoni*), as well as hybrids bred and reared in the lab: three F1 hybrids, three F2 hybrids, and three backcrosses (F1 x *L. vulgaris*). We assembled a raw msgVCF file containing only these samples, which initially had data on 807,389 sites across in total 7,135 targets. From this we extracted high quality SNPs as described before (except here we tolerated no missing data at all). The filtered msgVCF contained 532,332 SNPs from 6,866 targets in total and was used as input for *triangulaR*.

The *triangulaR* package uses SNPs from an input VCF file that are estimated to be species-diagnostic based on the parent species under a certain allele frequency threshold between ‘0’ and ‘1’, where ‘1’ equals a fixed difference between parental population (Wiens and Colella, 2024). Thus, the threshold of 1 is employed when searching for two consistent, diverged, homozygous

states in the parental populations. SNPs that pass this filter can then be used to calculate the hybrid index and heterozygosity for the samples analyzed. Considering that our parent populations did not comprise the actual parents of the hybrids included in our study and were represented by just three samples each, this complicated distinguishing species-diagnostic SNPs from SNPs that only appeared diagnostic by chance. Therefore, we only extracted SNPs that were 100% species-diagnostic (based on the information from the parental populations, by setting the allele frequency threshold to 1) *and* that were always heterozygous in the F1's. This extra functionality (i.e. filtering for heterozygosity in F1 hybrids) is currently not built into *triangulaR* (see the online tutorial; https://github.com/omys-omics/triangulaR), but we added this filtering option before conducting our calculations. We provide the customized R script in our GitHub repository (https://github.com/Wielstra-Lab/NewtCap_bioinformatics).

To assess the performance of NewtCap in the context of a conservation genetic study we conducted a Principal Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA) to determine the geographical origin of a captive bred population, which is presumed to originate from Cerkezköy in Turkey (Michael Fahrbach, pers. comm.). We used the R packages '*gdsfmt'* and '*SNPRelate'* (Zheng et al., 2012) on data from 24 *Triturus ivanbureschi* newts that originated from seven different wild populations and the captive population (i.e. three samples per population, see Table S1). Populations from the wild include both the glacial refugial area (Asia Minor and Turkish Thrace) and postglacially colonized area (the Balkans; Wielstra et al., 2017). We created a raw msgVCF file from these samples, which contained a total of 812,768 sites from across 7,135 targets. We extracted high quality SNPs in the same way as we did for the hybrid studies, and the filtered msgVCF that was used as input had a total of 486,891 SNPs from 5,774 targets. We conducted the hybrid and conservation genetic analyses in Rstudio (Team_R_Studio, 2021) using R v.4.1.2 (Team_R_Core, 2020).

Finally, we were also interested in the number of informative SNPs that we could find within Salamandridae species that are distantly related to *T. dobrogicus.* The most distantly related species for which we had more than one sample belonging to distinct populations in our dataset is *Chioglossa lusitanica* (Table S1), which is a true salamander, not a newt. Having more than one sample allows us to cross-compare these populations. Thus, we extracted the variant information of the two *C. lusitanica* samples from the overall msgVCF file –filtered for high quality SNPs allowing no missing genotypes– that was used to estimate the genetic divergence from *T. dobrogicus* (De Visser et al., in prep). We quantified the number of SNPs that were homozygous in one sample and heterozygous in the other sample – and *vice versa*. Additionally, we counted the number of SNPs that were homozygous in both samples, but for different alleles.

Results

Overall performance of NewtCap

In total we collected 149.0 GB of raw sequence data with on average 15,573,781 reads per sample (s.d. = 11,400,957). On average 42.6% (s.d. = 16.2%) of all reads were successfully mapped against the *Triturus* reference, and an average of 11.9% (s.d. = 10.6%) of all reads were marked as PCR duplicates. In the deduplicated BAM files, the mean read depth across all targets per sample varied widely, and ranged from 5.9 X (s.d. = 27.9) in *A. mexicanum* and 8.2 X (s.d. = 27.8) in *P. gorganensis* to 122.2 X (s.d. = 345) in one of the *Lissotriton* hybrids. The mean coverage of the sequence bases of all targets ranged from 36.1% (s.d. = 30.3%) in *P. gorganensis* and 36.4% (s.d. = 31.3%) in *A. mexicanum* to 98.31% in two *T. ivanbureschi* samples (s.d. = 7.2% and s.d. = 6.9%). On average, the mean read depth was 60.5 X (s.d. = 183.2) across all samples, and the overall mean coverage of the sequence bases was 92.0% (s.d. = 13.7%). Details can be found in Table S1.

We discovered significant correlations between the genetic divergence from *T. dobrogicus* and several performance variables. Firstly, as the divergence from *T. dobrogicus* increases, so does the amount of sequence data gathered, both as measured in GB in the raw FASTQ files and as measured in the total number of 'QC-passed' reads observed by SAMtools (Spearman's rank correlation test, $p<0.001$ for both, Table 1). We found no significant correlation between the percentage of reads marked as PCR duplicates by Picard and the level of divergence from *T. dobrogicus* ($p = 0.91$), however we did find a negative correlation between the percentage of reads that map against the *T. dobrogicus* reference sequences and the amount of divergence from *T. dobrogicus* (Spearman's rank correlation test, $p<0.001$, Table 1). Furthermore, there is a negative correlation between the coverage of reference targets and the level of divergence from *T. dobrogicus* - i.e. the mean read depth and the mean coverage of the sequence bases both drop as the divergence from *T. dobrogicus* increases (Pearson correlation test for the mean read depth, $p<0.001$ and Spearman's rank correlation test for the mean coverage of the sequence bases, $p<0.001$; Table 1). Lastly, with increased divergence from *T. dobrogicus*, more SNPs and INDELs are discovered by BCFtools in the eventual gVCF files (Spearman's rank correlation test, $p<0.001$ for both, Table 1).

Phylogenomics: Reconstruction of the Salamandridae phylogeny

The fully resolved and highly supported Salamandridae phylogeny resulting from Maximum Likelihood analysis of concatenated data in RAxML was based on 204,600 polymorphic SNPs (Fig. 1 and Fig. S1). The placement of the root on the branch connecting the newts and the clade containing the True Salamanders and *Salamandrina* was confirmed by our extended phylogeny

(based on 265,105 SNPs, Fig. S2) that included two non-Salamandridae salamander species (*A. mexicanum* and *P. gorganensis*), as well as a True Salamander species (*M. caucasica*).

Phylogeography: Fully resolved relationships of *Taricha*

The *Taricha* phylogeny that resulted from the concatenated data used by RAxML to perform a Maximum Likelihood analysis, was based on 9,730 polymorphic SNPs. The tree was fully resolved and highly supported (Fig. 2).

Hybridization studies: *Lissotriton* hybrids and backcrosses detected

After quality filtering, we identified 666 SNPs that we consider species-diagnostic (i.e. 'ancestry-informative') for *L. vulgaris* versus *L. montandoni* in the target capture data based on the genotypes of the six parental species samples and three F1 hybrids. Those SNPs enabled us to calculate the hybrid indices and interclass heterozygosity values in F2 and backcross ('Bx') hybrids, as visualized in a triangle plot (Fig. 3).

Conservation genetics: Separation of *Triturus* populations and *Chioglossa* subspecies

For the PCA and HCA analyses, the R calculations ended up being based on 9,135 bi-allelic SNPs. Along the first and second Principal Components of the PCA, the four wild populations from the postglacially colonized area cluster together, whereas the three wild populations from the glacial refugial area stand relatively apart Fig. 4A & B). The captive-bred individuals cluster closest to the Safaalan group, a population in Turkey just west of Istanbul, close to the presumed source locality Cerkezköy. The same pattern is observed in the HCA dendogram (Fig. 4C).

The check for high quality SNPs that display different genotypes in the two *C. lusitanica* samples originating from different subspecies, resulted in a list of over 10,000 polymorphic SNPs: we counted 8,029 SNPs for which one individual was homozygous, and the other individual was heterozygous, and we discovered an additional 2,301 informative SNPs where the individuals were both homozygous, but for alternate alleles.

Discussion

We introduce NewtCap, a target capture bait and reference set of 7,139 sequences applicable to salamandrids. We show that NewtCap works effectively across all main lineages within the

Salamandridae family. The target capture rates and mapping successes are influenced by the level of genetic divergence from *T. dobrogicus*, as anticipated, and we see that – within Salamandridae – more off-target regions are getting captured for more distantly related species. However, these influences appear to be minor and evidently do not hamper our downstream analyses: only for salamanders outside of the family Salamandridae does NewtCap provide data of insufficient quality. NewtCap thus proves to be a powerful tool to collect genomic data for Salamandridae studies regarding systematics (e.g. phylogenomics, phylogeography) and population genetics (e.g. hybrid studies, conservation genetics).

We encourage potential users to further tweak the NewtCap workflow. For instance, in terms of the laboratory protocol, researchers could consider; 1) replacing the C0t-1 blocker with a standard blocker, 2) reducing the hybridization time, and 3) including more individuals per capture reaction. In terms of bioinformatics, the lower mapping rate observed in Salamandridae species more distantly related to *T. dobrogicus* does not necessarily only reflect decreased enrichment efficiency due to genetic divergence; mapping rate is presumably also influenced by the reduced ability of the read mapper to align sequencing reads to more divergent reference sequences used in the bioinformatics pipeline (Bragg et al., 2016, Andermann et al., 2018). Users could explore; applying different mapping settings, using alternative mapper tools, or making use of (or generating) substitute reference sequences for read alignment (Schilbert et al., 2020, Andermann et al., 2019).

Our findings demonstrate that the current NewtCap protocol can effectively be applied to any member of the family Salamandridae – the crown of which is dated as far back as c. 100 MYA (Steinfartz et al., 2007, De Visser et al., in prep). Our Salamandridae phylogeny perfectly matches the topology of a transcriptome-based phylogeny (Rancilhac et al., 2021). Although a higher number of molecular markers does not necessarily result in a more accurate species tree (Mirarab et al., 2024, Zhang et al., 2021), independent RAxML analyses using different subsets of NewtCap data result in the same topology (with one notable exception, see; De Visser et al., in prep, France et al., in prep).

NewtCap has already recently been applied to study the systematics and taxonomy of certain modern European newts – including the genera; *Triturus* (Wielstra et al., 2019, Kazilas et al., 2024), *Lissotriton*, (Mars et al., 2024), and *Ommatotriton*, (Kalaentzis et al., 2024). However, we show that NewtCap allows for a (putative) genome-wide subsampling of thousands of markers for other Salamandridae lineages as well. Existing phylogenies of modern Asian newts (genera *Cynops*, *Paramesotriton*, and *Pachytriton*; Fig. 1) rely on a limited amount of molecular markers and frequently fail to recover genera within the clade as monophyletic groups, presumably due to the intricate biogeographical history of this clade (Veith et al., 2018, Zhang et al., 2008, Kieren et al., 2018). The New World Newts (genera *Taricha* and *Notophthalmus*; Fig. 1) have so far only been

studied based on mtDNA and allozyme data (Lawson and Kilpatrick, 2014, Gabor and Nice, 2004, Whitmore et al., 2013, Kuchta and Tan, 2006, Kuchta and Tan, 2005). Our *Taricha* phylogeny, the first one based on a considerable number of nuclear markers/SNPs, comprises a fully supported positioning of each of the four known species, with a topology that matches that of an existing mtDNA-based phylogeny (Tan and Wake, 1995, Tan, 1993). For the Asian members of the primitive newts (genera *Tylototriton* and *Echinotriton*; see Fig. 1), genetic resources have been scarce so far, which is why there is an outstanding call to conduct more extensive, genomic research in order to better understand the evolution and taxonomy of the (sub)species of these lineages (Dufresnes and Hernandez, 2022, Weisrock et al., 2006).

NewtCap performs well even for the sister clade of the newts, containing the 'True Salamanders' (*Salamandra, Chioglossa, Mertensiella, Lyciasalamandra*; see Fig. 1) and the spectacled salamanders (*Salamandrina*). First, NewtCap provides empirical support for the recent suggestion that *Salamandrina* represents the sister lineage to the True Salamanders instead of to all remaining salamandrids (Rancilhac et al., 2021). *Salamandrina* itself has so far only been studied with a relatively small number of markers (Mattoccia et al., 2011, Mattoccia et al., 2005, Canestrelli et al., 2014, Hauswaldt et al., 2014, Veith et al., 2009, Weisrock et al., 2006, Zhang et al., 2008). *Chioglossa* is a monotypic genus that is close to becoming threatened according to the IUCN Red List (IUCN_SSC_Amphibian_Specialist_Group, 2022). Conservationists generally identify monotypic taxa as 'evolutionarily unique', which helps justify elevated conservation imperatives (Liu et al., 2021, Vane-Wright et al., 1991, Vitt et al., 2023). We obtained c. ten thousand informative SNPs distinguishing two individuals belonging to different *C. lusitanica* subspecies (Sequeira et al., 2022), something that could aid conservation efforts. The sister genus of *Chioglossa*, *Mertensiella*, is monotypic as well, but so far few populations have been studied and only with mtDNA (Tarkhnishvili et al., 2000, Weisrock et al., 2001).

Next to accentuating the potential of NewtCap in the context of conservation genetics, we emphasize that the tool can also be used to identify the degree of interspecific gene flow and geographical structuring. For example, to test hypotheses about species status and historical biogeography. Introgressive hybridization is increasingly recognized as a source of adaptive variation in natural populations, (Fijarczyk et al., 2018, Li et al., 2016, Mallet, 2005), but also as a source of genetic pollution in case non-native and native individuals interbreed in nature (Quilodran et al., 2020, Wielstra et al., 2016, Simberloff, 2013, Dufresnes et al., 2015). We identify different hybrid classes of *Lissotriton* newts and are able to differentiate between genetically distinct groups of *T. ivanbureschi* – and determined to which wild population a known, captive population is genetically most similar. Such applications are valuable for guiding wildlife management practices

and conservation efforts that concern Salamandridae species both *in situ* and *ex situ* (Frankham et al., 2004, Ørsted et al., 2019, Weisrock et al., 2018). This is especially important considering that, out of all vertebrate groups, amphibians are facing the most drastic population declines and extinction rates observed in the Anthropocene (Stuart et al., 2004, Hayes et al., 2010, Sparreboom, 2014, Lucas et al., 2024, Martel et al., 2013). Overall, by providing in the range of thousands to hundreds of thousands of high-quality SNPs, NewtCap facilitates the molecular study of salamandrids whilst whole genome sequencing of the gigantic genomes of salamanders remains unattainable.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the European Research Council grant agreement No. 802759 and the Marie Skłodowska-Curie grant agreement No. 655487. This work was performed using the compute resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. Daniele Canestrelli, Andrea Chiocchio, Michael Fahrbach, Willem Meilink, and Frank Pasmans contributed samples. The tissue sample of *Notophthalmus viridescens* (SMF 99040) was donated by the Senckenberg Research Institute and Nature Museum. The *Taricha* tissue samples were collected following protocols for animal welfare approved by UCLA and permitted by California Department of Fish and Wildlife (SCP # 12430). We would like to thank Erik-Jan Bosch for making the scientific drawings of the salamanders used in Figure 1 under the open content License of Naturalis Biodiversity Center (© CC BY-NC-ND 4.0), Benjamin Wiens for providing support with *triangulaR* adjustments, and Dr. Peter Scott for providing input on bioinformatic analyses. The *Triturus ivanbureschi* geographical map cut-out (Figure 4A) was made thanks to OpenStreetMap and its contributors (® CC BY-SA 2.0), and we obtained the *Taricha* cartoon from Figure 2 copyright free via user 'clker-free-vector-images-3736' through Pixabay.com.

References

ALBERT, T. J., MOLLA, M. N., MUZNY, D. M., NAZARETH, L., WHEELER, D., SONG, X., RICHMOND, T. A., MIDDLE, C. M., RODESCH, M. J., PACKARD, C. J., WEINSTOCK, G. M. & GIBBS, R. A. 2007. Direct selection of human genomic loci by microarray hybridization. *Nature Methods,* 4**,** 903-5.

ALFARO, M. E., FAIRCLOTH, B. C., HARRINGTON, R. C., SORENSON, L., FRIEDMAN, M., THACKER, C. E., OLIVEROS, C. H., CERNY, D. & NEAR, T. J. 2018. Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. *Nature Ecology & Evolution,* 2**,** 688-696.

ANDERMANN, T., CANO, A., ZIZKA, A., BACON, C. & ANTONELLI, A. 2018. SECAPR-a bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments. *PeerJ,* 6**,** e5175.

ANDERMANN, T., TORRES JIMENEZ, M. F., MATOS-MARAVI, P., BATISTA, R., BLANCO-PASTOR, J. L., GUSTAFSSON, A. L. S., KISTLER, L., LIBERAL, I. M., OXELMAN, B., BACON, C. D. & ANTONELLI, A. 2019. A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project. *Frontiers in Genetics,* 10**,** 1407.

ANDREWS, K. R., GOOD, J. M., MILLER, M. R., LUIKART, G. & HOHENLOHE, P. A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews: Genetics,* 17**,** 81-92.

ANDREWS, S. 2010. *FastQC: A Quality Control Tool for High Throughput Sequence Data*.

ARNOLD, B., CORBETT-DETIG, R. B., HARTL, D. & BOMBLIES, K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology,* 22**,** 3179-90.

AUWERA, G. A. & O'CONNOR, B. D. 2020. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*, O'Reilly Media.

BABIK, W., MARSZALEK, M., DUDEK, K., ANTUNES, B., PALOMAR, G., ZAJAC, B., TAUGBOL, A. & PABIJAN, M. 2024. Limited evidence for genetic differentiation or adaptation in two amphibian species across replicated rural-urban gradients. *Evolutionary Applications,* 17**,** e13700.

BI, K., VANDERPOOL, D., SINGHAL, S., LINDEROTH, T., MORITZ, C. & GOOD, J. M. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data colection at moderate evolutionary scales. *BMC Genomics,* 13**,** 1471-2164.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-20.

BRAGG, J. G., POTTER, S., BI, K. & MORITZ, C. 2016. Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources,* 16**,** 1059-68.

BURGON, J. D., VENCES, M., STEINFARTZ, S., BOGAERTS, S., BONATO, L., DONAIRE-BARROSO, D., MARTINEZ-SOLANO, I., VELO-ANTON, G., VIEITES, D. R., MABLE, B. K. & ELMER, K. R. 2021. Phylogenomic inference of species and subspecies diversity in the Palearctic salamander genus *Salamandra*. *Molecular Phylogenetics and Evolution,* 157**,** 107063.

CALBOLI, F. C., FISHER, M. C., GARNER, T. W. & JEHLE, R. 2011. The need for jumpstarting amphibian genome projects. *Trends in Ecology & Evolution,* 26**,** 378-9.

CANESTRELLI, D., BISCONTI, R. & NASCETTI, G. 2014. Extensive unidirectional introgression between two salamander lineages of ancient divergence and its evolutionary implications. *Scientific Reports,* 4**,** 6516.

DA FONSECA, R. R., ALBRECHTSEN, A., THEMUDO, G. E., RAMOS-MADRIGAL, J., SIBBESEN, J. A., MARETTY, L., ZEPEDA-MENDOZA, M. L., CAMPOS, P. F., HELLER, R. & PEREIRA, R. J. 2016. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics,* 30**,** 3-13.

DANECEK, P., BONFIELD, J. K., LIDDLE, J., MARSHALL, J., OHAN, V., POLLARD, M. O., WHITWHAM, A., KEANE, T., MCCARTHY, S. A., DAVIES, R. M. & LI, H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience,* 10.

DAVEY, J. W., CEZARD, T., FUENTES-UTRILLA, P., ELAND, C., GHARBI, K. & BLAXTER, M. L. 2013. Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology,* 22**,** 3151-64.

DE VISSER, M. C., FRANCE, J., PAULOUSKAYA, O., BROWN, T., FAHRBACH, M., VAN DER PLOEG, C. & WIELSTRA, B. in prep. Conserved gene content and unique phylogenetic history characterize the 'bloopergene' underlying *Triturus*' balanced lethal system

DODSWORTH, S., POKORNY, L., JOHNSON, M. G., KIM, J. T., MAURIN, O., WICKETT, N. J., FOREST, F. & BAKER, W. J. 2019. Hyb-Seq for Flowering Plant Systematics. *Trends in Plant Science,* 24**,** 887-891.

DUFRESNES, C., DUBEY, S., GHALI, K., CANESTRELLI, D. & PERRIN, N. 2015. Introgressive hybridization of threatened European tree frogs (*Hyla arborea*) by introduced H. intermedia in Western Switzerland. *Conservation Genetics,* 16**,** 1507-1513.

DUFRESNES, C. & HERNANDEZ, A. 2022. Towards completing the crocodile newts' puzzle with all-inclusive phylogeographic resources. *Zoological Journal of the Linnean Society,* 197**,** 620-640.

FAIRCLOTH, B. C., BRANSTETTER, M. G., WHITE, N. D. & BRADY, S. G. 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources,* 15**,** 489-501.

FEATHERSTONE, L. A. & MCGAUGHRAN, A. 2024. The effect of missing data on evolutionary analysis of sequence capture bycatch, with application to an agricultural pest. *Molecular Genetics and Genomics,* 299**,** 11.

FIJARCZYK, A., DUDEK, K., NIEDZICKA, M. & BABIK, W. 2018. Balancing selection and introgression of newt immune-response genes. *Proceedings of the Royal Society B: Biological Sciences,* 285.

FITZPATRICK, B. M., MCCARTNEY-MELSTAD, E., JOHNSON, J. R. & SHAFFER, H. B. 2024. New evidence contradicts the rapid spread of invasive genes into a threatened native species. *Biological Invasions*.

FRANCE, J., BABIK, W., DUDEK, K., MARSZAŁEK, M. & WIELSTRA, B. 2024. Linkage mapping vs Association: A comparison of two RADseq-based approaches to identify markers for homomorphic sex chromosomes in large genomes. *Authorea Preprints*.

FRANCE, J., DE VISSER, M., PAULOUSKAYA, O., ARNTZEN, P., BABIK, W., CVIJANOVIĆ, M., IVANOVIĆ, A., SMITH, J. & VUČIĆ, T. in prep. Genomic evidence suggests the balanced lethal system in *Triturus* newts originated in an instantaneous speciation event.

FRANKHAM, R., BALLOU, J. D. & BRISCOE, D. A. 2004. A Primer of Conservation Genetics. 1-236.

FROST, D. R. 1985. *Amphibian species of the world*, A taxonomic and geographical reference. Lawrence, KS: Association of Systematic Collections.

GABOR, C. R. & NICE, C. C. 2004. Genetic variation among populations of eastern newts, *Notophthalmus viridescens*: a preliminary analysis based on allozymes. *Herpetologica,* 60**,** 373-386.

GNIRKE, A., MELNIKOV, A., MAGUIRE, J., ROGOV, P., LEPROUST, E. M., BROCKMAN, W., FENNELL, T., GIANNOUKOS, G., FISHER, S., RUSS, C., GABRIEL, S., JAFFE, D. B., LANDER, E. S. & NUSBAUM, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology,* 27**,** 182-9.

GOOD, J. M. 2011. Reduced representation methods for subgenomic enrichment and next-generation sequencing. *Molecular methods for evolutionary genetics***,** 85-103.

GREGORY, T. R. 2002. Genome size and developmental complexity. *Genetica,* 115**,** 131-146.

GREGORY, T. R. 2024. *Animal Genome Size Database* [Online]. Available: http://www.genomesize.com [Accessed].

GROVER, C. E., SALMON, A. & WENDEL, J. F. 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. *American Journal of Botany,* 99**,** 312-9.

GUO, Y., LONG, J., HE, J., LI, C.-I., CAI, Q., SHU, X.-O., ZHENG, W. & LI, C. 2012. Exome sequencing generates high quality data in non-target regions. *BMC genomics,* 13**,** 1-10.

HARVEY, M. G., SMITH, B. T., GLENN, T. C., FAIRCLOTH, B. C. & BRUMFIELD, R. T. 2016. Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Systematic Biology,* 65**,** 910-924.

HAUSWALDT, J. S., ANGELINI, C., GEHARA, M., BENAVIDES, E., POLOK, A. & STEINFARTZ, S. 2014. From species divergence to population structure: a multimarker approach on the most basal lineage of Salamandridae, the spectacled salamanders (genus *Salamandrina*) from Italy. *Molecular Phylogenetics and Evolution,* 70**,** 1-12.

HAYES, T. B., FALSO, P., GALLIPEAU, S. & STICE, M. 2010. The cause of global amphibian declines: A developmental endocrinologist's perspective. *Journal of Experimental Biology,* 213**,** 921-933.

HEYDUK, K., STEPHENS, J. D., FAIRCLOTH, B. C. & GLENN, T. C. 2016. Targeted DNA Region Re-sequencing. *Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing.*

HU, Q., CHANG, C., WANG, Q., TIAN, H., QIAO, Z., WANG, L., MENG, Y., XU, C. & XIAO, H. 2019. Genome-wide RAD sequencing to identify a sex-specific marker in Chinese giant salamander *Andrias davidianus*. *BMC Genomics,* 20**,** 415.

HUBBS, N. W., HURT, C. R., NIEDZWIECKI, J., LECKIE, B. & WITHERS, D. 2022. Conservation genomics of urban populations of Streamside Salamander (*Ambystoma barbouri*). *PLoS ONE,* 17**,** e0260178.

HUTTER, C. R., COBB, K. A., PORTIK, D. M., TRAVERS, S. L., WOOD, P. L., JR. & BROWN, R. M. 2022. FrogCap: A modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales. *Molecular Ecology Resources,* 22**,** 1100-1119.

IUCN_SSC_AMPHIBIAN_SPECIALIST_GROUP. 2022. *Chioglossa lusitanica. The IUCN Red List of Threatened Species 2022* [Online]. Available: https://dx.doi.org/10.2305/IUCN.UK.2022-1.RLTS.T4657A89698017.en [Accessed].

JIMENEZ-MENA, B., FLAVIO, H., HENRIQUES, R., MANUZZI, A., RAMOS, M., MELDRUP, D., EDSON, J., PALSSON, S., ASTA OLAFSDOTTIR, G., OVENDEN, J. R. & NIELSEN, E. E. 2022. Fishing for DNA? Designing baits for population genetics in target enrichment experiments: Guidelines, considerations and the new tool supeRbaits. *Molecular Ecology Resources,* 22**,** 2105-2119.

JONES, M. R. & GOOD, J. M. 2016. Targeted capture in evolutionary and ecological genomics. *Molecular Ecology,* 25**,** 185-202.

KALAENTZIS, K., KOSTER, S., ARNTZEN, J. W., BOGAERTS, S., FRANCE, J., FRANZEN, M., KAZILAS, C., LITVINCHUK, S. N., OLGUN, K. & DE VISSER, M. C. 2024. Phylogenomics Resolves the Ancient and Rapid Radiation of Banded Newts (Genus *Ommatotriton*). *[Available as SSRN Preprint]*.

KAZILAS, C., DUFRESNES, C., FRANCE, J., KALAENTZIS, K., MARTINEZ-SOLANO, I., DE VISSER, M. C., ARNTZEN, J. W. & WIELSTRA, B. 2024. Spatial genetic structure in European marbled newts revealed with target enrichment by sequence capture. *Molecular Phylogenetics and Evolution,* 194**,** 108043.

KHAN, R., BIFFIN, E., VAN DIJK, K. J., HILL, R. S., LIU, J. & WAYCOTT, M. 2024. Development of a Target Enrichment Probe Set for Conifer (REMcon). *Biology (Basel),* 13.

KIEREN, S., SPARREBOOM, M., HOCHKIRCH, A. & VEITH, M. 2018. A biogeographic and ecological perspective to the evolution of reproductive behaviour in the family Salamandridae. *Molecular Phylogenetics and Evolution,* 121**,** 98-109.

KNAUS, B. J. & GRUNWALD, N. J. 2017. vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources,* 17**,** 44-53.

KUCHTA, S. R. & TAN, A.-M. 2006. Lineage diversification on an evolving landscape: phylogeography of the California newt, *Taricha torosa* (Caudata: Salamandridae). *Biological Journal of the Linnean Society,* 89**,** 213-239.

KUCHTA, S. R. & TAN, A. M. 2005. Isolation by distance and post-glacial range expansion in the rough-skinned newt, *Taricha granulosa*. *Molecular Ecology,* 14**,** 225-44.

LAWSON, G. R. & KILPATRICK, E. S. 2014. Phylogeographic Patterns Among the Subspecies of *Notophthalmus viridescens* (Eastern Newt) in South Carolina. *Southeastern Naturalist,* 13**,** 444-455.

LEMMON, E. M. & LEMMON, A. R. 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics,* 44**,** 99-121.

LEWIS, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic biology,* 50**,** 913-925.

LI, G., DAVIS, B. W., EIZIRIK, E. & MURPHY, W. J. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Research,* 26**,** 1-11.

LI, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics,* 27**,** 2987-93.

LI, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & 1000_GENOME_PROJECT_DATA_PROCESSING_SUBGROUP 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LITVINCHUK, S. N., ROSANOV, J. M. & BORKIN, L. J. 2007. Correlations of geographic distribution and temperature of embryonic development with the nuclear DNA content in the Salamandridae (Urodela, Amphibia). *Genome,* 50**,** 333-342.

LIU, L., BOSSE, M., MEGENS, H. J., DE VISSER, M., M, A. M. G. & MADSEN, O. 2021. Genetic consequences of long-term small effective population size in the critically endangered pygmy hog. *Evolutionary Applications,* 14**,** 710-720.

LOU, R. N., JACOBS, A., WILDER, A. P. & THERKILDSEN, N. O. 2021. A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology,* 30**,** 5966-5993.

LOWRY, D. B., HOBAN, S., KELLEY, J. L., LOTTERHOS, K. E., REED, L. K., ANTOLIN, M. F. & STORFER, A. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources,* 17**,** 142-152.

LUCAS, P. M., DI MARCO, M., CAZALIS, V., LUEDTKE, J., NEAM, K., BROWN, M. H., LANGHAMMER, P. F., MANCINI, G. & SANTINI, L. 2024. Using comparative extinction risk analysis to prioritize the IUCN Red List reassessments of amphibians. *Conservation Biology***,** e14316.

MALLET, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology and Evolution,* 20**,** 229-237.

MAMANOVA, L., COFFEY, A. J., SCOTT, C. E., KOZAREWA, I., TURNER, E. H., KUMAR, A., HOWARD, E., SHENDURE, J. & TURNER, D. J. 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods,* 7**,** 111-8.

MARJANOVIĆ, D. & LAURIN, M. 2013. An updated paleontological timetree of lissamphibians, with comments on the anatomy of Jurassic crown-group salamanders (Urodela). *Historical Biology,* 26**,** 535-550.

MARS, J., KOSTER, S., BABIK, W., FRANCE, J., KALAENTZIS, K., KAZILAS, C., MARTÍNEZ-SOLANO, I., DE VISSER, M. C. & WIELSTRA, B. 2024. Phylogenomics Yields New Systematic and Taxonomical Insights for *Lissotriton* Newts, a Lineage with a Strong Legacy of Introgressive Hybridization. *[Available as SSRN Preprint]*.

MARTEL, A., SPITZEN-VAN DER SLUIJS, A., BLOOI, M., BERT, W., DUCATELLE, R., FISHER, M. C., WOELTJES, A., BOSMAN, W., CHIERS, K., BOSSUYT, F. & PASMANS, F. 2013. *Batrachochytrium salamandrivorans* sp. nov. causes lethal chytridiomycosis in amphibians. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 15325-15329.

MATTOCCIA, M., MARTA, S., ROMANO, A. & SBORDONI, V. 2011. Phylogeography of an Italian endemic salamander (genus *Salamandrina*): glacial refugia, postglacial expansions, and secondary contact. *Biological Journal of the Linnean Society,* 104**,** 903-992.

MATTOCCIA, M., ROMANO, A. & SBORDONI, V. 2005. Mitochondrial DNA sequence analysis of the spectacled salamander, *Salamandrina terdigitata* (Urodela: Salamandridae), supports the existence of two distinct species. *Zootaxa,* 995**,** 1–19-1–19.

MCCARTNEY-MELSTAD, E., MOUNT, G. G. & SHAFFER, H. B. 2016. Exon capture optimization in amphibians with large genomes. *Molecular ecology resources,* 16**,** 1084-1094.

MIRARAB, S., RIVAS-GONZALEZ, I., FENG, S., STILLER, J., FANG, Q., MAI, U., HICKEY, G., CHEN, G., BRAJUKA, N., FEDRIGO, O., FORMENTI, G., WOLF, J. B. W., HOWE, K., ANTUNES, A., SCHIERUP, M. H., PATEN, B., JARVIS, E. D., ZHANG, G. & BRAUN, E. L. 2024. A region of suppressed recombination misleads neoavian phylogenomics. *Proceedings of the National Academy of Sciences,* 121**,** e2319506121.

ØRSTED, M., HOFFMANN, A. A., SVERRISDÓTTIR, E., NIELSEN, K. L. & KRISTENSEN, T. N. 2019. Genomic variation predicts adaptive evolutionary responses better than population bottleneck history. *PLoS Genetics,* 15**,** 1-18.

ORTIZ, E. M. 2019. vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis.

PASMANS, F., BOGAERTS, S., WOELTJES, T. & CARRANZA, S. 2006. Biogeography of Neurergus strauchii barani Öz, 1994 and *N. s. strauchii* (Steindachner, 1887) (Amphibia: Salamandridae) assessed using morphological and molecular data. *Amphibia-Reptilia,* 27**,** 281-288.

POPLIN, R., RUANO-RUBIO, V., DEPRISTO, M. A., FENNELL, T. J., CARNEIRO, M. O., VAN DER AUWERA, G. A., KLING, D. E., GAUTHIER, L. D., LEVY-MOONSHINE, A., ROAZEN, D., SHAKIR, K., THIBAULT, J., CHANDRAN, S., WHELAN, C., LEK, M., GABRIEL, S., DALY, M. J., NEALE, B., MACARTHUR, D. G. & BANKS, E. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv,* 201178.

PORTIK, D. M., SMITH, L. L. & BI, K. 2016. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources,* 16**,** 1069-83.

QUEK, Z. B. R. & NG, S. H. 2024. Hybrid-Capture Target Enrichment in Human Pathogens: Identification, Evolution, Biosurveillance, and Genomic Epidemiology. *Pathogens,* 13.

QUILODRAN, C. S., MONTOYA-BURGOS, J. I. & CURRAT, M. 2020. Harmonizing hybridization dissonance in conservation. *Communications Biology,* 3**,** 391.

RANCILHAC, L., GOUDARZI, F., GEHARA, M., HEMAMI, M. R., ELMER, K. R., VENCES, M. & STEINFARZ, S. 2019. Phylogeny and species delimitation of near Eastern *Neurergus* newts (Salamandridae) based on genome-wide RADseq data analysis. *Molecular Phylogenetics and Evolution,* 133**,** 189-197.

RANCILHAC, L., IRISARRI, I., ANGELINI, C., ARNTZEN, J. W., BABIK, W., BOSSUYT, F., KÜNZEL, S., LÜDDECKE, T., PASMANS, F., SANCHEZ, E., WEISROCK, D., VEITH, M., WIELSTRA, B., STEINFARTZ, S., HOFREITER, M., PHILIPPE, H. & VENCES, M. 2021. Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. *Molecular Phylogenetics and Evolution,* 155.

ROBBEMONT, J., VAN VELDHUIJZEN, S., ALLAIN, STEVEN J. R., AMBU, J., BOYLE, R., CANESTRELLI, D., CATHASAIGH, É. Ó., CATHRINE, C., CHIOCCHIO, A., COGALNICEANU, D., CVIJANOVIĆ, M., DUFRESNES, C., ENNIS, C., GANDOLA, R., JABLONSKI, D., JULIAN, A., KRANŽELIĆ, D., LUKANOV, S., MARTÍNEZ-SOLANO, I., MONTGOMERY, R., NAUMOV, B., O'NEILL, M., NORTH, A., PABIJAN, M., PUSHENDORF, R., SALVI, D., SCHMIDT, B., SOTIROPOULOS, K., STANESCU, F., STANKOVIĆ, D., STAPLETON, S., ŠUNJE, E., SZABOLCS, M., VACHEVA, E., WILLIS, D., ZIMIĆ, A., FRANCE, J., MEILINK, WILLEM R. M., STARK, T., STRUIJK, RICHARD P. J. H., THEODOROPOULOS, A., DE VISSER, MANON C. & WIELSTRA, B. 2023. An extended mtDNA phylogeography for the alpine newt illuminates the provenance of introduced populations. *Amphibia-Reptilia,* 44**,** 347-361.

RODRÍGUEZ, A., BURGON, J. D., LYRA, M., IRISARRI, I., BAURAIN, D., BLAUSTEIN, L., GÖÇMEN, B., KÜNZEL, S., MABLE, B. K., NOLTE, A. W., VEITH, M., STEINFARTZ, S., ELMER, K. R., PHILIPPE, H. & VENCES, M. 2017. Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches. *Molecular Phylogenetics and Evolution,* 115**,** 16-26.

RUBIN, B. E., REE, R. H. & MOREAU, C. S. 2012. Inferring phylogenies from RAD sequence data. *PLoS ONE,* 7**,** e33394.

SAMBROOK, J. & RUSSELL, D. W. 2001. Molecular cloning: a laboratory manual. 3rd edition ed. London, United Kingdom: Coldspring-Harbour Laboratory Press.

SCHILBERT, H. M., REMPEL, A. & PUCKER, B. 2020. Comparison of Read Mapping and Variant Calling Tools for the Analysis of Plant NGS Data. *Plants (Basel),* 9.

SCHOTT, R. K., PANESAR, B., CARD, D. C., PRESTON, M., CASTOE, T. A. & CHANG, B. S. 2017. Targeted Capture of Complete Coding Regions across Divergent Species. *Genome Biology and Evolution,* 9**,** 398-414.

SCOTT, P., NAJAFI-MAJD, E., YıLDıRıM CAYNAK, E., GIDIŞ, M., KAYA, U. & SHAFFER, H. B. 2024. Phylogenomics Reveal Species Limits and Relationships in the Narrow-Range Endemic Lycian Salamanders. *[Available as SSRN Preprint]*.

SEQUEIRA, F., ARNTZEN, J. W., VAN GULIK, D., HAJEMA, S., DIAZ, R. L., WAGT, M. & VAN RIEMSDIJK, I. 2022. Genetic traces of hybrid zone movement across a fragmented habitat. *Journal of Evolutionary Biology,* 35**,** 400-412.

SESSIONS, S. K. 2008. Evolutionary cytogenetics in salamanders. *Chromosome Research,* 16**,** 183-201.

SIMBERLOFF, D. 2013. *Invasive species: what everyone needs to know*, Oxford University Press.

SINGHAL, S., GRUNDLER, M., COLLI, G. & RABOSKY, D. L. 2017. Squamate Conserved Loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources,* 17**,** e12-e24.

SMITH, J. J., TIMOSHEVSKAYA, N., TIMOSHEVSKIY, V. A., KEINATH, M. C., HARDY, D. & VOSS, S. R. 2019. A chromosome-scale assembly of the axolotl genome. *Genome Research,* 29**,** 317-324.

SPARREBOOM, M. 2014. Salamanders of the Old World: the salamanders of Europe, Asia and northern Africa. 431p.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-3.

STEINFARTZ, S., VICARIO, S., ARNTZEN, J. W. & CACCONE, A. 2007. A Bayesian Approach on Molecules and Behavior: Reconsidering Phylogenetic and Evolutionary Patterns of the Salamandridae with Emphasis on *Triturus* Newts. *Journal of experimental zoology. Part B, Molecular and developmental evolution,* 308B**,** 139-162.

STUART, S. N., CHANSON, J. S., COX, N. A., YOUNG, B. E., RODRIGUES, A. S. L., FISCHMAN, D. L. & WALLER, R. W. 2004. Status and Trends of Amphibian Declines and Extinctions Worldwide. *Science,* 306**,** 1783-1787.

SUN, C., SHEPARD, D. B., CHONG, R. A., LOPEZ ARRIAZA, J., HALL, K., CASTOE, T. A., FESCHOTTE, C., POLLOCK, D. D. & MUELLER, R. L. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biology and Evolution,* 4**,** 168-83.

TAN, A.-M. 1993. *Systematics, phylogeny and biogeography of the northwest American newts of the genus Taricha (Caudata: Salamandridae)*, University of California, Berkeley.

TAN, A.-M. & WAKE, D. B. 1995. MtDNA phylogeography of the California newt, *Taricha torosa* (Caudata, Salamandridae). *Molecular Phylogenetics and Evolution,* 4**,** 383-394.

TARKHNISHVILI, D. N., THORPE, R. S. & ARNTZEN, J. W. 2000. Pre-pleistocene refugia and differentiation between populations of the caucasian salamander (*Mertensiella caucasica*). *Molecular Phylogenetics and Evolution,* 14**,** 414-22.

TEAM_R_CORE 2020. RA language and environment for statistical computing, R Foundation for Statistical. *Computing*.

TEAM_R_STUDIO 2021. RStudio: integrated development environment for R.

TEASDALE, L. C., KOHLER, F., MURRAY, K. D., O'HARA, T. & MOUSSALLI, A. 2016. Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. *Molecular Ecology Resources,* 16**,** 1107-23.

VALBUENA-URENA, E., AMAT, F. & CARRANZA, S. 2013. Integrative phylogeography of *Calotriton* newts (Amphibia, Salamandridae), with special remarks on the conservation of the endangered Montseny brook newt (*Calotriton arnoldi*). *PLoS ONE,* 8**,** e62542.

VANE-WRIGHT, R. I., HUMPHRIES, C. J. & WILLIAMS, P. H. 1991. What to protect?—Systematics and the agony of choice. *Biological conservation,* 55**,** 235-254.

VEITH, M., BOGAERTS, S., PASMANS, F. & KIEREN, S. 2018. The changing views on the evolutionary relationships of extant Salamandridae (Amphibia: Urodela). *PLoS ONE,* 13**,** 1-16.

VEITH, M., STEINFARTZ, S., ZARDOYA, R., SEITZ, A. & MEYER, A. 2009. A molecular phylogeny of 'true' salamanders (family Salamandridae) and the evolution of terrestriality of reproductive modes. *Journal of Zoological Systematics and Evolutionary Research,* 36**,** 7-16.

VITT, P., TAYLOR, A., RAKOSY, D., KREFT, H., MEYER, A., WEIGELT, P. & KNIGHT, T. M. 2023. Global conservation prioritization for the Orchidaceae. *Scientific reports,* 13**,** 6718.

WEISROCK, D. W., HIME, P. M., NUNZIATA, S. O., JONES, K. S., MURPHY, M. O., HOTALING, S. & KRATOVIL, J. D. 2018. Surmounting the Large-Genome "Problem" for Genomic Data Generation in Salamanders. *Population Genomics: Wildlife.*

WEISROCK, D. W., MACEY, J. R., UGURTAS, I. H., LARSON, A. & PAPENFUSS, T. J. 2001. Molecular phylogenetics and historical biogeography among salamandrids of the "true" salamander clade: rapid branching of numerous highly divergent lineages in *Mertensiella luschani* associated with the rise of Anatolia. *Molecular Phylogenetics and Evolution,* 18**,** 434-448.

WEISROCK, D. W., PAPENFUSS, T. J., MACEY, J. R., LITVINCHUK, S. N., POLYMENI, R., UGURTAS, I. H., ZHAO, E., JOWKAR, H. & LARSON, A. 2006. A molecular assessment of phylogenetic relationships and lineage accumulation rates within the family Salamandridae (Amphibia, Caudata). *Molecular Phylogenetics and Evolution,* 41**,** 368-83.

WHITMORE, S. S., LOSEE, S., MEYER, L. & SPRADLING, T. A. 2013. Conservation genetics of the central newt (*Notophthalmus viridescens*) in Iowa: the importance of a biogeographic framework. *Conservation Genetics,* 14**,** 771-781.

WICKHAM, H. 2011. ggplot2. *Wiley interdisciplinary reviews: computational statistics,* 3**,** 180-185.

WIELSTRA, B., BURKE, T., BUTLIN, R. K. & ARNTZEN, J. W. 2017. A signature of dynamic biogeography: enclaves indicate past species replacement. *Proceedings of the Royal Society B: Biological Sciences,* 284.

WIELSTRA, B., BURKE, T., BUTLIN, R. K., SCHAAP, O., SHAFFER, H. B., VRIELING, K. & ARNTZEN, J. W. 2016. Efficient screening for 'genetic pollution' in an anthropogenic crested newt hybrid zone. *Conservation Genetics Resources,* 8**,** 553-560.

WIELSTRA, B., MCCARTNEY-MELSTAD, E., ARNTZEN, J. W., BUTLIN, R. K. & SHAFFER, H. B. 2019. Phylogenomics of the adaptive radiation of *Triturus* newts supports gradual ecological niche expansion towards an incrementally aquatic lifestyle. *Molecular Phylogenetics and Evolution,* 133**,** 120-127.

WIENS, B. J. & COLELLA, J. P. 2024. *triangulaR*: an R package for identifying AIMs and building triangle plots using SNP data from hybrid zones. *bioRxiv***,** 2024.03.28.587167.

YU, P. L., FULTON, J. C., HUDSON, O. H., HUGUET-TAPIA, J. C. & BRAWNER, J. T. 2023. Next-generation fungal identification using target enrichment and Nanopore sequencing. *BMC Genomics,* 24**,** 581.

ZAHARIAS, P., PANTE, E., GEY, D., FEDOSOV, A. E. & PUILLANDRE, N. 2020. Data, time and money: evaluating the best compromise for inferring molecular phylogenies of non-model animal taxa. *Molecular Phylogenetics and Evolution,* 142**,** 106660.

ZHANG, D., RHEINDT, F. E., SHE, H., CHENG, Y., SONG, G., JIA, C., QU, Y., ALSTROM, P. & LEI, F. 2021. Most Genomic Loci Misrepresent the Phylogeny of an Avian Radiation Because of Ancient Gene Flow. *Systematic Biology,* 70**,** 961-975.

ZHANG, P., PAPENFUSS, T. J., WAKE, M. H., QU, L. & WAKE, D. B. 2008. Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution,* 49**,** 586-97.

ZHENG, X., LEVINE, D., SHEN, J., GOGARTEN, S. M., LAURIE, C. & WEIR, B. S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics,* 28**,** 3326-8.

ZHOU, L. & HOLLIDAY, J. A. 2012. Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *BMC genomics,* 13**,** 1-12.

Conflict of Interest

All authors declare no conflict of interest.

Data accessibility

Supplementary Materials such as Table S1, the laboratory protocols, certain files required for some of the bioinformatics steps, and the FASTA file containing the sequences that were used for initial probe tiling, are publicly available through Zenodo (https://doi.org/10.5281/zenodo.13987888). The main bioinformatic steps and scripts are provided on the NewtCap GitHub pages (https://github.com/Wielstra-Lab/NewtCap_bioinformatics). Lastly, the raw sequencing reads used in this study have been submitted to the NCBI Sequence Read Archive (SRA) and are accessible via BioProject PRJNA1171613 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1171613).

Benefit-Sharing Statement

The samples used in this study are in compliance with national laws and the Nagoya Protocol. Moreover, as described above, the data are shared via suitable, biological databases.

Author Contributions

B.W., M.d.V., & J.F. conceived and designed the research. B.W., E.M., & H.B.S. initially designed the NewtCap probes for *Triturus*, and J.F. & B.W. further optimized the tool for broader use and higher efficiency. M.d.V., J.F., & A.T. performed the lab-work. M.d.V. conducted the pre-processing of the data, as well as all downstream analyses. G.B. contributed to the analyses of *Taricha*. M.d.V., B.W., & J.F. wrote the draft version of the manuscript, and all authors contributed to revising it.

Tables

Table 1: Correlation statistics for the relationship between NewtCap performance variables and the estimated amount of species divergence from **Triturus dobrogicus**[†]**.** *For each of the separate performance variable the appropriate statistical test - either the Spearmann's rank or the Pearson correlation test - was applied to determine whether the observed correlations were significant. Both the correlation coefficient (R) and its square value are provided, as well as the p-value. Results with a p-value lower than 0.00625 (including a Bonferroni correction, see Methods) are marked with an asterisk.*

Performance Variable	**Test applied**	**R**	**R^2**	**P-value**
Amount of raw data R1 + R2 (in GB)	Spearmann	0.610	0.372	1.6E-08*
Number of QC-passed reads	Spearmann	0.641	0.411	1.7E-09*
Percentage of reads mapped	Spearmann	-0.661	0.440	3.6E-10*
Percentage of PCR duplicates	Spearmann	0.015	<0.001	0.910
Mean read depth	Pearson	-0.597	0.357	3.8E-08*
Mean coverage of sequence bases	Spearmann	-0.918	0.842	2.2E-29*
Number of SNPs discovered	Spearmann	0.917	0.841	2.7E-29*
Number of INDELs discovered	Spearmann	0.899	0.807	2.4E-26*

Figures

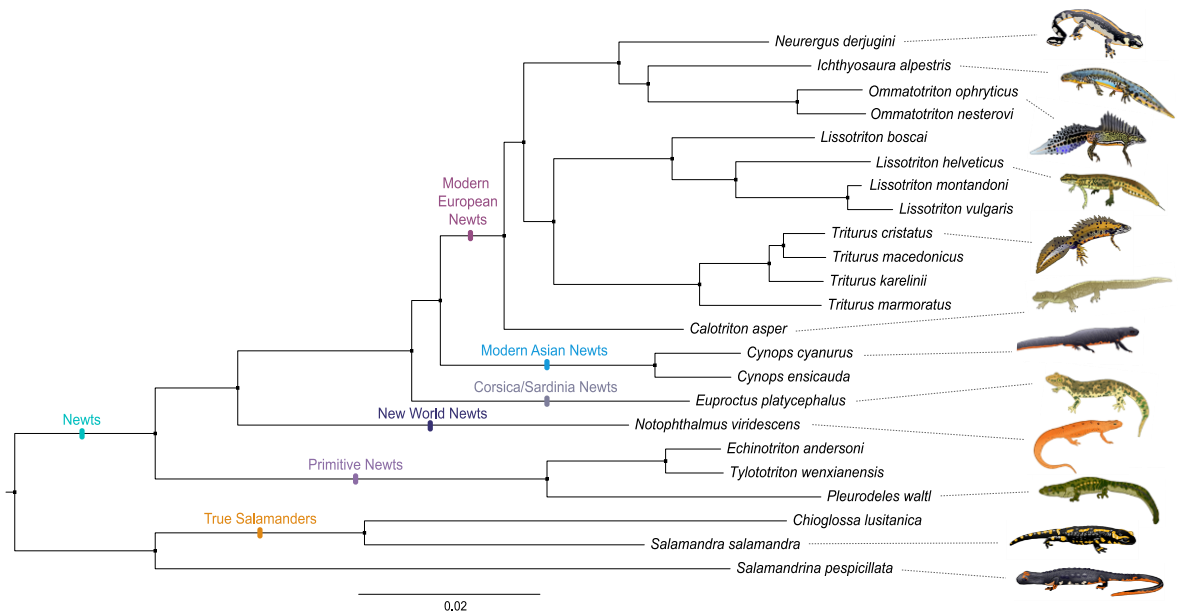

Figure 1: NewtCap-based phylogeny of the Salamandridae family. *The phylogeny is based on Maximum Likelihood inference of concatenated data of 204,600 informative SNPs using RAxML. Overall layout and clade labels conform to a previous transcriptome-based phylogeny (Rancilhac et al., 2021). The tree is rooted on the branch separating the newts and the clade containing the True Salamanders and* Salamandrina *(see also Fig. S1 for the same tree, but with original labels, and Fig. S2 for the additional, extended tree that includes* Ambstoma*,* Paradactylodon *and* Mertensiella *and confirms the root position adopted here). All nodes have a bootstrap support of 100%.*

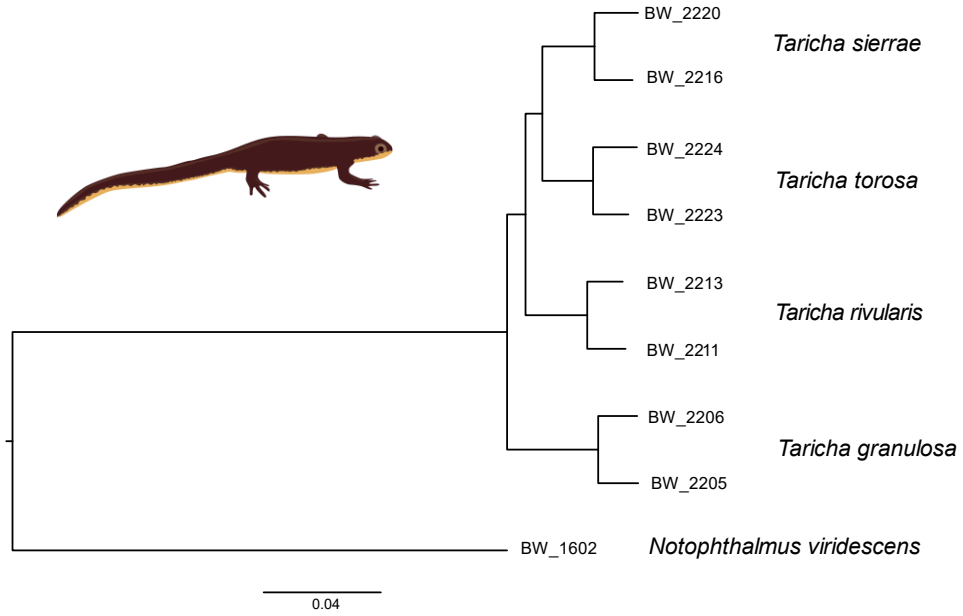

Figure 2: A* Taricha *phylogeny obtained through NewtCap-derived data. *The phylogeny is based on Maximum Likelihood inference of concatenated data of 9,730 informative SNPs using RAxML.* Notophthalmus *is used to root the tree. All nodes have a bootstrap support of 100% (not shown).*

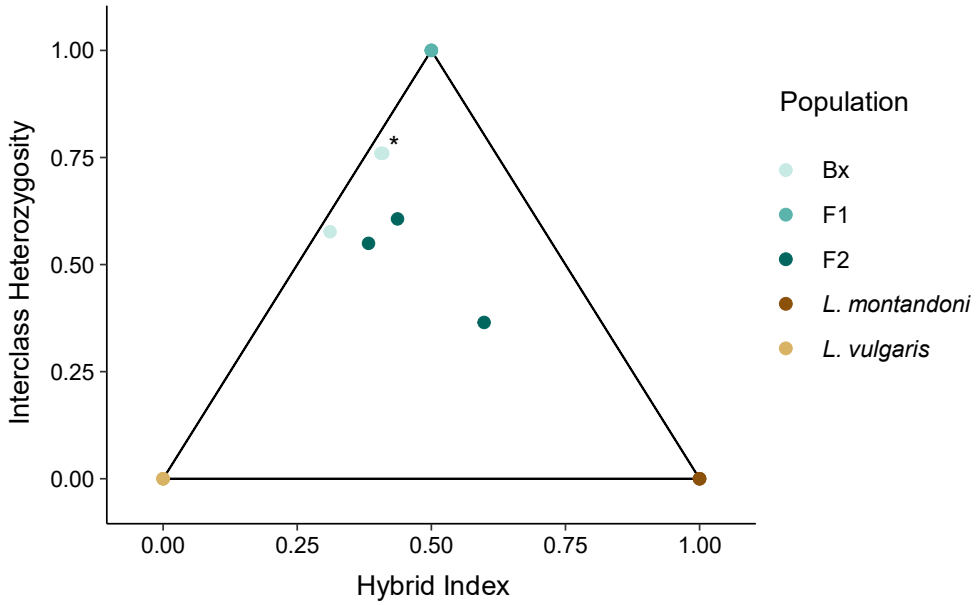

Figure 3: Triangle plot of different* Lissotriton *hybrid classes based on NewtCap-derived data. *The plot, based on 666 informative SNPs, shows the relationship between the hybrid index (the fraction of the alleles per individual that derived from each of the two parental species, also known as the ancestry) and the interclass heterozygosity (the fraction of the alleles per individual that is heterozygous for alleles from both parental species). The* L. vulgaris *individuals are in the bottom left corner, the* L. montandoni *individuals in the bottom right corner, and the F1 hybrid offspring in the top corner. The F2 and Bx ('backcross') hybrids are placed inside the triangle, with two Bx samples overlapping (marked with *).*

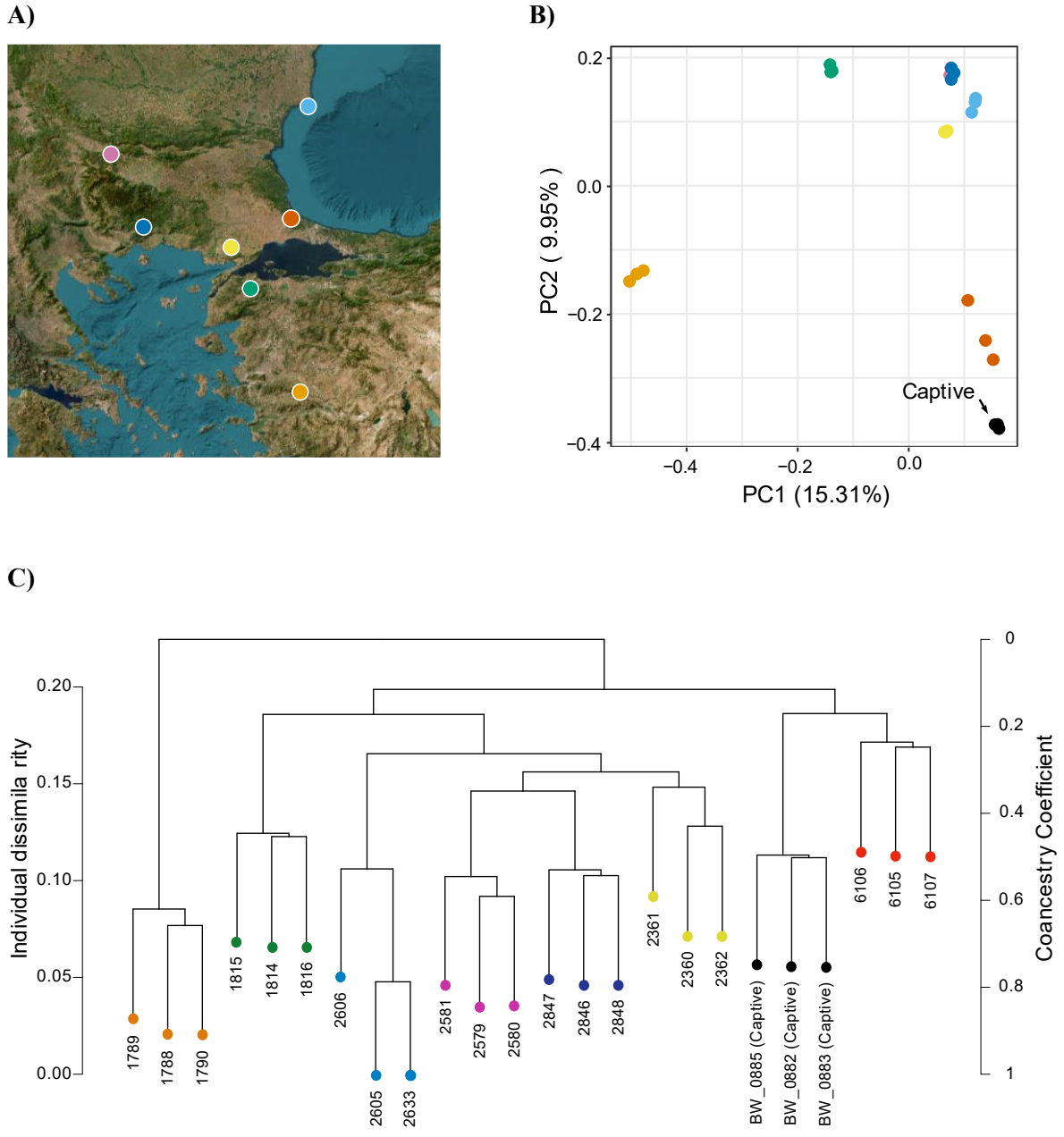

Figure 4: Genetic differentiation between wild and captive **Triturus ivanbureschi** ***populations based on NewtCap-derived data. (A)*** *The wild population localities (details in Table S1; four postglacial populations are represented by dark blue, light blue, yellow, and pink colors in Bulgaria, Greece and Turkey, and three populations from the glacial refugial area are represented by red, orange and green colors in Turkey).* ***(B)*** *A plot of the first versus the second Principal Component (PC) places the captive individuals closest to a population from just west of Istanbul.* ***(C)*** *The dendrogram produced by the HCA analysis, showing the Individual Dissimilarity as well as the Coancestry Coefficient, again shows that captive samples cluster with a population just west of Istanbul.*

Supplementary Figures

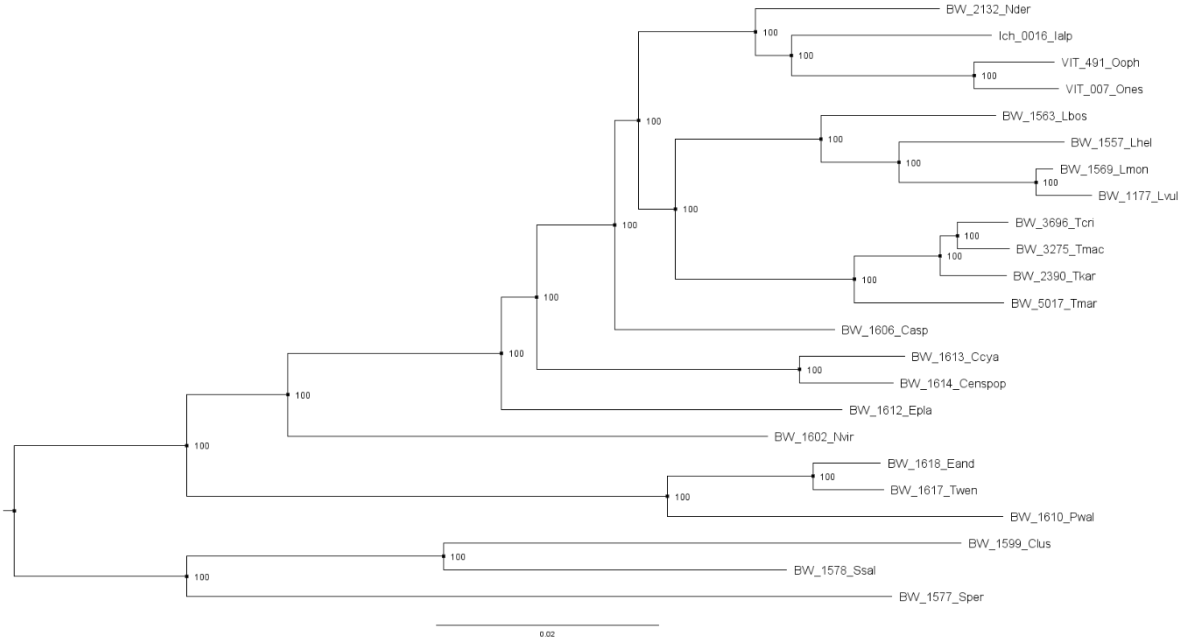

Figure S1: The raw, reconstructed NewtCap-based phylogeny of the Salamandridae family. *This is the same tree as provided in MS Fig. 1, but with original sample identifiers and bootstrap values. The tree is rooted on the branch separating the newts and the clade containing the True Salamanders and* Salamandrina *(see also Fig. S2, which confirms the root position adopted here).*

Figure S2: The raw, reconstructed NewtCap-based phylogeny of the Salamandridae family, including two non-salamandrids. *This tree is the result of an independent RAxML analysis that also includes a* Mertensiella caucasica *individual, and two non-salamandrid individuals: one* Ambystoma mexicanum *sample and one* Paradactylodon gorganensis *sample. The tree is based on 265,105 SNPs and shows sample identifiers and bootstrap values. The tree is rooted on the branch of* P. gorganensis*, which belongs to the Hynobiidae family and is more distantly related to Salamandridae than* A. mexicanum*, which belongs to the Ambystomatidae family (Marjanović and Laurin, 2013, Frost, 1985).*