

On the universality of target-enrichment baits for phylogenomic research

Silas Bossert  | Bryan N. Danforth

Department of Entomology, Cornell University, Ithaca, NY, USA

CorrespondenceSilas Bossert
Email: sb2346@cornell.edu**Funding information**

National Science Foundation, Grant/Award Number: DEB-1555905

Abstract

1. Capturing conserved genomic elements to shed light on deep evolutionary history is becoming the new gold standard for phylogenomic research. Ultraconserved elements are shared among distantly related organisms, allowing the capture of unprecedented amounts of genomic data of non-model taxa.
2. An underappreciated consequence of hybrid enrichment methods is the potential of introducing undetected DNA sequences from organisms outside the lineage of interest, facilitated through the high degree of conservation of the target regions. In this *in silico* study, we quantify ultraconserved loci using a data set of 400 published genomes. We utilized six newly designed UCE bait sets, tailored to various arthropod groups, and screened for shared conserved elements in all 242 currently published arthropod genomes. Additionally, we included a diverse set of other potential contaminating organisms, such as various species of fungi and bacteria.
3. Our results show that specific UCE bait sets can capture genomic elements from vastly divergent lineages, including human DNA. Nonetheless, our *in silico* modeling demonstrates that sufficiently strict bioinformatic processing parameters effectively filter out unintentionally targeted DNA from taxa other than the focus group. Lastly, we characterize all the 100 most widely shared UCE loci as highly conserved exonic regions.
4. We give practical recommendations to address contamination in data sets generated through targeted-enrichment.

KEYWORDS

contamination, phylogenomics, target enrichment, ultraconserved elements

1 | INTRODUCTION

Rapid advances of sequencing technologies facilitate the acquisition of DNA sequence data for phylogenetic research (Jennings, 2016; McCormack, Hird, Zellmer, Carstens, & Brumfield, 2013). An increasingly popular method for studying deep evolutionary divergences is the use of hybrid capture techniques to sequence conserved genomic anchor regions, such as Ultraconserved Elements (UCEs; Faircloth et al., 2012) and Anchor Enriched loci (AE; Lemmon, Emme,

& Moriarty Lemmon, 2012). Both are *reduced-representation* approaches—they rely on the extraction and analyses of a phylogenetically informative subset of the target organisms' genomes. Specifically designed nucleotide baits hybridize *in-solution* with targeted anchor sequences. These *anchors* are subsequently enriched and prepared for parallel sequencing. The target regions are conserved and shared among distantly related taxa, thereby overcoming the need for prior genomic information once a bait set is designed (Faircloth et al., 2012; Lemmon et al., 2012). This allows cost- and time-effective generation

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society

of genome-scale data for non-model organisms (Glenn & Faircloth, 2016).

We have only just begun to understand the nature of ultraconserved elements. The majority of the initially discovered UCEs from vertebrates were characterized as transcriptionally inactive elements, with a smaller proportion of conserved exonic regions (Bejerano et al., 2004). Among others, they were linked with diverse regulatory functions in splicing, transcription, enhancement and development (Bejerano et al., 2004; Dickel et al., 2014; Reneker et al., 2012; Woolfe et al., 2004). However, functional aspects of UCEs are almost exclusively studied in vertebrate systems, and their significance in invertebrate genomes is largely unknown.

Nonetheless, the application of UCEs for phylogenetic research in invertebrate systems is highly promising. The advantages of analysing hundreds to thousands of informative loci outcompete traditional Sanger sequencing (Blaimer et al., 2015), and open exciting avenues to overcome the persistent challenge of deep coalescences (McCormack et al., 2012; McCormack, Harvey, et al., 2013; Meiklejohn, Faircloth, Glenn, Kimball, & Braun, 2016). A major advance is the capability to recover informative sequence data from degraded starting material. Recent studies showed successful UCE capture from century old, dried museum specimens (Blaimer, Lloyd, Guillory, & Brady, 2016; McCormack, Tsai, & Faircloth, 2016), and even formalin-fixed material (Ruane & Austin, 2017). This renders the vast historical archives of dried, pinned, or ethanol preserved material in natural history museums available for use in phylogenomic analysis, including long-extinct species and irreplaceable specimens. This is an appealing advantage over the needs of RNA-seq, which relies on freshly collected specimens with intact mRNA (Yeates, Meusemann, Trautwein, Wiegmann, & Zwick, 2016).

A central aspect of hybrid capture techniques is the design of nucleotide bait sets that capture the targeted DNA *in-solution*. The principal UCE bait design workflow is based on genome–genome alignments of two or more representative genomes, and screens for 40–60 bp long identical anchor sequences (Faircloth, Branstetter, White, & Brady, 2015; Faircloth et al., 2012). A “hands-on” description of the workflow was recently published in this journal (Faircloth, 2017). The ability of the baits to hybridize *in-solution* is tested *in-silico* by aligning them to sequenced genomes. Hereby, a minimum of 80% sequence overlap and identity is used to predict good *in-solution* capture success (Branstetter, Longino, Ward, & Faircloth, 2017; Faircloth, 2017; Faircloth et al., 2015).

UCE bait sets are tailored to certain taxonomic groups, such as Hymenoptera (Branstetter, Longino, et al., 2017; Faircloth et al., 2015), arachnids (Starrett et al., 2016) or amniotes (Faircloth et al., 2012; Hosner, Faircloth, Glenn, Braun, & Kimball, 2016; Streicher & Wiens, 2016). However, we do not know to what extent these conserved sequences are shared among lineages outside the taxon of interest. More specifically, we do not know if these tailored UCE baits can capture significant amounts of contaminating sequences from organisms outside the target lineages. In this case study, we quantify these non-target UCE loci in a diverse set of organisms. We employed genomes of 400 different taxa and screened them

for sequences that match UCE baits of 6 newly designed bait sets, each targeting a separate arthropod lineage (Table 1). We conducted phylum-wide searches for these elements in all 242 sequenced, non-duplicate arthropod genomes that are publicly available on NCBI, as well as in 158 potentially contaminating organisms, such as humans, fungi, and bacteria. On the basis of our results, we evaluate the potential of non-target capture of UCEs *in-solution*. Lastly, we simulate the target capture of contaminating UCEs under varying bioinformatic parameters, and present cut-off values that exclude contamination while retaining a maximized amount of target UCEs.

2 | MATERIALS AND METHODS

We conducted phylum-wide screens for ultraconserved elements of six major arthropod lineages (Table 1) in all 242 arthropod genomes that were publicly available on NCBI at the beginning of March, 2017. We employed genomes of 212 hexapod species, covering 15 of 29 recognized insect orders, as well as 9 Crustacea, 20 Chelicerata, and the only sequenced myriapod taxon *Strigamia maritima* (Table S1). These lineages likely diverged as early as c. 570 million years ago (Misof et al., 2014). To evaluate the potential of including contaminating DNA during laboratory work, we searched for matching sequences in the genomes of 10 domesticated mammal species, such as cat, dog, rat, as well as two copies of the human genome. We further screened the genomes of 72 species of fungi that could be naturally associated with the extraction material, such as a wide range of different yeasts and moulds, as well as genera that have been identified as contaminants in ancient DNA studies (Austin, Ross, Smith, Fortey, & Thomas, 1997; Wang, Yan, & Jin, 1997). Lastly, we added 74 bacterial genomes that could in practice become part of the starting material of the UCE-enrichment process, such as insect gut symbionts, different strains of *Wolbachia*, as well as a range of human gut microbiota and typical bacterial contaminants in microbiological laboratories. In cases of multiple genomes of single taxa, we retained the genomes with the highest sequencing depth.

We first simulated the potential of a new Hymenoptera bait set (Branstetter, Longino, et al., 2017) to capture UCEs *in-solution*. We used LASTZ (Harris, 2007), as implemented in PHYLUCE (Faircloth, 2016),

TABLE 1 The UCE bait sets used in this study and the respective references

Targeted lineage	References
Arachnida	Starrett et al. (2016) and Faircloth (2017)
Coleoptera	Faircloth (2017)
Diptera	Faircloth (2017)
Hemiptera	Faircloth (2017)
Hymenoptera	Branstetter, Longino, et al., (2017)
Lepidoptera	Faircloth (2017)

to match each bait against each of the 400 genomes. We required a minimum sequence identity of 80% across an overlap region of 80% to score a positive hit. To explore the possibility of *in silico* filtering for contaminating UCEs, we then quantified the number of matches for each genome along a gradient from 65% to 95% sequence identity, while keeping the minimum overlap of 80%. Subsequently, we examined the relationship of genome size and capture success by plotting the amount of *in silico* matched UCEs as a function of the total sequence length of each genome. Lastly, we repeated the matching simulations with each of the other five bait sets (Table 1). We tested for significant positive associations between genome size and the number of UCE loci captured using standard linear regression methods implemented in R (R Development Core Team, 2016). The included genomes as well as their GenBank assembly accession numbers are listed in the appendix (Table S1).

3 | RESULTS AND DISCUSSION

Our results show that bait sets tailored to specific groups can effectively capture conserved genomic elements from organisms outside the targeted lineage. All six bait sets target hundreds of loci from all of the included insect orders, and consistently match several hundreds of elements in all other examined arthropod lineages (Figures 1 and 2). For three of the five tested sets, the amount of matching insect UCEs outside the targeted group is strongly predicted by genome size rather than relatedness (Figure 3). However, this trend is not present for the bait sets designed for Diptera and Lepidoptera. As the bait sets for all target groups were designed in a comparable manner (Faircloth, 2017), these differing trends are

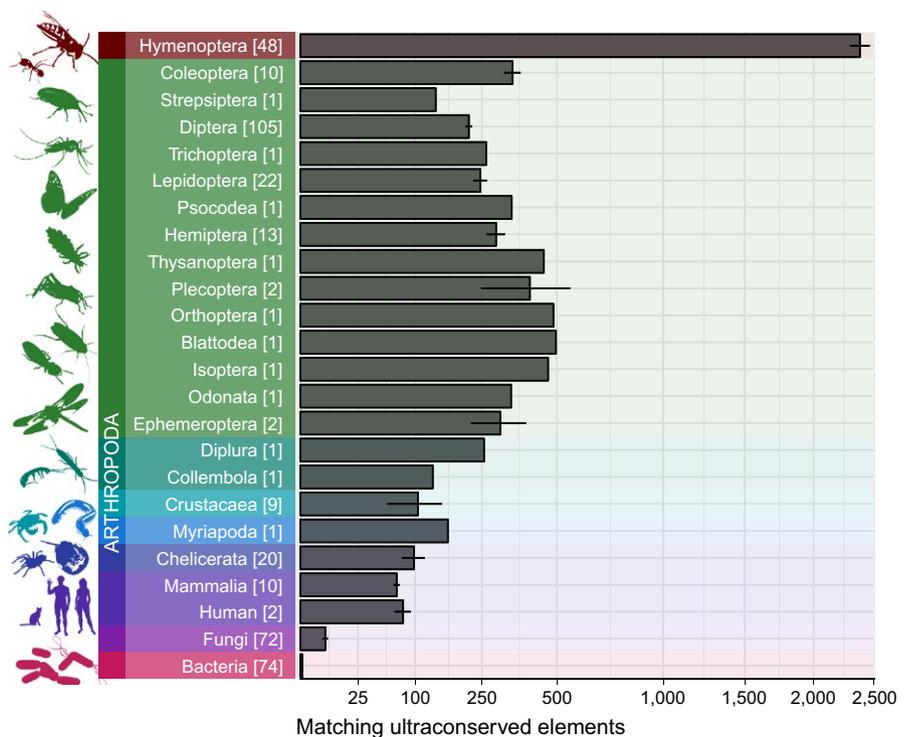
unexpected and further data exploration is required to clarify the underlying mechanisms that drive the different patterns.

Our results demonstrate that single bait sets can capture putative orthologous sequences from the entirety of the most diverse extant phylum Arthropoda, encompassing over 1,200,000 described taxa and about 80% of all known species of animals (Zhang, 2011). Even if none of the bait sets were specifically tailored towards a breadth of arthropod lineages, we identified loci that are universally shared throughout the phylum, representing highly promising candidate loci for a universal arthropod bait set.

We characterize the 100 most widely shared UCEs as exclusively conserved exons or partially exonic regions (Table S2). This sheds new light on their evolutionary significance. The initial discoveries of UCEs from vertebrates characterize them as predominantly non-coding sequences (Bejerano et al., 2004; Woolfe et al., 2004). In contrast, the gene fragments identified in this study code for essential proteins with various functions, among them transcription factors, motor proteins, membrane transport proteins, and various house-keeping genes. In addition, several baits matched highly conserved reading frames of yet uncharacterized proteins. These findings indicated that UCEs developed for arthropods are primarily composed of protein-coding genes that face strong purifying selection, and that their function differs from the initially identified non-coding vertebrate UCEs.

While it is promising that a single bait set captures orthologs from all included arthropod lineages, our findings also shed light on the potential for contamination. Contamination in phylogenomic datasets is not at all a new issue, and has been previously addressed in other pipelines, such as PRINSEQ (Schmieder & Edwards, 2011). However, we argue that contaminating UCEs are a previously overlooked issue

FIGURE 1 The number of conserved loci matching to the Hymenoptera-derived UCE bait set (Branstetter, Longino, et al., 2017). We used 400 genomes and scored positive hits based on 80% sequence overlap and identity. Putative orthologous sequences are shared among all included arthropod lineages. Taxon groups are organized by phylogenetic distance to Hymenoptera (according to Misof et al., 2014), and quotation marks indicate paraphyletic groupings. Values in brackets refer to the number of employed genomes. Error bars indicate 95% CI of sample means if $n > 1$



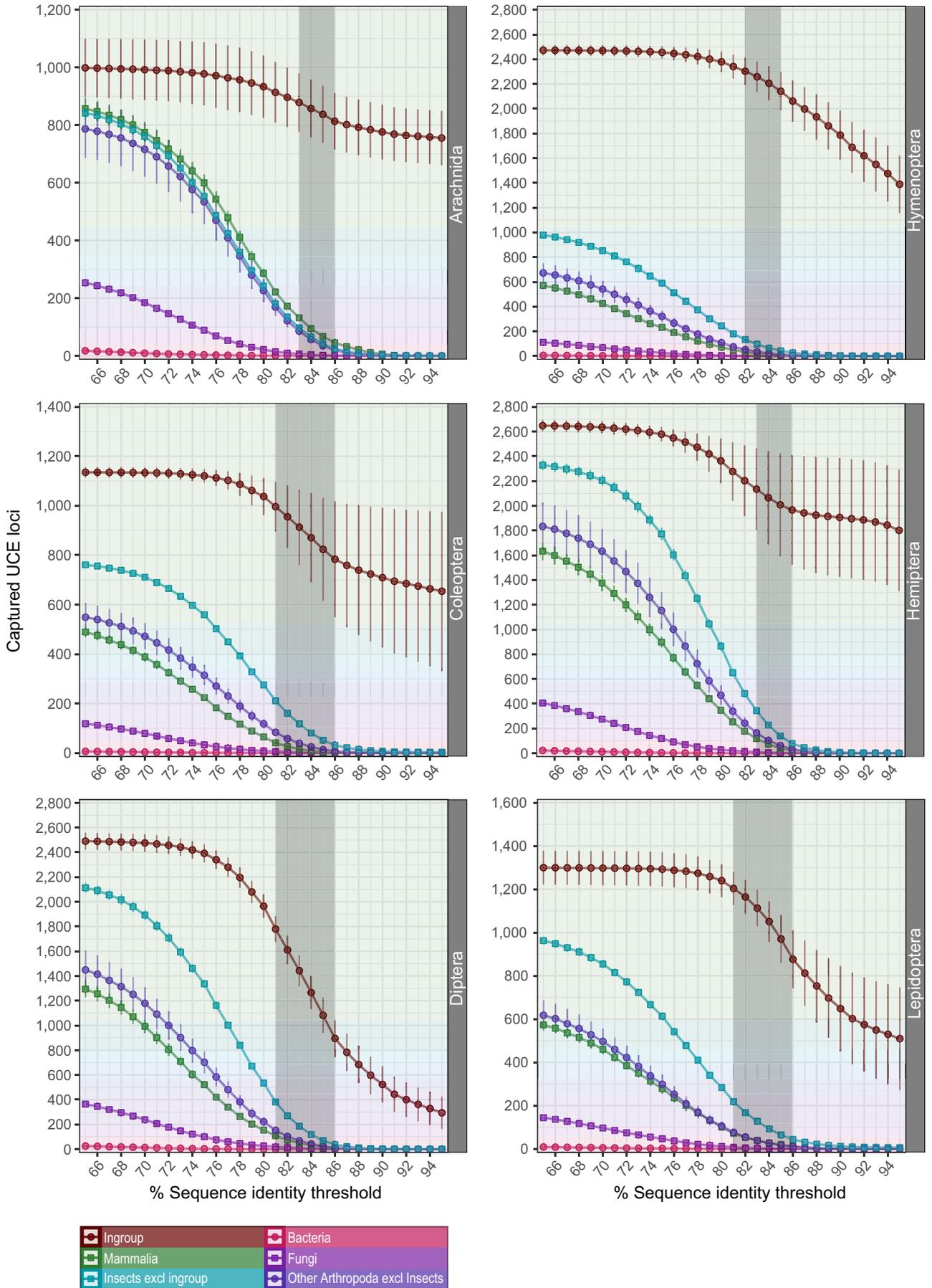


FIGURE 2 The simulated matching success of UCE baits to the employed genomes under varying conditions of sequence identity. The respective bait sets are indicated by the taxon names. Quotation marks refer to paraphyletic groups. Shaded areas indicate recommended cut-off thresholds. Dots show mean values and error bars indicate $\pm 2 \times SE$

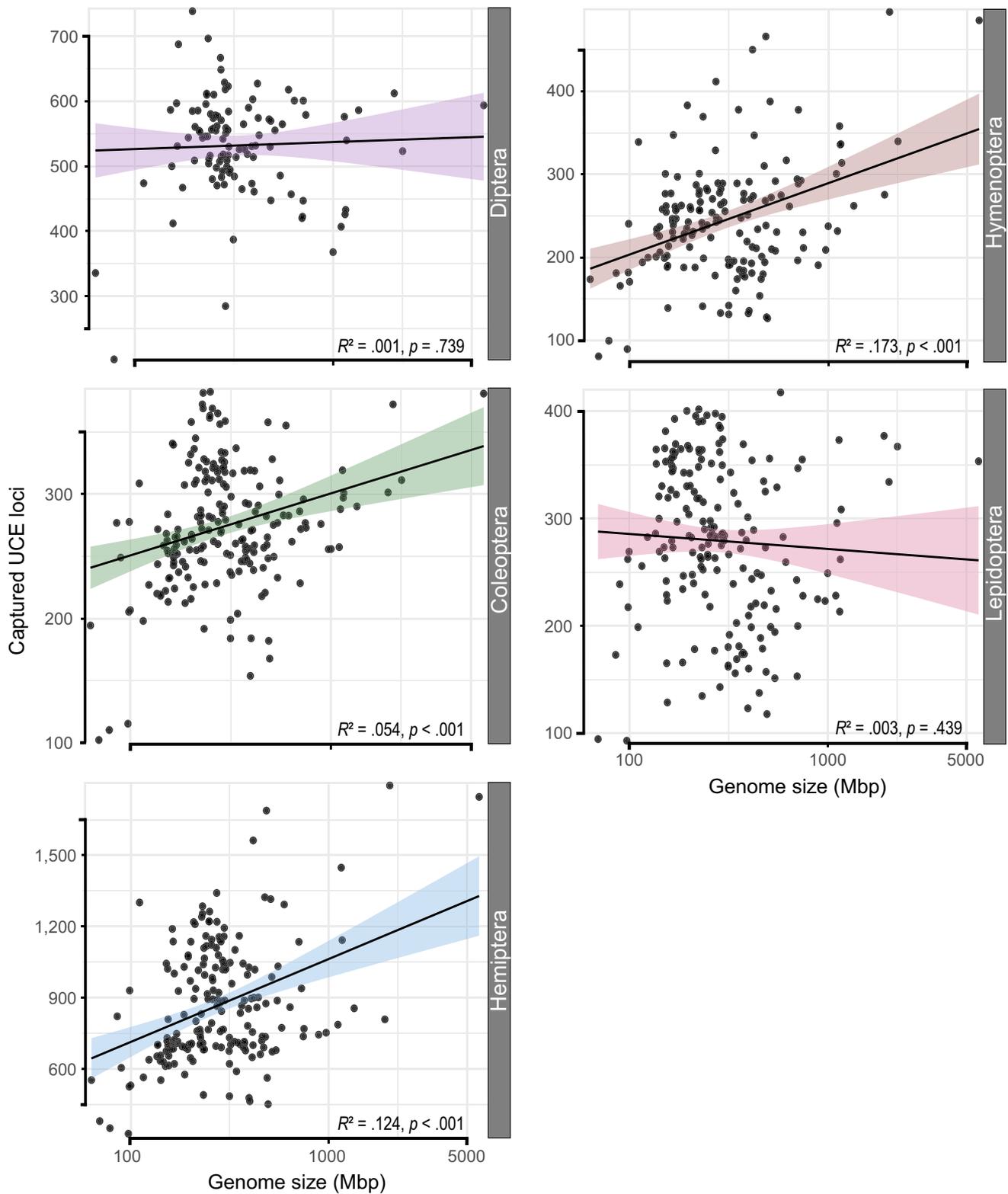


FIGURE 3 The relationship of genome size and matching UCEs among the included insects. The respective bait sets are indicated by the order names and ingroup taxa were excluded for each plot (i.e. the Hemiptera bait set was used on all insect genomes except for Hemipterans). Shaded areas show 95% CI of the linear regressions

of significant importance, as they can distort matrix processing and produce misleading estimates of organismal phylogenies. Below, we provide recommendations for addressing the issue of unintentionally targeted sequences when generating UCE datasets.

The three principle scenarios for contamination when using the UCE-targeted-enrichment methods are the following: (1) The same UCE locus is captured from the target taxon and the contaminant. If the consensus assemblies produce two different contigs for the same locus, the UCE software pipeline *PHYLUCE* (Faircloth, 2016) identifies these loci as duplicates and discards them. This does not lead to the introduction of the contaminating locus into the final matrix, but to the loss of the entire UCE. (2) The same locus is captured from both taxa, but the consensus assembly produces one chimeric contig. This contaminated UCE will be assigned to the target taxon. (3) The captured locus stems only from the contaminant and not from the target taxon. *PHYLUCE* will incorrectly assign the contaminating locus to the target.

Chimeric UCES have been primarily identified as a problem of low coverage sequencing (McCormack et al., 2016), and can thus be addressed with greater sequencing efforts. In contrast, increased sequencing will not overcome the problems caused by unintended sequence associations. A solution lies in the *in silico* identification of UCES that follows the actual wet-laboratory capture. Our simulation study shows that sufficiently strict matching parameters can effectively filter out loci from unintentionally included DNA (Figure 2). Minimum sequence identities between 82% and 85% similarity to the capture bait ensure the exclusion of contamination, while retaining sufficient capture rates for all included bait sets. As we employed full genomes of the tested contaminating organisms, we prove the exclusion to be efficient even with the highest possible levels of contamination. However, the capture success within lineages can vary significantly, as exemplified by the Hymenoptera bait set.

Specifically, the capture success for Hymenoptera decreases much more steeply for the parasitic aculeata (“Parasitica”), the paraphyletic Symphyta and the Vespidae (Figure 4).

Nonetheless, our data underline the need to minimize contamination in the first place. Careful lab practice for handling sensitive DNA samples requires thorough surface sterilization of dissecting tools, mounts, and tissues. This will help to overcome surface contamination of specimens handled by many standard collecting techniques used by entomologists, such as malaise trapping, pan trapping, collecting multiple specimens in a single kill jar, or pinning specimens by hand. Conducting laboratory work under a laminar flow hood will decrease the potential for introducing environmental DNA. In addition, a careful selection of tissue types could significantly lower the potential of contamination. This is particularly relevant for organisms with predatory feeding behaviours and potential contamination on appendages that are used to capture prey (Faircloth, B.C., pers. comm.). Non-destructive DNA extraction protocols (i.e. Gilbert, Moore, Melchior, & Worobey, 2007; Kanda, Pflug, Sproul, Dasenko, & Maddison, 2016) have the potential to contribute DNA from internal insect parasitoids or ectoparasites. Lastly, the simulated capture of mammalian loci shows the risk of introducing human DNA during wet-laboratory work. This holds true even if the bait set is tailored to distantly related target taxa.

Cross-contaminations are particularly difficult to identify if both organisms are closely related. Extensive BLAST searches (Altschul, Gish, Miller, Myers, & Lipman, 1990) of assembled contigs or individual reads could identify contamination from distantly related organisms. However, a reliable association of individual contigs to a species within a more closely related group would require a sequenced reference genome of the species, which makes

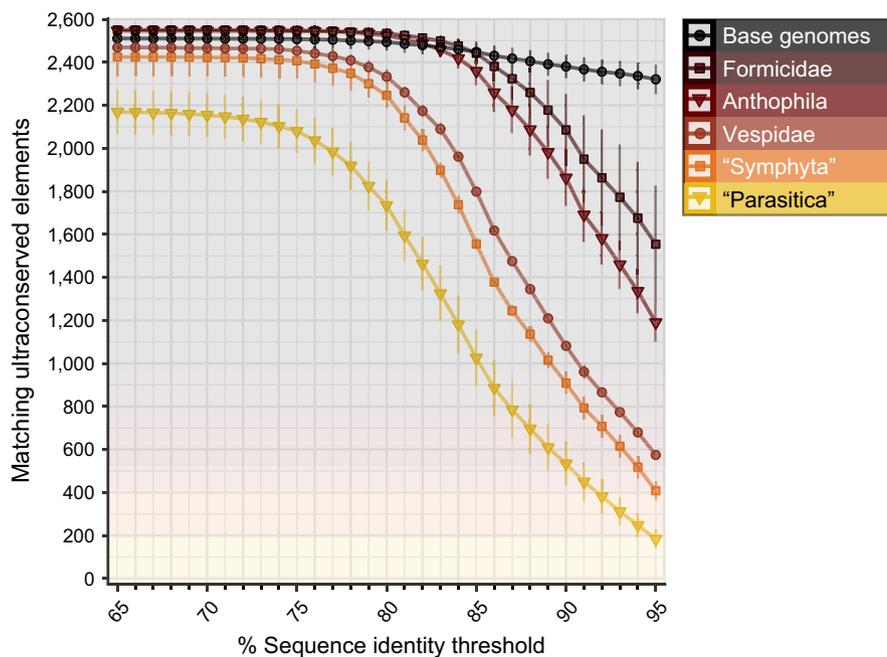


FIGURE 4 The simulated matching success of UCE baits among the examined hymenopteran lineages under varying conditions of sequence identity. The “Base genomes” were initially used to design the bait set (Branstetter, Longino, et al., 2017)

the initial UCE capture obsolete. An alternative bioinformatic approach to detect contaminated libraries is provided by the *assembly_extract_contigs_to_barcode* script, which is integrated in PHYLUCE. It allows the extraction of assembled mitochondrial contigs that correspond to the COI barcode region. These sequences can then be used to validate the presence of a single or multiple corresponding species. Given the detection of multiple divergent DNA barcodes, the entire sample can be excluded from further analyses.

Lastly, we argue that the issue of contamination is relevant to both of the currently most widely used target-enrichment strategies; the UCE approach (Faircloth et al., 2012) and Anchored Enrichment (AE; Lemmon et al., 2012). While we demonstrate the potential for contamination in the UCE method, we argue that it is entirely applicable for the AE strategy. The approaches differ in certain aspects, i.e., in the probe design workflow, different tiling strategies and the subsequent bioinformatic processing. However, central to both is the use of 120 bp long nucleotide baits for *in-solution* hybridizations with highly conserved DNA fragments (e.g. Fragoso-Martínez et al., 2017; Lemmon et al., 2012; Prum et al., 2015 for AE-based studies, and Branstetter, Danforth, et al., 2017; Starrett et al., 2016 for the UCE approach). The AE baits primarily target coding sequences and the most widely shared UCEs from this study were exclusively exonic or partially exonic. We therefore argue that both approaches target comparable, conserved exonic DNA.

ACKNOWLEDGEMENTS

We thank Brant C. Faircloth for the constructive feedback on the manuscript. This study was funded by a U.S. National Science Foundation grant to B.N.D., S.G. Brady, J.P. Pitts, and R. Ross (DEB-1555905).

COMPETING INTERESTS

The authors declare no competing interest.

AUTHORS' CONTRIBUTIONS

S.B. designed and conducted the study. S.B. and B.N.D. wrote the manuscript.

DATA ACCESSIBILITY

The scripts which were used to simulate the matching success of the UCE probe sets are available in a GitHub repository of the first author (https://github.com/MacGyverScripting/matching_uce_simulations) and are archived via Zenodo (<https://doi.org/10.5281/zenodo.1172056>).

ORCID

Silas Bossert  <http://orcid.org/0000-0002-3620-5468>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Austin, J. J., Ross, A. J., Smith, A. B., Fortey, R. A., & Thomas, R. H. (1997). Problems of reproducibility—does geologically ancient DNA survive in amber preserved insects? *Proceedings of the Royal Society B*, 264, 467–474. <https://doi.org/10.1098/rspb.1997.0067>
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, 304, 1321–1325. <https://doi.org/10.1126/science.1098119>
- Blaimer, B. B., Brady, S. G., Schultz, T. R., Lloyd, M. W., Fisher, B. L., & Ward, P. S. (2015). Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: A case study of formicine ants. *BMC Evolutionary Biology*, 15, 1–14. <https://doi.org/10.1186/s12862-015-0552-5>
- Blaimer, B. B., Lloyd, M. W., Guillory, W. X., & Brady, S. G. (2016). Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. *PLoS ONE*, 11, e0161531. <https://doi.org/10.1371/journal.pone.0161531>
- Branstetter, M. G., Danforth, B. N., Pitts, J. P., Faircloth, B. C., Ward, P. S., Buffington, M. L., ... Brady, S. G. (2017). Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Current Biology*, 27, 1019–1025. <https://doi.org/10.1016/j.cub.2017.03.027>
- Branstetter, M. G., Longino, J. T., Ward, P. S., & Faircloth, B. C. (2017). Enriching the ant tree of life: Enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods in Ecology and Evolution*, 8, 768–776. <https://doi.org/10.1111/2041-210X.12742>
- Dickel, D. E., Zhu, Y., Nord, A. S., Wylie, J. N., Akiyama, J. A., Afzal, V., ... Pennacchio, L. A. (2014). Function-based identification of mammalian enhancers using site-specific integration. *Nature Methods*, 11, 566–571. <https://doi.org/10.1038/nmeth.2886>
- Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32, 786–788. <https://doi.org/10.1093/bioinformatics/btv646>
- Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*, 8, 1103–1112. <https://doi.org/10.1111/2041-210X.12754>
- Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources*, 15, 489–501. <https://doi.org/10.1111/1755-0998.12328>
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61, 717–726. <https://doi.org/10.1093/sysbio/sys004>
- Fragoso-Martínez, I., Salazar, G. A., Martínez-Gordillo, M., Magallón, S., Sánchez-Reyes, L., Moriarty Lemmon, E., ... Granados Mendoza, C. (2017). A pilot study applying the plant anchored hybrid enrichment method to new world sages (*Salvia* subgenus *Calospatha*: Lamiaceae). *Molecular Phylogenetics and Evolution*, 117, 124–134. <https://doi.org/10.1016/j.ympev.2017.02.006>
- Gilbert, M. T. P., Moore, W., Melchior, L., & Worobey, M. (2007). DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS ONE*, 2, e272. <https://doi.org/10.1371/journal.pone.0000272>
- Glenn, T. C., & Faircloth, B. C. (2016). Capturing Darwin's dream. *Molecular Ecology Resources*, 16, 1051–1058. <https://doi.org/10.1111/1755-0998.12574>
- Harris, R. S. (2007). *Improved pairwise alignment of genomic DNA*. PhD thesis, The Pennsylvania State University

- Hosner, P. A., Faircloth, B. C., Glenn, T. C., Braun, E. L., & Kimball, R. T. (2016). Avoiding missing data biases in phylogenomic inference: An empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution*, 33, 1110–1125. <https://doi.org/10.1093/molbev/msv347>
- Jennings, W. B. (2016). *Phylogenomic data acquisition: Principles and practice*. Boca Raton, London, New York: CRC Press.
- Kanda, K., Pflug, J. M., Sproul, J. S., Dasenko, M. A., & Maddison, D. R. (2016). Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. *PLoS ONE*, 10, e0143929. <https://doi.org/10.1371/journal.pone.0143929>
- Lemmon, A. R., Emme, S. A., & Moriarty Lemmon, E. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, 61, 727–744. <https://doi.org/10.1093/sysbio/sys049>
- McCormack, J. E., Faircloth, B. C., Crawford, N. G., Gowaty, P. A., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Research*, 22, 746–754. <https://doi.org/10.1101/gr.125864.111>
- McCormack, J. E., Harvey, M. G., Faircloth, B. C., Crawford, N. G., Glenn, T. C., & Brumfield, R. T. (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE*, 8, e54848. <https://doi.org/10.1371/journal.pone.0054848>
- McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, 66, 526–538. <https://doi.org/10.1016/j.ympev.2011.12.007>
- McCormack, J. E., Tsai, W. L. E., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, 16, 1189–1203. <https://doi.org/10.1111/1755-0998.12466>
- Meiklejohn, K. A., Faircloth, B. C., Glenn, T. C., Kimball, R. T., & Braun, E. L. (2016). Analysis of a rapid evolutionary radiation using ultraconserved elements: Evidence for a bias in some multispecies coalescent methods. *Systematic Biology*, 65, 612–627. <https://doi.org/10.1093/sysbio/syw014>
- Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., ... Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346, 763–767. <https://doi.org/10.1126/science.1257570>
- Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Moriarty Lemmon, E., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 526, 569–573. <https://doi.org/10.1038/nature15697>
- R Development Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. <http://www.r-project.org/> [accessed 12 June 2017]
- Reneker, J., Lyons, E., Conant, G. C., Pires, J. C., Freeling, M., Shyu, C.-R., & Korkin, D. (2012). Long identical multispecies elements in plant and animal genomes. *PNAS*, 109, E1183–E1191. <https://doi.org/10.1073/pnas.1121356109>
- Ruane, S., & Austin, C. C. (2017). Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources*, 17, 1003–1008. <https://doi.org/10.1111/1755-0998.12655>
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27, 863–864. <https://doi.org/10.1093/bioinformatics/btr026>
- Starrett, J., Derkarabetian, S., Hedin, M., Bryson, R. W., McCormack, J. E., & Faircloth, B. C. (2016). High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. *Molecular Ecology Resources*, 17, 812–823. <https://doi.org/10.1111/1755-0998.12621>
- Streicher, J. W., & Wiens, J. J. (2016). Phylogenomic analyses reveal novel relationships among snake families. *Molecular Phylogenetics and Evolution*, 100, 160–169. <https://doi.org/10.1016/j.ympev.2016.04.015>
- Wang, H.-L., Yan, Z.-Y., & Jin, D.-Y. (1997). Reanalysis of published DNA sequence amplified from cretaceous dinosaur egg fossil. *Molecular Biology and Evolution*, 14, 589–591.
- Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., ... Elgar, G. (2004). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biology*, 3, e7. <https://doi.org/10.1371/journal.pbio.0030007>
- Yeates, D. K., Meusemann, K., Trautwein, M., Wiegmann, B., & Zwick, A. (2016). Power, resolution and bias: Recent advances in insect phylogeny driven by the genomic revolution. *Current Opinion in Insect Science*, 13, 16–23. <https://doi.org/10.1016/j.cois.2015.10.007>
- Zhang, Z.-Q. (2011). Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. In Z.-Q. Zhang (Ed.), *Animal biodiversity* (pp. 7–12). Zootaxa 3148. Auckland, New Zealand: Magnolia Press

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Bossert S, Danforth BN. On the universality of target-enrichment baits for phylogenomic research. *Methods Ecol Evol*. 2018;00:1–8. <https://doi.org/10.1111/2041-210X.12988>