Rapid recovery of nuclear and mitochondrial genes by genome skimming from Northern Hemisphere freshwater crayfish

Frederic Grandjean, Mun Hua Tan, Han Ming Gan, Yin Peng Lee, Tadashi Kawai, Robert J. Distefano, Martin Blaha, Angela J. Roles, Christopher M. Austin

DR. HAN MING GAN (Orcid ID : 0000-0001-7987-738X)

Received Date : 04-Nov-2016
Revised Date : 31-Jan-2017
Accepted Date : 11-Mar-2017
Article type : Original Manuscript

*** Corresponding author:**
Name: Han Ming Gan, PhD
Address: 3-3-20, School of Science, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500, Petaling Jaya, Selangor, Malaysia
Phone: (+603) 5514 6000 (ext. 61727)
Email: gan.han.ming@monash.edu

Rapid recovery of nuclear and mitochondrial genes by genome skimming from Northern Hemisphere freshwater crayfish

FREDERIC GRANDJEAN[1#], MUN HUA TAN[2,3,4#], HAN MING GAN[2,3,4*], YIN PENG LEE[2,3], TADASHI KAWAI[5], ROBERT J. DISTEFANO[6], MARTIN BLAHA[7], ANGELA J. ROLES[8], CHRISTOPHER M. AUSTIN[2,3,4]

 # Equal contribution

Northern crayfish evolution

FREDERIC GRANDJEAN *et al.*

[1] Laboratoire Ecologie et Biologie des Interactions, UMR CNRS 7267 Equipe Ecologie Evolution Symbiose, 5 rue Albert Turpin, Poitiers Cedex, France,
[2] School of Science, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway 47500, Petaling Jaya, Selangor, Malaysia
[3] Genomics Facility, Tropical Medicine and Biology Platform, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway 47500, Petaling Jaya, Selangor, Malaysia
[4] School of Life and Environmental Sciences, Deakin University, Burwood, Victoria 3125, Australia
[5] Fisheries Research Department, Wakkanai Fisheries Research Institute, Wakkanai, Hokkaido, Japan
[6] Missouri Department of Conservation, East Gans Road, Columbia, Missouri 65201, U.S.A.
[7] Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, University of South Bohemia in České Budějovice, Vodňany, Czech Republic
[8] Biology Department, Oberlin College, Oberlin, Ohio 44074, USA

Grandjean, F. (2016) Rapid recovery of nuclear and mitochondrial genes by genome skimming from Northern Hemisphere freshwater crayfish. *Zoologica Scripta*, 00, 000-000.

Molecular phylogenetics has benefited tremendously from the advent of next-generation sequencing, enabling quick and cost-effective recovery of whole mitogenomes via an approach referred to as “genome-skimming”. Recently, genome skimming has been utilised to recover highly-repetitive nuclear genes such as 18S and 28S ribosomal RNA genes that are useful for inferring deeper evolutionary relationships. To address some outstanding issues in the relationships among Northern Hemisphere freshwater crayfish (Astacoidea), we sequenced the partial genome of crayfish species from Asian, North American and European genera and report the successful recovery of whole mitogenome sequences in addition to three highly-repetitive nuclear genes namely, histone H3, 18S and 28S ribosomal RNA. Consistent with some previous studies using short mtDNA and nuclear gene fragments, phylogenetic analyses based on the concatenation of recovered mitochondrial and/or nuclear sequences recovered the Asian cambarid lineage as basal to all astacids and North American cambarids, which conflicts with the current taxonomic classification based on morphological and reproduction-related characters. Lastly, we show that complete H3, 18S and 28S ribosomal RNA genes can also be consistently recovered

from a diverse range of animal taxa, demonstrating the potential wide utility of genome skimming for nuclear markers.

Keywords: Genome skimming, mitogenomes, nuclear markers, Astacidae, Parastacidae, Cambaridae

Introduction

Estimation of evolutionary relationships among organisms using DNA sequence information is now an established part of comparative biology due to the PCR and automated Sanger sequencing revolution since the late 1990s (Avise et al. 2000; Hillis et al. 1996). Early in this phylogenetic revolution, most studies utilised sequences of single mitochondrial genes such as the 16S ribosomal RNA, cytochrome b, cytochrome oxidase I or nuclear genes especially the 18S ribosomal RNA. Recent studies have increasingly used nucleotide data from multiple mitochondrial genes and several nuclear genes, resulting in larger datasets commonly in the order of 5,000 to 10,000 bp. The development of next-generation sequencing (NGS) technology has led to rapidly declining costs for DNA sequencing and shows promise in producing datasets comprising of hundreds, if not thousands of loci or characters.

Assembling datasets using PCR-based methods presents several challenges especially with the increasing expectations that multiple loci are required for robust phylogenies. This amounts to thousands of base pairs and is often coupled with the need for adequate taxon samples, requiring up to and even in excess of 100 samples. These challenges become amplified with the use of museum-held specimens, which have many advantages in relation to supporting biodiversity-related and phylogenetic studies (Graham et al. 2004; McCormack et al. 2015; Suarez & Tsutsui 2004; Thomas

et al. 1990). Tissue samples from museum specimens are often limited in volume and are usually characterised by highly-degraded DNA requiring multiple rounds of short amplicon sequencing often with low success rate (Andersen & Mills 2012; Aznar-Cormano et al. 2015; Li et al. 2015; McCormack et al. 2015; Tin et al. 2014). The use of next-generation sequencing rather than Sanger sequencing can reduce the cost of this approach, often referred to as Targeted Amplicon Sequencing (Bybee et al. 2011). Nevertheless, this does not overcome the time and cost of multiple and failed PCR reactions which will be common for degraded samples, and also the bioinformatics workload of assembling and aligning data from multiple short fragments (Meimberg et al. 2016).

With high throughput sequencing, it is now possible to sequence the genome of eukaryotic organisms for a few thousand dollars in a matter of weeks (Goodwin et al. 2016). However, it is still costly and time-consuming to generate sufficient sequences for robust phylogenetic reconstruction, while also maximising taxon sampling. Currently, the two most popular methods for generating sizable phylogenomic datasets are: (1) the anchored hybrid enrichment approach (Lemmon et al. 2012; Ruane et al. 2015) and (2) the ultra-conserved element procedure (Faircloth et al. 2012; McCormack et al. 2015). Other methods are being developed to exploit museum samples with highly-degraded DNA but require whole genome resources for read mapping (Tin et al. 2014).

Alternatively, a simple, rapid, and low-cost method of rapidly assembling datasets of approximately 10-15 kbp is to use an NGS-based approach involving partial genome scans of samples, also referred to as genome skimming (Gan et al. 2014; Malé et al. 2014; Straub et al. 2012; Tan et al. 2015). Most animal genome skimming studies have focused on mitochondrial sequences that are present in many copies in the eukaryotic cell. Mitochondrial genomes have reduced intergenic elements making them straightforward to recover, assemble, and annotate using a suite of bioinformatics methods and pipelines (Bernt et al. 2013; Hahn et al. 2013; Malé et al. 2014), some of which also facilitate phylogenetic analysis (Tamura et al. 2013; Tan et al. 2015). A major drawback of this approach is that phylogenies based on mitochondrial sequences may not be reflective of the full evolutionary history of the organisms under study as represented by their nuclear genomes (Timm & Bracken-Grissom 2015).

In this regard, an important recent development is the discovery that repetitive nuclear genetic elements, predominately from the nuclear ribosomal cluster, can also be recovered by genome skimming. The data from these partial genome scans, often representing less than 1% of the genome, contain sufficient reads from repetitive nuclear genes to allow them to be routinely recovered for phylogenetic studies (Besnard et al. 2016; Dodsworth et al. 2015; Kocher et al. 2015; Kocher et al. 2014; Malé et al. 2014; Richter et al. 2015; Straub et al. 2012). In this study, we demonstrate the wide utility of genome skimming to 22 species of Northern and Southern Hemisphere freshwater crayfish. We show that, in addition to extracting the full mitogenomes for each species, it is also possible to recover the complete 18S rRNA, 28S rRNA and histone (H3) nuclear gene sequences from a fraction of a MiSeq run (approximately 800 Mbp output). All three of these nuclear genes are considered especially useful for establishing deeper level relationships as demonstrated by a number of studies on crustaceans, including freshwater crayfish (Bracken-Grissom et al. 2014; Bybee et al. 2011; Toon et al. 2010), that utilised information from these genes using PCR-based methods.

While freshwater crayfish have been subject to a number of molecular genetic studies that use conventional PCR-based approaches, there are still outstanding issues concerning phylogenetic relationships among major groups within each superfamily (Braband et al. 2006; Bracken-Grissom et al. 2014; Toon et al. 2010). One of the persistent issues in freshwater crayfish systematics is the unresolved phylogenetic placement of the Asian freshwater genus *Cambaroides*. Though this genus is taxonomically placed in the family Cambaridae, it is often recovered as sister to species of Astacidae or in a basal position rather as sister to other members of Cambaridae from North America. While several studies have used molecular data to study relationships among Northern Hemisphere crayfish species, these often have limitations with respect to taxon sampling (either limited or unbalanced) and number of molecular characters (Ahn et al. 2006; Braband et al. 2006; Bracken-Grissom et al. 2014), sometimes producing conflicting results. In this study, we assemble the complete mitochondrial genomes from ten species of European, North American and Asian crayfish and one lobster species. We also use genome skimming to recover complete or near-complete sequences of nuclear 18S and 28S RNA genes and the histone H3 gene from these samples plus data from 11 additional species of crayfish and lobsters. Our phylogenetic analyses show that the mitochondrial and nuclear trees

198 are fully congruent and the combined dataset produces trees with consistently high
199 nodal support that indicates the polyphyly of the Cambaridae. We also demonstrate
200 that our genome skimming approach recovers the same three nuclear genes from
201 samples of a number of major animal groups (e.g. Mammalia, Teleostei, Aves,
202 Mollusca, Arthropoda), suggesting this approach has wide utility for animal
203 molecular systematics.
204
205 **Material and Methods**
206
207 *Sampling and sequencing*
208
209 For this study, ten Northern Hemisphere freshwater crayfish samples
210 belonging to the superfamily Astacoidea and one lobster sample in the superfamily
211 Nephropoidea that do not have mitogenome representative sequences on NCBI were
212 acquired from various geographical locations (marked with '*' in Supplementary
213 Data 1), identified based on morphology and further validated with nucleotide
214 similarity searches against publicly available COI, 16S and 12S rRNA gene fragments
215 for the corresponding species (Supplementary Data 2). For *Cambaroides similis*,
216 whose mitogenome is already available on NCBI, a new additional sample of the
217 same species was collected from Korea (94% identity from a 810 bp alignment to the
218 *C. similis cox1* gene from NC_016925.1) to further scan for nuclear genes. For the
219 *Astacus* species, due to low mitogenome content in the muscle tissue, additional
220 isolates of each species were further enriched for mitochondria (Grandjean et al.
221 1997) and sequenced to recover the complete mitogenome. For all samples,
222 purification of ethanol-preserved tissues and partial whole genome sequencing on the
223 Illumina MiSeq (2 × 250 bp or 2 × 150 bp) was carried out as described in Gan et al.
224 (2014) at the Monash University Malaysia Genomics Facility.
225 Several species from other superfamilies, Parastacoidea (Southern Hemisphere
226 crayfish) and Nephropoidea (lobsters), were also included in this study for
227 comparative purposes or as outgroup species to the Northern Hemisphere crayfish
228 group. For these taxa, nuclear gene sequences or raw sequence read datasets were
229 recovered from various sources – existing mitogenome and nuclear sequences for
230 some species were obtained from NCBI (accession numbers cited in Supplementary

Data 1), whereas raw sequence reads for other species were available from previous mitogenome studies by our group (studies also cited in Supplementary Data 1).

Genome skimming for mitochondrial and nuclear sequences

Genome skimming was performed according to the workflow illustrated in Fig. 1. Sequences generated from the partial genome sequencing of each sample were initially pre-processed using Trimmomatic (Bolger et al. 2014) to remove adapters and low quality sequences (*Illumina clip 2:30:10, sliding window 4:20, leading: 3, trailing 3, min length 100*). The resulting quality-filtered reads were then assembled using: 1) IDBA-UD (Peng et al. 2012), an iterative *de novo* assembler for data with uneven sequencing coverage; or 2) MITObim (Hahn et al. 2013) for challenging assemblies through the provision of bait sequences to recruit reads for more localised assemblies of the mitogenome or specific nuclear genes.

Target sequences were identified from these assemblies through sequence type specific methods (e.g. mitochondrial, nuclear protein-coding genes, nuclear ribosomal RNA genes). Complete mitochondrial sequences were recovered for most samples from either *de novo* (IDBA-UD) or baited (MITObim) assemblies and annotated with MITOS (Bernt et al. 2013). Any recalcitrant gaps (i.e. more than one contig) were gap-closed through PCR using gap-bridging primers and Sanger sequencing. Nuclear ribosomal RNAs were predicted with RNAmmer (Lagesen et al. 2007) and the nuclear protein-coding gene (histone H3) was recovered through a BLASTn search (Altschul et al. 1990) against existing H3 sequences of related species. For histone H3, the start and stop coordinates were further refined with ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and translated with the *transeq* component provided by EMBOSS (Rice et al. 2000) to obtain the amino acid sequence.

The same genome skimming workflow was tested on reads sequenced from species representing a diversity of animal phyla, including representatives of the Mammalia, Arthropoda, Aves, Teleostei and Mollusca to evaluate the general applicability of our methods across different animal groups and tissue types. Specifically, sequence reads were obtained from nine other sequencing projects in our laboratory and three projects on NCBI's SRA database for species from a variety of animal Phyla and Classes and tissue sources (e.g. fin clips, liver, muscle, whole

organism). These sequence datasets were inspected for the presence of reads for the same three nuclear genes (18S, 28S, H3) recovered from this study of crayfish and lobster species.

Phylogenetic analyses

The construction of phylogenetic trees was carried out on seven different alignments (Datasets A to G, Table 1), consisting of various combinations of genes, sequence types (amino acid, aa, vs nucleotide, nt) and lengths. In analyses that utilise only mitochondrial gene sequences (13 protein-coding genes, 2 rRNAs), a total of 33 samples from the families Astacidae (7), Cambaridae (16), Parastacidae (6) and Nephropidae (4) were included. Datasets that included the nuclear genes (18S rRNA, 28S rRNA and histone H3) sampled fewer taxa (24), subject to the availability of these gene sequences on NCBI for species that were not sequenced in our laboratory (e.g. *Procambarus alleni, Procambarus fallax*).

Amino acid sequences of protein-coding genes (mitochondrial, H3) as well as nucleotide sequences of non-coding rRNAs (12S, 16S, 18S, 28S) were aligned with MAFFT (*mafft-linsi*) (Katoh & Standley 2013) and trimmed with trimAl (*automated1*) (Capella-Gutiérrez et al. 2009). Nucleotide sequences of protein-coding genes were aligned with TranslatorX (Abascal et al. 2010), which carries out nucleotide sequence alignment guided by amino acid translations followed by alignment trimming with Gblocks (Castresana 2000) implemented internally by the same program.

For phylogenetic analyses, alignments were concatenated for each of the seven datasets (Table 1) and supplied as partitioned alignments to IQ-TREE (Nguyen et al. 2014) for model testing and maximum-likelihood analysis, with node supports obtained with the Ultrafast bootstrap option (Minh et al. 2013). The same partitioned alignments were used for Bayesian inference using ExaBayes (Aberer et al. 2014). Four independent chains were run for a minimum of 5 million generations each, with 25% of initial samples as burn-in, and convergence of chains was determined when the average standard deviation of split frequencies (*asdsf*) fell below 1% indicating good convergence.

Topology testing

Topology testing was carried out using IQ-TREE (Nguyen et al. 2014) to evaluate the likelihood of alternate topologies (e.g. a monophyletic Cambaridae). The following tree topology tests were performed using Dataset F (Table 1), comprised of 18 genes (13 mitochondrial PCGs, 12S, 16S, 18S, 28S rRNA, H3):

I: ((Astacidae, Cambaridae-NA), Cambaridae-Asia), outgroups)
II: (Astacidae, (Cambaridae-NA, Cambaridae-Asia)), outgroups)
III: ((Astacidae, Cambaridae-Asia), Cambaridae-NA, outgroups)

The tree topology tests include the Shimodaira-Hasegawa test (Shimodaira & Hasegawa 1999) carried out using the RELL approximation (Kishino et al. 1990) based on 1,000 replicates and the approximately unbiased (AU) test (Shimodaira 2002).

Results

Genome skimming effectively recovers the mitogenome sequence and high copy number nuclear genes

An average of approximately 739.4 Mbp of raw sequence data per sample was generated from the freshwater crayfish and lobster libraries (Supplementary Data 1). Mitogenome sequences assembled for these species vary in size from 14,895 bp to 20,677 bp with AT content ranging from 67.9% to 73.1%. The typical 13 mitochondrial protein-coding genes, 22 transfer RNA genes and 2 ribosomal RNA genes (12S, 16S) are found in all Northern Hemisphere crayfish and lobster mitogenomes recovered in this study (marked with '*' in Supplementary Data 1). The organization of these genes in the mitogenomes of the ten Northern Hemisphere crayfish taxa we assembled is identical to the first sequenced species, *Procambarus clarkii*, which itself shows a large departure from the ground pancrustacean pattern (as represented by *Drosophila*, *Penaeus monodon* and the outgroup species, *Homarus americanus*). The gene order for the lobster *Metanephrops sibogae* is also aberrant compared to *H. americanus*, a result of multiple translocated protein-coding and tRNA genes. Most notably, *M. sibogae* possesses two control regions, each

approximately 2 kbp in length, resulting in a much longer mitochondrial genome size of 20,677 bp. The lengths of the 12S and 16S rRNA genes are generally shorter in the Northern Hemisphere crayfish compared to the Southern Hemisphere crayfish. Details such as coding regions, AT-content and intergenic lengths for each mitogenome are available in Supplementary Data 3.

Complete or near-complete sequences were recovered for the three nuclear genes 28S rRNA (4,144 - 5,391 bp), 18S rRNA (1,869 - 1,885 bp) and H3 (all 411 bp) from the same partial genome scan. Their degrees of similarity to available sequences on NCBI for the same species are detailed in Supplementary Data 4. Out of the total nuclear sequences contributed through this study, 38 gene sequences from 13 species are 'novel' (i.e. do not have any representation on NCBI). The remaining 28 gene sequences are highly similar to sequences held on NCBI based on local alignment, with average percent identities of 98.7% (28S), 99.4% (18S) and 98.9% (H3) for matching species. The lengths of the 18S rRNA sequences recovered in this study are comparable to those already available on public databases through PCR-based methods. However, the other nuclear sequences (28S, H3) obtained from genome skimming are much longer in length than those deposited on NCBI for crayfish and lobster species. Notably, the 28S rRNA gene sequences contributed in this study are almost double the length of their same-species counterpart available on NCBI. Also, the full length amino acid sequence of the histone H3 gene (411 bp) complete with start and stop codons was recovered, as opposed to the currently available partial H3 sequences that are mostly 333 bp or shorter. All recovered mitochondrial and nuclear sequences are available on NCBI at accession numbers listed in Supplementary Data 1.

Phylogenetic analyses and topology tests point to a polyphyletic Cambaridae

The nucleotide-based phylogenetic tree was generated from the longest alignment (16,211 bp, Dataset F; Fig. 2) with representative species from Parastacoidea (Southern Hemisphere crayfish) and Nephropoidea (lobsters) as outgroups. The focus of this study, the superfamily Astacoidea, is represented by species from two families, Astacidae (5 species) and Cambaridae (9 species). Maximal support is observed for most nodes in this clade of interest, except for the weaker ML support for the sister relationship between astacids and North American

cambarids (ultrafast bootstrap: 86, PP: 1.00). Both ML and BI trees inferred from this dataset imply a monophyletic Astacidae and a polyphyletic Cambaridae, with the split occurring between the North American cambarids (*Procambarus*, *Cambarus*, *Orconectes*) and the Asian cambarids (*Cambaroides*). Further topology testing (Fig. 3) shows Topology I (North American cambarids as sister taxa to astacids) as most likely, followed by Topology III (Asian cambarids as sister taxa to astacids). Both Topology I and Topology III support a polyphyletic Cambaridae. Topology II, containing a monophyletic Cambaridae, is rejected (P-value < 0.05 for tree tests in Fig. 3).

Nevertheless, tree topologies are variable depending on the dataset and the method used to infer phylogeny (Fig. 4A-G). While the most common topology is consistent with the tree in Fig. 2, other observed topologies mostly differ in the relationships among groups of the North American cambarids (Datasets B and C). The tree generated from Dataset G, which consists of only the nuclear 18S, 28S and H3 gene sequences (4,205 aligned sites), deviates from the other topologies. Its Bayesian tree does show a monophyletic Cambaridae but with only weak support (PP: 0.61), and is incongruent with the ML tree generated from the same alignment, which is similar to the other analyses and also fails to recover a monophyletic Cambaridae. Detailed ML and BI phylogenetic trees inferred from all datasets are available as Supplementary Data 5.

Generality of genome skimming for nuclear genes for animals

Out of the twelve tested animals, the 18S gene sequence was recovered from all species, whereas sequences from both 28S (partial or complete) and H3 were recovered from ten out of twelve species. Similarly, genome skimming successfully recovered substantially longer 28S (approximately 4 kbp) and H3 (411 bp) gene sequences in most cases compared to sequences available on NCBI (28S: 1.5 - 4 kbp, H3: 333 - 411 bp; Supplementary Data 4). The 18S, 28S and H3 sequences recovered for these species are available as Supplementary Data 6.

Discussion

Crayfish mitogenomes

This study increases the number of sequenced Northern Hemisphere crayfish mitogenomes from six to sixteen, substantially expanding the available resources for the family Cambaridae (*Procambarus*, *Cambaroides*, *Cambarus*, *Orconectes*) and Astacidae (*Astacus, Pacifastacus, Austropotamobius*). In addition, a new mitogenome for the lobster, *Metanephrops sibogae*, reveals an aberrant gene order for this group, but one identical to that recently described for *Metanephrops thomsoni* (Ahn et al. 2016). This is a surprising finding given that previous studies indicated that marine lobsters (*Homarus americanus* and *Enoplometopus*) possess a conserved mitogenome order that is common across the arthropods and is considered reflective of the primitive pancrustacean pattern (Boore et al. 1995; Shen et al. 2013).

Another equally surprising finding is the lack of mitogenome variation among Northern Hemisphere species, given the high frequency of novel mitogenome gene orders among Southern Hemisphere crayfish. No mitogenome gene rearrangements are apparent for the ten new mitogenomes provided from this study and the six from previous studies, all of which contributes to taxonomic sampling covering all families and the full geographic range of the superfamily. This is in stark contrast to the number and scale of mitogenome gene order rearrangements among Southern Hemisphere crayfish with most genera studied having distinct gene orders, including interspecific differences within the genus *Engaeus* (Lee et al. 2016; Tan et al. 2015).

The frequency of mitogenome rearrangements is not simply a function of divergence times. Based on the dated phylogeny of Bracken-Grissom et al. (2014), the *Engaeus* group of crayfish and its close relatives, containing significant rearrangements, diverged more recently (145.4 mya) than the Northern Hemisphere crayfish as a group (161.2 mya), which have none. Conversely, *Euastacus* and *Cherax*, which last shared a common ancestor approximately 200 mya have identical mitogenome gene orders. Thus, crayfish exhibit both extreme conservation and extreme lability of mitochondrial gene order, that is not a simple function of divergence time, an observation that invites further investigation on the dynamics and evolutionary drivers of mitogenome evolution in this group (Kilpert et al. 2012; Okajima & Kumazawa 2010; Poulsen et al. 2013).

Phylogenetic results and the status of the family Cambaridae

This study contributes to growing evidence suggesting that the family Cambaridae is non-monophyletic, but contradicts suggestions that the genus *Cambaroides* should be included within the family Astacidae. Several studies using a variety of morphological and molecular datasets from a range of genes and varying taxonomic sampling concur that North American cambarid species and Asian cambarid species (genus *Cambaroides*) do not share a common ancestor (Ahn et al. 2006; Braband et al. 2006; Bracken-Grissom et al. 2014; Bracken et al. 2009; Breinholt et al. 2009; Crandall et al. 2000; Porter et al. 2005; Rode & Babcock 2003). The phylogenetic position of Asian cambarid species and their taxonomic treatment within the superfamily Astacoidea remains controversial.

While most studies have supported the Asian cambarid lineage as the most basal within the Astacoidea, Bracken-Grissom et al. (2014) found the Asian cambarids and the astacids to be monophyletic (using a combination of morphological characters, three mitochondrial and three nuclear gene fragments and based on two samples of *Cambaroides japonicus*, single samples of *Astacus astacus* and *Austropotamobius torrentium,* and four samples of *Pacifastacus*). They suggested that the concept of the Astacidae should be expanded to include *Cambaroides*. Instead, our dataset strongly supports the *Cambaroides* lineage to be basal, based on our data from five astacid species, two *Procambarus*, two *Orconectes*, one *Cambarus* and four *Cambaroides* species consisting of both nuclear and mitochondrial genes.

A basal position for the Asian cambarid lineage requires a re-evaluation or re-interpretation of morphological and reproduction-related characters as either ancestral or convergent within the lineages as recovered in this study (Ahn et al. 2006; Braband et al. 2006). We suggest a family-level revision of the taxonomic classification of Northern Hemisphere crayfish that might consider placing the Asian cambarid crayfish in a new family, or placing all Northern Hemisphere crayfish in a single family, similar to the treatment of all Southern Hemisphere crayfish as members of the Parastacidae.

The utility of genome skimming for animal phylogenetics

This study demonstrates the utility of partial genome sequencing, also known as genome skimming, using the MiSeq NGS platform as a rapid and inexpensive approach to assemble substantial datasets to support phylogenetic studies. We used

our crayfish dataset to construct an alignment of 12,006 nucleotides from the mitochondrial genomes, which is now becoming a routine procedure for animal phylogenetic studies using NGS (Gan et al. 2014; Shen et al. 2013; Tan et al. 2015).

Less common is the use of sequences from nuclear genes that can also be recovered from the same partial genome scan used to assemble whole mitogenome sequences or to locate microsatellite markers for population genetic applications (Gan et al. 2014; Thai et al. 2016). Supplementary Data 7 summarises the only four recent studies we could find on animals that have reported nuclear genes recovered from NGS-based genome scans (Besnard et al. 2016; Kocher et al. 2015; Kocher et al. 2014; Richter et al. 2015). Genes and regions associated with the nuclear ribosomal cluster are the most common target, and these studies together with our data indicate that complete or almost complete gene sequences can be routinely recovered for the 18S and 28S genes from various animal groups including annelids, crustaceans, molluscs (Bivalvia and Gastropoda), and chordates (Aves, Chondrichthyes, Actinopterygii, Mammalia). Further, high copy number protein-coding genes can also be recovered. Our study is the first to report recovery of the histone H3 gene, for which the full amino acid sequence was retrieved for all our lobster and crayfish samples and ten of twelve species in our supplementary non-crustacean datasets (Table 2). It was also encouraging that other protein-coding genes can potentially be recovered from shotgun sequencing datasets (Besnard et al. 2016), especially as the phylogenetic utility of ribosomal nuclear genes has been called into question by some authors (Tsang et al. 2008).

We foresee exciting times ahead for the discovery and recovery of an increasing number of nuclear genes for phylogenetic analyses, given increasing use of NGS for partial genome sequencing for many animal samples plus the increasing number of whole genome sequences becoming available for a diversity of animal species. Further, we anticipate that animal systematics is entering a new era in which even more robust datasets can be assembled, maximising both taxon and gene sampling while minimising expense to an extent hitherto impossible (Richter et al. 2015; Straub et al. 2012).

Declaration of Interest

The authors report no conflicts of interest and the authors alone are responsible for the content and the writing of the paper.

Acknowledgement

Funding for this study was provided by the Monash University Malaysia Tropical Medicine and Biology Platform. MHT and HMG would like to acknowledge the Monash University Malaysia High Performance Computing infrastructure for the provision of computational resources.

Reference

Abascal F, Zardoya R, and Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* 38:W7-W13. 10.1093/nar/gkq291

Aberer AJ, Kobert K, and Stamatakis A. 2014. ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. *Molecular biology and evolution* 31:2553-2556. 10.1093/molbev/msu236

Ahn D-H, Kawai T, Kim S-J, Rho HS, Jung JW, Kim W, Lim BJ, Kim MS, and Min G-S. 2006. Phylogeny of northern hemisphere freshwater crayfishes based on 16S rRNA gene analysis. *Korean Journal of Genetics*.

Ahn D-H, Min G-S, Park J-K, and Kim S. 2016. The complete mitochondrial genome of the red-banded lobster *Metanephrops thomsoni* (Crustacea, Astacidea, Nephropidae): a novel gene order. *Mitochondrial DNA Part A* 27:2663-2664.

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. 1990. Basic local alignment search tool. *Journal of molecular biology* 215:403-410.

Andersen JC, and Mills NJ. 2012. DNA Extraction from Museum Specimens of Parasitic Hymenoptera. *PLoS One* 7:e45549. 10.1371/journal.pone.0045549

Austin CM, Tan MH, Lee YP, Croft LJ, Meekan MG, and Gan HM. 2016a. The complete mitogenome of the cow tail ray *Pastinachus atrus* (Macleay, 1883) (Elasmobranchii; Myliobatiformes; Dasyatidae). *Mitochondrial DNA Part A* 27:1372-1373. 10.3109/19401736.2014.947586

Austin CM, Tan MH, Lee YP, Croft LJ, Meekan MG, Pierce SJ, and Gan HM. 2016b. The complete mitogenome of the whale shark parasitic copepod *Pandarus*

rhincodonicus norman, Newbound & Knott (Crustacea; Siphonostomatoida; Pandaridae) – a new gene order for the copepoda. *Mitochondrial DNA Part A* 27:694-695. 10.3109/19401736.2014.913147

Avise JC, Nelson WS, Bowen BW, and Walker D. 2000. Phylogeography of colonially nesting seabirds, with special reference to global matrilineal patterns in the sooty tern (*Sterna fuscata*). *Molecular Ecology* 9:1783-1792. 10.1046/j.1365-294x.2000.01068.x

Aznar-Cormano L, Brisset J, Chan T-Y, Corbari L, Puillandre N, Utge J, Zbinden M, Zuccon D, and Samadi S. 2015. An improved taxonomic sampling is a necessary but not sufficient condition for resolving inter-families relationships in Caridean decapods. *Genetica* 143:195-205. 10.1007/s10709-014-9807-0

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, and Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69:313-319. http://dx.doi.org/10.1016/j.ympev.2012.08.023

Besnard G, Bertrand JAM, Delahaie B, Bourgeois YXC, Lhuillier E, and 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. 10.1111/bij.12494

Bolger AM, Lohse M, and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120. 10.1093/bioinformatics/btu170

Boore JL, Collins TM, Stanton D, and Daehler LL. 1995. Arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376:13.

Braband A, Kawai T, and Scholtz G. 2006. The phylogenetic position of the East Asian freshwater crayfish Cambaroides within the Northern Hemisphere Astacoidea (Crustacea, Decapoda, Astacida) based on molecular data Die phylogenetische Stellung der ostasiatischen Flusskrebsgattung Cambaroides innerhalb der Astacoidea (Crustacea, Decapoda, Astacida) basierend auf molekularen Daten. *Journal of Zoological Systematics and Evolutionary Research* 44:17-24. 10.1111/j.1439-0469.2005.00338.x

Bracken-Grissom HD, Ahyong ST, Wilkinson RD, Feldmann RM, Schweitzer CE, Breinholt JW, Bendall M, Palero F, Chan T-Y, Felder DL, Robles R, Chu K-H, Tsang L-M, Kim D, Martin JW, and Crandall KA. 2014. The Emergence of

570 Lobsters: Phylogenetic Relationships, Morphological Evolution and
571 Divergence Time Comparisons of an Ancient Group (Decapoda: Achelata,
572 Astacidea, Glypheidea, Polychelida). *Systematic Biology* 63:457-479.
573 10.1093/sysbio/syu008
574 Bracken HD, De Grave S, and Felder DL. 2009. Phylogeny of the infraorder Caridea
575 based on mitochondrial and nuclear genes (Crustacea: Decapoda). *Decapod*
576 *crustacean phylogenetics*:281-305.
577 Breinholt J, Pérez-Losada M, and Crandall KA. 2009. The timing of the
578 diversification of the freshwater crayfishes. *Decapod crustacean*
579 *phylogenetics*:343-356.
580 Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement
581 MJ, Udall JA, Wilcox ER, and Crandall KA. 2011. Targeted Amplicon
582 Sequencing (TAS): A Scalable Next-Gen Approach to Multilocus, Multitaxa
583 Phylogenetics. *Genome Biology and Evolution* 3:1312-1323.
584 10.1093/gbe/evr106
585 Capella-Gutiérrez S, Silla-Martínez JM, and Gabaldón T. 2009. trimAl: a tool for
586 automated alignment trimming in large-scale phylogenetic analyses.
587 *Bioinformatics* 25:1972-1973. 10.1093/bioinformatics/btp348
588 Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for
589 Their Use in Phylogenetic Analysis. *Molecular biology and evolution* 17:540-
590 552.
591 Crandall KA, Harris DJ, and Fetzner JW. 2000. The monophyletic origin of
592 freshwater crayfish estimated from nuclear and mitochondrial DNA
593 sequences. *Proceedings of the Royal Society of London B: Biological Sciences*
594 267:1679-1686. 10.1098/rspb.2000.1195
595 Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoël M, Weiss-
596 Schneeweiss H, and Leitch AR. 2015. Genomic Repeat Abundances Contain
597 Phylogenetic Signal. *Systematic Biology* 64:112-126. 10.1093/sysbio/syu080
598 Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, and Glenn
599 TC. 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers
600 Spanning Multiple Evolutionary Timescales. *Systematic Biology*.
601 10.1093/sysbio/sys004
602 Gan HM, Gan HY, Tan MH, Penny SS, Willan RC, and Austin CM. 2016a. The
603 complete mitogenome of the giant clam *Tridacna squamosa* (Heterodonta:

604 Bivalvia: Tridacnidae). *Mitochondrial DNA Part A* 27:3220-3221.
605 10.3109/19401736.2015.1007355
606 Gan HM, Schultz MB, and Austin CM. 2014. Integrated shotgun sequencing and
607 bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms
608 substantial gene rearrangements in Australian freshwater crayfishes. *BMC*
609 *Evol Biol* 14:19. 10.1186/1471-2148-14-19
610 1471-2148-14-19 [pii]
611 Gan HM, Tan MH, Lee YP, and Austin CM. 2016b. The complete mitogenome of the
612 Australian tadpole shrimp *Triops australiensis* (Spencer & Hall, 1895)
613 (Crustacea: Branchiopoda: Notostraca). *Mitochondrial DNA Part A* 27:2028-
614 2029. 10.3109/19401736.2014.974173
615 Gan HM, Tan MH, Lee YP, and Austin CM. 2016c. The complete mitogenome of the
616 river blackfish, *Gadopsis marmoratus* (Richardson, 1848) (Teleostei:
617 Percichthyidae). *Mitochondrial DNA Part A* 27:2030-2031.
618 10.3109/19401736.2014.974174
619 Gan HM, Tan MH, Thai BT, and Austin CM. 2016d. The complete mitogenome of
620 the marine bivalve *Lutraria rhynchaena* Jonas 1844 (Heterodonta: Bivalvia:
621 Mactridae). *Mitochondrial DNA Part A* 27:335-336.
622 10.3109/19401736.2014.892104
623 Goodwin S, McPherson JD, and McCombie WR. 2016. Coming of age: ten years of
624 next-generation sequencing technologies. *Nat Rev Genet* 17:333-351.
625 10.1038/nrg.2016.49
626 Graham CH, Ferrier S, Huettman F, Moritz C, and Peterson AT. 2004. New
627 developments in museum-based informatics and applications in biodiversity
628 analysis. *Trends in Ecology & Evolution* 19:497-503.
629 http://dx.doi.org/10.1016/j.tree.2004.07.006
630 Grandjean F, Souty-Grosset C, Raimond R, and Holdich D. 1997. Geographical
631 variation of mitochondrial DNA between populations of the white-clawed
632 crayfish *Austropotamobius pallipes*. *Freshwater Biology* 37:493-501.
633 10.1046/j.1365-2427.1997.00176.x
634 Hahn C, Bachmann L, and Chevreux B. 2013. Reconstructing mitochondrial genomes
635 directly from genomic next-generation sequencing reads—a baiting and
636 iterative mapping approach. *Nucleic Acids Research* 41:e129-e129.
637 10.1093/nar/gkt371

638 Hillis DM, Moritz C, Mable BK, and Olmstead RG. 1996. *Molecular systematics*:
639 Sinauer Associates Sunderland, MA.
640 Katoh K, and Standley DM. 2013. MAFFT multiple sequence alignment software
641 version 7: improvements in performance and usability. *Molecular biology and*
642 *evolution* 30:772-780.
643 Kilpert F, Held C, and Podsiadlowski L. 2012. Multiple rearrangements in
644 mitochondrial genomes of Isopoda and phylogenetic implications. *Mol*
645 *Phylogenet Evol* 64:106-117.
646 Kishino H, Miyata T, and Hasegawa M. 1990. Maximum likelihood inference of
647 protein phylogeny and the origin of chloroplasts. *Journal of Molecular*
648 *Evolution* 31:151-160. 10.1007/bf02109483
649 Kocher A, Guilbert É, Lhuillier É, and Murienne J. 2015. Sequencing of the
650 mitochondrial genome of the avocado lace bug *Pseudacysta perseae*
651 (Heteroptera, Tingidae) using a genome skimming approach. *Comptes Rendus*
652 *Biologies* 338:149-160. http://dx.doi.org/10.1016/j.crvi.2014.12.004
653 Kocher A, Kamilari M, Lhuillier E, Coissac E, Péneau J, Chave J, and Murienne J.
654 2014. Shotgun assembly of the assassin bug *Brontostoma colossus*
655 mitochondrial genome (Heteroptera, Reduviidae). *Gene* 552:184-194.
656 http://dx.doi.org/10.1016/j.gene.2014.09.033
657 Krzeminska U, Wilson R, Rahman S, Song BK, Gan HM, Tan MH, and Austin CM.
658 2016. The complete mitochondrial genome of the invasive house crow *Corvus*
659 *splendens* (Passeriformes: Corvidae). *Mitochondrial DNA Part A* 27:974-975.
660 10.3109/19401736.2014.926512
661 Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, and Ussery DW. 2007.
662 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic*
663 *Acids Research* 35:3100-3108. 10.1093/nar/gkm160
664 Lee YP, Gan HM, Tan MH, Lys I, Page R, Dias Wanigasekera B, and Austin CM.
665 2016. The complete mitogenome of the New Zealand freshwater crayfish
666 *Paranephrops planifrons* White 1842 (Crustacea: Decapoda: Parastacidae).
667 *Mitochondrial DNA Part A* 27:3333-3334. 10.3109/19401736.2015.1018209
668 Lemmon AR, Emme SA, and Lemmon EM. 2012. Anchored Hybrid Enrichment for
669 Massively High-Throughput Phylogenomics. *Systematic Biology*.
670 10.1093/sysbio/sys049

671 Li C, Corrigan S, Yang L, Straube N, Harris M, Hofreiter M, White WT, and Naylor
672 GJP. 2015. DNA capture reveals transoceanic gene flow in endangered river
673 sharks. *Proceedings of the National Academy of Sciences* 112:13302-13307.
674 10.1073/pnas.1508735112
675 Malé P-JG, Bardon L, Besnard G, Coissac E, Delsuc F, Engel J, Lhuillier E, Scotti-
676 Saintagne C, Tinaut A, and Chave J. 2014. Genome skimming by shotgun
677 sequencing helps resolve the phylogeny of a pantropical tree family.
678 *Molecular Ecology Resources* 14:966-975. 10.1111/1755-0998.12246
679 McCormack JE, Tsai WLE, and Faircloth BC. 2015. Sequence capture of
680 ultraconserved elements from bird museum specimens. *Molecular Ecology*
681 *Resources*:n/a-n/a. 10.1111/1755-0998.12466
682 Meimberg H, Schachtler C, Curto M, Husemann M, and Habel JC. 2016. A new
683 amplicon based approach of whole mitogenome sequencing for phylogenetic
684 and phylogeographic analysis: An example of East African white-eyes (Aves,
685 Zosteropidae). *Molecular Phylogenetics and Evolution* 102:74-85.
686 http://dx.doi.org/10.1016/j.ympev.2016.05.023
687 Minh BQ, Nguyen MAT, and von Haeseler A. 2013. Ultrafast Approximation for
688 Phylogenetic Bootstrap. *Molecular biology and evolution*.
689 10.1093/molbev/mst024
690 Nguyen L-T, Schmidt HA, von Haeseler A, and Minh BQ. 2014. IQ-TREE: A Fast
691 and Effective Stochastic Algorithm for Estimating Maximum-Likelihood
692 Phylogenies. *Molecular biology and evolution*. 10.1093/molbev/msu300
693 Okajima Y, and Kumazawa Y. 2010. Mitochondrial genomes of acrodont lizards:
694 timing of gene rearrangements and phylogenetic and biogeographic
695 implications. *BMC Evolutionary Biology* 10:141.
696 Peng Y, Leung HCM, Yiu SM, and Chin FYL. 2012. IDBA-UD: a de novo assembler
697 for single-cell and metagenomic sequencing data with highly uneven depth.
698 *Bioinformatics* 28:1420-1428. 10.1093/bioinformatics/bts174
699 Porter ML, Pérez-Losada M, and Crandall KA. 2005. Model-based multi-locus
700 estimation of decapod phylogeny and divergence times. *Molecular*
701 *Phylogenetics and Evolution* 37:355-369.
702 http://dx.doi.org/10.1016/j.ympev.2005.06.021
703 Poulsen J, Byrkjedal I, Willassen E, Rees D, Takeshima H, Satoh T, Shinohara G,
704 Nishida M, and Miya M. 2013. Mitogenomic sequences and evidence from

unique gene rearrangements corroborate evolutionary relationships of myctophiformes (Neoteleostei). *BMC Evolutionary Biology* 13:111.

Rice P, Longden I, and Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16:276-277. 10.1016/s0168-9525(00)02024-2

Richter S, Schwarz F, Hering L, Böggemann M, and 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. 10.1093/gbe/evv224

Rode AL, and Babcock LE. 2003. Phylogeny of Fossil and Extant Freshwater Crayfish and Some Closely Related Nephropid Lobsters. *JOURNAL OF CRUSTACEAN BIOLOGY* 23:418-435. doi:http://dx.doi.org/10.1163/20021975-99990351

Ruane S, Raxworthy CJ, Lemmon AR, Lemmon EM, and Burbrink FT. 2015. Comparing species tree estimation with large anchored phylogenomic and small Sanger-sequenced molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. *BMC Evolutionary Biology* 15:1-14. 10.1186/s12862-015-0503-1

Shen H, Braband A, and Scholtz G. 2013. Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. *Molecular Phylogenetics and Evolution* 66:776-789. http://dx.doi.org/10.1016/j.ympev.2012.11.002

Shimodaira H. 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology* 51:492-508. 10.1080/10635150290069913

Shimodaira H, and Hasegawa M. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular biology and evolution* 16:1114.

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, and Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99:349-364. 10.3732/ajb.1100335

Suarez AV, and Tsutsui ND. 2004. The Value of Museum Collections for Research and Society. *BioScience* 54:66-74. 10.1641/0006-3568(2004)054[0066:tvomcf]2.0.co;2

739 Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6:
740 Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and*
741 *evolution*. 10.1093/molbev/mst197
742 Tan MH, Gan HM, Schultz MB, and Austin CM. 2015. MitoPhAST, a new automated
743 mitogenomic phylogeny tool in the post-genomic era with a case study of 89
744 decapod mitogenomes including eight new freshwater crayfish mitogenomes.
745 *Molecular Phylogenetics and Evolution* 85:180-188.
746 http://dx.doi.org/10.1016/j.ympev.2015.02.009
747 Thai BT, Tan MH, Lee YP, Gan HM, Tran TT, and Austin CM. 2016.
748 Characterisation of 12 microsatellite loci in the Vietnamese commercial clam
749 *Lutraria rhynchaena* Jonas 1844 (Heterodonta: Bivalvia: Mactridae) through
750 next-generation sequencing. *Molecular Biology Reports* 43:391-396.
751 10.1007/s11033-016-3966-2
752 Thomas WK, Pääbo S, Villablanca FX, and Wilson AC. 1990. Spatial and temporal
753 continuity of kangaroo rat populations shown by sequencing mitochondrial
754 DNA from museum specimens. *Journal of Molecular Evolution* 31:101-112.
755 10.1007/bf02109479
756 Timm L, and Bracken-Grissom HD. 2015. The forest for the trees: evaluating
757 molecular phylogenies with an emphasis on higher-level Decapoda.
758 *JOURNAL OF CRUSTACEAN BIOLOGY* 35:577-592.
759 doi:http://dx.doi.org/10.1163/1937240X-00002371
760 Tin MM-Y, Economo EP, and Mikheyev AS. 2014. Sequencing Degraded DNA from
761 Non-Destructively Sampled Museum Specimens for RAD-Tagging and Low-
762 Coverage Shotgun Phylogenetics. *PLoS One* 9:e96793.
763 10.1371/journal.pone.0096793
764 Toon A, Pérez-Losada M, Schweitzer CE, Feldmann RM, Carlson M, and Crandall
765 KA. 2010. Gondwanan radiation of the Southern Hemisphere crayfishes
766 (Decapoda: Parastacidae): evidence from fossils and molecules. *Journal of*
767 *Biogeography* 37:2275-2290. 10.1111/j.1365-2699.2010.02374.x
768 Tsang LM, Ma KY, Ahyong ST, Chan TY, and Chu KH. 2008. Phylogeny of
769 Decapoda using two nuclear protein-coding genes: Origin and evolution of the
770 Reptantia. *Molecular Phylogenetics and Evolution* 48:359-368.
771 http://dx.doi.org/10.1016/j.ympev.2008.04.009
772

773

774 **Figure legends**

775

776 **Fig. 1.** Genome skimming workflow used to recover the mitogenome and high copy
777 number nuclear genes from partial genome scans.

778

779 **Fig. 2.** Phylogenetic relationships among Northern Hemisphere freshwater crayfish
780 inferred based on nucleotide alignment of Dataset F (Table 1) comprising of 13
781 mitochondrial PCG, 12S, 16S, 18S, 28S and H3 (16,211 sites). Topology shown was
782 obtained from Bayesian inference, with Ultrafast bootstrap values (from Maximum
783 likelihood analysis) and posterior probabilities indicated as support values at each
784 node. Square brackets '[ ]' indicate conflict in topology inferred by the two
785 phylogenetic methods.

786

787 **Fig. 3.** Evaluation of alternate tree topologies through topology testing based on
788 Dataset F (13 mt-pcg (nt) + 12S + 16S + 18S + 28S + H3).

789

790 **Fig. 4.** An overview of evolutionary relationships within Astacoidea (outgroups:
791 Parastacoidea and Nephropoidea). Tree topologies were constructed from each of the
792 seven datasets (Table 1) and numbers at the upper left corner of each tree indicate
793 dataset used for phylogenetic inference. Ultrafast bootstrap and/or posterior
794 probability values are used to show support at each node while coloured branches
795 highlight differences in topology between ML (left) and BI (right) trees.

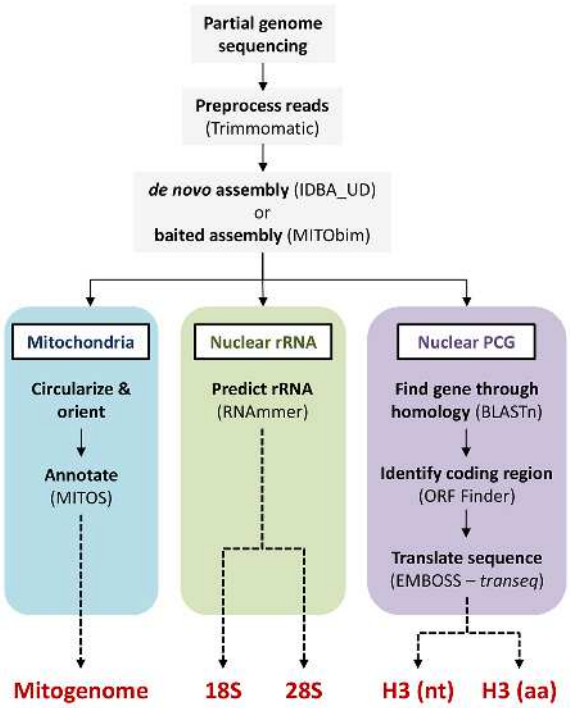
Table 1. Datasets used to construct alignments used in phylogenetic analyses. Trees inferred from these datasets are available in Supplementary Data 5.

Dataset	# Taxa	# Genes	Genes included	Alignment length (sites)
A	33	13	mt-pcg (aa)	3,657
B	33	15	mt-pcg (aa) + 12S + 16S	5,254
C	24	18	mt-pcg (aa) + 12S + 16S + 18S + 28S + H3	9,459
D	33	13	mt-pcg (nt)	10,449
E	33	15	mt-pcg (nt) + 12S + 16S	12,006
F	24	18	mt-pcg (nt) + 12S + 16S + 18S + 28S + H3	16,211
G	24	3	18S + 28S + H3	4,205

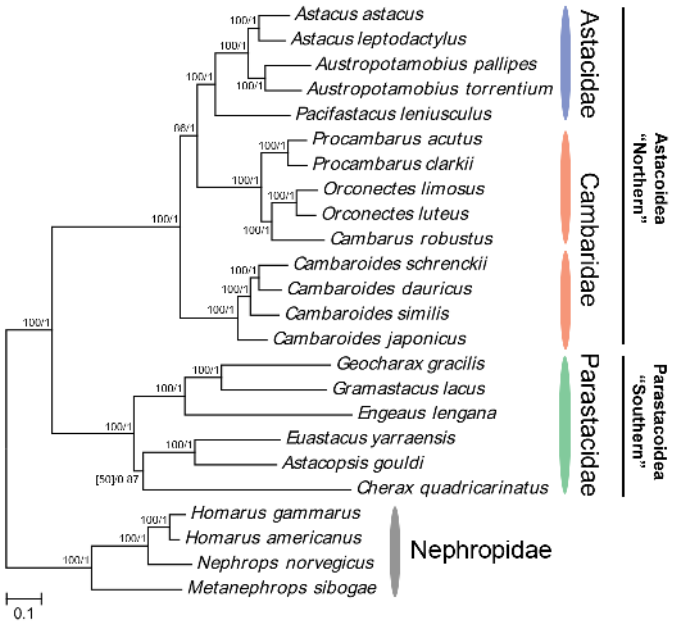
Table 2. Demonstration of the recovery of 28S rRNA, 18S rRNA and histone H3 sequences from performing genome skimming on sequence reads of animals from various taxonomic groups and tissue isolation sources. Gene sequences recovered for these animals are available in Supplementary Data 6.

Phylum	Class	Species	Tissue source	Sequence data	Recovered gene length (bp)		
					28S	18S	H3
Chordata	Actinopterygii	Gadopsis marmoratus[a]	fin clip	459 Mb	4492	1840	411
		Oryzias latipes[b]	SRA	1 Gb	4720	1842	411
	Aves	Corvus splendens[c]	liver	813 Mb	-	1822	411
	Chondrichthyes	Pastinachus atrus[d]	muscle	3.45 Gb	2699	1796	411
	Mammalia	Gallus gallus[e]	SRA	2 Gb	2065	1822	411
		Rattus norvegicus[f]	SRA	2 Gb	4803	1871	411
Mollusca	Bivalvia	Lutraria rhynchaena[g]	muscle	623 Mb	4201	1839	411
		Tridacna squamosa[h]	muscle	203 Mb	4314	1870	411
	Gastropoda	Babylonia areolata	muscle	61 Mb	4394	1828	411
Arthropoda	Branchiopoda	Triops australiensis[i]	whole	920 Mb	3988	1810	-
	Maxillopoda	Lepas anserifera	whole	425 Mb	4125	1870	411
		Pandarus rhincodonicus[j]	whole	480 Mb	-	1814	-

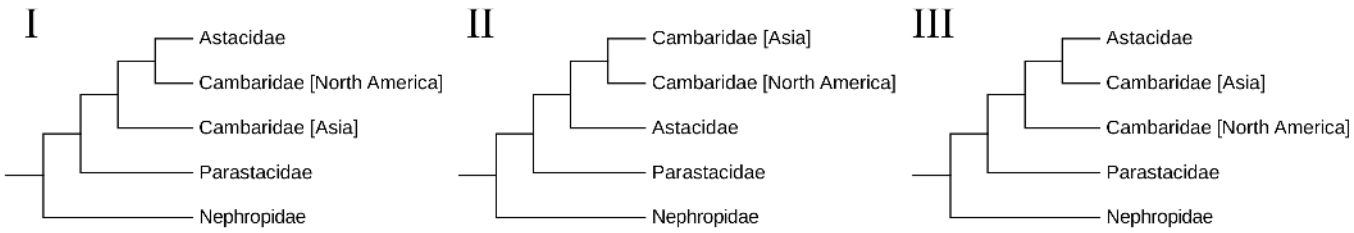
Raw reads were obtained from various internal projects and databases: a. Gan et al. (2016c); b. ERR110365 (SRA); c. Krzeminska et al. (2016); d. Austin et al. (2016a); e. SRR2131206 (SRA); f. ERR316506 (SRA); g. Gan et al. (2016d); h. Gan et al. (2016a); i. Gan et al. (2016b); j. Austin et al. (2016b).



zsc_12247_f1.tif



zsc_12247_f2.tif

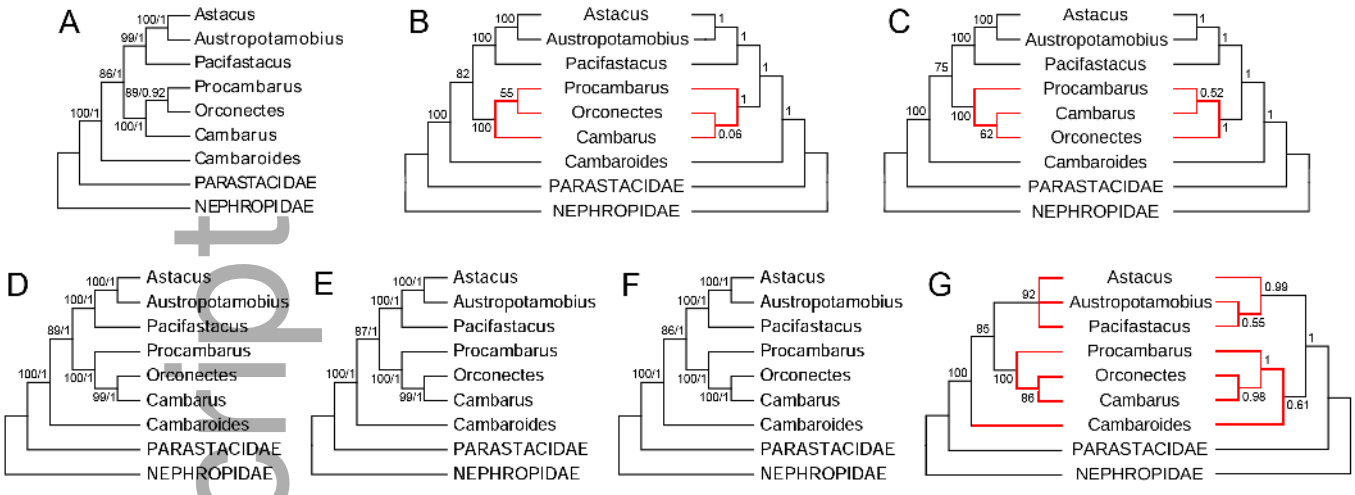


Tree	logL[1]	p-SH[2]	p-AU[3]
I	-155582.703	1.0000 (+)	0.5227 (+)
II	-155610.006	0.0350 (-)	0.0176 (-)
III	-155593.224	0.3300 (+)	0.2577 (+)

Plus signs (+) denote the 95% confidence sets
Minus signs (-) denote significant exclusion

[1] Tree log-likelihood
[2] P-value of Shimodaira-Hasegawa test (2000)
[3] P-value of approximately unbiased test (Shimodaira, 2002)

zsc_12247_f3.tif



zsc_12247_f4.tif

Author/s:
Grandjean, F;Tan, MH;Gan, HM;Lee, YP;Kawai, T;Distefano, RJ;Blaha, M;Roles, AJ;Austin, CM

Title:
Rapid recovery of nuclear and mitochondrial genes by genome skimming from Northern Hemisphere freshwater crayfish

Date:
2017-11-01

Citation:
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 (6), pp.718-728. https://doi.org/10.1111/zsc.12247.