

RESEARCH ARTICLE

Open Access

Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian *Eugongylus* group scincid lizards

Matthew C Brandley^{1,2*}, Jason G Bragg³, Sonal Singhal^{4,5}, David G Chapple⁶, Charlotte K Jennings^{4,5}, Alan R Lemmon⁷, Emily Moriarty Lemmon⁸, Michael B Thompson¹ and Craig Moritz^{3,9}

Abstract

Background: High-throughput sequencing using targeted enrichment and transcriptomic methods enables rapid construction of phylogenomic data sets incorporating hundreds to thousands of loci. These advances have enabled access to an unprecedented amount of nucleotide sequence data, but they also pose new questions. Given that the loci targeted for enrichment are often highly conserved, how informative are they at different taxonomic scales, especially at the intraspecific/phylogeographic scale? We investigate this question using Australian scincid lizards in the *Eugongylus* group (Squamata: Scincidae). We sequenced 415 anchored hybrid enriched (AHE) loci for 43 individuals and mined 1650 exons (1648 loci) from transcriptomes (transcriptome mining) from 11 individuals, including multiple phylogeographic lineages within several species of *Carlia*, *Lampropholis*, and *Saproscincus* skinks. We assessed the phylogenetic information content of these loci at the intergeneric, interspecific, and phylogeographic scales. As a further test of the utility at the phylogeographic scale, we used the anchor hybrid enriched loci to infer lineage divergence parameters using coalescent models of isolation with migration.

Results: Phylogenetic analyses of both data sets inferred very strongly supported trees at all taxonomic levels. Further, AHE loci yielded estimates of divergence times between closely related lineages that were broadly consistent with previous population-level analyses.

Conclusions: Anchored-enriched loci are useful at the deep phylogeny and phylogeographic scales. Although overall phylogenetic support was high throughout the Australian *Eugongylus* group phylogeny, there were nonetheless some conflicting or unresolved relationships, especially regarding the placement of *Pseudemoia*, *Cryptoblepharus*, and the relationships amongst closely-related species of Tasmanian *Niveoscincus* skinks.

Keywords: Anchored phylogenomics, Coalescence, Phylogeography, Skink, Transcriptome, Ultraconserved elements

Background

With the increasing access to genomic data enabled by high-throughput sequencing (HTS), phylogeneticists are applying phylogenomics to resolve deeper nodes in the tree of life (e.g., [1-3]). When used along with some form of genome reduction, HTS enables analysis of hundreds to thousands of loci across highly divergent organisms.

Especially promising approaches include comparative RNAseq (e.g., [4-6]) and targeted hybrid enrichment (also referred to as sequence capture) using conserved or highly conserved orthologous loci as targets [1,7-9]. To the extent that common target loci are employed across diverse groups, these enrichment approaches have the potential to produce highly connectable data across orthologous loci, as was the case for mtDNA and cpDNA in the past.

For molecular systematists working at the phylogeographic interface between population genetics and

* Correspondence: mbrandley@gmail.com

¹School of Biological Sciences, Heydon-Laurence Building A08, University of Sydney, Sydney, NSW 2006, Australia

²New York University - Sydney, The Rocks, NSW 2000, Australia

Full list of author information is available at the end of the article

phylogenetics, the utility of hybrid enrichment methods is not immediately obvious. Several methods have been used to generate data at low-moderate scales of phylogenetic divergence, including amplicon sequencing [10–12], various RAD-type methods (e.g., [13–15]), and target enrichment either using custom PCR-baits [16] or custom exon-capture designed from *de-novo* assembled transcriptomes [8]. Each of these has disadvantages – amplicon sequencing and RAD sequencing often generate very patchy matrices of loci by individuals, especially as divergence increases, and custom exon-capture requires more up-front investment. Custom PCR-bait capture has great potential for projects requiring tens of loci, but is not well suited where researchers aim to sequence hundreds of loci to improve phylogenetic precision. Directly mining transcriptomes (transcriptome mining or TM hereafter) is another source of potentially thousands of homologous loci, but phylogenetic analyses of mRNAs may be misled by recombination, especially in genes with exons separated by long introns. Moreover, transcriptome sequencing also requires very high quality, RNA-preserved tissue, and it remains expensive to sequence hundreds of individuals. Indeed, all phylogenomic methods require significant financial resources, but when factoring in the costs of reagents, sequencing, and labor versus the number of nucleotides sequenced, target enrichment type methods are far less expensive than traditional PCR-based methods [9].

Target enrichment using conserved targets – either “ultraconserved” elements (UCEs [17]) or conserved exons (anchored hybrid enrichment, AHE [1]) – could well yield useful data for phylogeographic inference, including historical demography, species delimitation and inference of divergence processes. However, given that the targets themselves are evolutionarily conserved, divergences and phylogenetic information content within and across closely related lineages could be low. In general, capture efficiency, and hence coverage, is highest across the relatively conserved target areas and decreases into the flanking sequences (e.g. intergenic regions or introns) that are expected to contain the most variability at this scale [9,17]. Further, given their high level of conservation, purifying selection could result in strongly biased estimates of population diversity, and thus of phylogenetic estimates via species trees or divergence parameters [18,19]. On the other hand, with the exception of genes that are highly conserved or consistently under strong adaptive or balancing selection, these issues may be less significant for loci mined from transcriptomes.

That conserved locus hybrid-capture might nonetheless have utility at a phylogeographic scale is suggested by a recent analysis of population divergence across multiple avian taxa from the Amazon basin that used

anchored enrichment of UCEs [17]. Similarly, AHE has provided good resolution of phylogeny and divergence histories across closely related species and lineages of chorus frogs (*Pseudacris*; Lemmon et al., unpub. data). However, using non-random sets of the genome for evolutionary inference has potential to bias estimates of divergence and historical demography because depending on how these loci are ascertained, they might show idiosyncratic patterns from the rest of the genome. For example, one recent study found that data from UCEs versus RAD-Seq yielded very different estimates of divergence history parameters among phylogeographic lineages [19].

Here we explore the information content of AHE loci at different taxonomic scales including across genera, congeneric species, and phylogeographic lineages of the Australian *Eugongylus* group skinks (Squamata: Scincidae). In doing so, we both infer phylogenies and divergence histories for this group of lizards. We also compare the performance of AHE to the only other current source of multi-locus data for the *Eugongylus* group: exons mined from transcriptomes.

The *Eugongylus* group is a species rich clade of lygosomine skinks that includes ~40 genera and 420 described species. The common ancestor of a clade consisting of 17 extant genera colonized Australia and adjacent landmasses (New Zealand, New Caledonia, South Pacific islands) from Asia in the Early to Middle Miocene (42–22 Mya [20–22]). Species from three of these genera – *Carlia*, *Lampropholis* and *Saproscincus* – have been subjects of detailed phylogeographic analyses in wet forests of eastern Australia [23–28] and of studies of the evolution of reproductive isolation among phylogeographic lineages [29]. Previous molecular phylogenetic analyses using relatively few loci have inferred strong support for relationships among species within these genera (e.g. [23,30]). However, relationships among genera in the *Eugongylus* group remain largely obscure and are evidently difficult to resolve using few loci (Shea et al., unpub. data). Moreover, *Niveoscincus* and *Pseudemoia* are viviparous, and species in these genera show a range of both placental complexity and variation in mother-to-embryo exchange [31–34]. Without a robust intergeneric phylogeny, it remains impossible to uncover the evolutionary processes that explain the diverse ecological, physiological, and reproductive traits among Australian *Eugongylus* group skinks.

Methods

We constructed two multi-locus data sets using from data collected from different sequencing methods. The first data set consists of 415 orthologous genes sequenced for 43 taxa using anchored hybrid enrichment [9]. The second data set includes 1650 exons from 1648 orthologous genes for 11 taxa that were mined from RNA transcriptomes.

Taxon sampling design

We evaluate the performance of AHE and TM at the intergeneric, interspecific, and intraspecific phylogenetic levels. Our nested sampling design includes 9 of 17 Australian *Eugongylus* group genera as well as two individuals per phylogeographic lineage for *Carlia rubrigularis*, *Lampropholis coggeri*, *La. robertsi*, and *Saproscincus basiliscus* (Additional file 1). As outgroup taxa, we sequenced *Lerista bougainvillii* of the *Sphenomorphus* group clade of lygosomine skinks, which shares a most recent common ancestor with the *Eugongylus* group 64–100 million years ago [22]. We used a single outgroup because this is as required for STEAC analyses (see below), and this makes the results directly comparable across methods. Inclusion of all three outgroups in RAxML analyses of the concatenated data set does not affect support for relationships of the ingroup (not shown). The final AHE data set includes 42 ingroup samples, and the TM data set includes 10 ingroup samples and a single outgroup in each data set (*Le. bougainvillii*).

Anchored enrichment data collection and processing

Data were collected following the methods of [9] through the Center for Anchored Phylogenomics at Florida State University. Briefly, each genomic DNA sample was sonicated to a fragment size of ~300-700 bp. Subsequently, library preparation and indexing were performed following a protocol modified from [35]. Indexed samples were then pooled at equal quantities (eight samples per pool), and enrichments were performed on each multi-sample pool using an Agilent Custom SureSelect kit, which contained probes designed for anchored loci from multiple vertebrate genomes (Vertebrate 1.0, described in [9]). After enrichment, the six reactions were pooled in equal quantities for sequencing on a single PE100 Illumina HiSeq2000 lane. Sequencing was performed at the QB3 Vincent J. Coates Sequencing Laboratory at the University of California, Berkeley.

Quality-filtered sequencing reads were processed following the methods described in [9]. In short, reads were demultiplexed (with no mismatches tolerated) and pairs were merged following [36]. Merged reads were then scanned for matches to the probe region sequences of *Anolis carolinensis* using the high sensitivity approach described in [9]. Reads matching to a reference sequence for each individual were then aligned to produce an assembly for each locus. The reads were aligned using a script written by ARL that: (1) sorts the reads by number of matches to the reference, (2) for each read in the sorted list (starting with best-matched read), the position maximizing the match to the previous read in the list is noted, (3) if the best matching position does not generate a 90% match (of at least 20 bp) the read is skipped, (4) the entire sorted list is repeatedly traversed until no additional reads have aligned during a traversal (source

code for all analyses in this study is available on Dryad doi:10.5061/dryad.g4mj2).

After assemblies were complete, consensus bases were called as follows: (1) all sites with less than 3-fold coverage and variant sites with less than 10-fold coverage were called as “N”, (2) invariant sites with coverage between 3-fold and 10-fold were called with the observed nucleotide, (3) for sites with greater than 9-fold coverage, the most common base was called unless the distribution of observed bases was unlikely to have arisen under a two allele model with equal allele frequencies. This likelihood was approximated using the equation $p = (1 - p_{\text{binom}}(n_{\text{Max}}, n, 0.5))$, with n equaling the number of unambiguous base calls and n_{Max} equaling the abundance of the most common base call. The most common base observed was called unless $p > 0.05$, suggesting that the two allele model could explain the diversity observed. In that case, an ambiguity code corresponding to all of the observed bases at that site was used in the consensus sequence. Sequences for each locus were aligned across species using Muscle version 3.8.31 under default settings [37]. Prior to phylogenetic analysis, each locus data set was trimmed to the second shortest sequence, and all gapped characters were removed.

Transcriptome assembly, identification, and mining

Methods for sequencing, assembly, and identification of the transcriptomes of *Le. bougainvillii*, *Pseudemoia entrecasteauxii*, *P. pagenstecheri*, and *P. spenceri* contigs were similar to [38]. Briefly, approximately 5 gigabases of 100 bp paired-end reads were generated using Illumina HiSeq 2000 from pregnant uterine (*Le. bougainvillii*, *P. entrecasteauxii*, and *P. pagenstecheri*) or brain tissue (*P. spenceri*) from a single individual of each species. Contigs were assembled for each species with Abyss v1.3.4 [39] using $kmer=69$, and otherwise default parameters. Contigs ≤ 100 bp were removed. To identify each contig, we used blastx to align them to the genomes of *Anolis* lizard, chicken, mouse, human, platypus, and wallaby genomes (Ensembl build 64).

For *Carlia*, *Lampropholis*, and *Saproscincus* species, data were collected and processed as described in [40]. Briefly, for each phylogeographic lineage, we collected 3.5 Gb of paired-end reads from each of five adults' livers. Reads were trimmed and merged using FLASH v1.0.2 [41] and Trimmomatic v0.16 [42]. We then generated an assembly across all individuals for that species using Trinity r2012-01-25 [43], and then removed redundancy across these assemblies using a custom merge script using blat, cd-hit-est, and cap3 [44-46]. The final transcriptome data sets included eleven taxa including species of *Carlia*, *Lampropholis*, *Pseudemoia*, *Saproscincus*, and *Le. bougainvillii* (outgroup).

Assessing phylogenetic history

To test the performance of both 415 AHE loci and 1650 TM loci, we performed phylogenetic analyses for different taxonomic levels including the (i) the full 43 sample data set (*Eugongylus* group + outgroup) to assess performance at the deep phylogenetic scale, and (ii) *Lampropholis* only and *Saproscincus* only data sets to test performance at the interspecific and phylogeographic scale. For each of the AHE loci and TM loci, we first calculated the most appropriate models of sequence evolution for each locus using MrAIC [47] choosing the model with the best AIC score amongst 24 possible models (GTR, SYM, HKY, K2P, F81, and JC69, with and without Γ and Γ parameters). We then estimated a maximum likelihood gene tree for each locus using PHYML [48] using the appropriate model selected by MrAIC.

We performed concatenated data and species tree phylogenetic analyses for the full taxon, *Lampropholis* + *S. basiliscus* N1 (outgroup), and *Saproscincus* + *La. coggeri* N1 (outgroup). Invariant loci were excluded and the final AHE data sets for the full taxon, *Lampropholis*, and *Saproscincus* data sets were 415, 414, and 412 loci, respectively. There are no invariant TM loci in all three data sets and the full 1650 loci were used. Maximum likelihood phylogenetic analyses of the concatenated data sets were performed using RAxML [49] assuming separate GTR + Γ models for each locus. We then estimated clade support using 1000 pseudoreplicates of bootstrap resampling with the same partition and model scheme as the single-tree analyses.

Simulated and empirical analyses have shown that, if there is strong incongruence amongst loci, analyses of concatenated data sets may infer a strongly supported, but incorrect trees [50–52]. We therefore performed additional species tree analyses using STEAC [53]. STEAC estimates a phylogeny by calculating the average coalescence times across multiple loci. The method is useful because it is computationally tractable with large multi-locus data sets. On the other hand, it assumes incongruence between the gene and species trees is due only to incomplete lineage sorting rather than other phenomena such as hybridization.

STEAC assumes that gene trees are estimated without error, so to incorporate estimation error for individual gene trees in our STEAC analyses, we used 1000 bootstrap trees estimated per locus rather than a single point estimate tree. Although our phylogenetic analyses of concatenated data used RAxML, we chose PHYML [48], an alternative maximum likelihood phylogenetic program, for this analysis to enable efficient scripting of the STEAC analytical pipeline. We first calculated the most appropriate models of sequence evolution for each locus using the same procedure as above. For each locus, we performed a 1000 pseudoreplicate bootstrap

analysis using these chosen models. We used STEAC to infer a species tree. We estimated support for this tree by calculating 1000 ‘bootstrap’ STEAC species trees, where each ‘bootstrap’ species trees draws (without replacement) one of the PHYML bootstrap trees for each locus (e.g., making use, in total, of 415,000 trees for the full AHE data set). Species tree analyses were performed using the phybase package in R [54]. Given the number of taxa and loci, use of Bayesian concordance analyses (e.g., BUCKy [55]) were not computationally feasible.

To evaluate the spatial distribution of variation across loci, and to enable comparison with results for ultraconserved elements (UCEs) [17], we assessed the distribution of phylogenetically informative characters (PICs)/ across the 415 AHE loci. All loci were aligned so that the 0th nucleotide corresponded to the middle of the aligned locus. Moving laterally along the loci in both directions, we plotted the number of total loci (max = 415) with a PIC site at that nucleotide position across all taxa. For example, if there were 69 loci with a PIC at the 4th nucleotide position, then we recorded a value of 69. Because the loci vary in length, we overlaid these plots on a distribution of locus lengths calculated by counting the number of loci with any character at a nucleotide position. Unlike the alignments used for the phylogenetic analyses, we did not remove gapped characters to avoid artificially increasing the number of PICs near the alignment origin due to shortening the alignments. We repeated these analyses using only the *Lampropholis* and *Saproscincus* data sets. Finally, we constructed a “normalized” plot for the full-taxon data set by dividing the number of PICs at each nucleotide position by the number of alignments of that specific length.

Inferring divergence histories

To further explore the utility of AHE markers for evolutionary inference, we inferred divergence history parameters for the five lineage-pairs for which we had prior estimates of divergence history from population samples of transcriptomes [28]. These lineage-pairs, representing different levels of divergence, were North and South populations of *C. rubrigularis* (N/S), Central and South *La. coggeri* (C/S), *La. N/C*, and *S. basiliscus* N/C. Additionally, we used comparative transcriptome data from *Le. bougainvillii*, *P. entrecasteauxii* and *P. pagenstecheri* (Griffith et al., unpublished data) for sequence divergence comparisons. To infer divergence history, we used the program 3s, which implements a coalescent model for isolation with migration [56]. 3s builds its inference across many genome loci sampled at one chromosome in each of the two focal lineages and an outgroup lineage. Previous work [57] has shown that

with single samples per taxon, migration rates and ancestral population sizes are non-identifiable but that estimates of divergence time (τ) are robust, so we focus our comparisons on divergence time.

To generate haplotypes for our 3s runs, we first took the raw reads generated from AHE data collection, trimmed them for quality and adaptors Trimmomatic v0.16 [42], merged them using FLASH v1.0.2 [41], and removed sequence duplicates. For a given lineage-pair, we defined one reference assembly, to which we mapped both ingroup lineages and the outgroup. We mapped cleaned reads to the previously-generated assemblies using Bowtie2 [58] and called SNP variants using Samtools default parameters [59]. We phased variants using GATK ReadBackedPhasing [60] and filtered any variant and non-variant sites with coverage less than 5X. The resulting, coverage-filtered haplotypes were used in all downstream analyses.

We calculated divergence between the lineage-pairs using several metrics. First, we calculated nuclear divergence for both AHE and TM sequences [61]. Second, we used previously calculated estimates for divergence time in years, which has been estimated for all lineage-pairs but *P. entrecasteauxii* and *P. pagenstecheri* [29]. Briefly, we had previously inferred these divergence times by sequencing transcriptomes for each lineage ($N = 5$) in the lineage-pair, identifying variants in the untranslated regions of the lineage-pair, constructing the two-dimensional site-frequency spectrum (2D-SFS) for these variants, and then using *dadi* [62] to fit an isolation-with-migration model to the 2D-SFS. Third, we used 3s to infer τ in coalescent units. For each lineage-pair, we used all available loci (details on loci number, sequence length, and outgroups available in Additional file 2: Table S1), running each analysis ten times to ensure convergence across runs. We compared these estimates of divergence history to those obtained previously, expecting that, although the values should not be absolutely similar due to differences in units and mutation rates, divergence estimates should be correlated.

Results

Sequence data

From the original target set of 512 loci, we successfully captured and sequenced 415 across all samples including the outgroups. After applying coverage criteria, trimming all loci to the length of the second shortest sequence and removing all internal gapped characters, the mean locus length was 534 bp with a minimum and maximum lengths of 250 bp and 1458 bp, respectively (Note that loci with trimmed alignments shorter than 250 bp were removed). Of the 221,792 total characters, only 0.36% of the data was missing in any species.

In the full 43 taxon data set, regions of the anchored-enriched (AHE) loci closer to the conserved anchor region ($\sim \pm 120$ bp) show slightly reduced numbers of PICs compared to more variable flanking regions (i.e. $> \pm 120$ bp in Figure 1a). The decline in the number of PICs moving beyond ± 200 bp corresponds to a decrease in the number of loci of that size (i.e., > 400 bp). This pattern is similar, but less extreme, to that seen in UCE loci [7]. When normalized by the number of loci represented, the same trend is seen in the central ± 200 bp, with stochastic scatter beyond that (Figure 1b). The spatial distributions of variation within *Lampropholis* and within *Saproscincus* are similar, albeit less pronounced due to the fewer overall PICs. Complete details on AHE loci are available in Additional file 3.

For the TM analyses, we identified 10,847 exons of 7233 loci that met the initial filtering process. However, not all taxa were present for all loci. From this initial data set, we identified 1650 exons from 1648 genes that could be aligned across all ten taxa. These exons ranged in length from 101 to 3249 bp (mean = 337). Because all but two of our loci are single exons, we expect that phylogenetic error due to recombination is minimal. We refer to these 1650 exons as loci (rather than exons) to simplify comparative discussion with the AHE loci. Loci were trimmed so that all nucleotide positions are present for all genes (i.e., no missing data). Complete details on TM loci are available in Additional file 4.

Phylogenetic information content and gene tree concordance

Across the full taxon AHE data set, and after excluding invariant loci, the mean number of PICs per locus was 52, but there was substantial variation across loci in information content with between 152 and 6 PICs per locus (Table 1). This variability is correlated with locus length ($R^2 = 0.19$; $P < 0.001$). Within genera and species, mean information content was reduced, with *Niveoscincus* having the fewest PICs per locus (1.6; Table 1).

The transcriptome mining (TM) data showed even stronger skew in information content (Table 1) and this was also correlated with locus length ($R^2 = 0.65$; $p < 0.001$). Although TM produced more usable loci than AHE (1650 vs 415), both the maximum number of PICs and average PICs per locus are higher in the AHE loci (mean $PIC_{AHE} = 52$; mean $PIC_{TM} = 19.2$; Table 1). To compare the data sets more directly, we limited our AHE data set to the 11 taxa also included in the TM data set. Analyzing these data, we find the mean number of PICs per locus dropped from 52 to 25. Despite this reduction, PIC_{AHE} remains higher than PIC_{TM} . We also trimmed the TM data to the 415 loci with the highest PICs to compare the most informative TM loci to the AHE loci. The mean number of PICs/locus in this data set is 41. To make a more direct

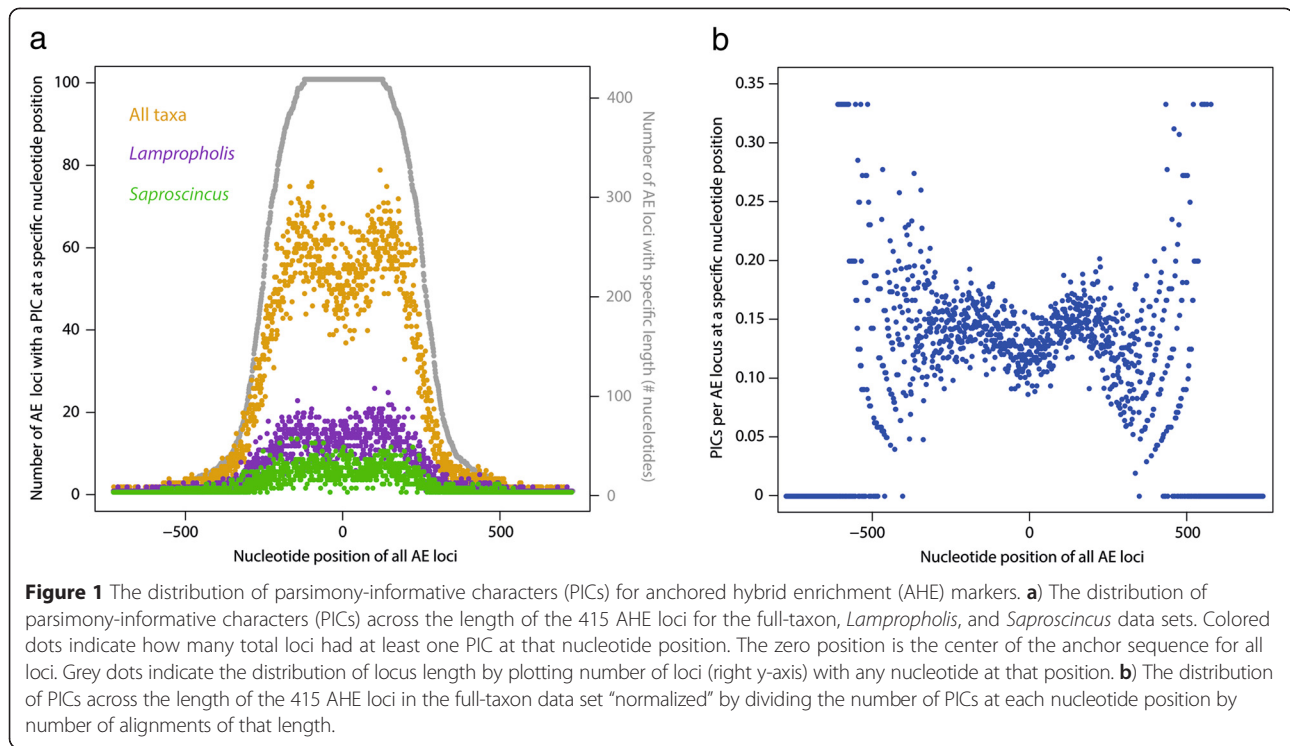


Table 1 Summary statistics for the anchored-enriched (AHE) loci and transcriptome mining (TM) exons at different taxonomic levels and excluding non-informative loci

	N samples	N loci	Locus length			PICs			% identity		
			Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
Anchored-enriched (AHE)											
All samples	44	415	1458	250	534	152	6	52.1	99.6	92.1	96.5
<i>Carlia</i> ¹	5	264	1458	250	541	36	1	2.3	99.0	94.1	99.0
<i>C. rubrigularis</i> ¹	4	253	1458	250	540	36	1	2.2	99.9	94.5	99.4
<i>Lampropholis</i>	13	414	1458	250	534	44	1	15.0	99.7	95.0	98.0
<i>La. coggeri</i> ¹	6	314	1458	250	543	13	1	3.1	99.9	97.8	99.4
<i>Niveoscincus</i> ¹	5	207	1458	250	538	8	1	1.6	99.7	96.7	98.9
<i>Saproscincus</i>	9	412	1458	250	535	45	1	6.7	99.9	90.1	98.4
<i>S. basiliscus</i> ¹	5	221	1458	250	555	13	1	2.1	99.9	95.6	99.5
Transcriptome mining (TM)											
All taxa	11	1650	3249	101	337	132	1	19.2	99.7	90.1	96.3
TM data pruned to number of AHE loci (highest PICs)											
All taxa	11	415	3249	163	617	152	23	41.3	99.1	95.3	90.1
TM data pruned to number of AHE loci (randomly sampled)											
All taxa	11	415	2480	103	337	120	1	19.2	99.6	90.8	96.3
AHE data pruned to TM taxon sampling											
All taxa	11	415	1458	250	534	96	1	25.4	99.7	91.8	96.3

¹Phylogenetic analyses were not performed on these data sets.

comparison, we randomly selected 415 loci from the 1650 TM loci 10,000 times and mean number of PICs/locus (19) is identical to the entire 1650 locus data set, as expected.

Phylogenetic inference

Analyses of the concatenated AHE loci using RAxML and using the STEAC species tree method were largely congruent, with high bootstrap support among phylogeographic lineages within species and among most congeners (except within *Niveoscincus*) (Figure 2). Monophyly of each genus with multiple species was strongly supported. By contrast, resolution among genera decreased towards the base of the tree, and there were two examples of strongly supported incongruence (bootstrap proportion [BP] > 90). In the RAxML analysis, the genus *Pseudemoia* forms a clade with all other Australian *Eugongylus* group skinks (BP = 100) exclusive of *Acritoscincus* (formerly *Bassiana*) and *Harrisoniascincus*; but it forms a clade with these two genera in the STEAC species tree (BP = 97). The genus *Cryptoblepharus* forms a clade with *Carlia*, *Lampropholis*, *Lygisaurus*, and *Saprosincus* (BP = 93) in the RAxML tree, whereas in the STEAC species tree this genus is sister to *Niveoscincus* (BP = 90).

Repeating the analyses at the congeneric to phylogeographic scale, there is no strongly supported incongruence amongst the concatenated RAxML tree and the STEAC species tree for either of the *Lampropholis* and *Saprosincus* datasets (Additional file 5: Figure S1), but clade support does differ at some nodes. Within *Lampropholis*, the RAxML and STEAC analyses strongly support the sister relationship of *La. coggeri* and *La. adonis*. Moreover, the relationships inferred among phylogeographic lineages within *La. coggeri* and within *S. basiliscus*, using either concatenation or species tree approaches accord with expectations from prior sequencing of multiple introns [24,28] (Figure 2).

To enable direct comparison between utility of AHE and TM at the phylogeographic scale, we trimmed the AHE data to just the taxa for which we had assembled transcriptomes. This included three lineages of *La. coggeri*, three closely related species of *Pseudemoia* and two lineages each for *S. basiliscus* and *C. rubrigularis* (Figure 3). Here, the inferred phylogenies were identical and completely resolved for both AHE and TM data sets.

Divergence histories

We inferred divergence for each of five lineage-pairs by comparing across nucleotide divergence, previously estimated using population sequencing, and the coalescent model implemented in 3s that samples a single chromosome per locus. Although AHE loci tended to have lower estimates than TM loci, nuclear divergence values

at AHE and transcriptome loci were highly correlated ($r = 0.97$; $P < 0.05$; Figure 4a). Log-likelihood tests comparing models of divergence without gene flow (Model 0) and with gene flow (Models 1 & 2) in 3s supported divergence with gene flow for each of the five contacts (Additional file 2: Table S2). This contrasts with previous population-level inference [28], which supported divergence without gene flow in the four contacts (*C. rubrigularis* N/S, *La. coggeri* N/C and C/S, *S. basiliscus* N/C) tested. Tau estimates from a population history with gene flow were correlated with nuclear divergence at the same loci ($r = 0.91$; $P < 0.05$; Figure 4b), suggesting that gene flow during divergence was likely minimal. Tau estimates showed reasonable, though not significant, concordance with previously estimated divergence times ($r = 0.85$; $P = 0.07$; Figure 4c).

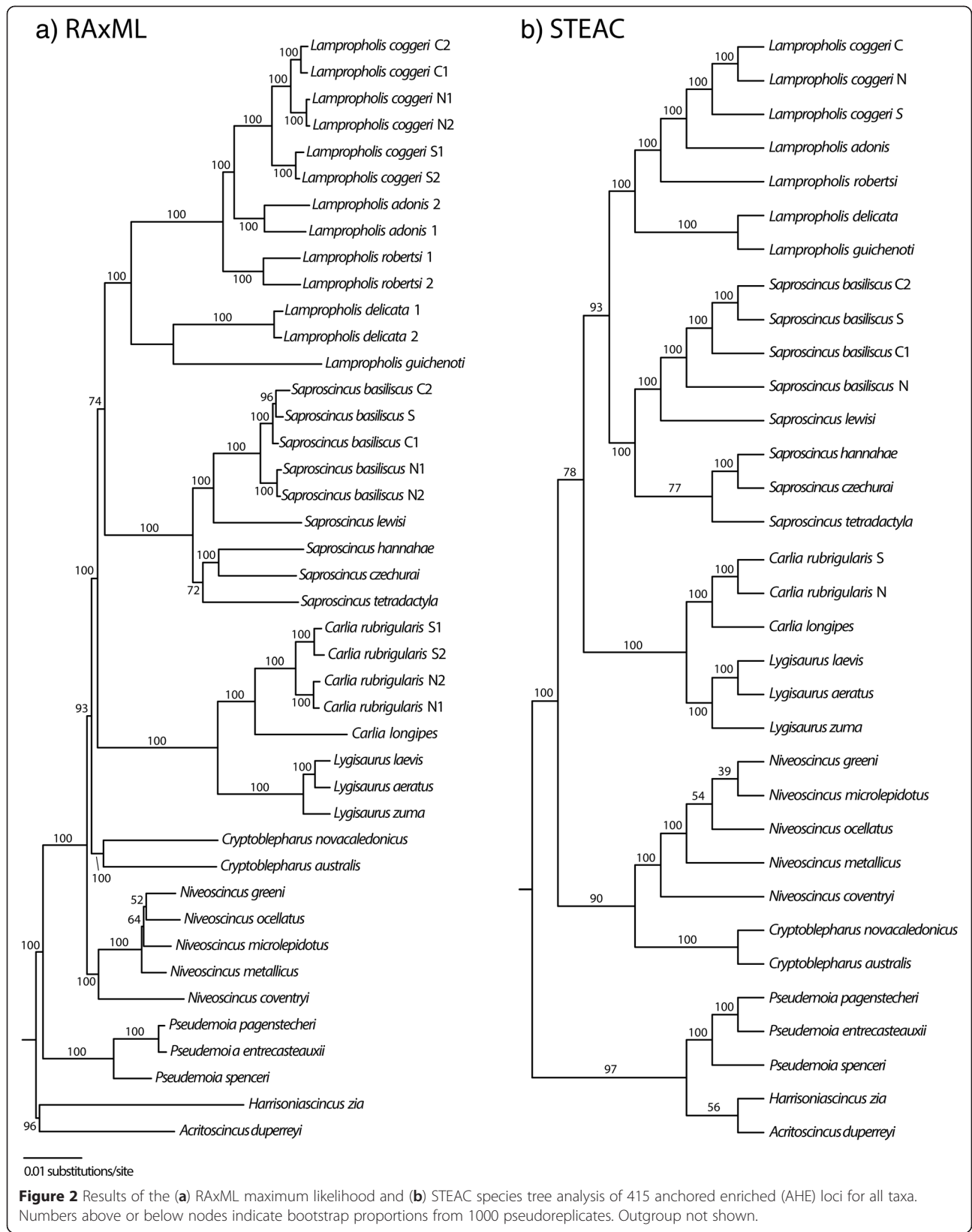
Discussion

Information content of phylogenomic data sets

We evaluated the phylogenetic information content of the anchored-enrichment capture approach (AHE) at multiple scales of divergence in Australian *Eugongylus* group skinks. By focusing on phylogeographic lineages that have been the focus of previous demographic inference (using multiple introns) and comparative transcriptome sequencing, we are able to gauge the performance of AHE loci against a broader sample of genes with varying rates of divergence.

Performance of anchored hybrid enrichment and transcriptome mining

When using conserved targets such as those employed in AHE, we expected to see low phylogenetic resolution among phylogeographic lineages and higher resolution towards the base of the tree, but we did not uncover this pattern. Rather, relationships among phylogeographic lineages and congeneric species were well resolved, except the Tasmanian *Niveoscincus*. This genus aside, the overall good performance of AHE loci across closely related taxa reflects the presence of reasonable levels of diversity within, as well as adjacent to, the “conserved” AE probe regions. However, ambiguity remains for some deeper relationships in the *Eugongylus* group and phylogenies inferred from concatenation and species tree approaches strongly supported alternative placements of *Cryptoblepharus* and *Pseudemoia*. A potential explanation for this conflict and/or lack of resolution is that current phylogenetic methods cannot adequately model the complex processes of molecular evolution inherent in such large data sets. However, the generally short and poorly resolved branches among genera suggest a rapid radiation subsequent to colonization of Australia (as suggested for *Carlia* [30]), in which case there is strong potential for confounding of phylogenetic signal due to



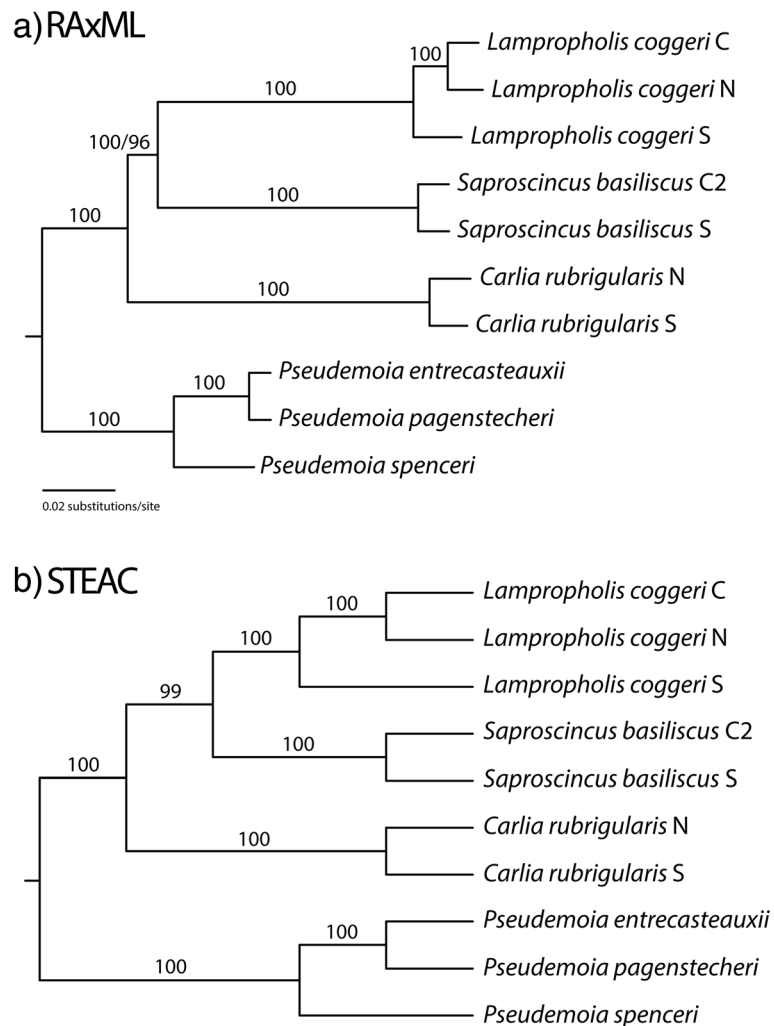


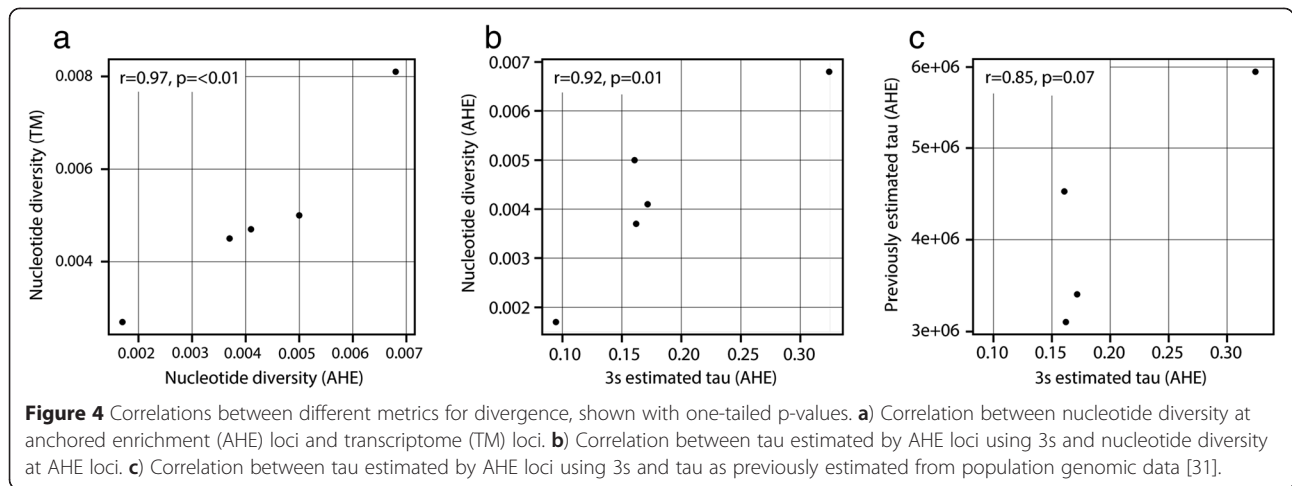
Figure 3 Results of the (a) RAXML and (b) STEAC analyses of all 1650 loci selected by our transcriptome mining (TM) analysis and the 415 anchored enriched (AHE) loci with taxa pruned the 10 ingroup and one outgroup taxa present in the TM data set. Numbers above the nodes in the RAXML tree indicate bootstrap proportions from 1000 pseudoreplicates. Clade support was identical between the TM and pruned AHE analyses, with the exception of the *Lampropholis* + *Saproscincus* clade that was supported by a bootstrap proportion of 100 and 96, respectively. We used PHYML maximum likelihood analyses to infer trees from 1000 bootstrap pseudoreplicates per locus. We used 1000 bootstrapped trees per locus (1,650,000 total trees) as input trees for the STEAC analysis, and numbers above or below the nodes indicate the proportion of times STEAC inferred that clade. Clade support was identical between the TM and pruned AHE analyses. Outgroup not shown.

incomplete lineage sorting across gene trees [63,64]. Because concatenation does not account for conflict due to incomplete lineage sorting, the STEAC tree is our preferred hypothesis of phylogenetic relationships of the Australian *Eugongylus* group species. However, this hypothesis should be subject to further testing with additional nucleotide data and broader sampling of taxa.

Anchored-enriched loci

The key and positive result is that AHE loci provided mostly high phylogenetic resolution among closely-related species and phylogeographic lineages. This parallels the reported utility of another targeted enrichment

approach – UCEs – for analyses of phylogeography [17]. Our primary test cases concerned phylogeographic lineages within three species of skinks from the Australian Wet Tropics rainforests (*Carlia rubrigularis*, *Lampropholis coggeri* and *Saproscincus basiliscus*) that have been the subject of multilocus analyses of phylogeography, historical demography and reproductive isolation in contact zones in the rainforests of north-east Australia [23-25,28,29]. The results of our analyses of the AHE loci add substantial support to previous analyses of 6–10 mostly intronic loci that found deep (~ Late Miocene to Pliocene) divergence and strong reproductive isolation between the two major lineages of *C. rubrigularis*



and between the southern (“S”) and central (“C”) lineages of *La. coggeri*, and more recent divergence and weaker isolation between the C and northern (“N”) lineages of *La. coggeri* and the N and C lineages of *S. basiliscus*. The southern “S” lineage of *S. basiliscus* represents a case of strong discordance between deep mtDNA divergence and shallow nDNA separation from the adjacent (but now isolated) central “C” lineage; ABC coalescent modeling inferred that this was due to recent range expansion of C and introgression [28].

Further bolstering AHE loci’s utility for phylogeographic studies, we found that AHE loci yield sequence divergence estimates strongly correlated with those from TM data, and inferred divergence histories reasonably consistent with previous population-level studies of intron sequences. Often, researchers use rapidly evolving markers like introns to infer divergence history, with the assumption that such markers have substitution rates that allow us to investigate recent history [65]. Here, we show that AHE loci, despite largely being conserved across deep phylogenetic scales, are sufficiently informative that they can be used to estimate such divergence histories. That said, we note that our inference deviated from previous population genomic analyses. First, the AHE loci supported divergence with gene flow where previous inference did not, and second, the correlations, while on trend, are weak. These deviations could be due to differences in model assumptions between inference methods – i.e., how sensitive a method is to recombination, expectations for how gene flow should affect coalescent trees and times – and/or such discrepancies can arise simply from differences in bioinformatic pipelines [66]?

The phylogeny inferred from AHE loci in this study is identical to that obtained using all 1650 loci derived from transcriptome mining. This was the case even for *Saproscincus* lineages for which the information content

of AHE loci and tree confidence factors were relatively low and noisy in relation to PIC scores. The stand-out exception is the poor resolution among closely-related species of Tasmanian *Niveoscincus*, evidently also a recent and rapid radiation (see also [67]). Ongoing improvements to the design of AHE targets, to increase locus length and information content by targeting adjacent exons and thus capturing intervening introns, will only serve to further increase their utility among closely related taxa (Lemmon and Lemmon, unpublished data).

Given the demonstrated strong performance of AHE loci for resolving a phylogeny at the phylogeographic scale of divergence and inferring divergence histories (and see also [17] for UCEs), along with the previous demonstrations of their utility at deeper phylogenetic scale (e.g., [9]), one could assume that these approaches will suffice for the vast majority of molecular systematics analyses. In particular, AHE loci exemplify *a priori* selection of loci to maximize both phylogenetic breadth of application and phylogenetic signal, both of which are emerging as key factors [1].

One caveat with the AHE (and UCE) approaches is the quality of alignments across non-coding regions adjacent to conserved anchor targets. By definition, conserved elements evolve slowly and therefore are expected to contribute few PICs at the shallow (e.g., phylogeographic) scale (Figure 1) as seen in UCE studies (Figure 2 in [17]). Although interspecific variation is lower close to the AHE target region compared to immediate flanking regions, there is nonetheless an intermediate level of variation as ~35% or more loci have at least one PIC close to the target region (Figure 1). As expected, the majority of PICs are from more variable flanking regions. On the other hand, sequencing coverage drops the farther from the anchor region, thereby increasing the possibility that variable sites are simply artifacts of sequencing

error and low coverage (Figure 1). However, one could largely mitigate this problem by trimming flanking data so that all or most taxa share the same set of characters, and removing all gapped characters as we did in this study.

There will remain situations where broader sampling of exons is desirable, either to attempt to resolve short internodes, whether shallow or deep in the tree, or to estimate historical demography and divergence processes. Phylogenetic and species-delimitation analyses of genome-scale data remain constrained by computational limits, though there are promising developments in inference methods that use independent SNPs across thousands of loci [68,69].

Transcriptome mining

Transcriptome mining can potentially provide 1000s of loci with many PICs (see Supp Info 2) and can incorporate transcriptomes for phylogenetic analysis that were otherwise collected for other gene expression projects (e.g., [38]). Exons do have the advantages of more secure alignment via open reading frames, a broad spectrum of variability, and potential to link the frameworks of molecular evolution and phylogenetics [70]. Further, the development of custom exon capture systems from *de novo* assembled transcriptomes is now relatively routine [8] and will become all the more informative when paired with genomes from reasonably closely-related species. Depending on the application, thousands of exons sequenced can be analysed *en masse* (e.g. for population genomic analyses via site frequency spectra), or loci with the highest information content and most coherent modes of sequence evolution can be selected for phylogenetic analyses.

On the other hand, several aspects currently make TM less feasible for large-scale phylogenetics. First, transcriptome sequencing requires high quality RNA, preferably from fresh tissue or preserved in an RNA preservative and stored at $\leq -80^{\circ}\text{C}$. This precludes ethanol-preserved tissues and older tissues that would otherwise preserve DNA. Second, transcriptome sequencing is still expensive when compared to AHE and exon-capture methods. Third, recombination between distant exons within single transcripts could violate assumptions of phylogenetic analyses, e.g., where all SNPs within a single locus are assumed to reflect a common gene tree history. This last phenomenon (termed a mixture tree in [1]) can be mitigated by using only a single exon from each gene.

Phylogeny of the Australian *Eugongylus* group skinks

We inferred phylogenies of the Australian *Eugongylus* group lizards with generally high clade support. When compared to other studies including *Eugongylus* group taxa, several relationships match our expectations such

as a clade including *Carlia* and *Lygisaurus* [30], the interrelationships of *Saproscincus basiliscus*, *S. czechurai*, *S. hannahae*, *S. lewisi*, and *S. tetradactyla* (the “northern lineage” of [71]), and a putative rapid radiation early in the Australian *Eugongylus* group’s history [21]. However, because there is no previous comprehensive molecular phylogenetic analysis of the group, most of our results are novel.

Despite using 415 loci (or subsets with the many PICs), some relationships remain unresolved or conflicting. The relationships of the Tasmanian species of *Niveoscincus* remain unknown. This lack of resolution likely represents a rapid radiation and that there is still insufficient amount of data (characters or taxa) to reconstruct the evolutionary history of these species. There is strong support for conflicting placements of both *Cryptoblepharus* and *Pseudemoia* between the concatenated data and STEAC analyses (Figures 3 and 4) reflecting the complex process of DNA evolution in these taxa as well. These conflicting or unresolved relationships continue to prohibit a complete understanding of placental evolution in *Niveoscincus* and *Pseudemoia* which exhibit a range of placental complexity from relatively simple placenta likely used for gas exchange and minimal maternal provisioning, to complex placentae with substantial maternal-to-embryo transfer [32,72]. Although TM performed well, it is not currently feasible to sequence transcriptomes for the remaining ~ 140 species of the group. Given the promising performance of AHE shown here, it is likely that further sampling across and within genera will do much to resolve the early phylogenetic history of the group.

Conclusions

Our study demonstrates that, overall, anchored-enriched loci are informative at the intraspecific phylogeographic scale as well as deeper in the tree of life. Transcriptome mining provides a wealth of informative loci, but unlike hybrid enrichment methods, is less practical when scaling up to 100 s of taxa unless obtained via a capture approach [8]. In practice, we suggest a mixed approach that uses standard systems such as AHE or UCE to obtain initial results, and then, if warranted, more extensive capture of exons or anonymous loci for more detailed analyses of rapid radiations, divergence histories and population genomics (including for museum specimens [73]). Alternatively, mixed capture kits can be designed to span the scales of divergence to be studied for various projects. Design and testing of such capture systems containing mixtures of anchored loci (centered in exons) and anonymous loci (centered in randomly-chosen parts of the genome) are already underway in numerous animal and plant clades. These systems produce a spectrum of loci that are informative at the deep phylogenetic, phylogeographic, and population genetic scales.

Availability of supporting data

The data sets and scripts supporting the results of this article are available in the Dryad repository, doi:10.5061/dryad.g4mj2.

Additional files

Additional file 1: Localities of specimens used in this analysis.

Additional file 2: Tables providing further information on demographic history inference. Table S1. Information about the anchored enrichment (AHE) sequence alignments used for inferring divergence history with the program 3s. Assembly used refers to the reference assembly to which all reads in the trio (the lineage-pair and the outgroup) were mapped. **Table S2.** Log likelihoods for the three models fit to data in the 3s program. M0 assumes an isolation model without migration, and M1 and M2 assume an isolation model with migration. Significant likelihood ratio tests (LRT) critical values are 2.71 for comparing model M1 to M0 and 5.99 for comparing model M2 to M0.

Additional file 3: Length, percent identity, and number of parsimony-informative characters (PICs) for the 415 anchored-enriched loci for the full taxon, *Lampropholis* only and *Saproscincus* only data sets.

Additional file 4: Length, percent identity, and number of parsimony-informative characters (PICs) for the 1650 transcriptome mined exons for the full taxon, *Lampropholis* only, and *Saproscincus* only data sets.

Additional file 5: Phylogenies for phylogeographic lineages within *Lampropholis* and *Saproscincus*. (A) The concatenated RAxML tree for *Lampropholis* lineages. (B) The STEAC species tree for *Lampropholis* lineages. (C) The concatenated RAxML tree for *Saproscincus* lineages. (D) The STEAC species tree for *Saproscincus* lineages. Outgroups not shown.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MCB and CM designed the research. MCB, CKJ, SS, ARL, and EML performed laboratory work. ARL and EML performed sequencing and data preparation. MCB, JGB, ARL, and SS wrote portions of the analytical pipeline. MCB, JG, and SS performed the analyses. MCB, DGC, SS, and MBT did field work to obtain necessary sampling. MCB and CM wrote the paper with input from JGB, DGC, CKJ, ARL, EML, and SS. All authors read and approved the final manuscript.

Acknowledgements

We thank L. Liu for assistance with STEAC analyses; D. Warren for assisting MCB with Perl scripting; J. Mai for computational assistance; Intersect Australia Ltd. for supercomputing resources; O. Griffith for access to unpublished *Pseudemoia* transcriptome data; S. Emme and H. Ralicki for assistance with AHE data collection; O. Griffith, M. Laird, A. Seago, J. Van Dyke, S. E. Williams and Moritz lab members for fieldwork assistance; and two anonymous reviewers for comments on the manuscript. MCB thanks S. Keogh and the Research School of Biology at the Australian National University for temporary accommodation. MCB was supported by the Discovery Early Career Researcher Award from the Australian Research Council (ARC; DE120101615), DGC by ARC DP0771913 and CM and JB by ARC LP110100104. Specimens were collected with permits from Queensland (WITK07252610), New South Wales (SL100401), South Australia (Q25988-1), and tissue collection was approved by the ANU (A2103/08), UC Berkeley (Moritz-R278) and University of Sydney Animal Ethics Committees (2010/1014 5409 and 2011/1013 5636).

Author details

¹School of Biological Sciences, Heydon-Laurence Building A08, University of Sydney, Sydney, NSW 2006, Australia. ²New York University - Sydney, The Rocks, NSW 2000, Australia. ³Research School of Biology and Centre for Biodiversity Analysis, The Australian National University, Canberra ACT 0200, Australia. ⁴Museum of Vertebrate Zoology, University of California, 3101

Valley Life Sciences Building, Berkeley, CA 94720, USA. ⁵Department of Integrative Biology, University of California, 3060 Valley Life Sciences Building, Berkeley, CA 94720, USA. ⁶School of Biological Sciences, Monash University, Clayton, Melbourne, VIC 3800, Australia. ⁷Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306, USA. ⁸Department of Biological Science, Florida State University, 319 Stadium Drive, PO Box 3064295, Tallahassee, FL 32306, USA. ⁹The Commonwealth Scientific and Industrial Research Organization Ecosystem Sciences Division, GPO Box 1700, Canberra, ACT 2601, Australia.

Received: 1 October 2014 Accepted: 24 February 2015

Published online: 11 April 2015

References

- Lemmon EM, Lemmon AR. High-throughput genomic data in systematics and phylogenetics. *Ann Rev Ecol Evol Syst.* 2013;44:99–121.
- Faircloth BC, Sorenson L, Santini F, Alfaro ME. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLoS ONE.* 2013;8:e65923.
- Johnson BR, Borowiec ML, Chiu JC, Lee EK, Atallah J, Ward PS. Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. *Curr Biol.* 2013;23:2565.
- Simon S, Narechania A, DeSalle R, Hadrys H. Insect phylogenomics: exploring the source of incongruence using new transcriptomic data. *Genome Biol Evol.* 2012;4:1295–309.
- Wen J, Xiong Z, Nie Z-L, Mao L, Zhu Y, Kan X-Z et al. Transcriptome sequences resolve deep relationships of the grass family. *PLoS ONE* 2013, e74394.
- Peters RS, Meusemann K, Petersen M, Mayer C, Wilbrandt J, Ziesmann T, et al. The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. *BMC Evol Biol.* 2013;14:52.
- Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol.* 2012;61:717–26.
- Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics.* 2012;13:403.
- Lemmon AR, Emme SA, Lemmon EM. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst Biol.* 2012;61:727–44.
- Bybee SM, Bracken-Grissom HD, Hermansen RA, Clement MJ, Crandall KA, Felder DL. Directed next generation sequencing for phylogenetics: an example using Decapoda (Crustacea). *Zool Anz.* 2011;250:497–506.
- O'Neill EM, Schwartz R, Bullock CT, Williams JS, Shaffer HB, Aguilar-Miguel X, et al. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Mol Ecol.* 2013;22:111–29.
- Barrow LN, Ralicki HF, Emme SA, Moriarty Lemmon E. Species tree estimation of North American chorus frogs (Hylidae: *Pseudacris*) with parallel tagged amplicon sequencing. *Mol Phylogenet Evol.* 2014;75:78–90.
- Miller MR, Atwood TS, Eames BF, Eberhart JK, Yan Y-L, Postlethwait JH, et al. RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biol.* 2007;8:R105.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE.* 2008;3:e3376.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE.* 2012;7:e37135.
- Peñalba JV, Smith LL, Tonione MA, Sass C, Hykin SM, Skipwith PL, et al. Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Mol Ecol.* 2014;14:1000–10.
- Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst Biol.* 2014;63:83–95.
- Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, et al. Ultraconserved elements in the human genome. *Science.* 2004;304:1321–5.

19. Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, Wilson RK, et al. Human genome ultraconserved elements are ultraselected. *Science*. 2007;317:915–5.
20. Smith SA, Sadlier RA, Bauer AM, Austin CC, Jackman T. Molecular phylogeny of the scincid lizards of New Caledonia and adjacent areas: evidence for a single origin of the endemic skinks of Tasmantis. *Mol Phylogenet Evol*. 2007;43:1151–66.
21. Chapple DG, Ritchie PA, Daugherty CH. Origin, diversification, and systematics of the New Zealand skink fauna (Reptilia: Scincidae). *Mol Phylogenet Evol*. 2009;52:470–87.
22. Skinner A, Hugall AF, Hutchinson MN. Lygosomine phylogeny and the origins of Australian scincid lizards. *J Biogeogr*. 2011;38:1044–58.
23. Moussalli A, Moritz C, Williams SE, Carnaval AC. Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeo-distribution models. *Mol Ecol*. 2009;18:483–99.
24. Bell RC, Parra JL, Tonione M, Hoskin CJ, Mackenzie JB, Williams SE, et al. Patterns of persistence and isolation indicate resilience to climate change in montane rainforest lizards. *Mol Ecol*. 2010;19:2531–44.
25. Dolman G, Moritz C. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). *Evolution*. 2006;60:573–82.
26. Chapple DG, Hoskin CJ, Chapple SNJ, Thompson MB. Phylogeographic divergence in the widespread delicate skink (*Lampropholis delicata*) corresponds to dry habitat barriers in eastern Australia. *BMC Evol Biol*. 2011;11:191.
27. Chapple DG, Chapple SNJ, Thompson MB. Biogeographic barriers in south-eastern Australia drive phylogeographic divergence in the garden skink, *Lampropholis guichenoti*. *J Biogeogr*. 2011;38:1761–75.
28. Singhal S, Moritz C. Testing hypotheses for genealogical discordance in a rainforest lizard. *Mol Ecol*. 2012;21:5059–72.
29. Singhal S, Moritz C. Reproductive isolation between phylogeographic lineages scales with divergence. *Proc R Soc B*. 2013;280:20132246.
30. Dolman G, Hugall AF. Combined mitochondrial and nuclear data enhance resolution of a rapid radiation of Australian rainbow skinks (Scincidae: *Carlia*). *Mol Phylogenet Evol*. 2008;49:782–94.
31. Adams SM, Biazik JM, Thompson MB, Murphy CR. Cyto-epitheliochorial placenta of the viviparous lizard *Pseudemoia entrecasteauxii*: a new placental morphotype. *J Morph*. 2005;264:264–76.
32. Thompson MB, Speake BK. A review of the evolution of viviparity in lizards: structure, function and physiology of the placenta. *J Comp Physiol B Biochem Syst Environ Physiol*. 2006;176:179–89.
33. Stewart JR, Thompson MB. Parallel evolution of placentation in Australian scincid lizards. *J Exp Zool B Mol Dev Evol*. 2009;312:590–602.
34. Stewart JR, Thompson MB. Placental ontogeny in Tasmanian snow skinks (Genus *Niveoscincus*) (Lacertilia: Scincidae). *J Morph*. 2009;270:485–516.
35. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*. 2011;6:pdb prot5448.
36. Rokyta DR, Lemmon AR, Margres MJ, Arnow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics*. 2012;13:312.
37. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and throughput. *Nucleic Acids Res*. 2004;32:1792–7.
38. Brandley MC, Young RL, Warren DL, Thompson MB, Wagner GP. Uterine gene expression in the live-bearing lizard, *Chalcides ocellatus*, reveals convergence of squamate reptile and mammalian pregnancy mechanisms. *Genome Biol Evol*. 2012;4:394–411.
39. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 2009;19:1117–23.
40. Singhal S. De novo transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. *Mol Ecol Resources*. 2013;13:403–16.
41. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011;27:2957–63.
42. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
43. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotech*. 2011;29:644–52.
44. Kent WJ. BLAT—The BLAST-Like Alignment Tool. *Genome Res*. 2002;12:656–64.
45. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.
46. Huang X, Madan A. CAP3: a DNA sequence assembly program. *I*. 1999;9:868–77.
47. Nylander JAA. MrAIC.pl, program distributed by the author. Evolutionary Biology Centre. Uppsala University.
48. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
49. Stamatakis A. RAxML-VI-HPC: maximum likelihood based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22:2688–90.
50. Kubatko LS, Degnan JH. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol*. 2007;56:17–24.
51. Belfiore NM, Liu L, Moritz C. Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). *Syst Biol*. 2008;57:294–310.
52. Weisrock DW, Smith SD, Chan LM, Biebouw K, Kappeler PM, Yoder AD. Mouse lemur phylogeny: an empirical demonstration of the effect of allele sampling in phylogenetics. *Mol Biol Evol*. 2012;29:1615–2630.
53. Liu L, Yu L, Pearl DK, Edwards SV. Estimating species phylogenies using coalescence times among sequences. *Syst Biol*. 2009;58:468–77.
54. Liu L, Yu L. Phybase: an R package for species tree analysis. *Bioinformatics*. 2010;26:962–3.
55. Larget BR, Kotha SK, Dewey CN, Ané C. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*. 2010;26:2910–1.
56. Zhu T, Yang Z. Maximum likelihood implementation of an isolation-with-migration model with three species for testing speciation with gene flow. *Mol Biol Evol*. 2012;29(10):3131–42.
57. Yang Z. A likelihood ratio test of speciation with gene flow using genomic sequence data. *Gen Biol Evol*. 2010;2:200–11.
58. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
60. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
61. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Mol Biol Evol*. 1992;9:678–87.
62. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 2009;5:e1000695.
63. Degnan JH, Rosenberg NA. Discordance of species trees with their most likely gene trees. *PLoS Genet*. 2006;2:e68.
64. Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol*. 2009;24:332–40.
65. Brito PH, Edwards SV. Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*. 2009;135:439–55.
66. Zwickl DJ, Stein JC, Wing RA, Ware D, Sanderson MJ. Disentangling Methodological and Biological Sources of Gene Tree Discordance on *Oryza* (Poaceae) Chromosome 3. *Syst Biol*. 2014;63:645–59.
67. Melville J, Swain R. Mitochondrial DNA-sequence based phylogeny and biogeography of the snow skinks (Squamata: Scincidae: *Niveoscincus*) of Tasmania. *Herpetol*. 2000;56:196–208.
68. Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol*. 2012;29:1917–32.
69. Mirarab S, Bayzid MS, Boussau B, Warnow T. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science*. 2014;29:1250463.
70. Cutter AD. Integrating phylogenetics, phylogeography and population genetics through genomes and evolutionary theory. *Mol Phylogenet Evol*. 2013;69:1172–85.

71. Moussalli A, Hugall AF, Moritz C. A mitochondrial phylogeny of the rainforest skink genus *Saproscincus*, Wells and Wellington (1984). *Mol Phylogenet Evol.* 2005;34:190–202.
72. Van Dyke JU, Brandley MC, Thompson MB. The evolution of viviparity: molecular and genomic data from squamate reptiles advance understanding of live birth in amniotes. *Reproduction.* 2014;147:R15–26.
73. Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. Unlocking the vault: next-generation museum population genomics. *Mol Ecol.* 2013;22:6018–32.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

