**Genomics reveals broad hybridization in deeply divergent Palearctic grass and water snakes (*Natrix* spp.)**

Yannis Schöneberg, Sven Winter, Oscar Arribas, Matteo Riccardo Di Nicola, Maya Master, John Benjamin Owens,

Michail Rovatsos, Wolfgang Wüster, Axel Janke, Uwe Fritz

Molecular Phylogenetics and Evolution

Supplementary Material 2

**Examining reference bias**

## Materials and Methods

### Phylogenetic Analyses

#### Genome Fragment Analysis and Dated Phylogeny

A genome fragment analysis, as described in Coimbra et al. (2021), was performed (scripts available at https://github.com/rtfcoimbra/Coimbra-et-al-2021\_CurrBiol). Reference-based assemblies were generated for all specimens (except *T. elegans*) by mapping the respective short-read data against the assembly of *T. elegans* using bwa mem v0.7.17 and samtools v1.12 (Langmead and Salzberg, 2012; Danecek et al., 2021). Mapping rates and qualities were evaluated with Qualimap v2.2.2 (Okonechnikov et al., 2016; Table 1). Repetitive regions in the assemblies were removed based on the *N. h. helvetica* assembly with "hard-masked" interspersed repeats using bedtools v2.30.0 and samtools v1.13 (Quinlan and Hall, 2010; Danecek et al., 2021). To generate the consensus sequences we first calculated the per base sequencing depth using Sambamba v0.8.1 (Tarasov et al., 2015) and called the consensus sequence using ANGSD v0.935 (Korneliussen et al., 2014). As the reference-based assemblies had the same reference, no additional alignment step was necessary. Sets of genome fragments with a length of 20 kb to 470 kb were created with a 50-kb-step size. An approximately unbiased (AU) tree topology test was run in IQ-tree v2.1.4 to determine the minimum fragment length with which alternative topologies can be significantly rejected (Shimodaira, 2002; Minh et al., 2020). For this analysis, a Maximum Likelihood (ML) tree of the entire genome represented the best tree topology. ASTRAL v5.7.7 (Zhang et al., 2018) was used to calculate a multispecies coalescent tree from 860 individual ML trees based on 170 kb long genome fragments. The individual ML trees were inferred using IQ-tree v2.1.4 (Minh et al., 2020). Internal branches were evaluated for support of alternative groupings by a quartet analysis (Sayyari et al., 2018).

We dated this topology using MCMCtree as implemented in PAML v4.9 (Yang, 2007) based on 430 genome fragments because MCMCtree could not handle the full amount of data. Each genome fragment was randomly sampled from the 860 genome fragments and 170 kb long. The large amount of data prevented the application of a Bayesian method. Two nodes were calibrated using hard lower and upper boundaries by setting both the left and right tail probabilities to 1e-300. The oldest *Natrix* fossil from Sardinia (Pliocene, 3.6 million years; Delfino et al., 2011) was included to constrain the divergence between the European mainland lineage represented by *N. h. helvetica* and the Sardinian subspecies *N. h. cetti* with a minimum of 3.6 million years and a maximum of 5.3 million years. The maximum corresponds to the end of the Miocene, coinciding with the termination of the Mediterranean salinity crisis and the interruption of the land bridge between Corsosardinia and mainland Italy (Krijgsman et al., 1999). The oldest record of a representative of the grass snake lineage, the extinct species *N. longivertebrata* (MN4 Mammal Neogene Zone 16.9-16.0 million years; Vasilyan et al., 2022), was used to calibrate the divergence between *N. tessellata* and the grass snake clade to a minimum of 16.0 Mya (Suppl. 1 Table S2). As the maximum for this node, the base of the Burdigalian (20.4 Mya), the respective Early Miocene age to which MN4 belongs, was chosen acknowledging that different geological ages are typically characterized by faunal breaks.

#### Network analysis and gene flow

SplitsTree v4.17.1 (Huson and Bryant, 2006) produced a phylogenetic network from ML trees of 860 genome fragments to examine the genome fragments for phylogenetic conflict. Then, we used PhyloNet v3.8.2 (Wen et al., 2018) and SNAQ as implemented in PhyloNetworks v0.15.3 (Solís-Lemus et al., 2017) on the individual GF trees. In the former analysis we tested the support of 0 to 6, and in the latter 0 to 10 reticulations. The results of the latter were depicted using PhyloPlots v1.0.0. Both analyses account for Incomplete lineage sorting (ILS). To further examine whether potential phylogenetic conflict is caused by gene flow, we ran D-statistics and a HyDe analysis (Blischak et al., 2018). For D-statistics, we first generated genotype likelihoods using ANGSD v0.935 (options ‑doMaf 2 -doMajorMinor 1 -doGeno 2 -doPlink 2 -doPost 1 -GL 1) and Linkage Disequilibrium (LD) was calculated with Plink 1.90b6.21 (Purcell et al., 2007; Korneliussen et al., 2014). Then, D-Statistics, commonly known as ABBA-BABA test, were computed using ANGSD (-doCounts 1 -useLast 1 -blockSize 1500000) to examine if incomplete lineage sorting or gene flow caused conflicting signals in the genome fragments. HyDe v0.4.3 was run on the concatenated reference-based assemblies and the reference genome of *T. elegans* using the standard settings (Blischak et al., 2018). Due to issues with ANGSD not calling the consensus sequence of a contig in 2 assemblies, we had to remove this contig from all assemblies. It consisted entirely of Ns in all *Natrix* assemblies and therefore did not contain any information.

## Results

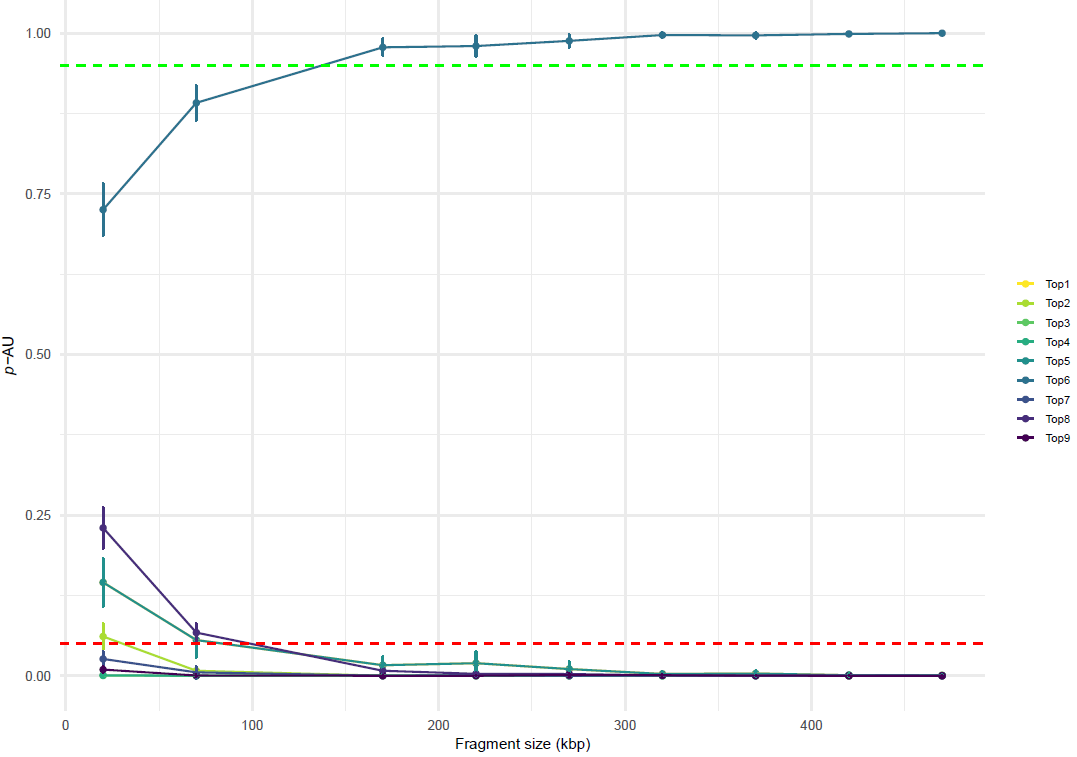
### Genome Fragment Analysis and Dated Phylogeny

The AU tree topology test for the best genome fragment length showed that 170 kbp is the shortest fragment length at which a topology can be significantly (p<0.05) differentiated from a set of alternative topologies (Fig. 1, Suppl. 1 Fig. S4). With this information, 1435 orthologous genome fragments of 170 kb length were created for each taxon. The estimated species tree shows the same topology as the one generated with the data mapped against the ingroup (Fig. 2). The dated phylogeny also resulted in divergence time estimates comparable to the ones using the data mapped against the ingroup (Fig. 2).

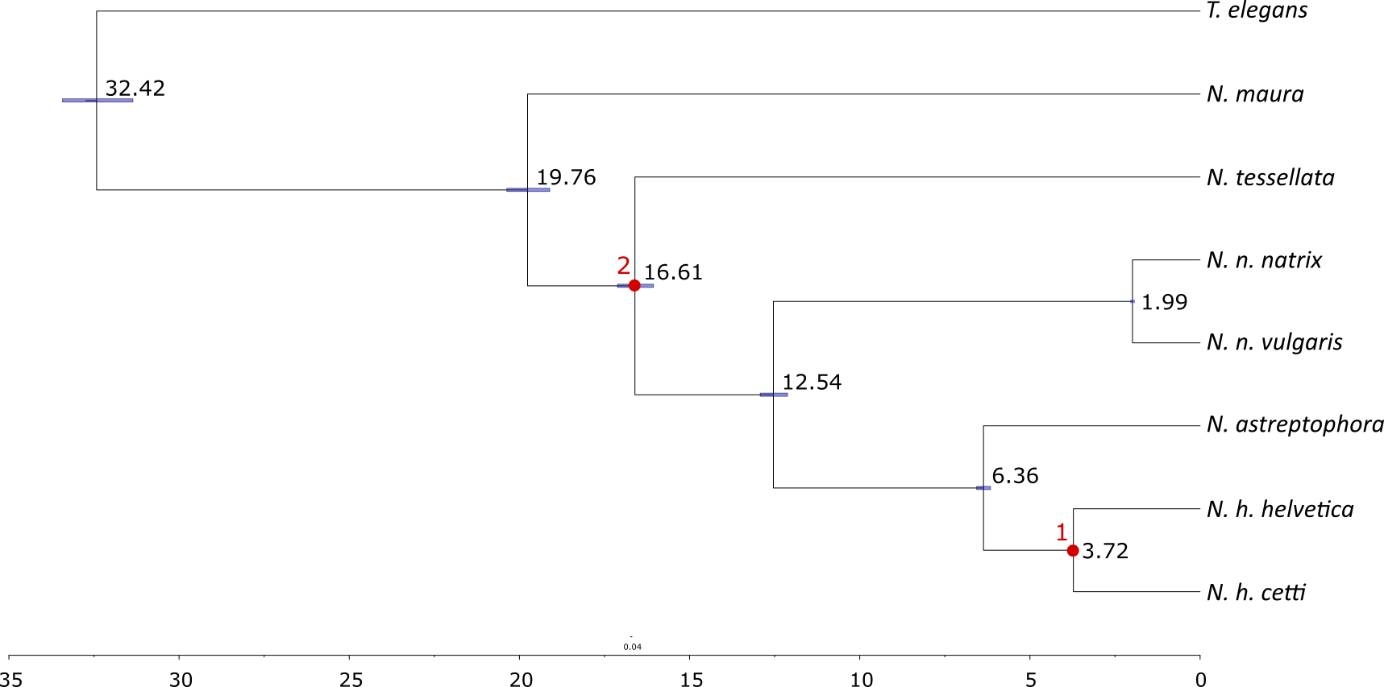
**Table 1**

Qualimap results for the data mapped against the outgroup Thamnophis elegans. All mapping statistics indicate a good mapping quality. All numbers were rounded to one decimal position.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Taxon | Coverage mean | Coverage std | GC [%] | Mapping quality mean | Insert size median [bp] | Mapped reads [%] |
| *N. astreptophora* | 27.8 | 428.7 | 41.6 | 39.3 | 340 | 91.2 |
| *N. h. helvetica* | 26.5 | 475.4 | 43.4 | 38.4 | 361 | 90.8 |
| *N. h. cetti* | 20.4 | 183.7 | 41.1 | 39.2 | 353 | 92.1 |
| *N. n. natrix* | 29.8 | 288.7 | 40.8 | 39.8 | 314 | 93.0 |
| *N. n. vulgaris* | 18.2 | 385.3 | 41.9 | 38.7 | 204 | 89.9 |
| *N. tessellata* | 21.1 | 642.3 | 40.9 | 39.8 | 306 | 92.7 |
| *N. maura* | 20.1 | 256.7 | 42.2 | 39.3 | 316 | 91.3 |
| *T. elegans* | 101.7 | 650.7 | 41.8 | 47.2 | 358 | 98.0 |



**Fig. 1.** Results of the AU test of nine different topologies for different fragment sizes be-tween 20 kb and 470 kb. This analysis tests whether one of the given topologies is supported significantly (95 % p-AU) over the others. In this case, Topology 6 was supported significantly at a fragment length of 170 kb or higher. We chose this length as fragment length for the genome fragment analysis since it was the shortest length at which one topology was significantly supported.

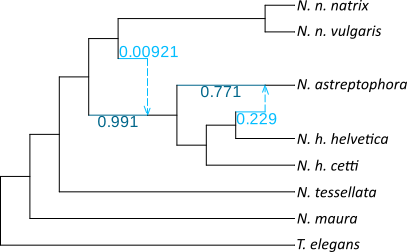


**Fig. 2.** Dated ML phylogeny using the sequence data from 430 genome fragments of 170 kb length in a MCMCTree analysis. The calibration point (1) for the split between N. h. helvetica and N. h. cetti was set to a minimum of 3.6 million years and a maximum of 5.3 million years. The second calibration point (2) for the divergence of the grass snake clade and N. tessellata was set to a minimum of 16.0 Mya and a maximum of 20.4 Mya. The fossil calibration points are marked with red dots. The error bars show the 95% confidence interval. The time bar is in million years. Values were rounded to the second decimal position.

### Gene Flow Analyses

The SNAQ-Analysis supports two reticulations, in contrast to three with the data mapped against the ingroup (i.e. between the ancestor of *N. n. natrix* + *N: n. vulgaris* and the ancestor of the remaining grass snake species; and between *N. h. helvetica* and *N. astreptophora*;Fig. 3; Table 2). PhyloNet did not support any reticulations (Table 3).

The HyDe analysis identified more species combinations with gene flow than the analysis using the ingroup mapped data. The only taxon very limited signal for hybridization was the island taxon *N. h. cetti*. Hybridization was only supported between it and *N. n. vulgaris* (Fig. 4; Table 4).



**Fig. 3.** The best-supported result from SNAQ set to have three reticulations. Analyses with a higher number of reticulations did not improve the log-likelihood. The reticulations are shown as alternative placements of the respective taxon.

Ein Bild, das Vieleck enthält.

Automatisch generierte Beschreibung

**Fig. 4.** Results of the HyDe analysis with the data mapped against the outgroup, indicating widespread hybridization among Natrix species. Red lines indicate significant evidence for hybridization between the connected taxa. Grey lines indicate species combinations where testing for hybridization was not possible due to the topology. Dotted lines link allopatric taxa; solid lines, sympatric taxa or parapatric taxa with abutting ranges with or without hybrid zones. The phylogeny is shown as underlying black tree.

**Table 2**

The pseudo-likelihood support values for the different number of reticulations tested in the SNAQ analysis. Adding more reticulations than three did not improve the support. Therefore, three reticulations were supported.

|  |  |
| --- | --- |
| # reticulations | -Ploglik |
| 0 | 60.0990298 |
| 1 | 14.8811329 |
| 2 | 12.0917635 |
| 3 | 12.5929193 |
| 4 | 12.8182492 |
| 5 | 12.6334497 |

**Table 3**

The pseudo-likelihood support values for the different number of reticulations tested in the PhyloNet analysis. Adding more reticulations than three did not improve the support. No reticulations are supported.

|  |  |
| --- | --- |
| # reticulations | -Log Lik |
| 0 | -11372.9854 |
| 1 | -11232.7673 |
| 2 | -11242.1047 |
| 3 | -11372.9865 |
| 4 | -11373.0055 |
| 5 | -11242.5167 |
| 6 | -11232.6771 |
| 7 | -11232.7062 |
| 8 | -11232.8301 |
| 9 | -11232.7473 |
| 10 | -11242.5055 |

**Table 4**

HyDe results using the data mapped against the outgroup. Only significant results are shown. This dataset identified more gene flow than the analysis using the data mapped to the ingroup. According to this analysis, the only taxon with limited signal for hybridization is the island subspecies N. h. cetti. Abbreviations: ast: N. astreptophora; hel: N. h. helvetica; hce: N. h. cetti; nat: N. n. natrix; nvu: N. n. vulgaris; mau: N. maura; tes: N. tessellata.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P1 | Hybrid | P2 | Z-score | Pvalue | Gamma | AAAA | AAAB | AABA | AABB | AABC | ABAA | ABAB | ABAC | ABBA | BAAA | ABBC | CABC | BACA | BCAA | ABCD |
| *ast* | *hel* | *nat* | 825.321971 | 0 | 0.66061555 | 181492091 | 27028226.4 | 16190229 | 18276540.3 | 13035342.9 | 23074778.2 | 13950127.1 | 13661364.8 | 22371537.5 | 403699480 | 43604251.2 | 13454199.5 | 26542419.5 | 35812845.5 | 12279699.54 |
| *hel* | *ast* | *tes* | 345.825501 | 0 | 0.60623507 | 186852087 | 20108345 | 17189036.8 | 15296639.3 | 13521064.6 | 17101941.1 | 13609315.5 | 12111131.3 | 16207096 | 416923440 | 31864835.9 | 12034200.7 | 26591834.3 | 28618198.2 | 11485440.63 |
| *ast* | *hel* | *hce* | 714.822328 | 0 | 0.74634017 | 177377135 | 31639821.2 | 16731551.8 | 18320956.9 | 13939416.9 | 22936269.1 | 14563773.9 | 14676075.5 | 25618486.5 | 395529420 | 50017251.1 | 14412074.3 | 27631328.4 | 35874147.5 | 13151246.04 |
| *ast* | *hel* | *mau* | 1471.3422 | 0 | 0.64650671 | 175665743 | 31050342.8 | 10990380.2 | 17775264.2 | 13323962.1 | 21489348.8 | 10120717.9 | 13663770.5 | 24120179 | 392263317 | 47878904.2 | 12374185.3 | 18595196.3 | 34281055.9 | 12013612.63 |
| *ast* | *nvu* | *hel* | 826.436096 | 0 | 0.3296043 | 217678283 | 56466478.1 | 33978371.1 | 54128620.8 | 66315154.8 | 66827117.8 | 40026162.7 | 78285893.9 | 46959725.6 | 490852350 | 93987609.9 | 66307436.4 | 68206153.9 | 108306909 | 66325285.25 |
| *nat* | *ast* | *tes* | 411.709382 | 0 | 0.17541591 | 177242248 | 18674029.4 | 16326743.8 | 22661145.9 | 14614591.4 | 26505620.1 | 13119640 | 13416659.7 | 15149429.2 | 394171895 | 29572427 | 13335715.3 | 25431407.8 | 43386788 | 12607874.33 |
| *nat* | *ast* | *hce* | 218.221447 | 0 | 0.75955468 | 169453349 | 28112135 | 23256716.4 | 20120729.4 | 15172495.5 | 24041868.1 | 18901235.8 | 15046447.7 | 22753554.8 | 376779634 | 44308297.8 | 15116142.9 | 36678950 | 38983734.3 | 13878169.25 |
| *nat* | *ast* | *mau* | 687.205229 | 0 | 0.71413391 | 164233088 | 32033515.7 | 19114875.5 | 19607724.9 | 15930805.7 | 22811549.2 | 15884356.1 | 15397041.3 | 25185857.9 | 365114993 | 49643944.5 | 14848392.3 | 30488996.7 | 37422055.2 | 14104357.33 |
| *ast* | *nvu* | *nat* | 1312.50988 | 0 | 0.55001812 | 206555716 | 67578442.9 | 34252238.5 | 52944521.8 | 67193535.1 | 65345165.3 | 40383696.4 | 79378522.3 | 55736940.2 | 464483635 | 111547242 | 67194748 | 68701911 | 105975407 | 67210018.63 |
| *tes* | *ast* | *hce* | 455.268774 | 0 | 0.84453754 | 173342964 | 30859992.4 | 16333801.7 | 15567566.3 | 14275634.5 | 18958133.7 | 13278566.7 | 14338816.3 | 25713376.6 | 386412430 | 49459273.2 | 15496907.8 | 25679302.2 | 30321596.3 | 13458782.29 |
| *tes* | *ast* | *mau* | 305.679779 | 0 | 0.90869268 | 169296221 | 32756814.5 | 13961601.4 | 13157988.4 | 12607906.3 | 15703265 | 11695626.8 | 12727047.5 | 26249084.3 | 377516435 | 51369782.5 | 12803085.1 | 22291920.9 | 25334835.2 | 11627647.62 |
| *ast* | *nvu* | *tes* | 374.710238 | 0 | 0.7139742 | 212072619 | 62053475.2 | 36558929.6 | 46498236.3 | 71308574.8 | 56951896.8 | 43356660.4 | 84774629.8 | 51198624.5 | 478128598 | 102416048 | 71311372.2 | 73385817.7 | 93016461.6 | 71319325.38 |
| *hce* | *ast* | *mau* | 1052.98365 | 0 | 0.59459895 | 160625423 | 32011547.5 | 19203196 | 22466975.1 | 16881478.3 | 26859831.7 | 16156465.8 | 16468712.4 | 25412046.6 | 357881377 | 50015788.6 | 15855635 | 30887967.2 | 43118218.1 | 15026374.88 |
| *ast* | *nvu* | *hce* | 1444.4579 | 0 | 0.56802716 | 202797283 | 71337424.4 | 33940941.3 | 53765788.7 | 66685230.5 | 66093863.9 | 40089035.5 | 78926130 | 58073423.7 | 457471448 | 116224616 | 66687674.8 | 68053023.4 | 107641550 | 66703171.63 |
| *ast* | *nvu* | *mau* | 729.180151 | 0 | 0.76811215 | 195923605 | 78207663.2 | 35659459.4 | 48669826.5 | 70051997.9 | 59420371.9 | 42351684.9 | 83326657.5 | 63280083.2 | 441872997 | 126605719 | 70063132.8 | 71535771.4 | 97393144.5 | 70072009.33 |
| *nat* | *hel* | *tes* | 1087.1995 | 0 | 0.44731966 | 179267829 | 24794835.2 | 15562998.5 | 21260634.4 | 13679728.1 | 25828985.1 | 13407576.7 | 13890574.6 | 19763560.1 | 397677571 | 38835127.9 | 13183812.9 | 25423473.5 | 41353672.3 | 12475568.71 |
| *nat* | *hel* | *hce* | 879.485022 | 0 | 0.62168895 | 174922409 | 30131567.2 | 18391603.1 | 20743850 | 14341764 | 25381479.9 | 15766611.6 | 14952960.4 | 23945844.3 | 388379559 | 46924836.9 | 14396125.7 | 30025577.5 | 40572611.8 | 13390028.21 |
| *nat* | *hel* | *mau* | 1829.79277 | 0 | 0.54497755 | 171995893 | 31250905.9 | 11447702.8 | 21693319.6 | 14914766 | 25643293.8 | 10315405.4 | 14720884.3 | 23942662.3 | 382363517 | 47522340.8 | 13546838 | 19096315.1 | 41715059.7 | 13183045.79 |
| *hel* | *nvu* | *nat* | 975.026506 | 0 | 0.68843478 | 214045214 | 70455404.8 | 33653176.7 | 47707689.6 | 65870517.5 | 57861490.2 | 39471831.6 | 77430887.5 | 57669790.3 | 480194083 | 115409842 | 65878663.8 | 67517954 | 95519495.8 | 65889793.21 |
| *tes* | *hel* | *hce* | 1095.66698 | 0 | 0.63994989 | 175001640 | 30625824.3 | 16012544.6 | 19973016.5 | 14116688.9 | 24841580.4 | 13927221.9 | 14916100 | 24672963.9 | 389118037 | 48092852.8 | 14666093.9 | 26337421.3 | 39222939.1 | 13371255.63 |
| *tes* | *hel* | *mau* | 1632.88845 | 0 | 0.58523729 | 173638944 | 29512609.7 | 10630260.2 | 18942659.7 | 13101729.9 | 22870878 | 9825823.06 | 13560663.1 | 22689835.6 | 386752093 | 45013978.5 | 12228846.5 | 17960321.4 | 36689084 | 11864108.58 |
| *hel* | *nvu* | *tes* | 574.745952 | 0 | 0.77561404 | 214257890 | 70240843.9 | 34472754.9 | 45387844 | 67364947.7 | 54778626.4 | 40556397.1 | 79423518.6 | 57256809.3 | 481447850 | 114563232 | 67375785.4 | 69189467.2 | 90844026 | 67384957 |
| *hce* | *hel* | *mau* | 1926.41871 | 0 | 0.48232717 | 168018172 | 31000873.1 | 11987800.5 | 24905695.2 | 15985559.8 | 30143289.3 | 10940060.1 | 15913221.6 | 23952149.5 | 374362855 | 47526244.8 | 14564139.8 | 20154938.4 | 48070200.8 | 14183605.29 |
| *hel* | *nvu* | *hce* | 1160.23885 | 0 | 0.68324544 | 209742308 | 74760220.5 | 33207868.3 | 48907353.3 | 65121929.3 | 59156023 | 38987259.3 | 76626650.3 | 60385085.8 | 472048651 | 120845761 | 65130898.5 | 66599470.7 | 97944594.5 | 65142490.96 |
| *mau* | *hel* | *nvu* | 150.305197 | 0 | 0.05777126 | 206168231 | 35916214.7 | 49177883.3 | 62897500.3 | 70195588.3 | 78325815.4 | 41119410.4 | 70175676.1 | 42454699.3 | 464551483 | 72102137 | 83111919.1 | 82271709.8 | 125804647 | 70196317.25 |
| *tes* | *nat* | *hce* | 192.66044 | 0 | 0.73001962 | 167312371 | 27326011.4 | 22978722.1 | 20209085.6 | 15011144.3 | 24817886.9 | 19139463.8 | 15200474.6 | 22031692.2 | 371068426 | 42831164.9 | 15161755.8 | 37005502.8 | 39342882 | 13890020.83 |
| *nat* | *tes* | *mau* | 418.759445 | 0 | 0.76282858 | 162563160 | 30666804.9 | 20095457.8 | 18923279.3 | 15913882.6 | 22050475.1 | 16662407.4 | 15260840.8 | 23934184.5 | 360643281 | 47128435 | 14793117.1 | 32114546.3 | 36004839.9 | 14037101.08 |
| *nat* | *nvu* | *tes* | 1291.10051 | 0 | 0.53498003 | 203559968 | 68332541.7 | 34892318.6 | 53745160.7 | 68521341.6 | 65464244.5 | 41062331.1 | 80788979.3 | 55653232.7 | 456333358 | 111391855 | 68522821.7 | 69977541 | 107564216 | 68537910.88 |
| *hce* | *nat* | *mau* | 456.395448 | 0 | 0.75650986 | 158168149 | 33572687 | 21549501.9 | 20944935.2 | 16509435 | 25509431.7 | 18358265 | 16629878.5 | 26394900.5 | 351123741 | 51933369.5 | 15890281 | 35110833.5 | 40408518.5 | 14991421.25 |
| *nat* | *nvu* | *hce* | 1471.57325 | 0 | 0.52419574 | 201608879 | 70284270.7 | 34335917 | 55312654.4 | 67512270 | 67273989.6 | 40415787 | 79627699.9 | 56827734.2 | 452813031 | 113739712 | 67513644.8 | 68835611.7 | 110728343 | 67528946.04 |
| *nat* | *nvu* | *mau* | 1140.10023 | 0 | 0.68523084 | 192430632 | 79460022.9 | 35448388.3 | 51898510.1 | 69806258.3 | 62906005.5 | 41856696.8 | 82547306.3 | 63717034.7 | 432146062 | 127509678 | 69815497.2 | 71079271.7 | 103869613 | 69826791.21 |
| *hce* | *tes* | *mau* | 817.540852 | 0 | 0.58964542 | 158839004 | 30535385.7 | 20351363 | 21939248.4 | 16970642.4 | 26266323.2 | 17090804.3 | 16438941.6 | 24057615.6 | 353103710 | 47301849.2 | 15888825.2 | 32816536.2 | 42003644 | 15055178.75 |
| *tes* | *nvu* | *hce* | 1483.48167 | 0 | 0.49377316 | 199588750 | 69436656 | 34517427.6 | 56661319.4 | 67893242 | 69294670.2 | 40695452.3 | 80197307.5 | 56268543.5 | 448762389 | 112615429 | 67892914.8 | 69201102.6 | 113431916 | 67909957.04 |
| *tes* | *nvu* | *mau* | 890.729714 | 0 | 0.69453127 | 193390694 | 75629720.8 | 36441409.7 | 50972238.6 | 71643041.6 | 61944005.1 | 43191574.8 | 85036883.9 | 60882139.8 | 434939629 | 121809278 | 71651336.5 | 73096040.9 | 101996323 | 71661579.96 |

## Discussion

The analyses using the data mapped to the outgroup differed compared to those using the *N. h. helvetica* genome as a reference. SNAQ supported only two reticulations in contrast to three with the ingroup as reference. However, we do not see this as evidence for ingroup effects. We used much shorter genome fragments than for the ingroup data, which resulted in far more trees. In general, phylogenetic analyses should not be affected by mapping bias, preferring one haplotype over the other (Günther and Nettelblad, 2019). Keeping in mind that SNAQ only analyses the GF trees, we put more confidence in the sequence-based analysis HyDe. This analysis identified way more hybridization using the outgroup as reference than with the ingroup as reference. We interpret this as reference bias as a result of using a rather distantly related outgroup (Prasad et al., 2022). There is the risk of experiencing mapping bias when using an ingroup as reference for suboptimal data (Günther and Nettelblad, 2019). However, we used high-quality DNA and accounted for these effects during generation of the reference-based assemblies. Therefore, mapping bias should be negligible compared to the bias introduced by mapping to a rather distantly related outgroup.

## References

Blischak, P.D., Chifman, J., Wolfe, A.D., Kubatko, L.S., 2018. HyDe: A Python package for genome-scale hybridization detection. Syst. Biol. 67, 821–829. https://doi.org/10.1093/sysbio/syy023

Coimbra, R.T.F., Winter, S., Kumar, V., Koepfli, K.-P., Gooley, R.M., Dobrynin, P., Fennessy, J., Janke, A., 2021. Whole-genome analysis of giraffe supports four distinct species. Curr. Biol. 31, 2929–2938.e5. <https://doi.org/10.1016/j.cub.2021.04.033>

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, giab008. <https://doi.org/10.1093/gigascience/giab008>

Delfino, M., Bailon, S., Pitruzzella, G., 2011. The Late Pliocene amphibians and reptiles from "Capo Mannu D1 Local Fauna" (Mandriola, Sardinia, Italy). Geodiversitas 33, 357–382. <https://doi.org/10.5252/g2011n2a10>

Günther, T., Nettelblad, C., 2019. The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genet. 15, e1008302. https://doi.org/10.1371/journal.pgen.1008302

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. <https://doi.org/10.1093/molbev/msj030>

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinform. 15, 356. https://doi.org/10.1186/s12859-014-0356-4

Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J., Wilson, D.S., 1999. Chronology, causes and progression of the Messinian salinity crisis. Nature 400, 652–655. https://doi.org/10.1038/23231

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. <https://doi.org/10.1038/nmeth.1923>

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Haeseler, A. von, Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32, 292–294. <https://doi.org/10.1093/bioinformatics/btv566>

Prasad, A., Lorenzen, E.D., Westbury, M.V., 2022. Evaluating the role of reference-genome phylogenetic distance on evolutionary inference. Mol. Ecol. Res. 22, 45–55. https://doi.org/10.1111/1755-0998.13457

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., Bakker, P.I.W. de, Daly, M.J., Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795

Quinlan, A.R., Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>

Sayyari, E., Whitfield, J.B., Mirarab, S., 2018. DiscoVista: Interpretable visualizations of gene tree discordance. Mol. Phylogenet. Evol. 122, 110–115. https://doi.org/10.1016/j.ympev.2018.01.019

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508. <https://doi.org/10.1080/10635150290069913>

Solís-Lemus, C., Bastide, P., Ané, C., 2017. PhyloNetworks: A package for phylogenetic networks. Mol. Biol. Evol. 34, 3292–3298. https://doi.org/10.1093/molbev/msx235

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., Prins, P., 2015. Sambamba: Fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Vasilyan, D., Čerňanský, A., Szyndlar, Z., Mörs, T., 2022. Amphibian and reptilian fauna from the early Miocene of Echzell, Germany. Foss. Rec. 25, 99–145. <https://doi.org/10.3897/fr.25.83781>

Wen, D., Yu, Y., Zhu, J., Nakhleh, L., 2018. Inferring phylogenetic networks using PhyloNet. Syst. Biol. 67, 735–740. https://doi.org/10.1093/sysbio/syy015

Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591. https://doi.org/10.1093/molbev/msm088

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19 (Suppl. 6), 153. https://doi.org/10.1186/s12859-018-2129-y