Lack of signal for the impact of venom gene diversity on speciation Mark A

Mark A Phuong, Michael E Alfaro, Gusti N Mahardika, Ristiyanti M Marwoto, Romanus Edy Prabowo, Thomas Von Rintelen, Philipp W. H. Vogt, Jonathan R Hendricks, Nicolas Puillandre

▶ **To cite this version:**

HAL Id: hal-02343430
https://hal.sorbonne-universite.fr/hal-02343430v1

Submitted on 2 Nov 2019

1   **Lack of signal for the impact of venom gene diversity on speciation**
2   **rates in cone snails**

3

4   Mark A Phuong[1], Michael E Alfaro[1], Gusti N Mahardika[2], Ristiyanti M Marwoto[3], Romanus

5   Edy Prabowo[4], Thomas von Rintelen[5], Philipp WH Vogt[5], Jonathan R Hendricks[6], Nicolas

6   Puillandre[7]

7

8   [1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA

9   90095, USA

10 [2]Animal Biomedical and Molecular Biology Laboratory, Faculty of Veterinary Medicine,

11 Udayana University Bali, Jl Sesetan-Markisa 6, Denpasar, Bali 80225, Indonesia

12 [3]Zoology Division (Museum Zoologicum Bogoriense), Research Center for Biology, LIPI,

13 Cibinong, Bogor, Indonesia

14 [4]Faculty of Biology, Jenderal Soedirman University, Purwokerto, Indonesia

15 [5]Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Berlin,

16 Germany

17 [6]Paleontological Research Institution, Ithica, NY 14850, USA

18 [7]Institut Systématique Evolution Biodiversité (ISYEB), Muséum national d'Histoire naturelle,

19 CNRS, Sorbonne Université, EPHE, 57 rue Cuvier, CP 26, 75005 Paris, France

20

21 **Corresponding author:**

22 M. A. Phuong

23 E-mail: markphuong@gmail.com

24

25 **Abstract**

26 Understanding why some groups of organisms are more diverse than others is a central goal in

27 macroevolution. Evolvability, or lineages' intrinsic capacity for evolutionary change, is thought

28 to influence disparities in species diversity across taxa. Over macroevolutionary time scales,

29 clades that exhibit high evolvability are expected to have higher speciation rates. Cone snails

30 (family: Conidae, >900 spp.) provide a unique opportunity to test this prediction because their

31 venom genes can be used to characterize differences in evolvability between clades. Cone snails

32 are carnivorous, use prey-specific venom (conotoxins) to capture prey, and the genes that encode

33 venom are known and diversify through gene duplication. Theory predicts that higher gene

34 diversity confers a greater potential to generate novel phenotypes for specialization and

35 adaptation. Therefore, if conotoxin gene diversity gives rise to varying levels of evolvability,

36 conotoxin gene diversity should be coupled with macroevolutionary speciation rates. We applied

37 exon capture techniques to recover phylogenetic markers and conotoxin loci across 314 species,

38 the largest venom discovery effort in a single study. We paired a reconstructed timetree using 12

39 fossil calibrations with species-specific estimates of conotoxin gene diversity and used trait-

40 dependent diversification methods to test the impact of evolvability on diversification patterns.

41 Surprisingly, did not detect any signal for the relationship between conotoxin gene diversity and

42 speciation rates, suggesting that venom evolution may not be the rate-limiting factor controlling

43 diversification dynamics in Conidae. Comparative analyses showed some signal for the impact

44 of diet and larval dispersal strategy on diversification patterns, though whether or not we

45 detected a signal depended on the dataset and the method. If our results remain true with

46 increased sampling in future studies, they suggest that the rapid evolution of Conidae venom

47   may cause other factors to become more critical to diversification, such as ecological opportunity

48   or traits that promote isolation among lineages.

49

50   **Introduction**

51   Why are some taxa more diverse than others? Species richness and phenotypic diversity are not

52   distributed evenly across the tree of life (Rabosky *et al.* 2013). For example, there exists over

53   10,000 species of birds, but their closest relatives (crocodiles and alligators) comprise only of 23

54   species. Differences in evolvability, or lineages' intrinsic capacity to adapt and diversify, is one

55   reason commonly used to explain these disparities (Wagner & Altenberg 1996; Yang 2001;

56   Jones *et al.* 2007; Pigliucci 2008; Losos 2010). Evolvability is thought to be determined by the

57   underlying genetic architecture of organisms – some genomes of organisms have a greater

58   propensity to generate variation that may be adaptive in the future (Wagner & Altenberg 1996;

59   Jones *et al.* 2007; Pigliucci 2008). For example, gene duplication increases evolvability – the

60   copied gene is free from the selective pressures of the original gene (Crow & Wagner 2006).

61   Mutation, selection, and drift can act on the copied gene, facilitating the possibility of new

62   phenotypes to arise; this shapes the extent that taxa can diversify and exploit resources (Crow &

63   Wagner 2006). Over long evolutionary time scales, clades that exhibit higher evolvability are

64   predicted to have increased species richness and diversification rates (Yang 2001).

65   Despite the ubiquity of this concept in macroevolutionary theory, few studies explicitly

66   test these predictions; this is possibly due to the difficulty of identifying genes responsible for

67   phenotype (Hoekstra & Coyne 2007). Past studies that have attempted to test the impact of

68   evolvability on diversification have produced mixed results (Santini *et al.* 2009; Soltis *et al.*

69   2009; Mayrose *et al.* 2011; Rabosky *et al.* 2013; Zhan *et al.* 2014; Tank *et al.* 2015; Malmstrøm

70   *et al.* 2016). For example, whole genome duplication events, which are hypothesized to increase

71   the genomic potential of organisms, have been documented to increase (Santini *et al.* 2009;

72   Soltis *et al.* 2009; Tank *et al.* 2015), decrease (Mayrose *et al.* 2011), and have no impact (Zhan *et*

73   *al.* 2014) on the long-term evolutionary success of clades. In another case, a positive correlation

74   between evolvability and speciation rates exist when measuring evolvability through

75   morphological proxies (Rabosky *et al.* 2013). One limitation of past research on this hypothesis

76   is the inability to tie genomic changes with ecological factors driving diversification patterns

77   (Robertson *et al.* 2017). Although gene duplication and whole genome duplication events can

78   increase the evolutionary capacity of organisms, genes that are ecologically relevant for

79   adaptation may not be readily available for selection to drive divergence.

80   Here, we study the relationship between evolvability and diversification in cone snails

81   (family, Conidae), a diverse group (> 900 spp.) of predatory marine gastropods. These snails

82   feed on either worms, molluscs, or fish by paralyzing their prey with a cocktail of venomous

83   neurotoxins (conotoxins, Duda & Palumbi 1999). Cone snail provides a unique opportunity to

84   test predictions of evolvability and diversification for the following reasons: first, cone snail

85   species share an ecologically relevant trait, venom. Conidae species are globally distributed in

86   tropical and subtropical regions, where >30 species can co-occur within the same habitat (Kohn

87   2001). High numbers of species hypothesized to be able to co-occur because species have

88   diversified to specialize on different prey using prey-specific conotoxins (Duda & Palumbi

89   1999). Second, venom genes are known and diversify through gene duplication (Duda &

90   Palumbi 2000; Kaas *et al.* 2010, 2012; Chang & Duda 2012). Diet specialization is thought to be

91   enabled by the rapid evolution of the genes that underlie conotoxins – estimated rates of gene

92   duplication and nonsynonymous substitutions rates for conotoxin genes are the highest across

93   metazoans (Duda & Palumbi 2000; Chang & Duda 2012). Therefore, conotoxin genes provide a

94   natural way to characterize differences in evolvability between clades.

95   We employ a sequence capture technique previously used in cone snails (Phuong &

96   Mahardika 2017) to recover phylogenetic markers and conotoxin genes from 314 described

97   species. We use the phylogenetic markers to reconstruct a time-calibrated phylogeny and

98   perform trait-dependent diversification analyses to test the impact of evolvability on

99   diversification patterns. We predict that clades with a greater number of conotoxin gene copies

100   should have higher speciation rates. In addition, we test other traits that may have an impact on

101   diversification patterns, including diet and larval dispersal strategy.

102

103   **Methods**

104

105   *Bait design*

106   We used a targeted sequencing approach to recover markers for phylogenetic inference and

107   obtain an estimate of conotoxin gene diversity from Conidae species. For the phylogenetic

108   markers, we identified loci using a previous Conidae targeted sequencing dataset (Phuong &

109   Mahardika 2017) and the Conidae transcriptome data from (Phuong *et al.* 2016). In the Conidae

110   targeted sequencing dataset, the authors generated a phylogeny using 5883 loci across 32 species

111   (Phuong & Mahardika 2017). For our sequencing experiment, we only retained loci that were

112   >180bp and were present in at least 26 out of 32 taxa with at least 10X coverage. We chose to

113   only include longer loci to increase confidence in identifying orthologous fragments in other

114   Conidae species. To identify additional phylogenetic markers from the transcriptome data

115   (Phuong *et al.* 2016), which consisted of venom duct transcriptomes from 12 species, we

116   performed the following:

117   (1) identified reciprocal best blast hits between the assembled transcriptome and the

118   *Lottia gigantea* protein reference (Simakov *et al.* 2013) using BLAST+ v2.2.31 (evalue =

119   1e-10). We also considered fragments that had their best hit to the protein reference, but

120   to a non-overlapping portion (<20% overlapping).

121   (2) mapped reads using bowtie2 v2.2.7 (Langmead & Salzberg 2012)

122   (3) removed duplicates using picard-tools v.2.1.1 (http://broadinstitute.github.io/picard)

123   (4) fixed assembly errors by calling single nucleotide polymorphisms (SNPs) using

124   samtools v1.3 and bcftools v1.3 (Li *et al.* 2009)

125   (5) aligned sequences per locus using mafft v7.222 (Katoh *et al.* 2005)

126   (6) calculated uncorrected pairwise distances within each locus for all possible pairwise

127   comparisons

128   (7) removed sequences if the uncorrected pairwise distance was greater than the 90th

129   percentile for those pair of species

130   (8) denoted exon boundaries by comparing the transcriptome sequences to the *Lottia*

131   *gigantea* genome reference (Simakov *et al.* 2013), retaining exons >180bp

132   For all retained phylogenetic markers, we also performed the following: (1) we generated an

133   ancestral sequence using FastML v3.1 (Ashkenazy *et al.* 2012) between a *Californiconus*

134   *californicus* sequence and another Conidae sequence that had the highest amount of overlap with

135   the *C. californicus* sequence (we generated these ancestral sequences to decrease the genetic

136   distances between the target sequence and the orthologous sequence from any Conidae species),

137   (2) removed sequences that had a GC content < 30% or > 70% because extreme GC contents can

138   reduce capture efficiency (Bi *et al.* 2012), (3) removed loci that contained repeats identified

139   through the RepeatMasker v4.0.6 web server (Smit *et al.* 2015), and (4) performed a self-blast

140   with the target sequences via blastn v2.2.31 (evalue = 1e-10) and removed loci that did not blast

141   to itself with sequence identity >90%. The final set of target loci for phylogenetic inference

142   included 1749 loci, with a total length of 470,435 bp.

143   To recover conotoxin loci, we targeted sequences generated from both the previous

144   targeted sequencing dataset (Phuong & Mahardika 2017) and the transcriptome dataset (Phuong

145   *et al.* 2016). For conotoxin sequences discovered from the targeted sequencing dataset (Phuong

146   & Mahardika 2017), we performed the following to generate our target sequences: (1) we

147   trimmed each sequence to only retain the coding region and included 100bp flanking the exon,

148   (2) merged sequences using cd-hit v4.6.4 (Li & Godzik 2006) at 95% sequence similarity to

149   reduce redundancy among conotoxin loci (3) masked repeats using the RepeatMasker v4.0.6 web

150   server (Smit *et al.* 2015), and (4) retained loci >120bp to ensure that the locus was longer than

151   our desired bait sequence length. We concatenated all sequences below 120bp to create a single,

152   chimeric sequence for capture. The final set of target sequences from the previous targeted

153   sequencing dataset consisted of 12,652 unique loci totaling 3,113,904 bp and a single

154   concatenated sequence representing 351 merged loci with a total length of 37,936 bp. We also

155   targeted conotoxin loci from the transcriptomes described in (Phuong *et al.* 2016) to obtain

156   conotoxin loci from gene superfamilies that were not targeted in (Phuong & Mahardika 2017) or

157   performed poorly. We performed the following to generate a set of conotoxin loci from the

158   transcriptome data: (1) we trimmed sequences from (Phuong *et al.* 2016) to only include the

159   coding region and 100bp of the untranslated regions (UTRs), (2) merged sequences using cd-hit

160   v4.6.4 (Li & Godzik 2006) at 97% sequence similarity to reduce redundancy among conotoxin

161   loci, and (3) masked repeats using the RepeatMasker v4.0.6 web server (Smit *et al.* 2015). This

162   filtered dataset contained 395 conotoxin loci with a total length of 171,317 bp.

163   We submitted the following datasets to MYcroarray (Ann Arbor, Michigan, USA) for

164   bait synthesis: (1) 1749 loci for phylogenetic inference, (2) 12652 conotoxin loci using data from

165   (Phuong & Mahardika 2017), (3) a single concatenated sequence using data from (Phuong &

166   Mahardika 2017), and (4) 395 additional conotoxin loci using transcriptome data from (Phuong

167   *et al.* 2016). We chose to synthesize a MYbaits-3 kit, which included 60,000 bait sequences to

168   accommodate all the targeted loci. Because our aim was to recover sequences from species

169   throughout Conidae, each bait sequence was 120bp in length, which increases the efficiency of

170   recovering divergent fragments. We used a 2X tiling density strategy (a new probe every 60bp)

171   across the sequences from datasets (1) and (2) and used a 4X tiling density strategy (a new probe

172   every 30bp) across datasets (3) and (4). We chose to increase the tiling density for datasets (3)

173   and (4) because the boundaries between exons were not denoted and we wanted to ensure

174   effective capture of the conotoxin loci. The set of probe sequences will be made available on

175   DRYAD following publication.

176

177   *Genetic samples, library preparation, hybridization, and sequencing*

178   We performed the targeted sequencing experiment across 362 samples representing both

179   described Conidae species and unique lineages/potential new species identified during routine

180   species verification using the mitochondrial locus (results not shown), CO1 (Table S1, Folmer *et*

181   *al.* 1994). We also sequenced *Bathyoma* sp. as an outgroup based on a recent molecular

182   phylogeny of the Conoideans, a clade of gastropods that includes Conidae (Table S1, Puillandre

183   *et al.* 2011). We obtained these genetic samples from two field expeditions in Indonesia and

184 Australia and from five museum collections (Table S1). We extracted DNA from tissue using the

185 EZNA Mollusc DNA kit (Omega Bio-Tek, Doraville, GA, USA). There was slight variation in

186 tissue preservation strategy among samples, with most tissues preserved directly in 95% ethanol

187 (Table S1). For 10 samples, tissue was not available but DNA was available from a previous

188 extraction. For these samples, we ran the DNA through the EZNA Mollusc DNA kit to purify the

189 DNA prior to library preparation. We extracted a minimum of 2000 ng per sample prior to

190 library preparation, when possible. We sheared DNA using a Biorupter UCD-200 (Diagenode)

191 when necessary and used a 1X bead purification protocol to ensure that the DNA fragments per

192 sample ranged from 250-600bp, centered on ~350bp. We aimed to generate libraries with longer

193 fragment sizes to ensure that we could recover exons containing the mature toxin region, which

194 are often only recoverable because they are flanking conserved regions that are targeted by our

195 bait design (Phuong & Mahardika 2017).

196   We prepared libraries for following the (Meyer & Kircher 2010) protocol with the

197 following modifications: (1) we started library preparation with at least 2000ng, rather than the

198 500ng suggested by the protocol to increase downstream capture efficiency, (2) we performed

199 1X bead clean-up for all enzymatic reactions and (3) generated dual-indexed libraries by

200 incorporating adapters with unique 7bp barcodes. We were able to re-use libraries for the 32

201 species sequenced in (Phuong & Mahardika 2017) and incorporated new indexes for these

202 samples.

203   We generated equimolar pools of 8 samples and hybridized probes with 2000ng of the

204 pooled DNA for ~24 hours. We substituted the adapter blocking oligonucleotides provided by

205 MYcroarray with custome xGen blocking oligonucleotides (Integrated DNA technologies). We

206 performed 3 independent post-capture amplifications using 12 PCR cycles and pooled these

207   products. We sequenced all samples across 5 Illumina HiSeq 4000 lanes with 100bp paired-end

208   reads. We multiplexed 80 samples per lane for the first 4 lanes and multiplexed the remaining 43

209   samples on the last lane. Sequencing was carried out at the Vincent J. Coates Genomics

210   Sequencing Laboratory at UC Berkeley. We note that our third lane containing 80 samples was

211   contaminated, with 65% of the reads belonging to corn DNA. We were able to resequence this

212   entire lane, resulting in overall increased sequencing effort for samples belonging to our third

213   lane (Table 1).

214

215   *Data filtration and initial assembly*

216   We filtered the raw read data as follows:

217   (1) we trimmed reads using Trimmomatic v0.36 under the following conditions: (a) we

218   used the ILLUMINACLIP option to trim adapters with a seed mismatch threshold of 2, a

219   palindrome clip threshold of 40, and a simple clip threshold of 15, (b) we performed

220   quality trimming used the SLIDINGWINDOW option with a window size of 4 and a

221   quality threshold of 20, (c) we removed reads below 36bp by setting the MINLEN option

222   to 36, and (d) we removed leading and trailing bases under a quality threshold of 15.

223   (2) we merged reads using FLASH v1.2.11 (Magoč & Salzberg 2011) with a min overlap

224   parameter of 5, a max overlap parameter of 100, and a mismatch ratio of 0.05.

225   (3) we removed low complexity reads using prinseq v0.20.4 (Schmieder & Edwards

226   2011) using the entropy method with a conservative threshold of 60.

227   We assembled the filtered read data using SPAdes v3.8.1 using default parameters and reduced

228   redundancy in the resultant assemblies with cap3 (Huang & Madan 1999) under default

229   parameters and cd-hit v4.6 (Li & Godzik 2006, sequence identity threshold = 99%).

230

Phylogenetic data processing and filtering

231

232   To associate assembled contigs with the target sequences for phylogenetic inference, we

233   used blastn v2.2.31 (word size = 11, evalue = 1e-10). For the set of target sequences that

234   originated from the transcriptome dataset, we redefined exon/intron boundaries using

235   EXONERATE v2.2.0 (Slater & Birney 2005) using the est2genome model because we found

236   that several predicted exons actually consisted of several smaller exons. For each sample, we

237   mapped reads using bowtie2 (very sensitive local and no discordant options enabled) to a

238   reference that contained only sequences associated with the targeted phylogenetic markers. We

239   marked duplicates using picard-tools v2.0.1 and masked all regions below 4X coverage and

240   removed the entire sequence if more than 30% of the sequence was below 4X coverage. We

241   called SNPs using samtools v1.3 and bcftools v1.3 and estimated average heterozygosity across

242   all contigs within a sample. We removed sequences if a contig had a heterozygosity value greater

243   than two standard deviations away from the mean.

244

245   *Conotoxin assembly, processing, and filtering*

246   Commonly used assembly programs are known to poorly reconstruct all copies of

247   multilocus gene families (Lavergne *et al.* 2015; Phuong *et al.* 2016). To address this issue, we

248   followed the conotoxin assembly workflow outlined in (Phuong & Mahardika 2017). Briefly, we

249   first mapped reads back to our assembled contigs using the 'very sensitive local' and no

250   discordant' options. Then, we identified conotoxins within our dataset by using blastn v2.2.31

251   (word size = 11, evalue = 1e-10) to associate our assembled contigs (from SPAdes) with

252   conotoxins we targeted in the bait design. We generated a set of unique conotoxin 'seed

253   sequences' (a short stretch [~100bp] of conotoxin-blasted sequence) using a combination of of

254   the pysam module (https://github.com/pysam-developers/pysam), cd-hit v4.6 (percent identity =

255   98%), cap3 (overlap percent identity cutoff = 99%), blastn v2.2.31 (word size = 11, evalue=1e-

256   10), and Tandem Repeats Finder v4.09 (Benson 1999, minscore = 12, maxperiod = 2). We

257   mapped reads to these seed sequences using bowtie2 v2.2.6 (very sensitive local and no

258   discordant options enabled) and built out the conotoxin sequences using the PRICE v1.2

259   algorithm, which uses an iterative mapping and extension strategy to build out contigs from

260   initial seed sequences (Ruby *et al.* 2013). We ran price on each seed sequence at 5 minimum

261   percentage identity (MPI) values (90%, 92%, 94%, 96%, 98%) with a minimum overlap length

262   value of 40 and a threshold value of 20 for scaling overlap for contig-edge assemblies. A

263   reassembled sequence was retained if it shared 90% identity with the original seed sequence and

264   we reduced redundancy by only retaining the longest sequence per seed sequence out of the 5

265   MPI assembly iterations. This approach is described in further detail in (Phuong & Mahardika

266   2017). We note that the final conotoxin sequences per sample consisted of exon fragments,

267   where each sequence represents a single conotoxin exon flanked by any adjacent noncoding

268   region.

269   We updated our conotoxin reference database because we targeted additional conotoxin

270   transcripts from (Phuong *et al.* 2016).We used blastn v2.2.31 (word size = 11, evalue =1e-10)

271   and EXONERATE v2.2.0 to define exon/intron boundaries for these additional conotoxin

272   transcripts and added them to our conotoxin reference database. The final conotoxin reference

273   database consisted of conotoxin sequences with the coding regions denoted and gene superfamily

274   annotated. We also annotated the conotoxin sequences for functional region (e.g., signal, pre,

275   mature, post) using blastn v2.2.31 (word size = 11, evalue = 1e-10) with a conotoxin reference

276   database that was previously categorized by functional region (Phuong & Mahardika 2017).

277   With the final conotoxin reference database, we performed blastn v2.2.31 (word size =

278   11, evalue = 1e-10) searches between the conotoxin reference and every sample's re-assembled

279   conotoxin sequences. We retained sequences if they could align across the entire coding region

280   of the reference sequence. We guessed the coding region for each retained sequence by aligning

281   the query sequence with the reference conotoxin using mafft v7.222 and denoting the coding

282   region as the region of overlap with the exon in the reference conotoxin. We fixed misassemblies

283   by mapping reads with bowtie2 (very sensitive local and no discordant options enabled, score

284   min = L, 70, 1) back to each conotoxin assembly and marked duplicates using picard-tools

285   v2.0.1. We masked regions below 5X coverage and discarded sequences if coverage was below

286   5X across the entire predicted coding region. To generate the final set of conotoxin sequences

287   per sample, we merged sequences using cd-hit v4.6.4 (percent identity = 98%, use local sequence

288   identity, alignment coverage of longer sequence = 10%, alignment coverage of short sequence =

289   50%).

290

291   *Targeted sequencing experiment evaluation*

292   We generated the following statistics to evaluate the overall efficiency of the capture

293   experiment: (1) we calculated the % reads mapped to our targets by mapping reads to a reference

294   containing all targets (both phylogenetic markers and conotoxin sequences) using bowtie2 v2.2.7

295   (very sensitive local and no discordant options enabled, score min = L, 70, 1), (2) we calculated

296   the % duplicates that were identified through the picard-tools, and (3) we calculated average

297   coverage across the phylogenetic markers and conotoxin sequences. We also evaluated the effect

298   of tissue quality (measured by the maximum fragment length of the extracted DNA sample via

299   gel electrophoresis) and genus (only on *Conus*, *Profundiconus*, and *Conasprella*, the three genera

300   with more than 1 sample included in this study) on these capture efficiency metrics using an

301   Analysis of Variance (ANOVA). To assess the effectiveness of conotoxin sequence recovery, we

302   compared our capture results with conotoxin diversity estimates from (Phuong & Mahardika

303   2017) and calculated the average change in those estimates.

304

305   *Phylogenetic inference*

306   In addition to the 362 samples that we sequenced in this study, we obtained sequences for

307   10 other species (Table 1). For two of these species, we used data from another targeted

308   sequencing study (Abdelkrim *et al.* unpublished). We used blastn (word size = 11, evalue = 1e-

309   10) to identify loci that were present in our phylogenetic marker reference. These sequences

310   were filtered under conditions similar to the filtering strategy applied to the phylogenetic

311   markers in this study. For the other eight species, we used data from venom duct transcriptomes

312   (Safavi-Hemami *et al.* unpublished). With these transcriptomes, we trimmed data using

313   trimmomatic v0.36 and merged reads using flash using parameters previously described above.

314   We assembled each transcriptome using Trinity v2.1.1 (Grabherr *et al.* 2011) reduced

315   redundancy in these transcriptomes with cap3 and cd-hit (percent identity = 99%). We used

316   blastn (word size = 11, evalue=1e-10) to associate contigs with the phylogenetic markers present

317   in our dataset. We used bowtie2 v2.2.7 (very sensitive local and no discordant enabled), samtools

318   v1.3, and bcftools 1.3 to map reads and call SNPs. We removed sequences if they were below

319   4X coverage for > 30% of the sequence and masked bases if they were below 4X coverage. We

320   also removed sequences if they had a heterozygosity value two standard deviations away from

321   the mean heterozygosity within a sample. We used to mafft v7.222 to align loci across a total of

322   373 samples.

323   We inferred phylogenies under both maximum likelihood (Stamatakis 2006) and

324   coalescent-based methods (Mirarab & Warnow 2015). We used RAxML v8.2.9 (Stamatakis

325   2006) to generate a maximum likelihood phylogeny using a concatenated alignment under a

326   GTRGAMMA model of sequence evolution and estimated nodal support via bootstrapping. We

327   generated the coalescent-based phylogeny using ASTRAL-II v5.5.9 (Mirarab & Warnow 2015)

328   with individual locus trees generated under default parameters in RAxML v8.2.9. We estimated

329   local posterior probabilities as a measure of branch support (Sayyari & Mirarab 2017). Due to

330   the underperformance of the capture experiment, we ran both phylogenetic analyses with loci

331   that had 80% of the taxa, 50% of the taxa, and 20% of the taxa. For each iteration, we removed

332   taxa that had > 90% missing data.

333

334   *Time calibration*

335   We estimated divergence times using a Bayesian approach with MCMCTree

336   implemented in PAML v4.9g (Yang 2007). Given the size of our alignments, we first estimated

337   branch lengths using baseml and then estimated divergence times using Markov chain Monte

338   Carlo (MCMC). We used a HKY85 + Γ substitution model and used an independent rates clock

339   model. We left all other settings on default. We performed two independent runs of the analysis

340   and checked for convergence among the runs. To account for uncertainty in branching order in

341   our phylogeny, we executed dating analyses across all trees generated from RAxML.

342   For time calibration, we applied a maximum constraint of 55 million years at the root of

343   Conidae, which corresponds with the first confident appearance of Conidae in the fossil record

344   (Kohn 1990). We assigned 12 additional fossils (Table S2, Fig. S1 (Duda Jr. *et al.* 2001;

345   Hendricks 2009, 2015, 2018)) to nodes throughout the phylogeny as minimum age constraints,

346   which MCMCtree treats as soft bounds on the minimum age (Yang 2007). Further information

347   on fossil placement on nodes can be found in the Supplementary. A recent paper showed that the

348   number of species in *Lautoconus* may be overestimated (Abalde *et al.* 2017). To account for

349   potential artificial inflation in the species richness of this clade, we artificially removed half the

350   unique species in *Lautoconus* from our dataset and ran all dating analyses and downstream

351   diversification analyses on this secondary dataset.

352

353   *Characterizing diversification patterns*

354   To visualize lineage accumulation patterns, we generated a log-lineage through time plot

355   using the R package APE (Paradis *et al.* 2018). We estimated diversification rates and identified

356   rate shifts using BAMM (Bayesian Analysis of Macroevolutionary Mixtures) (Rabosky 2014),

357   which uses reversible jump Markov chain Mone Carlo to explore potential lineage diversification

358   models. To account for non-randomness in species sampling across Conidae genera, we applied

359   generic-specific sampling fractions. Using the number of valid Conidae names on WoRMS as

360   estimates of total species diversity in each genus (Worms Editorial Board 2017), we applied a

361   sampling fraction of 32.1% to *Profundiconus*, 50% to *Lilliconus*, 100% to *Californiconus*, 16.7%

362   to *Pygmaeconus*, 28% to *Conasprella*, and 33.7% to *Conus*. We ran BAMM for 100 000 000

363   generations and assessed convergence by calculating ESS values. We analyzed and visualized

364   results using the R package BAMMtools (Rabosky *et al.* 2014).

365

366   *Trait dependent diversification*

367   We tested for the impact of evolvability (measured as conotoxin gene diversity) on

368   diversification patterns using two trait dependent diversification methods, focusing on the genus

369   *Conus*. We focused our hypothesis testing on *Conus* because conotoxin diversity is well-

370   characterized in this group (Phuong *et al.* 2016) and the sequence capture approach used in this

371   study likely represents uniform sampling in conotoxin gene diversity across the genus. This is in

372   contrast to other genera in Conidae, such as *Conasprella* or *Profundiconus*, where low conotoxin

373   diversity values are likely the result of poor knowledge of the venom repertoire of these genera

374   (Fig. S2)

375   First, we used BiSSE (binary state speciation and extinction, (Maddison *et al.* 2007))

376   implemeneted in the R package diversitree (FitzJohn 2012), which employs a maximum

377   likelihood approach to estimate the impact of a binary trait on speciation, extinction, and

378   transition rates between character states. We coded the conotoxin gene diversity data as 'low' or

379   'high' across several thresholds (i.e., 250, 300, 350, 400, 500, 550, or 600 estimated conotoxin

380   genes per species) and compared BiSSE models where speciation rates were allowed to vary or

381   remain equal between traits. We applied a sampling fraction of 33.7%, taking the maximum

382   number of *Conus* species to be the number of valid names on WoRMS (World Register of

383   Marine Species, (Worms Editorial Board 2017)). We determined the best-fitting model using

384   Akaike Information Criterion (AIC). Second, we used FiSSE (Fast, intuitive State-dependent

385   Speciation-Extinction analysis), a non-parametric statistical test that assesses the effects of a

386   binary character on lineage diversification rates (Rabosky & Goldberg 2017). We followed the

387   same coding strategy as in the BiSSE analyses to convert conotoxin gene diversity counts to

388   binary character states. Finally, we used STRAPP (Structured Rate Permutations on Phylogenies,

389   (Rabosky & Huang 2016) implemented in the R package BAMMtools (Rabosky *et al.* 2014).

390   STRAPP is a semi-parametric approach that tests for trait dependent diversification by

391   comparing a test statistic with a null distribution generated by permutations of speciation rates

392   across the tips of the phylogeny (Rabosky & Huang 2016). We generated the empirical

393   correlation (method = Spearman's rank correlation) between speciation rates and conotoxin gene

394   diversity and compared this test statistic with the null distribution of correlations generated by

395   permutations of evolutionary rates across the tree. We performed a two-tailed test with the

396   alternative hypothesis that there is a correlation between speciation rates and total conotoxin

397   gene diversity.

398   We also tested the impact of diet and larval dispersal strategy on diversification patterns.

399   Both piscivory and molluscivory is known to have evolved from the ancestral vermivory

400   condition in cone snails (Duda Jr. *et al.* 2001) and these diet transitions may be associated with

401   increased diversification rates due to access to new dietary niches. In addition, differing larval

402   dispersal strategies including long-lived larval stages (planktotrophy) and short-lived and/or

403   direct developing larvae (lecithotrophy) are hypothesized to impact long term diversification

404   patterns (Jablonski 1986). We coded diet as either vermivory, molluscivory, and piscivory using

405   natural history information from (Jiménez-Tenorio & Tucker 2013). We tested the impact of

406   speciation and extinction using MuSSE (multistate speciation and extinction, (FitzJohn 2012))

407   where speciation rates were allowed to vary or remained equal among traits. We excluded

408   species that were documented to feed on multiple diet types from this analysis. For larval type,

409   we used protoconch morphology from (Jiménez-Tenorio & Tucker 2013) to infer larval dispersal

410   strategy, where multispiral protoconchs were indicative of planktotrophic larvae. We tested the

411   impact of larval type on diversification patterns using BiSSE and FiSSE.

412

Results

Targeted sequencing data

We sequenced an average of 9,548,342 reads (range: 1,693,918 – 29,888,444) across the 363 samples (Table S1). After redefining exon/intron boundaries in the phylogenetic marker reference, we ultimately targeted 2210 loci. On average, we recovered 1388 of these loci per sample (range: 30 – 1849, Table S1) at an average coverage of 12.39X (range: 3.08X – 27.87X, Table S1). For the conotoxin dataset, each sequence we re-assembled contained a single conotoxin exon with any associated noncoding regions (referred to here as 'conotoxin fragments'). We recovered on average 3416 conotoxin fragments per sample (range: 74 – 11535 fragments, Table S1) at an average coverage of 32.3X (range: 5.06X – 65.77X, Table S1). When mapped to a reference containing both the phylogenetic markers and conotoxin genes, the % reads mapped to our targets was on average 14.86% (range: 0.7% - 38.07%, Table S1) and the average level of duplication was 47.47% (range: 22.89% - 89.06%, Table S1).

We found that genus had an impact on % mapped and % duplication, where non-*Conus* genera had lower % mapping and lower % duplication (Fig. S2). These differences likely occurred because conotoxin fragments were not easily recovered in these genera (ANOVA, p < 0.0001, Fig. S2). Genus did not have an impact on coverage or the number of phylogenetic markers recovered (ANOVA, p > 0.05, Fig. S2). We found that tissue quality, measured by the maximum fragment length visualized via gel electrophoresis, had a significant impact on the capture efficiency metrics (ANOVA, p < 0.0001, Fig. S3). DNA samples with strong genomic bands at the top of the gel tended to have higher % mapping, less % duplication, higher coverage, and a greater number of targets recovered (Fig. S3).

436   Our final conotoxin sequence dataset consists of exon fragments and we do not have

437   information on exon coherence (which exons pair together on the same gene). We were unable to

438   assemble full conotoxin genes because conotoxin introns are long (>1 kilobases, (Wu *et al.*

439   2013)) and exceed the average insert size of our sequencing experiment (~350bp). We recovered

440   fragments from all 58 gene superfamilies we targeted and obtained 159,670 sequences

441   containing some or all of the mature toxin region (Table S3). Total conotoxin gene diversity per

442   species (estimated by summing across all signal region exon fragments and sequences containing

443   the entire coding region) ranged from 5 to 1280 copies in *Conus*, 31 to 88 copies in

444   *Profundiconus*, and 7 to 164 in *Conasprella* (Table S1). Total conotoxin diversity was 311

445   copies for *Californiconus californicus*, 12 copies for *Pygmaeconus tralli*, and 30 copies for the

446   outgroup taxon, *Bathyoma sp* (Table S1). When compared to samples in (Phuong & Mahardika

447   2017), the average change (increase or decrease) in total conotoxin gene diversity was ~90 gene

448   copies (Table S4). If samples performed poorly in the number of phylogenetic markers

449   recovered, conotoxin gene diversity estimates tended to be lower in this study than in (Phuong &

450   Mahardika 2017) and vice versa (Fig. S4). The average absolute change in the number of

451   fragments recovered per gene superfamily by region was 3.7 for sequences containing the signal

452   region, 12.2 for the prepro region, 9.6 for the mature region, 48.9 for the post region, and 3.4 for

453   sequences containing the entire coding region (Table S5, Fig. S5). We note several key outliers:

454   the average absolute change in the number of fragments was 104.3 for the T gene superfamily

455   containing the prepro region, 210.4 for the O1 gene superfamily prepro region, 57.4 for the O1

456   gene superfamily mature region, 219.9 for the O2 gene superfamily mature region, and 1417 for

457   the T gene superfamily post region (Table S5, Fig. S5).

458

459   *Phylogeny*

460   The amount of missing data from the alignments was 15.4% when a minimum of 80% of

461   the taxa were present in each locus, 26.8% when 50% of the taxa were present, and 38.6% when

462   20% of the taxa were present. The number of loci retained in the alignment was 387 (107,011

463   bp) when a minimum of 80% of the taxa were present in each locus, 976 (237,027 bp) when 50%

464   of the taxa were present, and 1476 loci (336,557 bp) when 20% of the loci were present. Across

465   all methods and datasets, we recovered phylogenies with a moderate level of resolution (average

466   number of nodes resolved = 71.1%, range = 61.4 - 79.2%, Table S6). In general, as increased

467   amounts of sequence data was given to the phylogenetic programs, more nodes became resolved

468   (Table S6). While we recovered all 6 genera within Conidae with high confidence, relationships

469   among subgenera were less supported (bootstrap and PP = 100%, Fig. 1, Fig. S6, S7, S8).

470

471   *Divergence time estimation*

472   We found evidence for three major branching events during the Eocene: (1) a branching

473   event leading to *Profundiconus* (56.5 mya, CI = 46.3 – 65.3 mya, Fig. 1, S9), (2) a branching

474   event leading to *Conus* (54.7 mya, CI = 42.5 – 63.6 mya, Fig. 1, S9), and (3) a branching event

475   separating *Conasprella* and *Californiconus*, *Lilliconus*, and *Pygmaeconus* (46.0 mya, CI = 36.5 –

476   53.2 mya, Fig. 1, S9). The branching event leading to *Californiconus* occurred during the

477   Oligocene (26.1 mya, CI = 13.8 – 36.5 mya, Fig. 1, S9) and the split between *Lilliconus* and

478   *Pygmaeconus* occurred during the Miocene (17.8 mya, CI = 9.25 – 25.1 mya, Fig. 1, S9).

479

480   *Diversification patterns*

481 We found that most branching events within each genus began to occur in the Miocene

482 and continued until the present (Fig. 1). When analyzing the entire dataset, we found support for

483 diversification rate heterogeneity, where BAMM identified at least one rate shift across Conidae

484 (Fig. 1, S10). Across the 95% credible set of distinct shift configurations, BAMM detected an

485 increase in diversification rates on the branch leading to *Lautoconus*, a clade consisting mainly

486 of species endemic to the Cape Verde islands (Fig. 1, S10). However, when examining an

487 artificially reduced dataset consisting of half the species within *Lautoconus*, we detect no rate

488 shift or a decrease in diversification rates leading to the *Conus* clade (Fig. S11).

489

490 *Trait dependent diversification*

491 Across all thresholds for the BiSSE analysis, we found that diversification rates were not

492 influenced by conotoxin gene diversity. In all cases, the null model was either preferred (delta

493 AIC > 2, Table S7) or was indistinguishable from a model where speciation and extinction were

494 allowed to vary (delta AIC < 2, Table S7). Both the FiSSE and STRAPP analyses revealed that

495 speciation rates were not correlated with conotoxin gene diversity (p > 0.05). These results were

496 consistent across both the full dataset and the reduced dataset.

497 We found that diversification rates were not dependent on diet when analyzing the full

498 dataset (Table S8). However, in the reduced dataset, we found a signal for diet-dependent

499 speciation rates (delta AIC > 2, Table S8). We found that species with mollusk-feeding diets had

500 the fastest speciation rates (0.33), followed by piscivory (0.24), and vermivory (0.16). For the

501 larval dispersal trait, we found support for trait-dependent speciation rates in the full dataset

502 (delta AIC > 2, Table S9), where species with short-lived larvae had higher speciation rates (0.27

503   vs. 0.16). However, this result was not significant when examining the reduced dataset (Table

504   S9).

505

506

507   **Discussion**

508

509   *Capture results*

510   Our targeted sequencing experiment underperformed initial testing of this sequencing

511   method on cone snails (Phuong & Mahardika 2017). Although tissue quality impacted capture

512   metrics (Fig. S3), the % of reads mapping to our targets for even our best samples was ~30%

513   lower than expected (Phuong & Mahardika 2017). While it is difficult to determine the exact

514   cause of this depression in our capture statistics, we hypothesized that changes made in the bait

515   design between this study and (Phuong & Mahardika 2017) may have led to poorer capture

516   results. For example, we recovered an overabundance of conotoxin sequences containing the post

517   region from the T gene superfamily that has no clear co-variation pattern with phylogenetic

518   relatedness (Fig. S12), which likely indicates a large amount of non-specific binding due to

519   conotoxin misclassification. In the future, we suggest re-designing the baits to only include

520   sequences from only the most critical regions (signal region and mature region) to avoid non-

521   specific binding. Although overall capture efficiency statistics were low, the absolute change in

522   conotoxin diversity estimates per gene superfamily was generally minor (Table S5). Therefore,

523   we do not believe that total conotoxin diversity metrics were severely biased by the sequencing

524   method.

525

526   *Phylogenetic relationships*

527   Below, we discuss the results of our phylogenetic analyses, how the phylogenetic

528   relationships compare to past work, and their implications for Conidae taxonomy. Unless

529   otherwise noted, the results we highlight below have at least 90% bootstrap support in the

530   RAxML analyses and 90% posterior probabilities from the ASTRAL-II analyses (Figure S7, S8).

531   When present results on subgeneric relationships starting from the top of the tree shown in

532   Figure S6.

533   We recovered all six major deep lineages representing genera in Conidae that were

534   previously described in recent molecular phylogenetic studies using mtDNA (Puillandre *et al.*

535   2014a; Uribe *et al.* 2017), Fig. 1, S6, S7, S8). Specifically, we find strong support for

536   *Profundiconus*, *Californiconus*, *Lilliconus*, *Pygmaeconus*, *Conasprella*, and *Conus*, as separate

537   and distinct lineages. We also confirm the branching order of these six genera that were recently

538   described using mtDNA genomes (Uribe *et al.* 2017), with *Profundiconus* being sister to all

539   other genera, *Pygmaeconus* + *Lilliconus* sister to *Californiconus*, *Californiconus* + *Lilliconus* +

540   *Pygmaeconus* sister to *Conasprella*, and these four genera sister to *Conus*.

541   Based on the molecular phylogeny from three mtDNA genes, monophyletic groupings of

542   species from *Conasprella* were classified into several subgenera (Puillandre *et al.* 2014a; b). We

543   note several differences between past results and our study in the relationships among these

544   genera and their monophyly:

545   (1) *Ximeniconus* is sister to all other *Conasprella* in some trees, or we reconstructed a

546   polytomy at the base of *Conasprella*, which contrasts with *Conasprella* (*Kohniconus*)

547   *arcuata* recovered at the base of *Conasprella* in previous work (Puillandre *et al.* 2014a).

548   (2) *Kohniconus* is polyphyletic. In (Puillandre *et al.* 2014a), only a single species from

549   *Kohniconus* was included and we find evidence for the non-monophyly of *Kohniconus*

550   when we included the additional species, *C. centurio*. Given these results, we propose

551   that *C. emarginatus*, *C. delssertii*, and *C. centurio* be placed in the subgenus *Kohniconus*

552   and *C. arcuata* placed in a new subgenus.

553   (3) *Endemoconus* is paraphyletic. When including an additional species (*C. somalica*) not

554   sequenced in (Puillandre *et al.* 2014a), we find that *Endemoconus* is not monophyletic.

555   Based on these results, *C. somalica* should be transferred to *Conasprella*.

556 Within *Conus*, our results largely confirm previous findings that *C. distans* is sister to all other

557 *Conus* species and the relationships among subgenera remain tenuous and difficult to resolve

558 (Puillandre *et al.* 2014a). We note the following differences in subgenera relationships and

559 classification between our results and past work:

560   (1) We found support the sister relationship between *Turriconus* and *Stephanoconus*, which

561   has not been recovered in a previous study (Puillandre *et al.* 2014a).

562   (2) We found support for the monophyly of *Pyruconus* across our RAxML analyses, but not

563   our ASTRAL-II analyses. The monophyly of *Pyruconus* was not supported in (Puillandre

564   *et al.* 2014a).

565   (3) *C. trigonus* and *C. lozeti* were classified into the subgenus (*Plicaustraconus*) based on

566   morphological characters (Jiménez-Tenorio & Tucker 2013; Puillandre *et al.* 2014b). We

567   found this subgenus to be polyphyletic when sequence data was obtained.

568   (4) Similar to (Puillandre *et al.* 2014a), we found that *Textila* + *Afonsoconus* is sister to

569   *Pionoconus*. However, instead of the unsupported relationship of *Asprella* as sister to

570   these three subgenera, we found support for *Gastridium* as the sister group.

571   (5) We found support for the sister relationship between *Asprella* and *Phasmoconus*, which

572   conflicts with the unsupported relationship shown (Puillandre *et al.* 2014a), where these

573   subgenera branch in different parts of the phylogeny.

574   (6) We find support for the following successional branch order: *Tesselliconus*,

575   *Plicaustraconus*, *Eugeniconus*, and *Conus*. We found that *Conus* is sister to *Leptoconus*,

576   *Darioconus*, and *Cylinder*, but the relationships among these three subgenera remained

577   unresolved. This conflicts with (Puillandre *et al.* 2014a) as *Cylinder* was paraphyletic,

578   whereas in our results with increased sampling of *Eugeniconus*, *Cylinder* became

579   monophyletic.

580   (7) We did not find strong support for the subgenus *Calibanus*, contrasting with previous

581   work (Puillandre *et al.* 2014a). In our results, we found that *C. thalassiarchus* and *C.*

582   *furvus* were not sister to each other, or their relationship resulted in an unresolved

583   polytomy. Additional investigation into the subgeneric status of these two species.

584   (8) *C. sanderi* was classified into its own subgenus (*Sandericonus*) based on morphological

585   characters (Jiménez-Tenorio & Tucker 2013; Puillandre *et al.* 2014b). Here, when

586   sequence data were obtained, we found it nested within *Dauciconus*. Therefore, we

587   synonymize *Sandericonus* with *Dauciconus* because *C. sanderi* is the type species for

588   *Sandericonus*.

589   (9) *C. granulatus* was classified into its own subgenus (*Atlanticonus*) based on

590   morphological characters (Jiménez-Tenorio & Tucker 2013; Puillandre *et al.* 2014b).

591   Here, we found that it was nested within *Dauciconus*. No other species within this

592   subgenus have been sequenced up until this point. Therefore, we synonymize

593   *Atlanticonus* with *Dauciconus* because *C. granulatus* is the type species for *Atlanticonus*.

594   (10)   Two species (*C. pergrandis* and *C. moncuri*) sequenced in this study were placed

595   into the subgenus *Elisaconus* (Puillandre *et al.* 2014b). Our results do not support the

596   monophyly of *Elisaconus*, as the sister relationship between *C. moncuri* and *C.*

597   *pergrandis* was not supported in 5/6 trees. Additional data is required to classify *C.*

598   *moncuri* and *C. pergrandis* into the appropriate subgenus.

599   (11)   *C. cocceus* was placed into *Floraconus* based on morphological characters in

600   (Puillandre *et al.* 2014b). With sequence data, we found that it was actually nested within

601   *Phasmoconus*. Therefore, we transfer *C. cocceus* to the subgenus, *Phasmoconus*.

602

603   Classification within Conidae is known to be highly unstable (Jiménez-Tenorio & Tucker 2013;

604   Puillandre *et al.* 2014a; b; Puillandre & Tenorio 2018). Although the phylogeny presented here

605   improved understanding of subgeneric relationships and monophyly of subgenera, resolving

606   relationships within Conidae still remains a significant challenge. Given the underperformance of

607   our capture experiment (Table S1), it is unclear if the reason for the moderate power in resolving

608   relationships is due to insufficient data/incomplete data or due to short internal branches during

609   the origination of Conidae subgenera that are extremely difficult to resolve. Overall, our results

610   suggest that both additional data and increased sampling of Conidae species are reasonable

611   pursuits to continue attempting to resolve the phylogeny and classification of this family of

612   marine snails.

613

614   *Timing of diversification*

615   The timing of splits between major are largely congruent with past estimates from a study

616   using mtDNA genomes (Uribe *et al.* 2017), Fig. 1, S9). However, our age estimates for the

617   branching events between *Californiconus, Lilliconus*, and *Pygmaeconus* are much younger

618   (occurring across the Oligocene into the Miocene) than previous estimates (occurring across the

619   Eocene into the Oligocene, (Uribe *et al.* 2017), Fig. 1, S9). This discrepancy may have been

620   caused by differences in fossil calibration, as we included many more fossils in this study

621   compared to previous studies. The Conidae fossil record and analyses of several molecular

622   phylogenetic studies suggest a major radiation of *Conus* during the Miocene (Kohn 1990; Duda

623   Jr. *et al.* 2001; Uribe *et al.* 2017). While we noted that many branching events within *Conus*

624   ocurred during the Miocene into the present, we did not detect an increase in diversification on

625   the branch leading to the origin of *Conus* (Fig. 1, S10, S11). This is congruent with

626   diversification rates estimated from the fossil record (Kohn 1990), suggesting that the

627   accumulation of species during the Miocene may have been a function of an increased number of

628   lineages present rather than an increase in diversification rates. The number of species we

629   included in the subgenus *Lautoconus* had an impact on the BAMM diversification analyses. On

630   the full dataset, BAMM detected an increase in diversification rates leading to *Lautoconus* (Fig.

631   1, S10), a known and documented radiation of cone snails (Duda & Rolán 2005; Cunha *et al.*

632   2005). However, when we remove half the species in response to recent work suggesting

633   taxonomic inflation in this subgenus (Abalde *et al.* 2017), we do not detect the same shift.

634   Rather, there is partial support for no shift across Conidae, or a slight decrease in diversification

635   rates leading to *Conus* (Fig. S11). These results suggest that the original diversification analyses

636   and identified radiation of *Lauotoconus* may have been due to taxonomic inaccuracies biasing

637   the diversification analyses results, rather than a true radiation. What is even more striking about

638   these results is that we found minimal diversification rate heterogeneity across Conidae, despite

639   the expansive species richness across this group. It is unclear whether this signal is real, or due to

640   other technical artifacts. For sample, although we included over 300 species in this study, this

641   only represents ~30% of the total diversity in this group and may have hindered our ability to

642   effectively estimate diversification rates. Similarly, new Conidae species are continually

643   described, with over 100 species described over the last few years (Worms Editorial Board

644   2017). Therefore, our inability to estimate the number of living taxa may have weakened our

645   ability to test the impact of diversification on this group.

646

647   *Speciation rates and conotoxin gene diversity*

648   Contrary to macroevolutionary expectations, we were unable to detect any relationship

649   between speciation rates and conotoxin gene diversity across all trait dependent diversification

650   analyses (Fig. 1, S11, Table S7). Even when performing the analyses with BiSSE, a method in

651   recent years that has become the subject of criticism due to high false positive rates (Abosky

652   2017; Rabosky & Goldberg 2017), our analyses did not detect an impact of conotoxin gene

653   diversity on diversification rates (Table S7). These results may have been expected, given that

654   we found minimal levels of diversification rate heterogeneity in Conidae (at minimum, one shift,

655   Fig. 1, S10, S11). As discussed previously, taxonomic instability in this group may have

656   hindered our efforts to estimate past historical diversification patterns. However, we did find

657   some signal for the impact of diet and larval dispersal strategy on diversification rates when

658   using the BiSSe and MuSSE methods (Table S8, S9). Further work is needed to be fully

659   confident in this signal given high false positive rates in these methods (Abosky 2017; Rabosky

660   & Goldberg 2017) and given that our results depended on which dataset was used.

661   What is remarkable about these results is the lack of any signal on the impact of venom

662   gene diversity on diversification rates in cone snails, even as we found some signal for trait-

663   dependent diversification in other Conidae characters. If this lack of signal is real, several

664   biological factors may explain this decoupling between conotoxin gene diversity and speciation

665   rates. A critical assumption in *Conus* biology is that ecological diversification driven by diet

666   specialization is a major factor governing diversification dynamics in cone snails (Duda &

667   Palumbi 1999; Duda Jr. *et al.* 2001). Past studies have shown that cone snail venom repertoires

668   track their dietary breadth, providing a link between diet and venom evolution (Phuong *et al.*

669   2016; Phuong & Mahardika 2017). However, it is unclear whether or not the relationship

670   between diet and venom evolution leads to ecological speciation due to divergence in prey

671   preference. Ecological speciation is often difficult to detect in marine ecosystems and long-term

672   diversification patterns may be better explained by traits that limit dispersal and promote

673   isolation (Bowen *et al.*). Another possibility is that conotoxin phenotypic divergence may not be

674   the rate-limiting factor in prey specialization and divergence (Duda Jr. *et al.* 2001). Conotoxin

675   genes are under continuous positive selection and gene duplication that allow venom components

676   to change rapidly in response to the environment (Duda & Palumbi 1999; Duda Jr. *et al.* 2001;

677   Chang & Duda 2012; Phuong & Mahardika 2017). This persistent evolutionary change in the

678   venom cocktail suggests that perhaps venom evolution is not necessarily the factor limiting

679   dietary shifts among species and ultimately, speciation among taxa. Ecological opportunity is

680   hypothesized as a necessary component for diversification (Losos 2010) and may be a more

681   critical factor limiting Conidae diversification. Indeed, evidence from the fossil record and past

682   Conidae molecular phylogenetic studies indicate a concentration of lineage formation during the

683   Miocene (Kohn 1990; Duda Jr. *et al.* 2001; Uribe *et al.* 2017), a period that is coincident with the

684   formation of coral reefs in the Indo-Australian Archipelago (Cowman & Bellwood 2011). Our

685   results also show a concentration of branching events during this period as well, though we do

686   not detect a shift in diversification rates (Fig. 1, S10, S11). Overall, our results point to increased

687   taxonomic sampling and a holistic approach to investigating factors shaping diversification in

688   Conidae for future work.

689   Venom evolution is assumed to be a key innovation that led to the evolutionary success

690   of venomous animal lineages (Pyron & Burbrink 2011; Sunagar *et al.* 2016) and a large body of

691   work is devoted towards understanding how venom evolves and responds to the environment

692   over time (Kordis & Gubensek 2000; Wong & Belov 2012; Casewell *et al.* 2013). However, the

693   impact of venom evolution on higher-level diversification patterns is rarely tested. Here, we

694   examined the effect of variation in the adaptive capacity of venom across Conidae species and

695   found it had no influence on macroevolutionary diversification patterns. Although we do not

696   detect a strong signal of conotoxin gene diversity shaping speciation rates in Conidae, it does not

697   refute the importance of venom evolution in adaptation and prey specialization as venom may be

698   necessary, but not sufficient, to promote speciation (Duda *et al.* 2009; Safavi-Hemami *et al.*

699   2015; Chang & Duda 2016; Phuong *et al.* 2016; Phuong & Mahardika 2018). Future work in

700   other venomous animal systems may shed light on whether or not the ability to adapt to different

701   prey through venom evolution translates to the long-term evolutionary success of taxa.

702

703   **Data availability**

704   Raw read data will be made available at the National Center for Biotechnology Information

705   Sequence Read Archive. Bait sequences, conotoxin sequences, scripts, and final datasets used for

706   analyses will be uploaded onto Dryad following publication.

707

708   **Acknowledgements**

778

Figure 1. Time calibrated maximum likelihood phylogeny of the cone snails. Phylogeny was

estimated in RAxML using a concatenated alignment of loci and was calibrated using 13 fossils

placed at nodes throughout the tree. Only loci with at least 20% of the taxa present were included

in the alignment. Colors across the phylogeny show instantaneous diversification rates and are

averaged across all rate models sampled from a BAMM analysis. Warmer colors indicate higher

speciation rates. Log-lineage through time plot is shown below the phylogeny. First four

columns shown next to tip represent the following from left to right: presence of vermivory

(blue), presence of molluscivory (blue), presence of piscivory (blue), larval type (planktotrophy:

light gray, lecithotrophy, blue), and missing data is represented as dark gray. Bars are shown at

tips depicting variation conotoxin gene diversity across the phylogeny. If bar is not shown, data

is not available or were excluded from downstream diversification analyses. Histogram on the

bottom right shows variation in conotoxin gene diversity. Abbreviations: Plio. = Pliocene; Pleis.

= Pleistocene.

Table S1. Sample information and capture efficiency metrics. We first list the species name

listed in all phylogenetic analyses ("species") and the accepted taxonomic classification in the

WoRMS database at the genus, subgenus, and species level ("WoRMS genus", "WoRMS

subgenus", "WoRMS species"). We then list the specimen ID ("ID"), the collection source

("Collection"), the year the sample was collected ("Year Collected"), how the sample was

preserved ("Preservation type"), and the country the sample originated from ("Country"). We

then list "Gelsize", or the largest fragment size visualized via gel electrophoresis as a way to

measure tissue quality. Values can be either "g" for genomic band, or 1500, 1000, or 500 for

bands beginning at 1500bp, 1000bp, or 500bp. We list the data collection method ("Data

801 collection method"), the number of reads sequenced ("# of reads sequenced") and several capture

802 efficiency metrics. Finally, we list the total estimated number of conotoxin genes per species

803 ("Conotoxin gene diversity").

804

805 **Table S2.** Information on fossils used for calibration. We list the fossil taxon ("Fossil Species"),

806 the genus ("Clade assignment"), extant species related to the fossil ("Compared with"), the

807 formation ("Formation"), the age of the fossil ("Age"), and its citation ("Reference").

808

809 **Table S3.** Table showing the number of conotoxin sequences recovered per species for each

810 gene superfamily. Within each gene superfamily, conotoxin sequences were categorized based

811 on whether they contained the entire coding region or mostly the signal, prepro, mature, or post

812 regions.

813

814 **Table S4.** Comparison of conotoxin gene diversity estimates between this study and (Phuong &

815 Mahardika 2017). These represent comparisons between technical replicates (capture experiment

816 was performed on the same libraries in both studies).

817

818 **Table S5.** Comparison of conotoxin gene diversity estimates between this and (Phuong &

819 Mahardika 2017), broken down by gene superfamily. Within each gene superfamily, conotoxin

820 sequences were categorized based on whether they contained the entire coding region or mostly

821 the signal, prepro, mature, or post regions. These represent comparisons between technical

822 replicates (capture experiment was performed on the same libraries in both studies).

823

824 **Table S6.** Number of nodes resolved depending on the amount of missing data and the tree

825 inference method. Phylogenetic trees were inferred using either RAxML or ASTRAL-II. The "%

826 taxa per locus was" the percent of samples needed per locus in order to retain the locus for

827 phylogenetic inference.

828

829 **Table S7.** Venom gene diversity BiSSE AIC results. "Dataset" represents whether the full

830 dataset was used or the reduced dataset. "Threshold" represents the conotoxin gene diversity

831 value used to decide between "high" and "low" conotoxin diversity. Values above the threshold

832 value were categorized as "high" and values below were categorized as "low". "AIC – variable

833 rates" shows AIC values for a model where speciation and extinction rates were allowed to vary

834 depending on a trait. "AIC – equal rates" represents AIC values for the null model, where rates

835 were not allowed to vary by trait.

836

837 **Table S8.** Diet BiSSE AIC results. Model values were generated under a variable rates model

838 (where speciation was allowed to vary) or under an equal rates model (speciation rates across

839 trait states were equal).

840

841 **Table S9.** Larval dispersal type BiSSE AIC results. Model values were generated under a

842 variable rates model (where speciation was allowed to vary) or under an equal rates model

843 (speciation rates across trait states were equal).

844

845 **Figure S1.** Node placement of fossils. Numbers correspond to node placement justification in

846 the supplementary information on node assignment. Tree was generated from a RAxML analysis

847   of a concatenated alignment where loci were kept if at least 20% of species was present in the

848   locus. Best tree is shown and erroneous and intraspecific tips were pruned.

849

850   **Figure S2.** Boxplots showing impact of phylogeny (categorized by Conidae genus) on capture

851   efficiency metrics. Graph title shows resultant P value from ANOVA analyses.

852

853   **Figure S3.** Boxplots showing impact of tissue quality (estimated by maximum DNA fragment

854   lengths assessed via gel electrophoresis) on capture efficiency metrics. Categories are either "g"

855   for genomic band, or 1500, 1000, or 500 for bands beginning at 1500bp, 1000bp, or 500bp.

856   Graph title shows resultant P value from ANOVA analyses.

857

858   **Figure S4.** Scatterplot showing relationship between the number of phylogenetic markers

859   recovered and the change in total conotoxin gene diversity between this study and (Phuong &

860   Mahardika 2017). Results showed a positive relationship between the two parameters, suggesting

861   that if a sample performed poorly in the capture experiment, it performed poorly in recovering

862   data across all loci (phylogenetic loci or conotoxin loci).

863

864   **Figure S5.** Histograms showing absolute change in conotoxin sequence diversity per gene

865   superfamily between this study and (Phuong & Mahardika 2017). Graphs are partitioned by

866   conotoxin functional region, where sequences were categorized based on whether they contained

867   the entire coding region or mostly the signal, prepro, mature, or post regions. On average,

868   estimates of conotoxin diversity per gene superfamily varied slightly.

869

870   **Figure S6.** Maximum likelihood phylogeny inferred using RAxML, where 20% of the taxa

871   needed to be present within a locus to be included in the final concatenated alignment. The six

872   major genera are colored and subgenera re noted for *Conasprella* and *Conus*.

873

874   **Figure S7.** Phylogenies inferred through the coalescent-based method, ASTRAL-II. Individual

875   loci were inferred under default paramenters in RAxML. Nodes are collapsed when posterior

876   probailities are <90%. Trees are colored and labeled by genus. We varied the level of missing

877   data for each ASTRAL run, where we only retained loci if (a) 80% of taxa had sequences, (b)

878   50% had sequences, and (c) 20% of taxa had sequences.

879

880   **Figure S8.** Maximum likelihood phylogenies generated using a concatenated alignment. Nodes

881   are collapsed when boostrap support values are <90%. Trees are colored and labeled by genus.

882   We varied the level of missing data for each RAxML run, where we only retained loci for the nal

883   concatenated alignment if (a) 80% of taxa had sequences, (b) 50% had sequences, and (c) 20% of

884   taxa had sequences.

885

886   **Figure S9.** Maximum likelihood phylogeny dated with 13 fossil node calibrations in MCMCtree.

887   95% confidence intervals shown at nodes. The final concatenated alignment consisted of loci

888   where 20% of the taxa needed to be present within the locus to be included.

889

890   **Figure S10.** 95% credible set of distinct shift configurations from BAMM for the full dataset.

891   Each graph is labeled by the posterior probability of each shift configuration. Warmer, red colors

892   represent faster speciation rates than cooler, blue colors. We note that in all shift configurations,

893   there is a shift in diversification rates in the clade leading to *Lautoconus*.

894

895   **Figure S11.** 95% credible set of distinct shift configurations from BAMM for the reduced

896   dataset. Each graph is labeled by the posterior probability of each shift configuration. Warmer,

897   red colors represent faster speciation rates than cooler, blue colors. We note that in all shift

898   configurations, there is a shift in diversification rates in the clade leading to *Lautoconus*.

899

900   **Figure S12.** Diversity estimates for the A gene superfamily signal region and the T gene

901   superfamily post region. Estimates are plotted next to the RAxML phylogeny where 20% of taxa

902   had sequences in each locus.

903

904   Abalde S, Tenorio MJ, Afonso CML *et al.* (2017) Phylogenetic relationships of cone snails

905   endemic to Cabo Verde based on mitochondrial genomes. , 1–19.

906   Abosky DALR (2017) Model Inadequacy and Mistaken Inferences of Trait-Dependent

907   Speciation. , **64**, 340–355.

908   Ashkenazy H, Penn O, Doron-faigenboim A *et al.* (2012) FastML : a web server for probabilistic

909   reconstruction of ancestral sequences. , **40**, 580–584.

910   Benson G (1999) Tandem Repeats Finder: a program to analyse DNA sequences. *Nucleic Acids*

911   *Res.*, **27**, 573–578.

912   Bi K, Vanderpool D, Singhal S *et al.* (2012) Transcriptome-based exon capture enables highly

913   cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC*

914   *Genomics*, **13**, 403.

915   Bowen BW, Rocha LA, Toonen RJ, Karl SA The origins of tropical marine biodiversity.

916   Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG (2013) Complex cocktails: the

917   evolutionary novelty of venoms. *Trends in ecology & evolution*, **28**, 219–29.

918   Chang D, Duda TF (2012) Extensive and continuous duplication facilitates rapid evolution and

919   diversification of gene families. *Molecular biology and evolution*, **29**, 2019–29.

920   Chang D, Duda TF (2016) Age-related association of venom gene expression and diet of

921   predatory gastropods. *BMC Evolutionary Biology*, **16**, 27.

922   Cowman PF, Bellwood DR (2011) Coral reefs as drivers of cladogenesis: expanding coral reefs,

923   cryptic extinction events, and the development of biodiversity hotspots. *Journal of*

924   *evolutionary biology*, **24**, 2543–2562.

925   Crow KD, Wagner GP (2006) What is the role of genome duplication in the evolution of

926   complexity and diversity? *Molecular biology and evolution*, **23**, 887–92.

927   Cunha RL, Castilho R, Rüber L, Zardoya R (2005) Patterns of cladogenesis in the venomous

928   marine gastropod genus Conus from the Cape Verde islands. *Systematic biology*, **54**, 634–

929   50.

930   Duda TF, Chang D, Lewis BD, Lee T (2009) Geographic variation in venom allelic composition

931   and diets of the widespread predatory marine gastropod Conus ebraeus. *PloS one*, **4**, e6245.

932   Duda Jr. TF, Kohn AJ, Palumbi SR (2001) Origins of diverse feeding ecologies within Conus, a

933   genus of venomous marine gastropods. *Biological Journal of the Linnean Society*, **73**, 391–

934   409.

935   Duda TF, Palumbi SR (1999) Molecular genetics of ecological diversification: duplication and

936   rapid evolution of toxin genes of the venomous gastropod Conus. *Proceedings of the*

937   *National Academy of Sciences of the United States of America*, **96**, 6820–3.

938   Duda TF, Palumbi SR (2000) Evolutionary diversification of multigene families: allelic selection

939   of toxins in predatory cone snails. *Molecular biology and evolution*, **17**, 1286–93.

940   Duda TF, Rolán E (2005) Explosive radiation of Cape Verde Conus, a marine species flock.

941   *Molecular ecology*, **14**, 267–72.

942   FitzJohn RG (2012) Diversitree : comparative phylogenetic analyses of diversification in R.

943   *Methods in Ecology and Evolution*, **3**, 1084–1092.

944   Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of

945   mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.

946   *Molecular marine biology and biotechnology*, **3**, 294–9.

947   Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-

948   Seq data without a reference genome. *Nature biotechnology*, **29**, 644–52.

949   Hendricks JR (2009) *The Genus Conus (Mollusca:Neogastropoda) in the Plio-Pleistocene of the*

950   *Southeastern United States*. Paleontological Research Institution, Ithaca, NY.

951   Hendricks JR (2015) Glowing Seashells : Diversity of Fossilized Coloration Patterns on Coral

952   Reef-Associated Cone Snail (Gastropoda : Conidae) Shells from the Neogene of the

953   Dominican Republic. *Plos One*, **10**, 1–59.

954   Hendricks JR (2018) Diversity and preserved shell coloration patterns of Miocene Conidae

955   (Neogastropoda) from an exposure of the Gatun Formation, Colón Province, Panama.

956   *Journal of Paleontology*, **In press**.

957   Hoekstra HE, Coyne JA (2007) The Locus of Evolution: Evo Devo and the Genetics of

958   Adaptation. *Evolution*, **61**, 995–1016.

959   Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research*, **9**,

960   868–877.

961   Jablonski D (1986) Larval ecology and macroevolution in marine invertebrates. *Bulletin of*

962   *Marie Science*, **39**, 565–587.

963   Jiménez-Tenorio M, Tucker JK (2013) *Illustrated catalog of the Living Cone snails*. MdM

964   Publishing, Wellington, FL.

965   Jones AG, Arnold SJ, Bürger R (2007) The mutation matrix and the evolution of evolvability.

966   *Evolution*, **61**, 727–45.

967   Kaas Q, Westermann J-C, Craik DJ (2010) Conopeptide characterization and classifications: an

968   analysis using ConoServer. *Toxicon*, **55**, 1491–509.

969   Kaas Q, Yu R, Jin A-H, Dutertre S, Craik DJ (2012) ConoServer: updated content, knowledge,

970   and discovery tools in the conopeptide database. *Nucleic acids research*, **40**, D325-30.

971   Katoh K, Kuma K-I, Toh H, Miyata T (2005) MAFFT version 5: Improvement in accuracy of

972   multiple sequence alignment. *Nucleic Acids Research*, **33**, 511–518.

973   Kohn AJ (1990) Tempo and mode of evolution in conidae. *Malacologia*, **32**, 55–67.

974   Kohn AJ (2001) Maximal species richness in Conus: diversity, diet and habitat on reefs of

975   northeast Papua New Guinea. *Coral Reefs*, **20**, 25–38.

976   Kordis D, Gubensek F (2000) Adaptive evolution of animal toxin multigene families. *Gene*, **261**,

977   43–52.

978   Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**,

979   357–9.

980   Lavergne V, Harliwong I, Jones A *et al.* (2015) Optimized deep-targeted proteotranscriptomic

981   profiling reveals unexplored Conus toxin diversity and novel cysteine frameworks.

982   *Proceedings of the National Academy of Sciences*, **112**.

983   Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein

984   or nucleotide sequences. *Bioinformatics*, **22**, 1658–9.

985   Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and

986   SAMtools. *Bioinformatics*, **25**, 2078–9.

987   Losos JB (2010) Adaptive radiation, ecological opportunity, and evolutionary determinism. *The*

988   *American naturalist*, **175**, 623–639.

989   Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation

990   and extinction. *Systematic biology*, **56**, 701–10.

991   Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome

992   assemblies. *Bioinformatics*, **27**, 2957–63.

993   Malmstrøm M, Matschiner M, Tørresen OK *et al.* (2016) Evolution of the immune system

994   influences speciation rates in teleost fishes. *Nature Genetics*, **48**.

995   Mayrose I, Zhan SH, Rothfels CJ *et al.* (2011) Recently Formed Polyploid Plants Diversify at

996   Lower Rates. , **333**.

997   Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed

998   target capture and sequencing. *Cold Spring Harbor protocols*, **2010**,

999   doi:10.1101/pdb.prot5448.

1000   Mirarab S, Warnow T (2015) ASTRAL-II : coalescent-based species tree estimation with many

1001   hundreds of taxa and thousands of genes. *Bioinformatics*, **45**, 44–52.

1002   Paradis E, Claude J, Strimmer K (2018) APE : Analyses of Phylogenetics and Evolution in R

1003   language. , **20**, 289–290.

1004   Phuong M, Mahardika G (2017) Targeted sequencing of venom genes from cone snail genomes

1005   reveals coupling between dietary breadth and venom diversity. *biorxiv*.

1006   Phuong MA, Mahardika GN (2018) Targeted Sequencing of Venom Genes from Cone Snail

1007   Genomes Improves Understanding of Conotoxin Molecular Evolution. , 1–15.

1008 Phuong MA, Mahardika GN, Alfaro ME (2016) Dietary breadth is positively correlated with

1009   venom complexity in cone snails. *BMC Genomics*, **17**, 401.

1010 Pigliucci M (2008) Is evolvability evolvable? *Nature reviews. Genetics*, **9**, 75–82.

1011 Puillandre N, Bouchet P, Duda TF *et al.* (2014a) Molecular phylogeny and evolution of the cone

1012   snails (Gastropoda, Conoidea). *Molecular phylogenetics and evolution*, **78**, 290–303.

1013 Puillandre N, Duda TF, Meyer C, Olivera BM, Bouchet P (2014b) One, four or 100 genera? A

1014   new classification of the cone snails. *Journal of Molluscan Studies*, 1–23.

1015 Puillandre N, Kantor YI, Sysoev a. *et al.* (2011) The dragon tamed? A molecular phylogeny of

1016   the Conoidea (Gastropoda). *Journal of Molluscan Studies*, **77**, 259–272.

1017 Puillandre N, Tenorio MJ (2018) Molluscan Studies. , 200–210.

1018 Pyron RA, Burbrink FT (2011) EXTINCTION , ECOLOGICAL OPPORTUNITY , AND THE

1019   ORIGINS OF GLOBAL SNAKE DIVERSITY. , 163–178.

1020 Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-

1021   dependence on phylogenetic trees. *PloS one*, **9**, e89543.

1022 Rabosky DL, Goldberg EE (2017) FiSSE : A simple nonparametric test for the effects of a binary

1023   character on lineage diversification rates. , 1432–1442.

1024 Rabosky DL, Grundler M, Anderson C *et al.* (2014) BAMMtools : an R package for the analysis

1025   of evolutionary dynamics on phylogenetic trees. , 701–707.

1026 Rabosky DL, Huang H (2016) A Robust Semi-Parametric Test for Detecting Trait-Dependent

1027   Diversification. *Systematic biology*, **65**, 181–193.

1028 Rabosky DL, Santini F, Eastman J *et al.* (2013) Rates of speciation and morphological evolution

1029   are correlated across the largest vertebrate radiation. *Nature communications*, **4**, 1–8.

1030   Robertson FM, Gundappa MK, Grammes F *et al.* (2017) Lineage-specific rediploidization is a

1031   mechanism to explain time-lags between genome duplication and evolutionary

1032   diversification. , 1–14.

1033   Ruby JG, Bellare P, Derisi JL (2013) PRICE: software for the targeted assembly of components

1034   of (Meta) genomic sequence data. *G3*, **3**, 865–80.

1035   Safavi-Hemami H, Gajewiak J, Karanth S *et al.* (2015) Specialized insulin is used for chemical

1036   warfare by fish-hunting cone snails. *Proceedings of the National Academy of Sciences*, **112**,

1037   1743–1748.

1038   Santini F, Harmon LJ, Carnevale G, Alfaro ME (2009) Did genome duplication drive the origin

1039   of teleosts? A comparative study of diversification in ray-finned fishes. *BMC evolutionary*

1040   *biology*, **9**, 194.

1041   Sayyari E, Mirarab S (2017) Fast Coalescent-Based Computation of Local Branch Support from

1042   Quartet Frequencies Article Fast Track. , **33**, 1654–1668.

1043   Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets.

1044   *Bioinformatics*, **27**, 863–864.

1045   Simakov O, Marletaz F, Cho S-J *et al.* (2013) Insights into bilaterian evolution from three

1046   spiralian genomes. *Nature*, **493**, 526–31.

1047   Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence

1048   comparison. *BMC bioinformatics*, **6**, 31.

1049   Smit AFA, Hubley R, Green P (2015) RepeatMasker Open-4.0.

1050   Soltis DE, Albert V a, Leebens-Mack J *et al.* (2009) Polyploidy and angiosperm diversification.

1051   *American journal of botany*, **96**, 336–48.

1052   Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with

1053   thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

1054   Sunagar K, Casewell NR, Varma S *et al.* (2016) Deadly Innovations: Unraveling the Molecular

1055   Evolution of Animal Venoms. , 1–27.

1056   Tank DC, Eastman JM, Pennell MW *et al.* (2015) Nested radiations and the pulse of angiosperm

1057   diversification : increased diversification rates often follow whole genome duplications.

1058   Uribe JE, Puillandre N, Zardoya R (2017) Phylogenetic relationships of Conidae based on

1059   complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **107**, 142–151.

1060   Wagner GP, Altenberg L (1996) Perspective: Complex Adaptations and the Evolution of

1061   Evolvability. *Evolution*, **50**, 967–976.

1062   Wong ESW, Belov K (2012) Venom evolution through gene duplications. *Gene*, **496**, 1–7.

1063   Worms Editorial Board (2017) World Register of Marine Species.

1064   Wu Y, Wang L, Zhou M *et al.* (2013) Molecular evolution and diversity of Conus peptide toxins,

1065   as revealed by gene structure and intron sequence analyses. *PloS one*, **8**, e82495.

1066   Yang AS (2001) Modularity, evolvability, and adaptive radiations: a comparison of the hemi-

1067   and holometabolous insects. *Evolution & development*, **3**, 59–72.

1068   Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and*

1069   *evolution*, **24**, 1586–91.

1070   Zhan SH, Glick L, Tsigenopoulos CS, Otto SP, Mayrose I (2014) Comparative analysis reveals

1071   that polyploidy does not decelerate diversification in fish. , **27**, 391–403.

1072