Journal of Crustacean Biology Advance Access published 23 April 2021

The Crustacean Society

Journal of Crustacean Biology

Journal of Crustacean Biology (2021) 41(2), 1–7. doi:10.1093/jcbiol/ruab009

More from less: Genome skimming for nuclear markers for animal phylogenomics, a case study using decapod crustaceans

Mun Hua Tan[1,2,3,4], Han Ming Gan[1,2], Heather Bracken-Grissom[6], Tin-Yam Chan[7], Frederic Grandjean[8] and Christopher M. Austin[1,2,5]

[1]*Centre of Integrative Ecology, School of Life and Environmental Sciences Deakin University, Geelong, VIC, Australia;*
[2]*Deakin Genomics Centre, Deakin University, Geelong, VIC, Australia;*
[3]*School of BioSciences, Bio21 Institute, University of Melbourne, Melbourne, VIC, Australia;*
[4]*Department of Microbiology and Immunology, University of Melbourne, Bio21 Institute, Melbourne, VIC, Australia;*
[5]*Genomics Facility, Tropical Medicine and Biology Platform, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway 47500, Petaling Jaya, Malaysia;*
[6]*Institute of Environment, Department of Biological Sciences, Florida International University, North Miami, FL 33181, USA;*
[7]*Institute of Marine Biology and Center of Excellence for the Oceans, National Taiwan Ocean University, 2 Pei-Ning Road, Keelung 202301, Taiwan; and*
[8]*Laboratoire Ecologie et Biologie des Interactions, UMR CNRS 7267 Equipe Ecologie Evolution Symbiose, 5 rue Albert Turpin, 86073 Poitiers Cedex, France*

Correspondence: M.H. Tan; e-mail: munhua.tan@unimelb.edu.au

(Received 4 December 2020; accepted 8 March 2021)

ABSTRACT

Low coverage genome sequencing is rapid and cost-effective for recovering complete mitochondrial genomes for crustacean phylogenomics. The recovery of high-copy-number nuclear genes, including histone H3, 18S and 28S ribosomal RNAs, is also possible using this approach based on our research with freshwater crayfishes (Astacidea). We explored the potential of genome skimming (GS) to recover additional nuclear genes from shallow sequencing projects using decapod crustaceans. Using an *in silico*-baited approach, we recovered three additional core histone genes (H2A, H2B, and H4) from our low-coverage decapod dataset (99 species, 69 genera, 38 families, 10 infraorders). Phylogenetic analyses using various combinations of mitochondrial and nuclear genes for the entire decapod dataset and a subset of 40 species of crayfishes showed that the evolutionary rates for different classes of genes varied widely. A very high level of congruence was nevertheless found between trees from the six nuclear genes and those derived from the mitogenome sequences for freshwater crayfish. These findings indicate that nuclear genes recovered from the same genome skimming datasets designed to obtain mitogenomes can be used to support more robust and comprehensive phylogenetic analyses. Further, a search for additional intron-less nuclear genes identified several high-copy-number genes across the decapod dataset, and recovery of NaK, PEPCK, and GAPDH gene fragments is possible at slightly elevated coverage, suggesting the potential and utility of GS in recovering even more nuclear genetic information for phylogenetic studies from these inexpensive and increasingly abundant datasets.

Key Words: freshwater crayfishes, histones, mitochondria, NGS, phylogenetics, ribosomal genes

INTRODUCTION

Outputs from second-generation sequencing (SGS) platforms enable the rapid and inexpensive assembly of genomes of various organisms (Li *et al.*, 2009; Colbourne *et al.*, 2011; Austin *et al.*, 2015, 2017; Tan *et al.*, 2018a). While this supports global collaborative efforts aiming to sequence an ever-increasing diversity of species for tackling large-scale and ambitious biodiversity assessment and phylogenomic projects (Consortium i.K., 2013; Lewin *et al.*, 2018), the problem of generating sufficient data for comprehensive phylogenetic and biodiversity-related studies can

still be challenging for certain groups of organisms with large and repetitive genomes.

Consequently, an increasing number of laboratories are using SGS to generate partial genome sequences at low cost (Gan *et al.*, 2014), but with sufficient coverage to consistently recover highly repetitive genes (Gan *et al.*, 2018; Tan *et al.*, 2018c) for potentially large numbers of individuals for phylogenetic analyses. These studies have mostly focused on the recovery of genes from plastid genomes that are highly abundant in cells. While these genes are useful for taxonomic purposes, they can either be rapidly or slowly evolving and therefore limited in their power to consistently recover relationships over a range of evolutionary depths. Further, evolutionary relationships from mitochondrial genes have the limitation that they are linked and therefore represent just a single gene tree that may not reflect the true evolutionary history (Timm & Bracken-Grissom, 2015).

A relatively recent development that can help address these limitations is the discovery that repetitive nuclear genetic elements, predominately from the nuclear ribosomal cluster, can also be recovered from partial genome scans (i.e., genome skimming). The data derived by genome skimming (GS), a term used for shallow sequencing datasets designed to recover high-copy fractions of a genome (Straub *et al.*, 2012), contain numerous reads from repetitive nuclear genes (Straub *et al.*, 2012; Malé *et al.*, 2014; Dodsworth, 2015; Govindarajulu *et al.*, 2015; Richter *et al.*, 2015; Zimmer & Wen, 2015; Besnard *et al.*, 2016). We have shown that the consistent recovery of genes from the nuclear ribosomal cluster (18S and 28S) as well as a protein-coding gene (histone H3) is possible (Grandjean *et al.*, 2017; Tan *et al.*, 2018c) in freshwater crayfishes.

Our study further explores the potential of GS using other decapod crustaceans, a group with typically large genomes ranging from 1 to 40 Gbp (http://www.genomesize.com), as further evidence of the concept. Expanding upon Grandjean *et al.* (2017) and Tan *et al.* (2018c), we investigated the use of GS to recover non-traditional repetitive nuclear genes from a diverse range of decapod species (99 species, 69 genera, 38 families, 10 infraorders) and demonstrate the consistent recovery of three nuclear histone genes (H2A, H2B, and H4), which were included in phylogenetic analyses at the infraordinal level within Decapoda and for a subset of samples representing freshwater crayfishes (3 families, 14 genera. 40 species). We also further evaluated the potential of GS by recovering several nuclear intron-less genes as well as finding known informative genes (NaK, PEPCK) (Chu *et al.*, 2016) and GAPDH with slightly increased coverage.

MATERIALS AND METHODS

A more detailed description of methods can be found in Supplementary material S1 Data.

Genome skimming for four nuclear histones, 18S and 28S ribosomal RNA genes

We used low coverage raw Illumina sequence data obtained from previous projects (Tan *et al.*, 2015, 2017, 2018b, 2018c, 2019) for 99 decapod species from 10 infraorders within the suborder Pleocyemata, with the exclusion of the sister suborder Dendrobranchiata (Supplementary material S2 Data). Genome skimming to extract histone (H2A, H2B, H3, H4) and ribosomal RNA (18S, 28S) genes was performed as described by Grandjean *et al.* (2017). Briefly, genes were recovered as contigs from either the *de novo* assembly or baited assemblies using sequences from closely-related organisms. Open-reading frames (ORF) were predicted from contigs with *getorf* from EMBOSS v6.6.0 (Rice *et al.*, 2000), followed by the identification of target genes based on sequence homology with blastp v2.6.0+ (Altschul *et al.*, 1990).

Phylogenetic analyses and phylogenetic informativeness profiles

The MitoPhAST v3 pipeline (Tan *et al.*, 2015) uses the IQ-TREE v1.5.5 (Nguyen *et al.*, 2015) program to construct maximum-likelihood phylogenetic trees based on datasets consisting of protein-coding genes (PCG) and ribosomal RNAs (rRNA) from mitochondrial (mito) and/or nuclear (nuc) origins: 1) 13 mito PCGs, 2) 2 mito rRNAs, 3) 4 nuc histones, 4) 2 nuc rRNAs, 5) 13 mito PCGs + 2 mito rRNAs, 6) 4 nuc histones + 2 nuc rRNAs, 7) 13 mito PCGs + 2 mito rRNAs + 4 nuc histones + 2 nuc rRNAs. Analyses used nucleotides as the histone genes show minimal variation at the amino- acid level across all infraorders. Further, using a dataset of only protein-coding genes, we constructed an ultrametric tree (Bouckaert *et al.*, 2014) and used PhyDesign (López-Giráldez & Townsend, 2011) to obtain profiles of phylogenetic informativeness for each gene.

Preliminary scan for intron-less nuclear genes

A set of 640 intron-less nuclear genes was downloaded from the Intronless Gene Database (IGD) (Louhichi *et al.*, 2011) as target gene candidates to simplify the search strategy, since these lack exon-intron structures. Using open-reading frames predicted from assemblies (≥ 50 amino acid characters), single-copy orthologs of intron-less genes were identified with OrthoFinder v2.2.7 (Emms & Kelly, 2015).

Genome skimming applied on datasets of variable sequence depths

Data from a crayfish (*Cherax quadricarinatus* (von Martens, 1868); Tan *et al.*, 2020), with an estimated genome size of 5 Gbp, was subsampled to generate multiple short read datasets at sequencing depths of 0.1×, 0.5×, 1×, and 2×. Using OrthoFinder v2.2.7 (Emms & Kelly, 2015), these datasets were scanned for: 1) intron-less genes, 2) the housekeeping gene GAPDH, and 3) NaK and PEPCK typically applied to multiple decapod phylogenetic analyses (Tsang *et al.*, 2008, 2014; Chu *et al.*, 2016).

RESULTS

Genome skimming recovers additional histone genes

Raw sequences used are available on the Sequence Read Archive (SRA) under BioProject PRJNA485382 (https://www.ncbi.nlm.nih.gov/bioproject/485382). Genome skimming recovered three additional histone genes (H2A, H2B, H4) as well as the same H3, 18S, and 28S rRNA genes recovered by Grandjean *et al.* (2017) for all decapod species with minor exceptions (Supplementary material S2 Data, S3 Data). Only partial 28S rRNA genes were recovered, but with large gaps, for five species, and so these were excluded from further analyses and only partial histone H1 were obtained for 25 samples. The average sequence lengths for each recovered histone gene corresponded to expected full-length sizes. Per-gene coverage analysis showed an average of ~100× for the mitogenome but were generally much higher for the nuclear genes (~300×). Figure 1A illustrates pairwise sequence identity for each mitochondrial and nuclear gene, showing an overall higher sequence divergence for mitochondrial genes.

Identity matrices in Figure 1B and Supplementary material S4 Data indicate overall highly conserved protein sequences for all four histones, with slightly higher variability observed for the H2A and H2B genes. Histone sequences are more divergent (average identities 83.0% to 85.4%) at the nucleotide level, with greater levels of differentiation at higher taxonomic levels. By contrast, pairwise identities of COX1 at the amino acid level show obvious sequence differences (average identity 84.8%) with much lower similarities using nucleotide characters (average identity 75.0%).

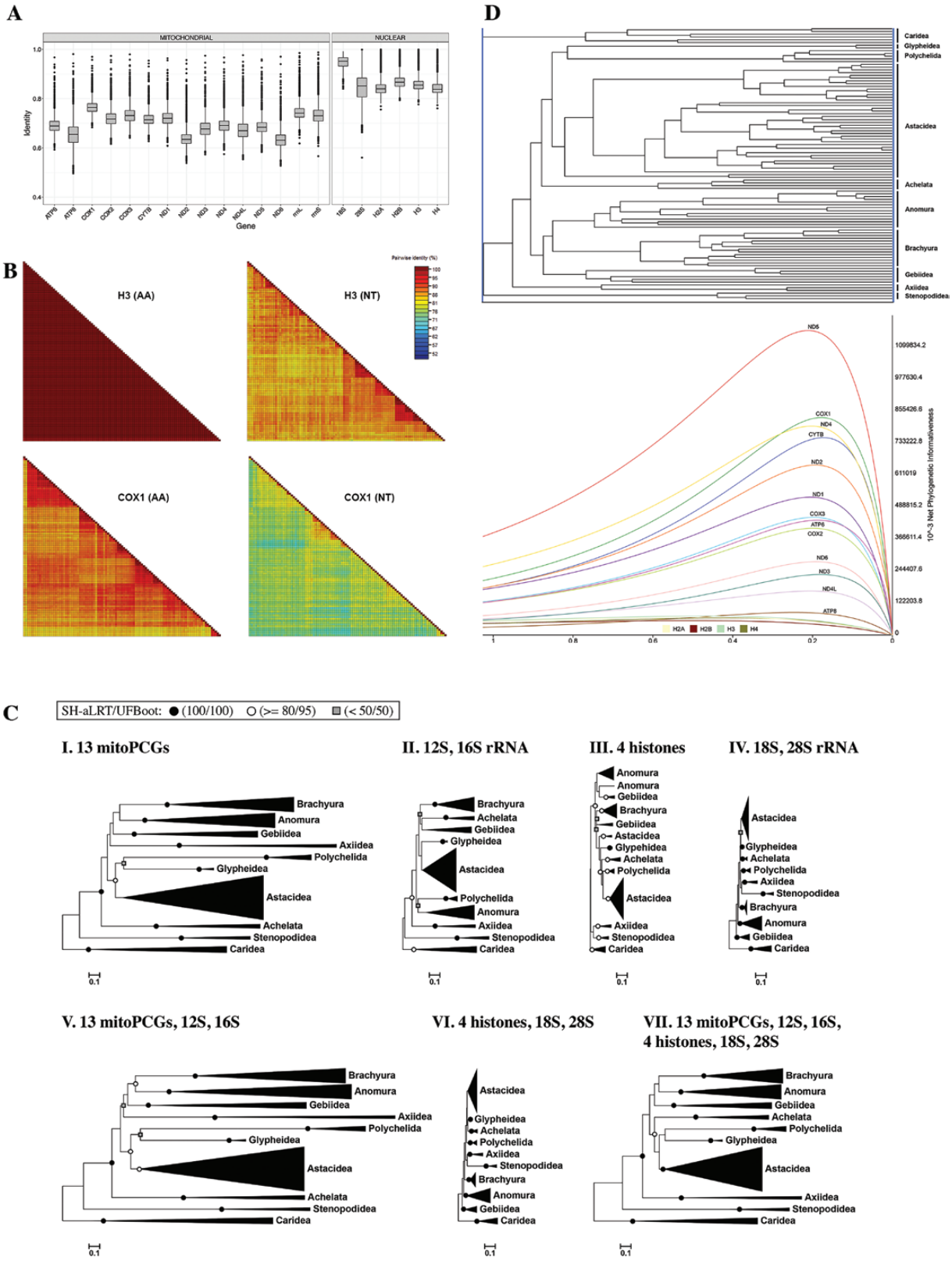

Figure 1. Sequence identities and phylogenetic analyses with mitochondrial and nuclear genes. Based on 99 decapod species, a higher overall divergence level is observed for mitochondrial genes than for nuclear genes (**A**). Matrices show pairwise sequence identities for the nuclear histone H3 compared to the mitochondrial COX1 at nucleotide and amino acid levels (**B**). Collapsed into decapod infraorders, branch lengths of every maximum-likelihood tree are adjusted to the same scale. High SH-aLRT and UFBoot support values are indicated with closed circles (100/100) and open circles (≥80/≥ 95). Unlabelled nodes have mid-level support values whereas weak nodes are labelled with grey rectangles (< 50 for either SH-aLRT or UFBoot) (**C**). Phylogenetic informativeness profiles of 13 mitochondrial protein-coding genes and 4 nuclear histone genes (**D**).

Phylogenetic analyses

Maximum-likelihood phylogenetic trees from the seven datasets are shown in Fig. 1C, with clades collapsed into infraorders. Rooted with infraorder Caridea, tree topologies vary depending on the dataset. All trees generally showed good nodal support for each infra-ordinal clade represented by high nodal support values (SH-aLRT ≥ 80%, UFBoot ≥ 95%) (Fig. 1C). By contrast, nodes at the deeper level are mostly unsupported, especially in trees based on histones and/or nuclear rRNAs. Trees based on these nuclear-only datasets also have shorter deep internodes whereas trees based on mitochondrial datasets have relatively longer internodes and even longer branch lengths (Supplementary material S5 Data). The fast-evolving mitochondrial genes generally yielded the highest informativeness in all phylogenetic informativeness (PI) analyses, whereas all four histone genes show the lowest utility for most of the tested time scale, slightly exceeding ATP8 in informativeness at deeper evolutionary divergence times (Fig. 1D; Supplementary material S6 Data).

By contrast, relationships within infraorder Astacidea, consisting of 40 crayfish species, generally show high levels of congruence between the mitogenome and nuclear datasets (Supplementary material S7 Data). The 28S and 18S data were generally more conserved, but did show some significant heterogeneity in evolutionary rates within and between lineages. The greatest congruence was between the combined nuclear dataset and the mitochondrial datasets (shares 33/40 nodes with tree VII), with both superfamilies, all genera, and several major clades all recovered in the independent analyses. An analysis of the combined data for freshwater crayfishes, represents one of the most comprehensive estimates of evolutionary relationships for this group based on taxon and gene sampling, comprising 12,120 bp of mitochondrial data and 6,629 bp of nuclear data (Tree VII; Supplementary material S7 Data).

Scan for intron-less genes detects presence of high copy number genes

Figure 2 displays the 45 intron-less genes detected in at least 10 species, with those in the majority of datasets indicated by long horizontal stretches of black squares and marked with red arrows. Histograms show the sequence data (Gbp) and the estimated coverage (×) for each sample (from 0.02× to 0.83×). Species from which these gene fragments were recovered belong to different orders and have varying genome sizes. Most of these are genes often present in high copy numbers such as beta-actin-like protein (ACTBL2) or highly repetitive transposable elements including Piggybac (PGBD4) and various Tigger (TIGD1, TIGD7) subfamily genes.

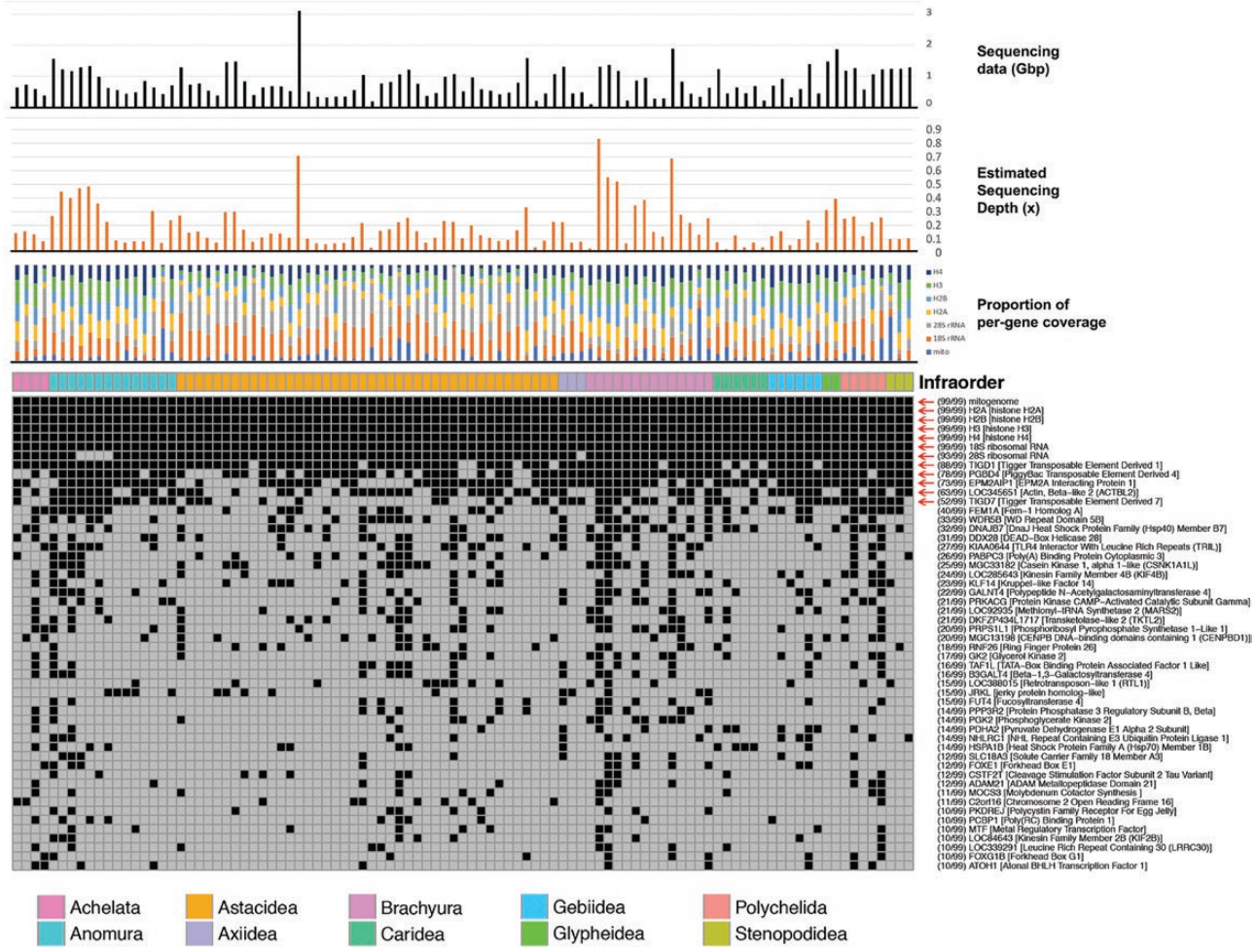

Figure 2. Presence of four nuclear histone genes, 18S rRNA, 28S rRNA, and 45 other intron-less protein-coding genes in 99 partial genome scans of decapods. Black and grey squares represent the presence and absence of a gene, respectively. Shown are single-copy orthologs detected for 45 intron-less genes in at least 10 decapod datasets (in parentheses). Red arrows highlight genes detected in > 50% of datasets. Sequence depths are approximated based on genome sizes from the Animal Genome Size Database (http://www.genomesize.com). Genome sizes of different species from the next-closest genus, family, or infraorder are averaged to provide the next best estimate of a genome size in species lacking genome size information.

Increased sequence depths positively impact the number of nuclear loci recovered

More intron-less genes were detected at elevated coverages in crayfish samples, with relatively better performance for the former (Table 1). For this dataset, partial NaK, PEPCK and GAPDH nuclear protein-coding genes were also detected in datasets with at least 1× coverage.

DISCUSSION

The generated nucleotide datasets for 99 decapod species were previously used for the recovery of mitogenomes, and in addition, we here demonstrate the consistent recovery of two nuclear rRNA genes and four histone genes across all major decapod groups. Three of the histone genes have scant representation in public databases (H2A, H2B, H4) and have not been previously reported from any GS study to our knowledge.

Many researchers are unaware of the treasure trove of potential genomic markers already available for their species or taxa of interest, be it in the form of public sequence databases (Leinonen *et al.*, 2011) or as in-house generated datasets. While the higher throughput capacity of sequencing has enabled targeting genomic loci through sequencing-based approaches such as anchored hybrid enrichment (e.g., Wolfe *et al.*, 2019) to yield large numbers of informative loci, these are still relatively expensive and require prior knowledge of genomic information to determine the targeted regions. Conversely, the genome skimming approach can be applied to existing raw genomic sequence data previously generated for small-scale mitogenome or microsatellite marker projects. Several protein-coding genes commonly used for phylogenetic studies can also be recovered from higher-depth data. More importantly, the application of GS transcends the evolution of sequencing platforms, permitting the recycling of data generated not only from the more recent platforms (e.g., NovaSeq, HiSeq, MiSeq) but also older ones (e.g., Illumina GAIIx, SOLiD, Roche 454) (Straub *et al.*, 2012; Besnard *et al.*, 2016; Grandjean *et al.*, 2017).

While GS seems promising for retrieving new genomic markers, it is necessary to identify the "sweet spot" at which sequence coverage can be minimised while accommodating for differences in genome size, the age or source of genetic material (e.g., tissue type, preservation method), systematic biases in sequencing (Aird *et al.*, 2011; Tilak *et al.*, 2018) and the relative abundance of mitochondria and nuclear copies among samples (Barazzoni *et al.*, 2000; Herbst *et al.*, 2017). Genome sizes for decapod crustaceans have been estimated in the range of 1 to 40 Gbp (http://www.genomesize.com) and the total amount of data available for decapod species in our study ranges from 93 Mbp to 3.4 Gbp, with an average of 807 Mbp, providing < 1× coverage for each species. Independent of the data volume, we observed variable levels of mapping coverage for the mitogenome and nuclear markers, and while generally higher sequence depths were observed for nuclear histones and rRNA relative to the mitogenome, several datasets show contrary patterns.

Table 1. Recovery of nuclear gene fragments from the crayfish *Cherax quadricarinatus* (~ 5 Gbp genome) dataset at variable sequencing depths.

Depth	Sequence data	Intron-less	NaK	PEPCK	GAPDH
0.1×	0.5 Gbp	4	-	-	-
0.5×	2.5 Gbp	18	-	-	-
1.0×	5 Gbp	58	✓	✓	✓
2.0×	10 Gbp	87	✓	✓	✓
Max bp of gene recovered:			372	1458	885

The subsampling of sequence data for the crayfish *C. quadricarinatus* dataset, with a known genome size, has allowed the assessment of coverage on the recovery of different sets of genes (Table 1). Nuclear intron-less genes were easily detected in extremely low coverage datasets, with this number increasing with high sequence volume, with NaK, PEPCK, and GAPDH genes consistently recovered in datasets of at least 1× coverage. We thus recommend sequence coverage of 2 to 3× as the optimal range for the GS approach, subject to adjustments according to the species and target genes of interest. The GS method could benefit in the process of recovering genes, moving away from the use of seed sequences to other signature profiles and functional domains for identification, especially when targeting highly divergent genes.

It is frequently argued that mitochondrial genes are inappropriate when used as the sole marker for inferring evolutionary relationships (Ballard & Whitlock, 2004; Hurst & Jiggins, 2005; Wolfe *et al.*, 2019), with many studies now including several nuclear markers (Tsang *et al.*, 2008, 2014; Schultz *et al.*, 2009; Bracken-Grissom *et al.*, 2013, 2014; Grandjean *et al.*, 2017). Our finding of a high degree of congruence between trees from mitochondrial sequences and nuclear sequences, especially from the expanded range of histone genes for the freshwater crayfish dataset is thus noteworthy and bodes well for the greater use of genome skimming for phylogenetic and potentially molecular taxonomic studies. A caveat is that it will still be important to understand better the most appropriate taxonomic levels for maximum phylogenetic utility. This caveat is illustrated by the finding that the addition of nuclear-based information at the infraorder level gave incongruent topologies and extremely short internodes, recalcitrant to any clear resolution of deeper relationships. This result was further corroborated by the PI profiles of the four histone genes showing low utility across the whole tested time scale relative to that of mitochondrial protein-coding genes, regardless of codon positions considered. Our results also indicate that the recovered nuclear genes vary in their utility. The 18S and 28S gene sequences, commonly used in phylogenetics using traditional PCR methods, are the most conserved, but can show idiosyncratic levels of divergence in some species and lineages, leading to heterogeneity in branch lengths. Achieving consistent alignment is also problematics as nucleotide variation can vary widely along these genes with insertions and deletion common. By contrast, the histone genes, being protein coding, are straightforward to align and show more even levels of divergence across the crayfish tree.

Several studies have explored the use of various markers to bring us closer to a resolved phylogeny (i.e., Shen *et al.*, 2013; Tan *et al.*, 2015, 2018c; Wolfe *et al.*, 2019). While the histones and the 18S and 28S genes recovered by GS in our study make available important additional markers for phylogenetic studies that can be combined with mitochondrial markers recovered from the same samples, they are insufficient for shedding new light on decapod relationships. Our results nevertheless indicate the potential for recovering additional protein-coding markers useful for studies of deep relationships among animal groups. With slightly elevated sequence coverage, GS can likely routinely recover additional protein-coding genes such as NaK and PEPCK, known to be informative in resolving decapod phylogenetic relationships (Tsang *et al.*, 2008, 2014; Chu *et al.*, 2016).

As the cost of sequencing continues to plummet, the prospect of GS as an effective tool for generating datasets for phylogenetics are increasingly brighter. Most of the samples we used were less than 1× coverage at an average cost of less than US$300 per sample. Using costing based on a NovaSeq (Deakin Genomics Centre, Deakin University, Geelong, Australia) we estimate that over 5 Gb of data can now be generated for less than US$150 per sample and that with increased coverage, will consistently yield a greater range of nuclear markers that can be routinely used for phylogenetics. This observation also raises the prospects of a more

sophisticated form of DNA barcoding to that currently based mostly on the mitochondrial COI gene (Bohmann *et al.*, 2020).

SUPPLEMENTARY MATERIAL

Supplementary material S1 Data. Description of methods.

Supplementary material S2 Data. Gene length and coverage for four nuclear histones, 18S and 28S ribosomal RNA genes recovered from 99 species from ten decapod infraorders.

Supplementary material S3 Data. Sequence files for ribosomal RNA and histone genes.

Supplementary material S4 Data. Pairwise identity matrices for COX1 and histone genes.

Supplementary material S5 Data. Phylogenetic trees inferred from different combination of genes.

Supplementary material S6 Data. Phylogenetic informativeness profiles obtained for alignments with only codon positions 1 + 2 or 3.

Supplementary material S7 Data. Phylogenetic trees with ultrafast bootstrap support values, inferred from different combinations of genes (expanded Astacidea clade).

ACKNOWLEDGEMENTS

This work was supported by the Monash University Malaysia Tropical Medicine and Biology Platform and by the Deakin Genomics Centre, Deakin University. We also thank the anonymous reviewers whose comments helped improve and clarify the manuscript. This is contribution 249 from the Coastlines and Oceans Division in the Institute of Environment at Florida International University. The authors declare that they have no conflict of interest.

REFERENCES

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C. & Gnirke, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, **12**: R18 [doi: 10.1186/gb-2011-12-2-r18].

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.

Austin, C.M., Tan, M.H., Croft, L.J., Hammer, M.P. & Gan, H.M. 2015. Whole genome sequencing of the Asian arowana (*Scleropages formosus*) provides insights into the evolution of ray-finned fishes. *Genome Biology and Evolution*, **7**: 2885–2895.

Austin, C.M., Tan, M.H., Harrisson, K.A., Lee, Y.P., Croft, L.J., Sunnucks, P., Pavlova, A. & Gan, H.M. 2017. De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. *GigaScience*, **6**: 1–6 [doi: 10.1093/gigascience/gix063].

Ballard, J.W.O. & Whitlock, M.C. 2004. The incomplete natural history of mitochondria. *Molecular Ecology*, **13**: 729–744.

Barazzoni, R., Short, K.R. & Nair, K.S. 2000. Effects of aging on mitochondrial DNA copy number and cytochromec oxidase gene expression in rat skeletal muscle, liver, and heart. *Journal of Biological Chemistry*, **275**: 3343–3347.

Besnard, G., Bertrand, J.A.M., Delahaie, B., Bourgeois, Y.X.C., Lhuillier, E. & Thébaud, C. 2016. Valuing museum specimens: high-throughput DNA sequencing on historical collections of New Guinea crowned pigeons (Goura). *Biological Journal of the Linnean Society*, **117**: 71–82.

Bohmann, K., Mirarab, S., Bafna, V. & Gilbert, M.T.P. 2020. Beyond DNA barcoding: The unrealized potential of genome skim data in sample identification. *Molecular Ecology*, **29**: 2521–2534.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology*, **10**: e1003537 [doi: 10.1371/journal.pcbi.1003537].

Bracken-Grissom, H.D., Cannon, M.E., Cabezas, P., Feldmann, R.M., Schweitzer, C.E., Ahyong, S.T., Felder, D.L., Lemaitre, R. & Crandall, K.A. 2013. A comprehensive and integrative reconstruction of evolutionary history for Anomura (Crustacea: Decapoda). *BMC Evolutionary Biology*, **13**: 128 [doi: 10.1186/1471-2148-13-128].

Bracken-Grissom, H.D., Ahyong, S.T., Wilkinson, R.D., Feldmann, R.M., Schweitzer, C.E., Breinholt, J.W., Bendall, M., Palero, F., Chan, T.Y., Felder, D.L., Robles, R., Chu, K.-H., Tsang, L.-M., Kim, D., Martin, J.W. & Crandall, K.A. 2014. The emergence of lobsters: Phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (Decapoda: Achelata, Astacidea, Glypheidea, Polychelida). *Systematic Biology*, **63**: 457–479.

Chu, K.H., Tsang, L.M, Ma, T.W., Chan, T.Y. & Ng, P.K.L. 2016. Decapod phylogeny: What can protein-coding genes tell us? In: *Decapod crustacean phylogenetics* (J.W. Martin, K.A. Crandall & D.L. Felder, eds). *Crustacean Issues*, **18**: 101–112.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Cáceres, C.E., Carmel, L., Casola, C., Choi, J.H., Detter, J.C., Dong, Q., Dusheyko, S., Eads, B.D., Fröhlich, T. Geiler-Samerotte, K.A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E.V., Kültz, D., Laforsch, C., Lindquist, E., Lopez, J., Manak, J.R., Muller, J., Pangilinan, J., Patwardhan, R.P., Pitluck, S., Pritham, E.J., Rechtsteiner, A., Rho, M., Rogozin, I.B., Sakarya, O., Salamov, A., Schaack, S., Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D., Tu, H., Vos, H., Mei Wang, Yuri I Wolf, Hideo Yamagata, Takuji Yamada, Yuzhen Ye, Joseph R Shaw, Justen Andrews, Teresa J Crease, Haixu Tang, Lucas, S.M., Robertson, H.M., Bork, P., Koonin, E.V., Zdobnov, E.M., Grigoriev, I.V., Lynch. M., & Boore. J.L. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science*, **331**: 555 [doi: 10.1126/science.1197761].

Consortium i.K. 2013. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, **104**: 595–600.

Dodsworth, S. 2015. Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science*, **20**: 525–527.

Emms, D.M. & Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, **16**: 157 [doi: 10.1186/s13059-015-0721-2].

Gan, H.M., Schultz, M.B. & Austin, C.M. 2014. Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evolutionary Biology*, **14**: 19 [doi: 10.1186/1471-2148-14-19].

Gan, H.M., Tan, M.H., Lee, Y.P., Schultz, M.B., Horwitz, P., Burnham, Q. & Austin, C.M. 2018. More evolution underground: Accelerated mitochondrial substitution rate in Australian burrowing freshwater crayfishes (Decapoda: Parastacidae). *Molecular Phylogenetics and Evolution*, **118**: 88–98.

Govindarajulu, R., Parks, M., Tennessen, J.A., Liston, A. & Ashman, T.-L. 2015. Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *American Journal of Botany*, **102**: 544–554.

Grandjean, F., Tan, M.H., Gan, H.M., Lee, Y.P., Kawai, T., Distefano, R.J., Blaha, M., Roles, A.J. & Austin, C.M. 2017. Rapid recovery of nuclear and mitochondrial genes by genome skimming from Northern Hemisphere freshwater crayfish. *Zoologica Scripta*, **46**: 718–728.

Herbst, A., Widjaja, K., Nguy, B., Lushaj, E.B., Moore, T.M., Hevener, A.L., McKenzie, D., Aiken, J.M. & Wanagat, J. 2017. Digital PCR quantitation of muscle mitochondrial DNA: Age, fiber type, and mutation-induced changes. *The Journals of Gerontology: Series A, Biological Sciences and Medical Sciences*, **72**: 1327–1333.

Hurst, G.D.D. & Jiggins, F.M. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, **272**: 1525 [doi: 10.1098/rspb.2005.3056].

Leinonen, R., Sugawara, H. & Shumway, M., on behalf of the International Nucleotide Sequence Database C. 2011. The Sequence Read Archive. *Nucleic Acids Research*, **39**: D19–D21.

Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., Goldstein. M.M., Grigoriev, I.V., Hackett, K.J., Haussler, D.J., Jarvis, E.D., Johnson, W.E., Patrinos, A., Richards, S.,

Castilla-Rubio, J.C., van Sluys, M.-A., Soltis, P.S., Xu, X., Yang, H. & Zhang, G. 2018. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United tates of America*, **115**: 4325 [doi: 10.1073/pnas.1720115115].

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., Zhang, Z., Zhang, Y., Wang, W., Li, J., Wei, F., Li, H., Jian, M., Li, J., Zhang, Z., Nielsen, R., Li, D., Gu, W., Yang, Z., Xuan, Z., Ryder, O.A., Leung, F.C.-C., Zhou, Y., Cao, J., Sun, X., Fu, Y., Fang, X., Guo, X., Wang, B., Hou, R., Shen, F., Mu, B., Ni, P., Lin, R., Qian, W., Wang, G., Yu, C., Nie, W., Wang, J., Wu, Z., Liang, H., Min, J., Wu, Q., Cheng, S., Ruan, J., Wang, M., Shi, Z., Wen, M., Liu, B., Ren, X., Zheng, H., Dong, D., Cook, K., Shan, G., Zhang, H., Kosiol, C., Xie, X., Lu, Z., Zheng, H., Li, Y., Steiner, C.C., Lam, T.T.-Y., Lin, S., Zhang, Q., Li, G., Tian, J., Gong, T., Liu, H., Zhang, D., Fang, L., Ye, C., Zhang, J., Hu, W., Xu, A., Ren, Y., Zhang, G., Bruford, M.W., Li, Q., Ma, L., Guo, Y., An, N., Hu, Y., Zheng, Y., Shi, Y., Li, Z., Liu, Q., Chen, Y., Zhao, J., Qu, N.., Zhao, S., Tian, F., Wang, X., Wang, H., Xu, L., Liu, X., Vinar, T., Wang, Y., Lam, T.-W.,Yiu, S .-M., Liu, S., Zhang, H., Li, D., Huang, Y., Wang, X., Yang, G., Jiang, Z., Wang, J., Qin, N., Li, L., Li, J., Bolund, L., Kristiansen, K., Wong, G.K.-S. Olson, M., Zhang, X., Li, S., Yang, H.,Wang, J. & Wang, J. 2009. The sequence and de novo assembly of the giant panda genome. *Nature*, **463**: 311 [doi: 10.1038/nature08696].

López-Giráldez, F. & Townsend, J.P. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology*, **11**: 152 [doi: 10.1186/1471-2148-11-152].

Louhichi, A., Fourati, A. & Rebaï, A. 2011. IGD: A resource for intronless genes in the human genome. *Gene*, **488**: 35–40.

Malé, P.-J.G., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., Lhuillier, E., Scotti-Saintagne, C., Tinaut, A. & Chave, J. 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Molecular Ecology Resources*, **14**: 966–975.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**: 268–274.

Rice, P., Longden, I. & Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, **16**: 276–277.

Richter, S., Schwarz, F., Hering, L., Böggemann, M. & Bleidorn, C. 2015. The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). *Genome Biology and Evolution*, **7**: 3443–3462.

Schultz, M.B., Smith, S.A., Horwitz, P., Richardson, A.M.M., Crandall, K.A. & Austin, C.M. 2009. Evolution underground: A molecular phylogenetic investigation of Australian burrowing freshwater crayfish (Decapoda: Parastacidae) with particular focus on *Engaeus Erichson*. *Molecular Phylogenetics and Evolution*, **50**: 580–598.

Shen, H., Braband, A. & Scholtz, G. 2013. Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. *Molecular Phylogenetics and Evolution*, **66**: 776–789.

Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C. & Liston, A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*, **99**: 349–364.

Tan, M.H., Gan, H.M., Schultz, M.B. & Austin, C.M. 2015. MitoPhAST, a new automated mitogenomic phylogeny tool in the post-genomic era with a case study of 89 decapod mitogenomes including eight new freshwater crayfish mitogenomes. *Molecular Phylogenetics and Evolution*, **85**: 180–188.

Tan, M.H., Gan, H.M., Lee, Y.P., Poore, G.C.B. & Austin, C.M. 2017. Digging deeper: new gene order rearrangements and distinct patterns of codons usage in mitochondrial genomes among shrimps from the Axiidea, Gebiidea and Caridea (Crustacea: Decapoda). *PeerJ*, **5**: e2982 [doi: 10.7717/peerj.2982].

Tan, M.H., Austin, C.M., Hammer, M.P., Lee, Y.P., Crof, L.J. & Gan, H.M. 2018a. Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *GigaScience*, **7**: 1–6 [doi: 10.1093/gigascience/gix137].

Tan, M.H., Gan, H.M., Lee, Y.P., Grandjean, F., Croft, L.J. & Austin, C.M. 2020. A giant genome for a giant crayfish (*Cherax quadricarinatus*) with insights into cox1 pseudogenes in decapod genomes. *Frontiers in Genetics*, **11**: 201 [doi: 10.3389/fgene.2020.00201].

Tan, M.H., Gan, H.M., Dally, G., Horner, S., Moreno, P.A.R., Rahman, S. & Austin, C.M. 2018b. More limbs on the tree: mitogenome characterisation and systematic position of 'living fossil' species *Neoglyphea inopinata* and *Laurentaeglyphea neocaledonica* (Decapoda: Glypheidea: Glyp heidae). *Invertebrate Systematics*, **32**: 448–456.

Tan, M.H., Gan, H.M., Lee, Y.P., Bracken-Grissom, H., Chan, T.-Y., Miller, A.D. & Austin, C.M. 2019. Comparative mitogenomics of the Decapoda reveals evolutionary heterogeneity in architecture and composition. *Scientific Reports*, **9**: 10756 [doi: 10.1038/s41598-019-47145-0].

Tan, M.H., Gan, H.M., Lee, Y.P., Linton, S., Grandjean, F., Bartholomei-Santos, M.L., Miller, A.D. & Austin, C.M. 2018c. ORDER within the chaos: Insights into phylogenetic relationships within the Anomura (Crustacea: Decapoda) from mitochondrial sequences and gene order rearrangements. *Molecular Phylogenetics and Evolution*, **127**: 320–331.

Tilak, M.-K., Botero-Castro, F., Galtier, N. & Nabholz, B. 2018. Illumina library preparation for sequencing the GC-rich fraction of heterogeneous genomic DNA. *Genome Biology and Evolution*, **10**: 616–622.

Timm, L. & Bracken-Grissom, H.D. 2015. The forest for the trees: Evaluating molecular phylogenies with an emphasis on higher-level Decapoda. *Journal of Crustacean Biology*, **35**: 577–592.

Tsang, L.M., Ma, K.Y., Ahyong, S.T., Chan, T.Y. & Chu, K.H. 2008. Phylogeny of Decapoda using two nuclear protein-coding genes: Origin and evolution of the Reptantia. *Molecular Phylogenetics and Evolution*, **48**: 359–368.

Tsang, L.M., Schubart, C.D., Ahyong, S.T., Lai, J.C.Y., Au, E.Y.C., Chan, T.-Y., Ng, P.K.L. & Chu, K.H. 2014. Evolutionary history of true crabs (Crustacea: Decapoda: Brachyura) and the origin of freshwater crabs. *Molecular Biology and Evolution*, **31**: 1173–1187.

Wolfe, J.M., Breinholt, J.W., Crandall, K.A., Lemmon, A.R., Lemmon, E.M., Timm, L.E., Siddall, M.E. & Bracken-Grissom, H.D. 2019. A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. *Proceedings of the Royal Society B: Biological Sciences*, **286**: 20190079 [doi: 10.1098/rspb.2019.0079].

Zimmer, E.A. & Wen, J. 2015. Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *Journal of Systematics and Evolution*, **53**: 371–379.