Draft genomes and genomic divergence of two Lepidurus tadpole shrimp species (Crustacea, Branchiopoda, Notostraca)

Draft genomes and genomic divergence of two *Lepidurus* tadpole shrimp species (Crustacea, Branchiopoda, Notostraca)

Castrense Savojardo[1,†] **Andrea Luchetti**[2,†] **Pier Luigi Martelli**[1] **Rita Casadio**[1] **Barbara Mantovani**[2]

[1]Biocomputing Group, Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

[2]Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

Correspondence
Andrea Luchetti, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy.
Email: andrea.luchetti@unibo.it

Funding information
Canziani funding

Abstract

Crustaceans of the order Notostraca (Branchiopoda) are distributed worldwide and are known for the remarkable morphological stasis between their extant and Permian fossil species. Moreover, these crustaceans show relevant ecological traits and a wide range of reproductive strategies. However, genomic studies on notostracans are fairly limited. Here, we present the genome sequences of two notostracan taxa, *Lepidurus arcticus* and *Lepidurus apus lubbocki*. Taking advantage of the small genome sizes (~0.11 pg) of these taxa, genomes were sequenced for one individual per species with one run on the Illumina HiSeq X platform. We finally assembled 73.2 Mbp (*L. arcticus*) and 90.3 Mbp (*L. apus lubbocki*) long genomes. Assemblies cover up to 84% of the estimated genome size, with a gene completeness >97% for both genomes. In total, 13% 16% of the assembled genomes consist of repeats, and based on read mapping, *L. apus lubbocki* shows a significantly lower transposable element content than *L. arcticus*. The analysis of 2,376 orthologous genes indicates an ~7% divergence between the two *Lepidurus* taxa, with a nucleotide substitution rate significantly lower than that of *Daphnia* taxa. K_a/K_s analysis suggests purifying selection in both branchiopod lineages, raising the question of whether the low substitution rate of *Lepidurus* is correlated with morphological conservation or is linked to specific biological traits. Our analysis demonstrates that, in these organisms, it is possible to obtain high quality draft genomes from single individuals with a relatively low sequencing effort. This result makes *Lepidurus* and Notostraca interesting models for genomic studies at taxonomic, ecological and evolutionary levels.

KEYWORDS
Arthropoda, genome sequencing, living fossils, tadpole shrimps

1 | INTRODUCTION

The order Notostraca (Branchiopoda) comprises globally distributed crustaceans, commonly known as tadpole shrimps, ascribed to the two monophyletic genera *Triops* Schrank, 1803, and *Lepidurus* Leach, 1816 (Longhurst, 1955).

Tadpole shrimps can be found mainly in freshwater temporary ponds, where they are adapted to live. In fact, the drought resistant, diapausing eggs of these shrimps will hatch only in favourable conditions, and since the eggs do not all hatch simultaneously, they constitute a cyst bank composed of eggs from different generations and genotypes over time (Brendonck, 1996; Brendonck & De Mesteer, 2003). Although the ecological factors triggering hatching are still poorly understood, egg banks hold a biodiversity that could enable

[†]These authors contributed equally to this work.

environmental variations to be overcome, even if the present climate change poses new challenges to the resilience of this adopted strategy (Brendonck & De Mesteer, 2003; Pinceel, Buschke, Weckx, Brendonck, & Vanschoenwinkel, 2018).

Despite the ephemerality and the unpredictable nature of their habitat (Brendonck, 1996), Notostraca are also known for their ancient origin. In fact, the earliest notostracan, *Strudops goldenbergi*, dates back to the upper Devonian (approximately 420 360 million years ago; Lagebro et al., 2015), with clear representatives of extant genera reported from the Permian period (approximately 300 250 million years ago): *Triops cancriformis permiensis* (recently elevated to the species status as *T. permiensis* Gand, Garric, & Lapeyrie, 1997 [stat. nov.]; Korn, Rabet, Ghate, Marrone, & Hundsdoerfer, 2013) and *Lepidurus occitaniacus* (Gand et al., 1997). Notostraca fossil and extant species demonstrate remarkable morphological stasis, only differing by a few minor characters, leading to the controversial epithet of "living fossils" (Fryer, 1988; Mathers, Hammond, Jenner, Hanfling, & Gomez, 2013). The morphological conservation of Notostraca, however, is strikingly counterbalanced by their high variability in reproductive strategies, which range from bisexuality (either gonochoric, hermaphroditic or androdioecious) to unisexuality (parthenogenesis), even in different populations of the same species (e.g., *Triops cancriformis* or *Lepidurus arcticus*; Lakka, 2015; Mantovani, Cesari, Luchetti, & Scanabissi, 2008; Zierold, Hanfling, & Gómez, 2007 and references therein). The absence of clear diagnostic sexual characters further adds to the problem of understanding the reproductive mode of Notostraca taxa in field populations, and ultrastructural analyses may also be inconclusive. For example, *Lepidurus apus lubbocki* is described as gonochoric (Longhurst, 1955), but ultrastructural analyses showed that males are sterile, raising the hypothesis of parthenogenetic reproduction (Scanabissi & Mondini, 2002; Wingstrand, 1978).

Overall, Notostraca species share ecological and biological features that make these species an interesting framework for evolutionary studies. Moreover, among Arthropoda, Notostraca include species with very small genomes; the average C value of these genomes is 0.11 pg (~107.5 Mb), and the smallest *C* value belongs to *L. arcticus* (~88.0 Mb, 0.09 pg; Jeffery, 2015). Therefore, these organisms appear to be particularly suitable for comparative genomics and population genomics studies, especially because, from a phylogenetic point of view, the class Branchiopoda in general is within a hot spot of the Arthropoda tree of life. In fact, the relationships among (a) Branchiopoda, Remipedia, Cephalocarida and Hexapoda (Meusemann et al., 2010; Schwentner, Combosch, Nelson, & Giribet, 2017; von Reumont, Jenner, Wills, & Dell'ampio, E., Pass, G., Ebersberger, I., Misof, B., 2012), (b) Crustacea classes (Jenner, 2010) and (c) Branchiopoda orders (deWaard et al., 2006) are still highly debated.

Only a few genomic resources are available for Branchiopoda and are mainly focused on well known *Daphnia* species (*Daphnia* Genomics Consortium; Colbourne et al., 2011). We therefore initiated the genomic characterization of Notostraca taxa, starting with *L. arcticus* and *L. apus lubbocki*. *L. arcticus* is distributed in arctic regions (Brtek & Thiéry, 1995) and is the only notostracan species living in permanent lakes (Fryer, 1988). On the other hand, *L. apus lubbocki*, recently suggested to be elevated to a specific rank as *L. lubbocki* (Korn et al., 2013; Mantovani, Cesari, & Scanabissi, 2009), is distributed in South Italy, the Middle East and North Africa (Brtek & Thiéry, 1995). Molecular phylogenetic studies have indicated the latter taxon is the sister clade of the remaining *Lepidurus* species, while pointing to a derived status for *L. arcticus* (Korn et al., 2013; Mathers et al., 2013; Vanschoenwinkel et al., 2012).

Therefore, we report here on the genome sequencing and assembly of both taxa with the aim of producing resources that are useful for the study of taxonomy, phylogeny, ecology of Notostraca and for a better understanding of the reproductive biology of this ancient animal group.

2 | MATERIALS AND METHODS

2.1 | DNA isolation and genome sequencing

Individuals of *L. arcticus* and *L. apus lubbocki* were sampled in Thjórsárver (Iceland) and Castel Porziano (Italy), respectively. High molecular weight genomic DNA was extracted from single specimens using a DNA extraction kit (STRATEC) after dissection for gut removal. Dissection was performed on samples submerged in 95% ethanol with a microscalpel blade and tweezers under a stereomicroscope, and the whole gut was removed cautiously. A single library with an insert size of 350 bp (2 × 150 bp paired end reads) was produced using the Illumina TruSeq DNA PCR free library kit from a single female individual for each species. Whole genome sequencing was carried out at Macrogen Inc. (South Korea) on the high throughput Illumina HiSeq X platform.

2.2 | Genome assembly

Analysis of the heterozygosity of *L. arcticus* and *L. apus lubbocki* was performed using JELLYFISH v.2.2.10 (Marçais & Kingsford, 2011) with a k mer size of 21 bp and default parameters. The resulting k mer count histograms were then submitted to the GENOMESCOPE web server (Vurture et al., 2017; accessed on June 2018) to estimate heterozygosity.

Before assembly, raw fragment libraries of *L. arcticus* and *L. apus lubbocki* were stripped of adapters and quality trimmed (preserving read pairing) using TRIMMOMATIC v.0.35 (Bolger, Lohse, & Usadel, 2014), with minimum quality set to 20 and minimum length set to 36. Genome assemblies were performed using the ABYSS assembler v.1.5.2 (Simpson et al., 2009) based on *de Brujin* graphs. Several preliminary assemblies were obtained by using default parameters and varying the k mer size (between 32 96 bp) used by the algorithm to guide the construction of the assembly *de Brujin* graph. Scaffolding was carried out with the internal ABYSS module. The different assemblies were therefore evaluated using the QUAST genome quality assessment tool (v.3.1; Gurevich, Saveliev, Vyahhi, & Tesler, 2013) and compared according to N50 statistics.

The completeness of the assembled genomes was assessed by checking for the presence of conserved representative genes. For this purpose, we used BUSCO v. 2 (Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015), with the Arthropoda orthologue set, and CEGMA (Parra, Bradnam, Ning, Keane, & Korf, 2009), with the Core Eukaryotic Gene (CEG) orthologue set, implemented on the GVOLANTE online platform v. 1.0 (https://gvolante.riken.jp/; Nishimura, Hara, & Kuraku, 2017; accessed on June 2018).

2.3 | Genome annotation

Repeat content was assessed with de novo searches. Transposable elements (TEs) were searched using REPEATMODELER v. 1.0 (Smit & Hubley, 2008 2015) and LTR FINDER v. 1.06 (Xu & Wang, 2007) with default parameters. Final TE libraries were used as a database for REPEATMASKER v. 4.0 (Smit et al., 2013 2015) to calculate the relative abundance of TE families and the repeat landscape. The TE content in each *Lepidurus* genome was further estimated by mapping sequence reads on TE libraries with BOWTIE2 (Langmead & Salzberg, 2012): the proportion of mapped reads was used as a proxy for TE copy number. It has been suggested that this method helps avoid possible biases in TE content estimates that may occur when working on the assembly (Bast et al., 2016). The tandem repeat content was assessed with TANDEM REPEAT FINDER v. 4.07b (Benson, 1999) using default search parameters.

After repeat masking, protein coding genes were predicted with AUGUSTUS v. 2.5 (Stanke et al., 2006) by adopting two *Lepidurus* species specific gene models.

This procedure consisted of three steps: (a) the selection of two reliable training sets of gene loci in both the *L. arcticus* and *L. apus lubbocki* genomes, (b) the training of AUGUSTUS and (c) the final gene prediction with the newly generated gene models.

To build the training sets, we identified putative gene loci by running AUGUSTUS on each genome, using available gene models for *Daphnia magna*. To select the most reliable gene loci, protein sequences translated from detected genes were further aligned with the complete proteome of *D. magna* (downloaded from NCBI, accessed in September 2017). To this aim, we used the BLASTP algorithm with an e value threshold of $1e^{-4}$. We then retained the *L. arcticus* and *L. apus lubbocki* proteins that mapped to at least one *D. magna* protein with a sequence identity >90% and a coverage for both sequences >70%. On this basis, training sets for *L. arcticus* and *L. apus lubbocki* were constructed by selecting 500 random nonredundant loci (sharing a sequence similarity <30%). Training was performed with three runs of optimization and default parameters and generated two statistical models specific to *L. arcticus* and *L. apus lubbocki* that were used for ab initio gene prediction on the two genomes.

2.4 | Orthologous gene identification and phylogenetics

Orthologue groups were identified with ORTHOMCL (Li, Stoeckert, & Roos, 2003) by comparing the predicted genes with eight well annotated arthropod genomes and the corresponding predicted proteomes, all available from the NCBI database (accessed on September 2017): *D. magna* (GCA 001632505.1), *Daphnia pulex* (GCA 900092285.1), *Orchesella cincta* (GCA 001718145.1), *Zootermopsis nevadensis* (GCA 000696155.1), *Tribolium castaneum* (GCA 000002335.3), *Hyalella azteca* (GCA 000764305.2), *Limulus polyphemus* (GCA 000517525.1) and *Parasteatoda tepidariorum* (GCA 000365465.2). Selected proteomes, along with the *L. arcticus* and *L. apus lubbocki* predicted proteins, were combined into a single protein database and cross compared (all *vs* all) using BLASTP with an *e* value threshold of $1e^{-5}$. The BLASTP output was then provided as input to the ORTHOMCL algorithm (Li et al., 2003), which identified clusters of orthologous proteins.

For the phylogenomic analysis, we selected 432 core orthologue clusters that included one protein for each considered taxon (i.e., cluster size = 10). Proteins were aligned using MAFFT v. 7.205 (Katoh & Standley, 2013), with automatic detection of parameter set, and concatenated in a single super alignment. Before the phylogenomic analysis, ambiguous/noisy amino acid positions were removed with GBLOCKS v. 0.91b (Castresana, 2000; available at https://molevol.cmima.csic.es/castresana/Gblocks server.html, with all the options for less stringent block selection; accessed on June 2018). A maximum likelihood tree was computed with PHYML v. 3.0 (Guindon & Guascuel, 2003) using the LG + G + I substitution model, with nodal support calculated after 100 bootstrap replicates.

Genetic divergence between *Lepidurus* and *Daphnia* species pairs was evaluated by comparing orthologous genes at the nucleotide and amino acid levels. Both protein coding sequences and translated amino acid sequences were aligned on a codon basis using MACSE v. 1.02 (Ranwez, Harispe, Delsuc, & Douzery, 2011).

Natural selection estimation was carried out by calculating K_a/K_s values from orthologue pairs using KAKS CALCULATOR v. 0.1.1 (Wang, Zhang, Zhang, Zhu, & Yu, 2010); only pairs showing significant Fisher's test results were considered, and among these, only comparisons with $K_s > 0.001$ and $K_s < 2$ were retained to avoid saturation and/or biased calculation issues. Although a $K_a/K_s > 1$ is considered indicative of positive selection, examples of positive selection have been found with the less conservative threshold of $K_a/K_s > 0.5$ (Swanson, Wong, Wolfner, & Aquadro, 2004; Tang & Wu, 2006). We therefore considered both K_a/K_s values.

3 | RESULTS

3.1 | Genome assemblies

Overall, 84.6 and 53.6 Gbp of raw sequences were generated for *L. arcticus* and *L. apus lubbocki*, respectively (Table 1).

The 21 bp k mer analysis on raw reads estimated heterozygosities of 0.03% and 0.32% for *L. arcticus* and *L. apus lubbocki*, respectively.

After trimming, 557,702,002 and 352,774,850 clean paired reads remained for *L. arcticus* and *L. apus lubbocki*, respectively (Table 1). Considering the estimated genome sizes (Jeffery, 2015), the

TABLE 1 Results of libraries sequencing and cleaning

Species	Total read bases (Gb[a])	Number of reads	After trimming
Lepidurus arcticus	84.62	560,411,340	557,702,002
Lepidurus apus lubbocki	53.56	354,684,728	352,774,850

[a]Giga base pairs

theoretical maximum genome coverage ranges from 925 × (*L. arcticus*) to 478 × (*L. apus lubbocki*).

The best *L. arcticus* genome assembly was obtained using a k mer of 80 bp, has a total length of 74.4 Mb and comprises 14,809 contigs and 7,167 scaffolds (Table 2; Supporting Information Figure S1). The contig and scaffold N50 values are 52.2 and 97.9 kb, respectively. When scaffolds shorter than 1,000 bp are filtered out, 3,160 scaffolds are obtained with an N50 of 118.9 kb and a total size of 73.2 Mb. The assembly of *L. arcticus* corresponds to approximately 83.1% of the genome size, which is estimated to be 88.0 Mb (Jeffery, 2015). Similarly, the best assembly of *L. apus lubbocki* (Table 2; Supporting Information Figure S1) was obtained with a k mer of 64 bp, has a total length of 94.3 Mb and comprises 35,647 contigs and 20,738 scaffolds. The total size of the assembled genome is 95.2 Mb (scaffold level) with contig and scaffold N50 values of 15.2 kb and 40.2 Mb, respectively. When the 8,001 scaffolds longer than 1,000 bp are considered, the assembled genome length is 90.3 Mb, corresponding to 84.0% of the genome (assuming an estimated size of ~107.5 Mb; Jeffery, 2015). The N50 value increases to 43.4 kb.

The results indicate a high degree of completeness for both genome assemblies (Table 3). Depending on the use of BUSCO or CEGMA, 98.4% 99.2% of the core selected genes have a complete match with an orthologue in *L. arcticus*; similarly, these percentages for *L. apus lubbocki* range from 97.8% to 99.6%.

TABLE 2 Global statistics on *Lepidurus arcticus* and *Lepidurus apus lubbocki* genome assemblies

Assembly level	*L. arcticus*			*L. apus lubbocki*		
	Number	N50 (kb[a])	Total Size (Mb[b])	Number	N50 (kb[a])	Total Size (Mb[b])
Contig	14,809	73.9	74.4	35,647	15.2	94.3
Scaffold	7,167	116.3	74.7	20,738	40.2	95.2
Scaffold ≥1.0 kb	3,160	118.9	73.2	8,001	43.4	90.3

[a]Kilo base pairs. [b]Mega base pairs.

3.2 | Repeat content analysis

The analysis pipeline identified 291 and 300 interspersed repeat elements in *L. arcticus* and *L. apus lubbocki*, respectively; 64.6% and 61.3% were not successfully classified by REPEATMODELER in any of the known TE families. Overall, REPEATMASKER analysis indicates that *L. apus lubbocki* has a slightly larger proportion of repeats than *L. arcticus* (Table 4); the total fractions of interspersed repeats in the two genomes are 13.03% and 12.26%, respectively. Moreover, *L. apus lubbocki* showed an increase in tandem repeats (Table 4). Regarding TEs, *L. apus lubbocki* has a higher proportion of DNA and LTR elements than *L. arcticus*.

For a more accurate estimation of TE content, we also considered the proportion of sequencing reads mapping to TE libraries. The obtained data indicated that 4.1% of *L. arcticus* reads and 3.5% of *L. apus lubbocki* reads mapped to the respective TE library (Figure 1a), and the difference was statistically significant (Wilcoxon test, $p < 0.001$).

Then, we examined the evolutionary dynamics of *Lepidurus* interspersed repeats through repeat landscapes. In this analysis, relative repeat abundances are plotted against the Kimura genetic divergence (Kimura, 1980) of each repeat copy *versus* the consensus sequence of its family; the less the divergence is, the more recent the transposition event was. This analysis indicates there was a single wave of TE expansion in both genomes, with peaks at 7% and 12% divergence in *L. arcticus* and *L. apus lubbocki*, respectively (Figure 1b).

3.3 | Orthologous gene analysis

Overall, we identified 10,718 and 16,383 protein coding genes for *L. arcticus* and *L. apus lubbocki*, respectively. On average, the *L. arcticus* genes contain 6.76 introns and 7.71 exons, with average lengths of 269.3 bp and 232.7 bp, respectively. The *L. apus lubbocki* genes have, on average, 6.65 introns and 6.0 exons, with average lengths of 257.5 and 252.7 bp, respectively.

When considering all arthropods included in the analysis, approximately 24% (2,602) of the 10,718 proteins identified in *L. arcticus* were included in an orthologous cluster, while 8,116 proteins were not clustered (Figure 2a). A slightly lower proportion is observed in *L. apus lubbocki*, which has 3,334 clustered proteins out of 16,383 total proteins (Figure 2a).

The maximum likelihood tree obtained with 432 orthologous proteins (cluster size = 10) shows that *Lepidurus* taxa form a monophyletic cluster in a sister relationship with the *Daphnia* species clade

TABLE 3 Assessment of genomes completeness

Species	BUSCO			CEGMA		
	Complete[a] (%)	Complete + Partial[b] (%)	Missing[c] (%)	Complete[a] (%)	Complete + Partial[b] (%)	Missing[c] (%)
Lepidurus arcticus	1,049 (98.41)	1,054 (98.87)	12 (1.13)	246 (99.19)	247 (99.60)	1 (0.40)
Lepidurus apus lubbocki	1,043 (97.84)	1,050 (98.50)	16 (1.50)	247 (99.60)	248 (100.00)	0 (0.00)

[a]Number (and percentage) of core reference genes with a complete match with a predicted gene. [b]Number (and percentage) of core reference genes with a complete or partial match with a predicted gene. [c]Number (and percentage) of core reference genes without any match among predicted genes.

TABLE 4 Repeat content in assembled genomes

Repeat family	*Lepidurus arcticus, %*	*Lepidurus apus lubbocki, %*
DNA	1.81	2.77
RC/Helitron	--	0.17
LTR	1.98	2.23
LINE	0.43	0.31
SINE	0.38	0.33
Unclassified	7.66	7.22
Total TE	12.26	13.03
Tandem repeats	0.59	2.95
Total	12.85	15.98

(Figure 2b). Moreover, Branchiopoda (*Lepidurus*+*Daphnia*) has a sister relationship with Hexapoda taxa, and the other crustacean *H. azteca* is the sister taxon of the Branchiopoda + Hexapoda clade (Figure 2b).

To evaluate the extent of genomic divergence between the two *Lepidurus* taxa, we analysed the sequence variability of the 2,376 shared orthologous genes. The nucleotide variability per gene ranges from 1.28% to 77.4%, with a median divergence of 7.31% (Figure 3a), and most of the nucleotide substitutions occur at the 3rd codon position (Figure 3b).

Considering the species split occurred an estimated ~65 million years ago (Mathers et al., 2013), the overall nucleotide substitution rate is 5.63×10^{-4} substitutions/site per million years per lineage. Amino acid sequences show approximately the same level of divergence, with a median divergence of 8.47% and an estimated replacement rate of 6.52×10^{-4} substitutions/site per million years (Figure 3c; Supporting Information Table S1). The same analysis on the 3,597 orthologues of the two available *Daphnia* species indicates a nucleotide substitution rate of 7.05×10^{4} substitutions/site per million years per lineage and an amino acid replacement rate of 6.60×10^{4} substitutions/site per million years (Figure 3c; Supporting Information Table S1). Substitution rate estimates for nucleotide sequences differ significantly between *Lepidurus* and *Daphnia* (Wilcoxon test, $p < 0.001$; Figure 3c). In contrast, the amino acid replacement rate is not significantly different (Wilcoxon test, $p = 0.133$).

The evaluation of selective pressures on 2,193 *Lepidurus* spp. orthologues indicate a median $K_a/K_s = 0.227$; all calculated K_a/K_s values are <1, and 143 gene pairs showed $K_a/K_s \geq 0.5$ (Figure 3d). As a comparison, the same analysis on 2,989 *Daphnia* orthologues results in a median $K_a/K_s = 0.104$ with only 25 gene pairs with $K_a/K_s \geq 0.5$ (Supporting Information Figure S2).

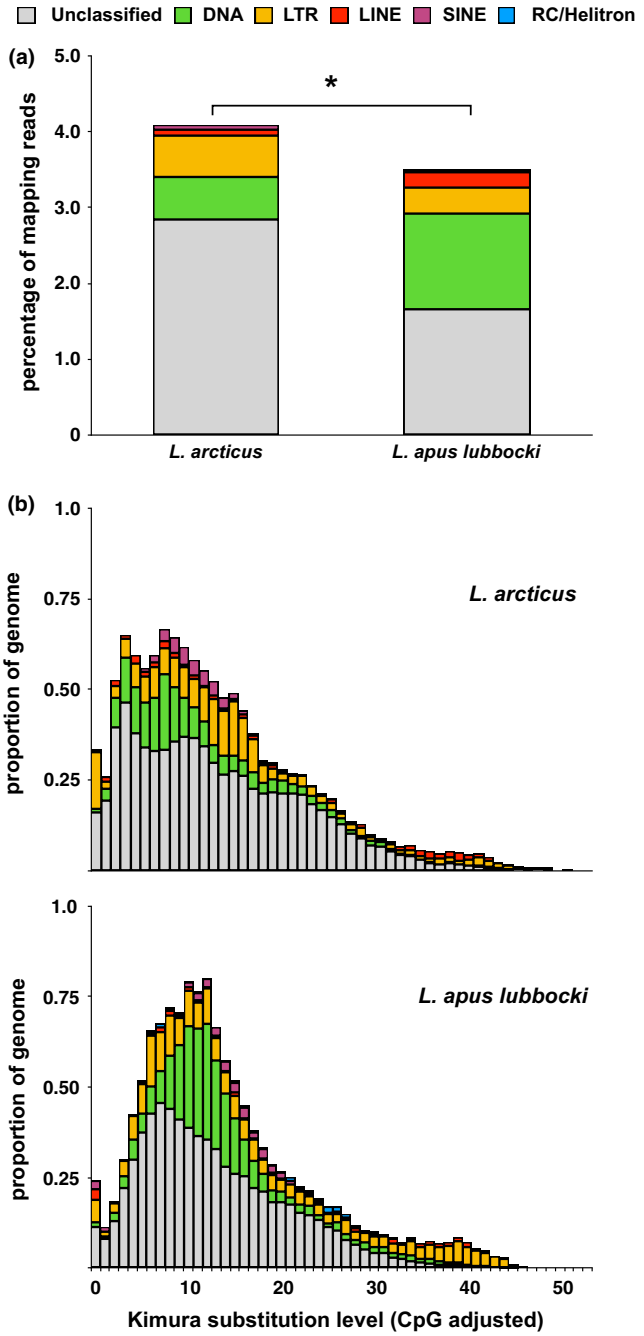

FIGURE 1 Transposable element (TE) analysis. (a) Percentage of reads mapping on TE libraries of *Lepidurus arcticus* and *Lepidurus apus lubbocki* (*$p < 0.001$). (b) TE landscapes. Each bin represents the 1% Kimura divergence of each transposable element copy from the relative family consensus sequence [Colour figure can be viewed at wileyonlinelibrary.com]

4 | DISCUSSION

Here, we report on the assembly and analysis of two *Lepidurus* draft genomes: *L. arcticus* and *L. apus lubbocki*. These branchiopods belong to the order Notostraca, a group that includes taxa with the smallest genome sizes among arthropods. We obtained two draft assemblies with a high completeness, covering up to 83% 84% of the estimated genome size (average *C* value = 0.11 pg ≈107.5 Mb). The completeness estimation based on gene content, which was calculated with two different methods, gave a very high percentage (98% 100%) of core gene presence. These are very interesting results, considering that these genomes were obtained by single organism, short insert library sequencing. Clearly, the low heterozygosity observed in the

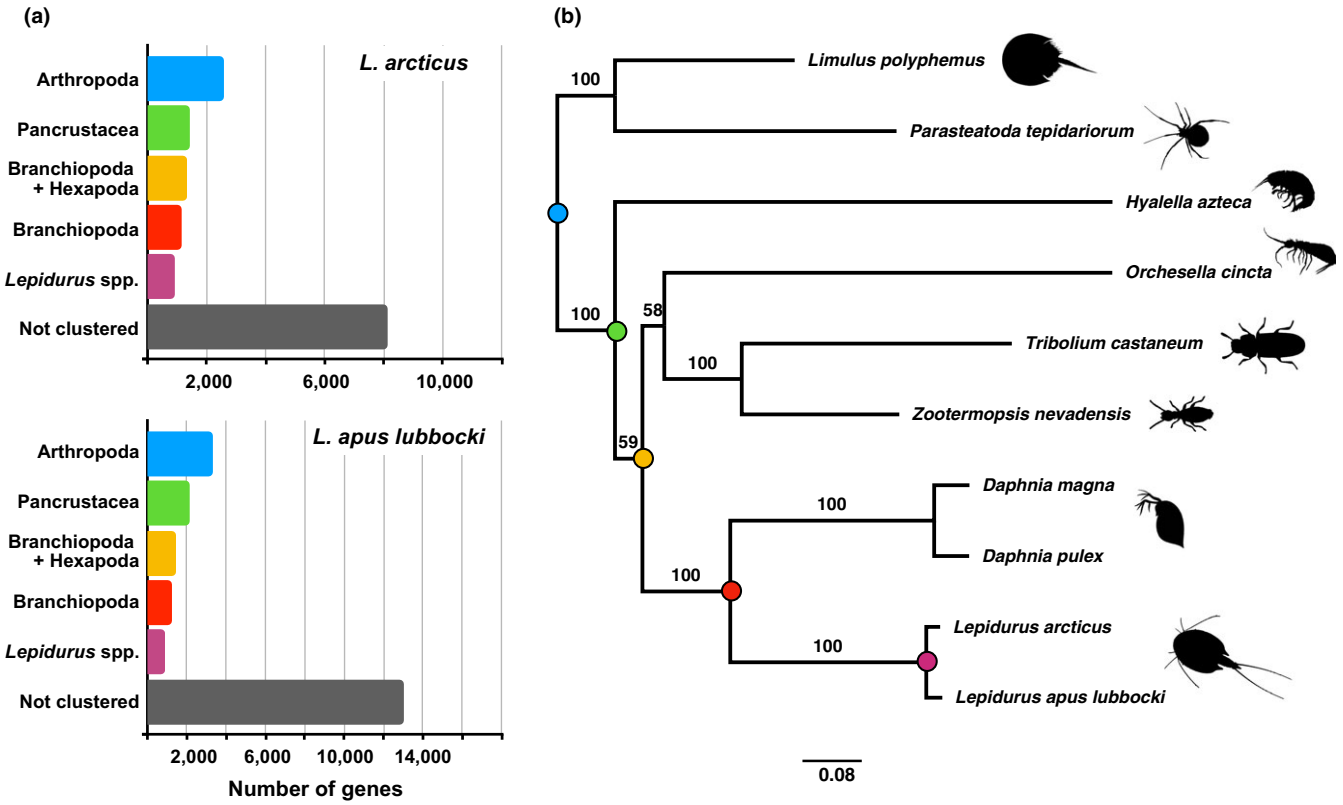

FIGURE 2 Orthologous gene analysis. (a) Taxonomic distribution of orthologous genes. (b) Maximum likelihood tree (lnL = 1,502,204.92) built on 432 orthologous proteins. The colour of squares at nodes corresponds to the taxonomic colour codes of the bins in the panel (a) bar plots; numbers above branches indicate bootstrap values [Colour figure can be viewed at wileyonlinelibrary.com]

sequencing of the two taxa, which is possibly linked to their peculiar reproductive mode (putative parthenogenesis, hermaphroditism), helped to obtain high quality draft genomes. As a comparison, the published *D. magna* and *D. pulex* genomes, which were produced with a higher sequencing effort than the genomes in this study, showed completeness percentages of 96.3% 94.7% and 97.7% 98.8%, respectively, with the same core gene search (*Daphnia* Genomics Consortium, unpublished; Colbourne et al., 2011). *Lepidurus* and *Daphnia* genomes have widely different sizes; the latter has an estimated size of 0.23 0.34 pg (≈225 332 Mb; Jalal, Wojewodzic, Laane, & Hessen, 2013), which is three times higher than the genome size estimated for *Lepidurus* taxa (Jeffery, 2015). Therefore, it is not surprising that, in terms of gene content, we obtained comparable high quality genomes with a lower sequencing effort. Notably, this opens the possibility of using Notostraca species as models for population genomic studies, as the depth of genome coverage is crucial for proper analyses (Ellegren, 2014); for this purpose, the possibility of sequencing the genomes of single individuals at high coverage with a single run is clearly an advantage.

The genomes sequenced in this study may also represent a good framework for studies on reproductive biology and associated genomic variations. In fact, while *L. arcticus* appears to be able to reproduce via either bisexual or parthenogenetic reproduction (Longhurst, 1955; Lakka, 2015; Wojtasik & Brylka Wolk, 2010), the Italian *L. apus lubbocki* population shows males with nonfunctional sperm: it was, therefore, suggested that *L. apus lubbocki* reproduces through parthenogenesis (Scanabissi & Mondini, 2002). It has been repeatedly suggested that parthenogenetic reproduction is involved in, for example, TE proliferation; in particular, genomes of parthenogenetic taxa are expected to accumulate more TEs than those of the bisexual relatives of these taxa because the low effectiveness of recombination may lead to a reduction in the TE insertion elimination rate (Nuzhdin & Petrov, 2003). However, empirical studies on bisexual and parthenogenetic lineages have given contrasting results. Studies on *Bacillus* stick insects have indicated that some parthenogenetic lineages have a relatively high TE load (Bonandin, Scavariello, Luchetti, & Mantovani, 2014; Bonandin, Scavariello, Mingazzini, Luchetti, & Mantovani, 2017), while analyses have shown less insertions in obligately parthenogenetic *Daphnia* lineages than in cyclically parthenogenetic *Daphnia* lineages (Valizadeh & Crease, 2008). Recent genomewide studies have highlighted similar TE contents in bisexual and parthenogenetic taxa, including *D. pulex* lineages (Bast et al., 2016; Kraaijeveld et al., 2012). The present analysis, carried out with the same method described in Bast et al. (2016, i.e., read mapping as a proxy for TE content), indicated that the TE content of the putatively parthenogenetic *L. apus lubbocki* is significantly lower than that of *L. arcticus*. These preliminary results, therefore, are in line with the hypothesis that TE lineages with lower proliferating activity were selected in parthenogenetic organisms to favour the survival of both host organisms and TE lineages (Bonandin et al.,

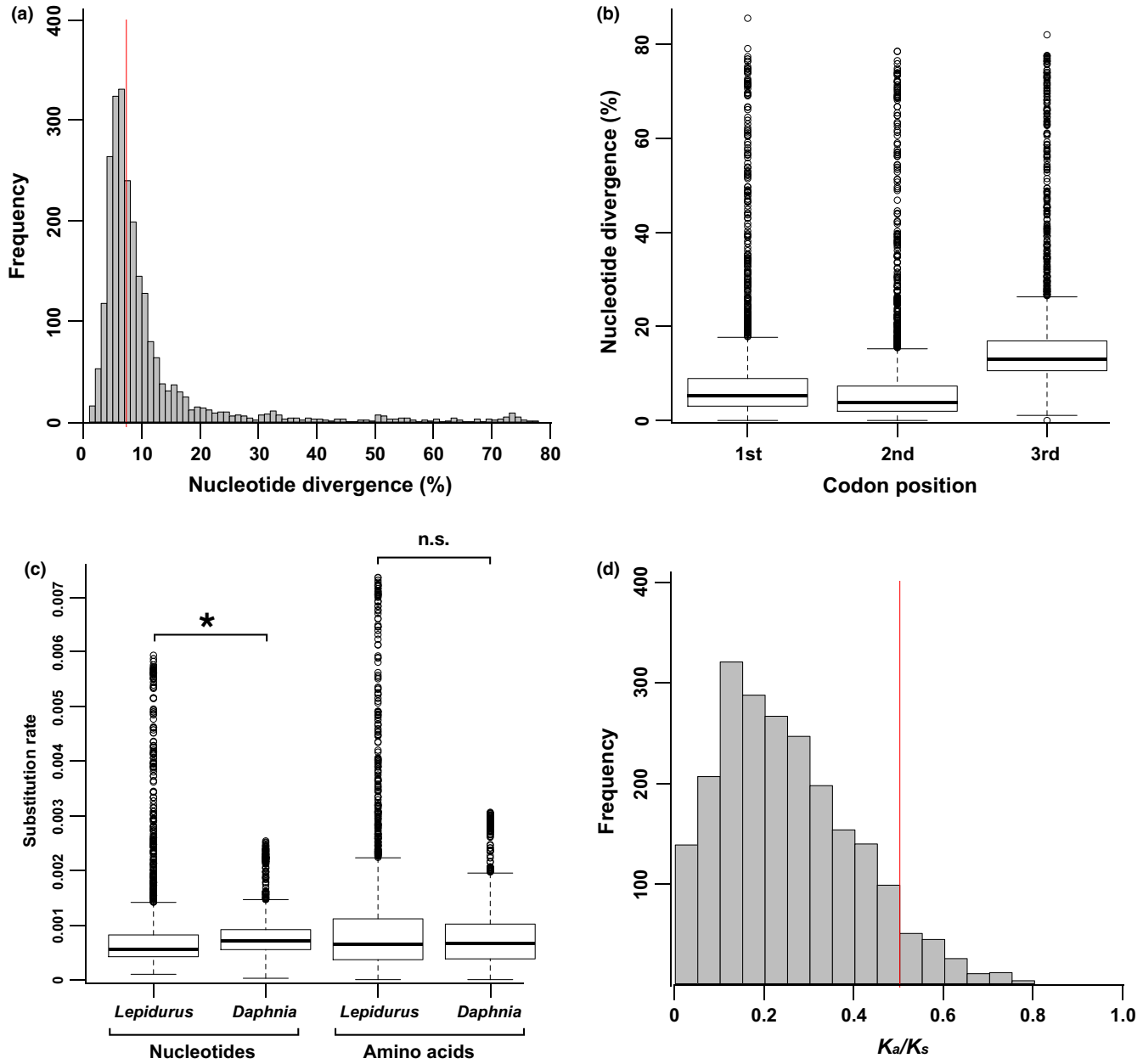

FIGURE 3 Orthologous gene sequence divergence. (a) Distribution of sequence divergence based on 2,376 *Lepidurus* orthologous genes. The red vertical line indicates the median value. (b) Sequence divergence per codon position. (c) Comparisons of nucleotide and amino acid substitution rates between *Lepidurus* and *Daphnia* species orthologues (*$p < 0.001$). (d) Distribution of K_a/K_s values calculated for *Lepidurus* orthologues; the red line indicates the 0.5 threshold [Colour figure can be viewed at wileyonlinelibrary.com]

2017; Bull, Molineux, & Rice, 1991; Wright & Finnegan, 2001). However, it could be interesting to evaluate TE content variation at the population level, considering polymorphic insertions and the relative frequency of these insertions within populations/taxa. Notwithstanding the significant difference in the TE content of *L. apus lubbocki* and *L. arcticus*, we detected similar TE proliferation dynamics in these taxa; in fact, we observed a single proliferation wave in both genomes. Moreover, the two waves apparently occurred at almost the same time. We can tentatively date the two waves by using the scored substitution rate at the nearly neutral 3rd codon position (Supporting Information Table S1) and by assuming a strict molecular clock, as described in Lander et al. (2001) and Luchetti, Plazzi, and Mantovani (2017). Although these results should be taken with caution due to potential limitations of the above described method, we estimated that the two waves occurred ~69 118 million years ago, before the two species split (Mathers et al., 2013), and during the Middle Late Cretaceous, a period characterized by a global warming phase (Frakes, Probst, & Ludwig, 1994). Links among TE proliferation, speciation and climatic fluctuation have already been hypothesized; genomic changes induced by TE activity may trigger speciation by promoting adaptation to changing environmental conditions (Belyayev, 2014). This, together with the presently obtained data,

could suggest that the increased TE activity due to climatic changes was involved in the diversification of the extant *Lepidurus* lineage.

A peculiar feature of Notostraca taxa is the remarkable morphological stasis since 250 Myr, with extant species appearing to be morphologically very similar to fossilized taxa. Our analysis of *Lepidurus* species divergence indicates that the nucleotide substitution rate of these species is significantly lower than that of *Daphnia* species, possibly indicating remarkable genetic conservation in the *Lepidurus* taxa. However, the amino acid replacement rates of *Daphnia* and *Lepidurus* are not significantly different. Although a formal selection test was not carried out, both *Lepidurus* and *Daphnia* species show low K_a/K_s values (mostly <0.5), suggesting that the genes of both taxa are mostly under purifying selection. Therefore, the low substitution rate observed in *Lepidurus* cannot be merely explained by purifying selection. A relationship between low substitution rates and morphological stasis has been suggested many times, especially for organisms identified as "living fossils"; for example, coelacanth species appears to evolve more slowly than other vertebrates (Nikaido et al., 2013), even though data on TE dynamics suggest the contrary (Naville, Chalopin, Casane, Laurenti, & Volff, 2015). Although a correlation between low substitution rates and morphological stasis cannot be excluded, the observed data may also be explained by an accelerated substitution rate within the *Daphnia* lineage. Increasing the genome sampling among brachiopods will likely better elucidate this issue. It is likely that differences in body sizes, generation times, metabolic rates and reproductive biology play a role in the different substitution rates observed between *Lepidurus* and *Daphnia* species.

The sequencing of genomes for two *Lepidurus* taxa represents the first step towards a wider study aiming to reveal the evolutionary dynamics of Notostraca and the class Branchiopoda, for which the only currently available genomes are those of two *Daphnia* water flea species. The small size of the *Lepidurus* genomes allowed draft genomes with high coverages and, regarding the genic content, a level of completeness >97% to be obtained with a single Illumina HiSeq X run per individual per species. Overall, the qualities of the presently analysed genomes are comparable to those obtained of the *D. magna* (*Daphnia* Genomics Consortium, unpublished) and *D. pulex* (Colbourne et al., 2011) genomes.

The ease with which the small genomes of Notostraca species can be sequenced with a good quality makes these organisms particularly suitable for population genomic studies. Moreover, the remarkable morphological conservation, range of reproductive strategies and peculiar ecological features of Notostraca make this order a suitable framework for ecological and evolutionary studies.

ACKNOWLEDGEMENTS

This work was supported by Canziani funding to BM. We wish to thank Franca Scanabissi for collecting *Lepidurus apus lubbocki* and Gísli Már Gíslason for helping in the *L. arcticus* collection. The *Daphnia magna* sequence data were produced by The Center for Genomics and Bioinformatics at Indiana University and distributed via wFleaBase in collaboration with the *Daphnia* Genomics Consortium. We would like to thank the Editor and two anonymous reviewers, whose comments and suggestions helped to substantially improve the manuscript.

AUTHORS' CONTRIBUTION

A.L. and B.M. designed the study; A.L. collected the samples, isolated the genomic DNA, performed the repeat finding, annotation and analysis, carried out genetic diversity and phylogenomic analyses and wrote the manuscript; C.S. assembled the genomes, carried out gene annotation and orthologue cluster analysis and wrote the manuscript. P.L.M., R.C. and B.M. supervised the analyses and helped write the manuscript.

DATA ACCESSIBILITY

Raw reads data are available in the NCBI SRA database under the Accession no. SRP130041. Genome assemblies, gene annotations and repeat libraries are available on Figshare under the https://doi.org/10.6084/m9.figshare.6729074.

REFERENCES

Bast, J., Schaefer, I., Schwander, T., Maraun, M., Scheu, S., & Kraaijeveld, K. (2016). No accumulation of transposable elements in asexual arthropods. *Molecular Biology and Evolution*, *33*, 697 706. https://doi.org/10.1093/molbev/msv261

Belyayev, A. (2014). Burst of transposable elements as an evolutionary driving force. *Journal of Evolutionary Biology*, *27*, 2573 2584.

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*, 573 580. https://doi.org/10.1093/nar/27.2.573

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114 2120. https://doi.org/10.1093/bioinformatics/btu170

Bonandin, L., Scavariello, C., Luchetti, A., & Mantovani, B. (2014). Evolutionary dynamics of R2 retroelement and insertion inheritance in the genome of bisexual and parthenogenetic *Bacillus rossius* populations (Insecta Phasmida). *Insect Molecular Biology*, *23*, 808 820.

Bonandin, L., Scavariello, C., Mingazzini, V., Luchetti, A., & Mantovani, B. (2017). Obligatory parthenogenesis and TE load: *Bacillus* stick insects and the R2 non-LTR retrotransposon. *Insect Science*, *24*, 409 417.

Brendock, L., & De Meester, L. (2003). Egg banks in freshwater zooplankton: Evolutionary and ecological archives in the sediment. *Hydrobiologia*, *491*, 65 84. https://doi.org/10.1023/A:1024454905119

Brendonck, L. (1996). Diapause, quiescence, hatching requirements: What we can learn from large freshwater branchiopods (Crustacea: Branchiopoda: Anostraca, Notostraca, Conchostraca). *Hydrobiologia*, *320*, 85 97. https://doi.org/10.1007/BF00016809

Brtek, J., & Thiéry, A. (1995). The geographic distribution of the European branchiopods (Anostraca, Notostraca, Spinicaudata, Laevicaudata). *Hydrobiologia*, *298*, 263 280. https://doi.org/10.1007/BF00033821

Bull, J. J., Molineux, I. J., & Rice, W. R. (1991). Selection of benevolence in a host-parasite system. *Evolution*, *45*, 875 882. https://doi.org/10.1111/j.1558-5646.1991.tb04356.x

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and*

Evolution, *17*, 540 552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., … Boore, J. L. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science*, *331*, 555 561. https://doi.org/10.1126/science.1197761

deWaard, J. R., Sacherova, V., Cristescu, M. E. A., Remigio, E. A., Crease, T. J., & Hebert, P. D. N. (2006). Probing the relationships of the branchiopod crustaceans. *Molecular Phylogenetics and Evolution*, *39*, 491 502. https://doi.org/10.1016/j.ympev.2005.11.003

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*, 51 63. https://doi.org/10.1016/j.tree.2013.09.008

Frakes, L. A., Probst, J.-L., & Ludwig, W. (1994). Latitudinal distribution of paleotemperature on land and sea from early Cretaceous to middle Miocene. *Comptes Rendus De L'académie Des Sciences Paris*, *318*, 1209 1218.

Fryer, G. (1988). Studies on the functional morphology and biology of the Notostraca (Crustacea: Branchiopoda). *Philosophical Transactions of the Royal Society London, Series B, Biological Sciences*, *321*, 27 124. https://doi.org/10.1098/rstb.1988.0091

Gand, G., Garric, J., & Lapeyrie, J. (1997). Biocénoses à triopsidés (Crustacea, Branchiopoda) du Permien du bassin de Lodève (France). *Geobios*, *30*, 673 700. https://doi.org/10.1016/S0016-6995(97)80157-X

Guindon, S., & Gascuel, O. (2003). PhyML: A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*, 696 704. https://doi.org/10.1080/10635150390235520

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*, 1072 1075. https://doi.org/10.1093/bioinformatics/btt086

Jalal, M., Wojewodzic, M. W., Laane, C. M. M., & Hessen, D. O. (2013). Larger *Daphnia* at lower temperature: A role for cell size and genome size? *Genome*, *56*, 511 519.

Jeffery, N. W. (2015). Genome size diversity and evolution in the Crustacea. PhD Thesis, University of Guelph.

Jenner, R. A. (2010). Higher-level crustacean phylogeny: Consensus and conflicting hypotheses. *Arthropod Structure and Development*, *39*, 143 153. https://doi.org/10.1016/j.asd.2009.11.001

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772 780. https://doi.org/10.1093/molbev/mst010

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*, 111 120. https://doi.org/10.1007/BF01731581

Korn, M., Rabet, N., Ghate, H. V., Marrone, F., & Hundsdoerfer, A. K. (2013). Molecular phylogeny of Notostraca. *Molecular Phylogenetics and Evolution*, *69*, 1159 1171.

Kraaijeveld, K., Zwanenburg, B., Hubert, B., Vieira, C., de Pater, S., van Alphen, J. J. M., … de Knijff, P. (2012). Transposon proliferation in an asexual wasp. *Molecular Ecology*, *21*, 3898 3906.

Lagebro, L., Gueriau, P., Hegna, T. A., Rabet, N., Butler, A. D., & Budd, G. E. (2015). The oldest notostracan (Upper Devonian S trud locality, Belgium). *Palaeontology*, *58*, 497 509. https://doi.org/10.1111/pala.12155

Lakka, H.-K. (2015). Description of the male *Lepidurus arcticus* (Branchiopoda: Notostraca) and the potential role of cannibalism in defining male form and population sex-ratio. *Journal of Crustacean Biology*, *35*, 319 329. https://doi.org/10.1163/1937240X-00002324

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860 921.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357 359. https://doi.org/10.1038/nmeth.1923

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*, 2178 2189. https://doi.org/10.1101/gr.1224503

Longhurst, A. R. (1955). A review of the Notostraca. *Bulletin of the British Museum of Natural History*, *D3*, 1 57.

Luchetti, A., Plazzi, F., & Mantovani, B. (2017). Evolution of two short interspersed elements in *Callorhinchus milii* (Chondrichthyes, Holocephali) and related elements in sharks and the coelacanth. *Genome Biology and Evolution*, *9*, 1406 1417. https://doi.org/10.1093/gbe/evx094

Mantovani, B., Cesari, M., Luchetti, A., & Scanabissi, F. (2008). Mitochondrial and nuclear DNA variability in the living fossil *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca). *Heredity*, *100*, 496 505. https://doi.org/10.1038/hdy.2008.3

Mantovani, B., Cesari, M., & Scanabissi, F. (2009). Molecular taxonomy and phylogeny of Italian *Lepidurus* taxa (Branchiopoda: Notostraca). *Italian Journal of Zoology*, *76*, 358 365.

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*, 764 770. https://doi.org/10.1093/bioinformatics/btr011

Mathers, T. C., Hammond, R. L., Jenner, R. A., Hanfling, B., & Gomez, A. (2013). Multiple global radiations in tadpole shrimps challenge the concept of 'living fossils'. *PeerJ*, *1*, e62.

Meusemann, K., von Reumont, B. M., Simon, S., Roeding, F., Strauss, S., Kück, P., … Misof, B. (2010). A phylogenomic approach to resolve the arthropod tree of life. *Molecular Biology and Evolution*, *27*, 2451 2464. https://doi.org/10.1093/molbev/msq130

Naville, M., Chalopin, D., Casane, D., Laurenti, P., & Volff, J.-N. (2015). The coelacanth: Can a "living fossil" have active transposable elements in its genome? *Mobile Genetic Elements*, *5*, 55 59. https://doi.org/10.1080/2159256X.2015.1052184

Nikaido, M., Noguchi, H., Nishihara, H., Toyoda, A., Suzuki, Y., Kajitani, R., … Okada, N. (2013). Coelacanth genomes reveal signatures for evolutionary transition from water to land. *Genome Research*, *23*, 1740 1748. https://doi.org/10.1101/gr.158105.113

Nishimura, O., Hara, Y., & Kuraku, S. (2017). gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*, *33*, 3635 3637. https://doi.org/10.1093/bioinformatics/btx445

Nuzhdin, S. V., & Petrov, D. A. (2003). Transposable elements in clonal lineages: Lethal hangover from sex. *Biological Journal of the Linnean Society*, *79*, 33 41. https://doi.org/10.1046/j.1095-8312.2003.00188.x

Parra, G., Bradnam, K., Ning, Z., Keane, T., & Korf, I. (2009). Assessing the gene space in draft genomes. *Nucleic Acids Research*, *37*, 289 297. https://doi.org/10.1093/nar/gkn916

Pinceel, T., Buschke, F., Weckx, M., Brendonck, L., & Vanschoenwinkel, B. (2018). Climate change jeopardizes the persistence of freshwater zooplankton by reducing both habitat suitability and demographic resilience. *BMC Ecology*, *18*, 2. https://doi.org/10.1186/s12898-018-0158-z

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. P. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One*, *6*, e22594. https://doi.org/10.1371/journal.pone.0022594

Scanabissi, F., & Mondini, C. (2002). A survey of the reproductive biology in Italian branchiopods. Part B. The male gonad of *Lepidurus apus lubbocki* Brauer, 1873 (Notostraca). *Hydrobiologia*, *486*, 273 278.

Schwentner, M., Combosch, D. J., Nelson, J. P., & Giribet, G. (2017). A phylogenomic solution to the origin of insects by resolving crustacean-hexapod relationships. *Current Biology*, *27*, 1818 1824. https://doi.org/10.1016/j.cub.2017.05.040

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and

annotation completeness with single-copy orthologs. *Bioinformatics*, *31*, 3210 3212. https://doi.org/10.1093/bioinformatics/btv351

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, *19*, 1117 1123. https://doi.org/10.1101/gr.089532.108

Smit, A. F. A., & Hubley, R. (2008 2015). *RepeatModeler Open-1.0*. Retrieved from https://www.repeatmasker.org

Smit, A. F. A., Hubley, R., & Green, P. (2013 2015). *RepeatMasker Open-4.0*. Retrieved from http://www.repeatmasker.org.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: *Ab initio* prediction of alternative transcripts. *Nucleic Acids Research*, *4*, W435 W439. https://doi.org/10.1093/nar/gkl200

Swanson, W. J., Wong, A., Wolfner, M. F., & Aquadro, C. F. (2004). Evolutionary expressed sequence tag analysis of drosophila female reproductive tracts identifies genes subjected to positive selection. *Genetics*, *168*, 1457 1465. https://doi.org/10.1534/genetics.104.030478

Tang, H., & Wu, C. I. (2006). A new method for estimating nonsynonymous substitutions and its applications to detecting positive selection. *Molecular Biology and Evolution*, *23*, 372 379. https://doi.org/10.1093/molbev/msj043

Valizadeh, P., & Crease, T. J. (2008). The association between breeding system and transposable element dynamics in *Daphnia pulex*. *Journal of Molecular Evolution*, *66*, 643 654. https://doi.org/10.1007/s00239-008-9118-0

Vanschoenwinkel, B., Pinceel, T., Vanhove, M. P. M., Denis, C., Jocque, M., Timms, B. V., & Brendonck, L. (2012). Toward a global phylogeny of the "living fossil" crustacean order of the Notostraca. *PLoS One*, *7*, e34998. https://doi.org/10.1371/journal.pone.0034998

von Reumont, B. M., Jenner, R. A., Wills, M. A., Dell'ampio, E., Pass, G., Ebersberger, I., … Misof, B. (2012). Pancrustacean phylogeny in the light of new phylogenomic data: Support for Remipedia as the possible sister group of Hexapoda. *Molecular Biology and Evolution*, *29*, 1031 1045. https://doi.org/10.1093/molbev/msr270

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, *33*, 2202 2204. https://doi.org/10.1093/bioinformatics/btx153

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., & Yu, J. (2010). KaKs Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*, *8*, 77 80. https://doi.org/10.1016/S1672-0229(10)60008-3

Wingstrand, K. G. (1978). Comparative spermatology of the Crustacea Entomostraca. 1. Subclass Branchiopoda. *Det Kongelige Danske Videnskabernes Selskabs Skrifter*, *22*, 1 66.

Wojtasik, B., & Brylka-Wolk, M. (2010). Reproduction and genetic structure of a freshwater crustacean *Lepidurus arcticus* from Spitsbergen. *Polish Polar Research*, *31*, 33 44.

Wright, S., & Finnegan, D. (2001). Genome evolution: Sex and transposable elements. *Current Biology*, *11*, R296 R299.

Xu, Z., & Wang, H. (2007). LTR FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, *35*, W265 W268. https://doi.org/10.1093/nar/gkm286

Zierold, T., Hanfling, B., & Gómez, A. (2007). Recent evolution of alternative reproductive modes in the 'living fossil' *Triops cancriformis*. *BMC Evolutionary Biology*, *7*, 161. https://doi.org/10.1186/1471-2148-7-161

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Savojardo C, Luchetti A, Martelli PL, Casadio R, Mantovani B. Draft genomes and genomic divergence of two *Lepidurus* tadpole shrimp species (Crustacea, Branchiopoda, Notostraca). *Mol Ecol Resour*. 2019;19:235 244. https://doi.org/10.1111/1755-0998.12952