



OPEN ACCESS

Edited by:

Gabriel Luz Wallau,
Aggeu Magalhães Institute
(IAM), Brazil

Reviewed by:

Geyner Cruz,
Universidade de Pernambuco, Brazil
Lukas Kratochvil,
Charles University, Czechia
Adriana Ludwig,
Carlos Chagas Institute (ICC), Brazil

***Correspondence:**

Jelena Mlinarec
jelena@biol.pmf.hr

†These authors have contributed
equally to this work and share
first authorship

†ORCID:

Višnja Besendorfer
orcid.org/0000-0001-9706-4921
Ivana Maguire
orcid.org/0000-0001-7456-8449
Carola Greve
orcid.org/0000-0003-4993-1378
Jelena Mlinarec
orcid.org/0000-0002-2627-5374
Frederick Grandjean
orcid.org/0000-0002-8494-0985
Ljudevit Luka Boštjančić
orcid.org/0000-0001-8941-9753

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 29 September 2020

Accepted: 21 December 2020

Published: 21 January 2021

Citation:

Boštjančić LL, Bonassin L, Anušić L,
Lovrenčić L, Besendorfer V, Maguire I,
Grandjean F, Austin CM, Greve C,
Hamadou AB and Mlinarec J (2021)
The *Pontastacus leptodactylus*
(Astacidae) Repeatome Provides
Insight Into Genome Evolution and
Reveals Remarkable Diversity of
Satellite DNA.
Front. Genet. 11:611745.
doi: 10.3389/fgene.2020.611745

The *Pontastacus leptodactylus* (Astacidae) Repeatome Provides Insight Into Genome Evolution and Reveals Remarkable Diversity of Satellite DNA

Ljudevit Luka Boštjančić^{1†}, Lena Bonassin^{1†}, Lucija Anušić¹, Leona Lovrenčić², Višnja Besendorfer^{1†}, Ivana Maguire^{2†}, Frederic Grandjean^{3†}, Christopher M. Austin⁴, Carola Greve^{5†}, Alexander Ben Hamadou⁵ and Jelena Mlinarec^{1*†}

¹ Division of Molecular Biology, Department of Biology, University of Zagreb, Zagreb, Croatia, ² Division of Zoology, Department of Biology, University of Zagreb, Zagreb, Croatia, ³ Laboratoire Ecologie Biologie des Interactions-UMR CNRS 7267, University of Poitiers, Poitiers, France, ⁴ Centre of Integrative Ecology, School of Life and Environmental Sciences Deakin University, Geelong, VIC, Australia, ⁵ LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt, Germany

Pontastacus leptodactylus is a native European crayfish species found in both freshwater and brackish environments. It has commercial importance for fisheries and aquaculture industries. Up till now, most studies concerning *P. leptodactylus* have focused onto gaining knowledge about its phylogeny and population genetics. However, little is known about the chromosomal evolution and genome organization of this species. Therefore, we performed clustering analysis of a low coverage genomic dataset to identify and characterize repetitive DNA in the *P. leptodactylus* genome. In addition, the karyogram of *P. leptodactylus* ($2n = 180$) is presented here for the first time consisting of 75 metacentric, 14 submetacentric, and a submetacentric/metacentric heteromorphic chromosome pair. We determined the genome size to be at ~ 18.7 gigabase pairs. Repetitive DNA represents about 54.85% of the genome. Satellite DNA repeats are the most abundant type of repetitive DNA, making up to $\sim 28\%$ of the total amount of repetitive elements, followed by the Ty3/Gypsy retroelements ($\sim 15\%$). Our study established a surprisingly high diversity of satellite repeats in *P. leptodactylus*. The genome of *P. leptodactylus* is by far the most satellite-rich genome discovered to date with 258 satellite families described. Of the five mapped satellite DNA families on chromosomes, PISAT3-411 co-localizes with the AT-rich DAPI positive probable (peri)centromeric heterochromatin on all chromosomes, while PISAT14-79 co-localizes with the AT-rich DAPI positive (peri)centromeric heterochromatin on one chromosome and is also located subterminally and intercalary on some chromosomes. PISAT1-21 is located intercalary in the vicinity of the (peri)centromeric heterochromatin on some chromosomes, while PISAT6-70 and PISAT7-134 are located intercalary on some *P. leptodactylus* chromosomes. The FISH results reveal amplification of interstitial telomeric

repeats (ITRs) in *P. leptodactylus*. The prevalence of repetitive elements, especially the satellite DNA repeats, may have provided a driving force for the evolution of the *P. leptodactylus* genome.

Keywords: FISH, genome size, interstitial telomeric repeats, karyotype, narrow-clawed crayfish, (peri)centromeric heterochromatin

INTRODUCTION

Freshwater crayfish constitute a monophyletic group of over 640 described species, arranged into four families: Astacidae, Cambaridae, Cambaroididae, and Parastacidae (Crandall and De Grave, 2017). These species are distributed across all but the Antarctic continent, the Indian subcontinent, and African mainland, with centers of diversity in the southeastern Appalachian Mountains in the North America and southeastern Australia (Crandall and Buhay, 2008). The Northern (Astacidae, Cambaroididae, and Cambaridae) and Southern (Parastacidae) hemisphere families form deeply divergent reciprocally monophyletic clades (Bracken-Grissom et al., 2014). The crayfish species of the family Astacidae belong to four genera of which *Pacifastacus* is native to North America, while *Astacus*, *Pontastacus*, and *Austropotamobius* are native to the European continent (Crandall and De Grave, 2017). In the last decades numbers and sizes of native European crayfish populations have been in decline due to climate change, degraded water quality, negative anthropogenic pressure on freshwater habitats, and the introduction of alien invasive crayfish species and their pathogens (e.g., *Aphanomyces astaci*) (Holdich et al., 2009; Kouba et al., 2014). One of the native European crayfish species is *Pontastacus leptodactylus* (Eschscholtz, 1823), found both in freshwater and brackish environments with a nowadays distribution encompassing Europe, eastern Russia, and the Middle East (Kouba et al., 2014). Up till now, the majority of studies on this species have focused on morphology, phylogeny and population genetics (Maguire and Dakić, 2011; Akhan et al., 2014; Maguire et al., 2014; Gross et al., 2017; Khoshkholgh and Nazari, 2019). Analyses of phylogenetic relationships among *P. leptodactylus* populations, using mtDNA, revealed three well-supported divergent lineages; one distributed in Europe (Croatia, Bulgaria, Poland, and Turkey) (European lineage *sensu* Maguire et al., 2014), another in Asia (Armenia, Russia) (Asian lineage *sensu* Maguire et al., 2014), and the third endemic to Turkey (Clade III *sensu* Akhan et al., 2014). While genomic information has started to accumulate for North American and Australian species (Gutekunst et al., 2018; Tan et al., 2020; Van Quyen et al., 2020), so far few studies have focused on cytogenetic and genome organization of European freshwater crayfish species (Mlinarec et al., 2011, 2016), and therefore the general aim of this study was to increase knowledge on genome evolution and diversity focusing on repetitive DNAs in *P. leptodactylus*.

The majority of animal and plant genomes contain a substantial portion of repetitive DNA, collectively referred to as the repeatome of a species, which is considered largely responsible for genome size variation. The repeatome is comprised of dispersed (DNA transposons and retrotransposons) and tandemly arranged sequences (such as nuclear ribosomal RNA genes and satellite DNAs) (Garrido-Ramos, 2017). Satellite DNAs (satDNAs) are organized in large tandem arrays of highly repetitive non-coding short sequences. SatDNAs are one of the most rapidly evolving DNAs in the genome (Garrido-Ramos, 2017). Their evolution is mainly marked by amplification and homogenization processes (both decreasing divergence) and point mutations (increasing divergence) (Ruiz-Ruano et al., 2019). Considering the differences in the size of the repeating units, satDNAs are classified into microsatellites (repeat units <10 bp), minisatellites (repeat units in the range 10–100 bp), and conventional satellites (repeat units larger than 100 bp) (Garrido-Ramos, 2017). Conventional satellites are found specifically at pericentromeric and subtelomeric locations of the chromosomes, but might be found occupying interstitial positions of the chromosomes constituting heterochromatin segments (HSs) (Garrido-Ramos, 2017). The satDNAs perform functions in the regulation of gene expression and play an important structural role in the vital functions including among others, chromosome segregation and the preservation of genetic material (Blackburn, 2005; Louis and Vershinin, 2005; Riethman et al., 2005; Kuo et al., 2006).

The characterization of repetitive DNAs from poorly characterized genomes or species lacking a reference genome can be a challenging task (Ávila Robledillo et al., 2018). Up to now, only a few satDNAs have been reported in crustaceans, mainly using traditional methods such as centrifugation through sequential CsCl gradients (Chambers et al., 1978; Wang et al., 1999). Today, repetitive DNAs can now be analyzed more easily owing to the recent advances in next generation sequencing (NGS) and high-throughput *in silico* analysis of the information contained in the NGS reads (Weiss-Schneeweiss et al., 2015; Ruiz-Ruano et al., 2019). Development of the RepeatExplorer software tool allows for *de novo* repeat identification using analyses of short sequences, randomly sampled from the genome (Novák et al., 2010, 2013). The Tandem Repeat Analyzer (TAREAN) further improved the RepeatExplorer pipeline allowing for the automatic identification and reconstruction of monomer sequences for each satDNA family in the species, collectively referred to as satellitome (Novák et al., 2017).

Decapod crustaceans present an attractive study model due to the existence of polyploidy, a large quantity of AT-rich HSs as well as the adaptation to a broad range of environments (Mlinarec et al., 2011; Martin et al., 2015; Tan et al., 2019). However, the

Abbreviations: FISH, Fluorescence *in situ* hybridization; HOR, higher order structures; HSs, heterochromatin segments; ITRs, interstitial telomeric repeats; LTRs, long terminal repeats; rDNA, ribosomal DNA; SFs, superfamilies; SSRs, simple sequence repeats; TEs, transposable elements.

majority of crustaceans have been poorly investigated at the genomic and cytogenomic level (Tan et al., 2020; Van Quyen et al., 2020). To a large extent, this is reflective of the fact that decapod crustaceans, and freshwater crayfish in particular, have a low mitotic index, a high diploid chromosome number, small chromosomes, and highly repetitive genomic elements (Tan et al., 2004, 2019, 2020; Mlinarec et al., 2011; Gutekunst et al., 2018; Van Quyen et al., 2020). Therefore, cytogenetic studies on freshwater crayfish species are rare, often limited to the report of chromosome number and structure, with very few reports on molecular cytogenetics (Tan et al., 2004; Indy et al., 2010; Scalici et al., 2010; Mlinarec et al., 2011, 2016; Kostyuk et al., 2013; Salvadori et al., 2014) (Table 1).

Keeping in mind the lack of research in the field for European freshwater crayfish, this study aims to: (i) identify and characterize repetitive sequences in the *P. leptodactylus* genome in order to get better insight into genome organization and evolution of this species, and (ii) analyze the chromosomal distribution patterns of major tandem repetitive DNA families to contribute with the chromosome organization and evolution. In addition, *COI* barcoding was used to place the samples used in this study within the context of patterns of diversity to determine the phylogenetic placement of *P. leptodactylus* individuals from Lake Maksimir.

MATERIALS AND METHODS

Samples and DNA Extraction

Seven individuals (four males and three females) of narrow-clawed crayfish *Pontastacus leptodactylus* (Eschscholtz, 1823) were collected from the Third Maksimir Lake (Zagreb, Croatia); 45.82972°N 16.02056°E.

One pereopod from each individual was removed and stored in 96% ethanol at 4°C until DNA extraction. Genomic DNA was isolated from muscle tissue using the GenElute Mammalian Genomic DNA Miniprep kit (Sigma-Aldrich, St. Louis, MO) following the manufacturer's protocol and stored at -20°C.

DNA Barcoding and Phylogenetic Network Reconstruction

Mitochondrial cytochrome oxidase subunit I (*COI*) barcode region was amplified and sequenced from genomic DNA of two individuals taken from Lake Maksimir using primer pairs LCO-1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO-2198 (5'-TAAACTTCAGGGTGACCAAAAATCA-3') described in Folmer et al. (1994). PCR reaction conditions and purification of PCR product followed the protocols described in Maguire et al. (2014). Sequencing of purified PCR products was performed by Macrogen Inc. (Amsterdam, Netherlands). Phylogenetic analysis included a total of 129 *COI* gene sequences of which 127 were downloaded from GenBank (accession KX279350), while the other two were obtained from Lake Maksimir individuals obtained in this study (Supplementary Table 1). Sequences were edited using SEQUENCHER 5.4.6 (Gene Codes Corporation, Ann Arbor, MI, USA) and aligned using MAFFT (Katoh and Standley, 2013). Sequences were collapsed to unique *COI* haplotypes using the software DnaSP 6.12.03 (Rozas et al., 2017). A

median joining network was constructed on *COI* haplotype dataset using PopArt (Bandelt et al., 1999) to visualize non-hierarchical haplotype relationships and their geographical distribution. Sites containing ambiguities were excluded from network reconstruction. This approach is recommended as a standard for cytogenetic studies as it links karyotypes with DNA barcodes (Lukhtanov and Iashenkova, 2019).

Flow Cytometry Analysis

The genome size was estimated following a flow cytometry protocol with propidium iodide-stained nuclei described in Hare and Johnston (2011). Different tissue (tail muscle, vascular tissue, and gills) of -80°C frozen adult samples of *P. leptodactylus* and neural tissue of the internal reference standard *Acheta domesticus* (female, 1C = 2Gb) was each mixed and chopped with a razor blade in a petri dish containing 2 ml of ice-cold Galbraith buffer. The suspension was filtered through a 42-µm nylon mesh and stained with the intercalating fluorochrome propidium iodide (PI, Thermo Fisher Scientific) and treated with RNase II A (Sigma-Aldrich), each with a final concentration of 25 µg/ml. The mean red PI fluorescence of stained nuclei was quantified using a Beckman-Coulter CytoFLEX flow cytometer with a solid-state laser emitting at 488 nm. Fluorescence intensities of 5000 nuclei per sample were recorded. We used the CytExpert 2.3 software for histogram analyses. The total quantity of DNA in the sample was calculated as the ratio of the mean fluorescence signal of the 2C peak of the stained nuclei of the crayfish sample divided by the mean fluorescence signal of the 2C peak of the stained nuclei of the reference standard times the 1C amount of DNA in the reference standard. Three individuals were scored to produce biological replicates. For one individual we prepared different tissues to make sure that we have not used polyploid tissue. The genome size is reported as 1C, the mean amount of DNA in Mb in a haploid nucleus.

Next Generation Sequencing, Data Pre-processing, and Clustering Analysis

Raw Illumina pair-end reads 150 bp long obtained from low coverage DNA-seq experiments on *Pontastacus leptodactylus* are available from the European Nucleotide Archive (NGS run accession: SRR7698976). After the quality filtering (quality cut-off value: 10 according to Novák et al., 2020b; percent of bases in sequence that must have quality equal to/higher than the cut-off value: 95 and filtered against a customized database containing *P. leptodactylus* mitochondrial sequences), the reads were subjected to similarity-based clustering analysis using RepeatExplorer2 (Novák et al., 2010, 2013). We used a subset of reads (2 × 125,000) representing coverage of 0.002×. Genome coverage was calculated as follows: coverage = (r × l)/g, where r corresponds to number of reads used in our analysis, l to read length and g to haploid genome size of *P. leptodactylus*. The clustering was performed using the default settings of 90% similarity over 55% of the read length. To confirm the results obtained through the RepeatExplorer pipeline, reconstruction of monomer sequences of individual satellite DNA families was performed using TAREAN analysis, specific for identification of satellite DNA repeats (Novák et al., 2017).

TABLE 1 | Chromosomal and cytogenetic characteristics of freshwater crayfish species of families Astacidae, Cambaridae, and Parastacidae.

Family/species	2n	45S rDNA	Karyotype formula	IHSs	Reference
Family Astacidae					
<i>Astacus astacus</i>	176	2	52m+35sm+1a	22 pairs of chromosomes	Mlinarec et al., 2011
<i>Pontastacus leptodactylus</i> ^{*(1)}	180	2		6 pairs of chromosomes	Mlinarec et al., 2011
<i>Pontastacus leptodactylus</i> ^{*(2)}	180	1	75m+14sm+1sm/m	10 pairs of chromosomes	This study
<i>Austropotamobius torrentium</i>	176	2	76m+11sm+1a		Mlinarec et al., 2016
<i>Austropotamobius pallipes</i>	176	2	76m+11sm+1a		Mlinarec et al., 2016
<i>Pacifastacus leniusculus</i>	376				Niiyama et al., 1962
Family Cambaroididae					
<i>Cambaroides japonicus</i>	194				Komagata and Komagata, 1992
Family Cambaridae					
<i>Procambarus clarkii</i>	188	2			Salvadori et al., 2014
<i>Procambarus llamasii</i>	120		120t		Indy et al., 2010
<i>Procambarus digueti</i>	102		35M+15m+1st		Diupotex Chong et al., 1997
<i>Procambarus alleni</i>	188				Martin et al., 2015
<i>Procambarus fallax</i>	184				Martin et al., 2015
<i>Procambarus virginialis</i> ^{*(3)}	276		171m+39sm+3st+63t		Martin et al., 2015
<i>Faxonius virilis</i>	200				Fasten, 1914
<i>Faxonius immunis</i>	208				Fasten, 1914
Family Parastacidae					
<i>Cherax destructor</i>	188		70m+42sm+48st+28t		Scalici et al., 2010
<i>Cherax quadricarinatus</i>	200		33m+25sm+14st+28t		Tan et al., 2004

2n, diploid chromosome number; 45SrDNA, number of 45S rDNA loci; IHSs, number of homologous chromosome pairs possessing probable interstitial heterochromatic segments. m, metacentric; ms, metacentric-submetacentric; sm, submetacentric; a, acrocentric; t, telocentric; st, subtelocentric chromosome. Species classification according to Crandall and De Grave (2017).

^{*(1)} Euroasian lineage.

^{*(2)} Asian lineage I.

^{*(3)} autotriploid species.

Repeat Classification

Repeat cluster classification of the top 0.01% clusters identified in comparative analysis was implemented in RepeatExplorer through which similarity searches with DNA and protein databases. After *de novo* identification of contigs that make up repetitive elements in RepeatExplorer, contigs were further classified using two homology-based approaches applied in LTRClassifier (Monat et al., 2016), specific for LTR retrotransposons, and Censor (Jurka et al., 1996) for all repetitive elements. This was followed by manual examination of individual clusters graph shapes, similarity searches using BLASTN and BLASTX against public databases (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), inspection for the presence of sub-repeats using program dotmatcher (<https://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher>) with parameters specific to individual monomer length (10% of length as window size and sequence specific similarity cut off), for the final manual annotation and quantification of repeats.

Putative satellite repeats were identified based on the properties of cluster graphs obtained by similarity-based clustering of low coverage genome sequencing Illumina reads, as implemented in the TAREAN pipeline (Novák et al., 2017). All satellite repeats with an abundance exceeding 0.1% of the *P. leptodactylus* genome were subjected to detailed sequence

analysis (Supplementary Table 2). This analysis focused on AT content, genomic abundance, and presence of telomeric (TTAGG)_n repeats and detection of sequence similarities (Supplementary Table 2). Individual satellite DNA clusters were further classified into the satellite groups via h-CD-HIT-EST (Fu et al., 2012) in two consecutive runs, with sequence identity cut-off set at 90% followed by 80% cut-off. Algorithm parameters were kept at default value. Furthermore, we classified tandem repeats as minisatellites (10–100 bp) and conventional satellites (>100 bp) depending on the monomer size (Garrido-Ramos, 2017).

To explore the relation between the repeat length and the %GC of satellite DNA we first performed Shapiro–Wilk's test to access the normality of both length and the %GC variable. Because the length variable did not follow normal distribution, we used non-parametric Spearman's rank correlation test to access correlation between two variables. Bioinformatic and statistical analysis were conducted in the R software environment (R Core Team, 2016).

Primer Design, PCR Amplification, and Cloning of Satellite DNA Families

From the *P. leptodactylus* reference monomers, outward facing primers were designed (Table 2, Supplementary Figure 1).

TABLE 2 | Characterization of selected satellite DNA families in *P. leptodactylus*.

Satellite family	Genome abundance (%)	Monomer (bp)	GC (%)	Localization	Primer Sequence (5'-3')
PISAT1-21	10.91	21	47.62	Interstitial	F-AGTTTCAATCGTCCCTGCTG R-TCAGCAGGGACGATTGAAAC
PISAT3-411	1.29	411	26.76	(Peri)centromeric	F-TGTCTATTTCCGTATATTGTAATGA R-ATCAACCATTTCATTTCGTTTC
PISAT6-70	0.40	70	35.71	Interstitial	F-GACATGTTTACATTAGACTTGTGA R-TATATGTGCCTGCAAGGTAAGT
PLSAT7-134	0.35	134	29.10	Interstitial	F-GGCAAGCCCAATTGGGTCTGA R-TCCGTAACGAAAGTAGAC
PLSAT14-79	0.17	79	44.30	Subtelomeric, interstitial, (peri)centromeric, the whole arm	F-GGTCAGTAAGCTATTGTGTGT R-CAACCTATGGAAGTTATTAAGG

Repeat unit lengths, G+C content (%), abundances (%), chromosomal localization, and primer pairs used for satellite repeat amplification.

Specific primer pairs have been used for amplification of satellite DNA probes for FISH. All PCRs were performed using GoTaq[®] Green Master Mix (Promega, Madison, WI, USA): 1X GoTaq[®] Green Master Mix, 10 pmol of each primer (Macrogen, Amsterdam, The Netherlands) and 1 µl of template DNA (16 ng), in a 50 µl final reaction volume. PCR program consisted of 35 cycles, each with 1 min denaturation at 95°C, 10 s annealing at 56°C, 1 min extension at 72°C, and a final extension of 20 min.

The sequences of the amplified monomers were verified by cloning of the PCR product into pGEM-T Easy vector according to the manufacturer's instruction (Promega, Madison, WI, USA). Amplicons were extracted and purified using ReliaPrep[™] DNA Clean-Up and Concentration System and cloned into pGEM-T Easy vector according to the manufacturer's instruction (Promega, Madison, WI, USA). The individual clones (from one to four per sample) were sequenced by Macrogen (Amsterdam, The Netherlands).

Preparation of Chromosome Spreads, Chromosome Measurements, and Idiogram Reconstruction

Four adult males ($m = 17.01, 16.10, 32.21, \text{ and } 16.27$ g) were used for the cytogenetic study. Chromosome spreads were prepared according to the method described in Mlinarec et al. (2011). Individual chromosomes in karyotype were measured using LEVAN plug-in (Sakamoto and Zacaro, 2009) for the program ImageJ (Schneider et al., 2012) to obtain the relative chromosomal length (RCL) data. RCL were then imported into the RIdiogram package (Hao et al., 2020) of R programming environment for the ideogram reconstruction. Idiogram was further modified in the Inkscape vector graphics software (Inkscape Project, 2020) to include the 45S rDNA and DAPI-positive bands.

Fluorescence *in situ* Hybridization (FISH)

The 2.4 kb *Hind*III fragment of the partial 18S rDNA and ITS1 from *Cucurbita pepo*, cloned into the pUC19 vector, was used as the 45S rDNA probe (Torres-Ruiz and Hemleben, 1994). Telomeric DNA was generated by PCR amplification in the absence of template using primers (TTAGG)₄ and (CCTAA)₄

according to Ijdo et al. (1991). Probes used to map satDNAs in the chromosomes were DNA fragments cloned into the plasmid vector. Plasmids containing the monomer sequence were directly labeled with either Aminoallyl-dUTP-Cy3 (Jena Bioscience GmbH, Jena, Germany) or Green-dUTP (Abbott Molecular Inc., USA) using Nick Translation Reagent Kit according to the manufacturer's instructions (Abbott Molecular Inc., USA) with some modifications: Plasmid DNA (700 ng) was labeled in a total volume of reaction of 25 µl using 2.5 µl of enzyme mixture for 6 h at 15°C. FISH was performed according to Mlinarec et al. (2019) with slight modification: chromosome preparations were denatured at 72°C for 5 min after applying the hybridization mix. The preparations were mounted in Dako Fluorescence Mounting Medium (Dako North America Inc., USA) and stored at 4°C overnight. Signals were visualized and photographs captured using an Olympus BX51 microscope, equipped with a cooled CCD camera (Olympus DP70). Single channel images were overlaid and contrasted using Adobe Photoshop 6.0 with only those functions that apply to the whole image. An average of 10 well-spread metaphases was analyzed per each individual.

Accession Codes

Cloned sequences of satellite repeats were deposited in GenBank under accession numbers MW044674 for PISAT1-21, MW044678 for PISAT3-411, MW044675 for PISAT6-70, MW044677 for PISAT7-134, and MW044676 for PISAT14-79. *COI* gene sequences were deposited in GenBank under accession numbers MW045515 for Hap1 and MW045516 for Hap2.

RESULTS

DNA Barcoding and Phylogenetic Network Reconstruction

Phylogenetic tree was constructed to place samples used in this study within the context of patterns of diversity across the range of *P. leptodactylus*. Final alignment consisted of *COI* barcode sequences 487 bp long and included 91 unique haplotypes from across 10 countries. Haplotype relatedness and geographical haplotype distribution is presented in the

Supplementary Figure 2. Three distinct lineages were observed in the median joining network, separated by 8–24 mutational steps. DNA barcoding showed that the samples from the lake Maksimir (Zagreb, Croatia) belong to the Asian lineage *sensu* Maguire et al. (2014) and formed two haplotypes (Hap 1 and Hap 2) closely related to haplotypes from Armenia.

***Pontastacus leptodactylus* Karyotype and Genomic Organization of 45S rDNA and Telomeric (TTAGG)_n Repeats**

The karyogram of *P. leptodactylus* ($2n = 180$) is presented here for the first time (Figures 1A,B). The karyotype consists of 75 metacentric, 14 submetacentric, and 1 submetacentric/metacentric heteromorphic chromosome pair. Thus, the proposed diploid formula is $2n = 75m + 14sm + 1sm/m$. The probable HSs revealed after DAPI staining were found in the (peri)centromeric region of all chromosome pairs as well as in the intercalary regions of 10 chromosome pairs. FISH performed with the 45S rDNA probe revealed two signals positioned on the entire longer arm of the submetacentric/metacentric heteromorphic chromosome pair (Figures 1A,B). Chromosome size and morphology of each chromosome pair within the complement is presented in Supplementary Table 3, while idiogram with position DAPI-positive bands and 45S rDNA loci is presented in Supplementary Figure 3.

FISH experiments using the probe (TTAGG)_n revealed strong and consistent signals in the terminal ends of both chromosomal arms of all *P. leptodactylus* chromosomes. The telomeric probe also hybridized to interstitial regions (ITRs) of eight chromosome pairs (Figures 1C,D). The ITR signals were of different sizes and intensity and the majority of ITR signals were more intense than the signals in the terminal chromosome ends. All ITRs were devoid of microscopically recognizable heterochromatic regions and did not co-localize with 45S rDNA loci.

***Pontastacus leptodactylus* Repeatome Characterization and Identification of Tandem Repeats**

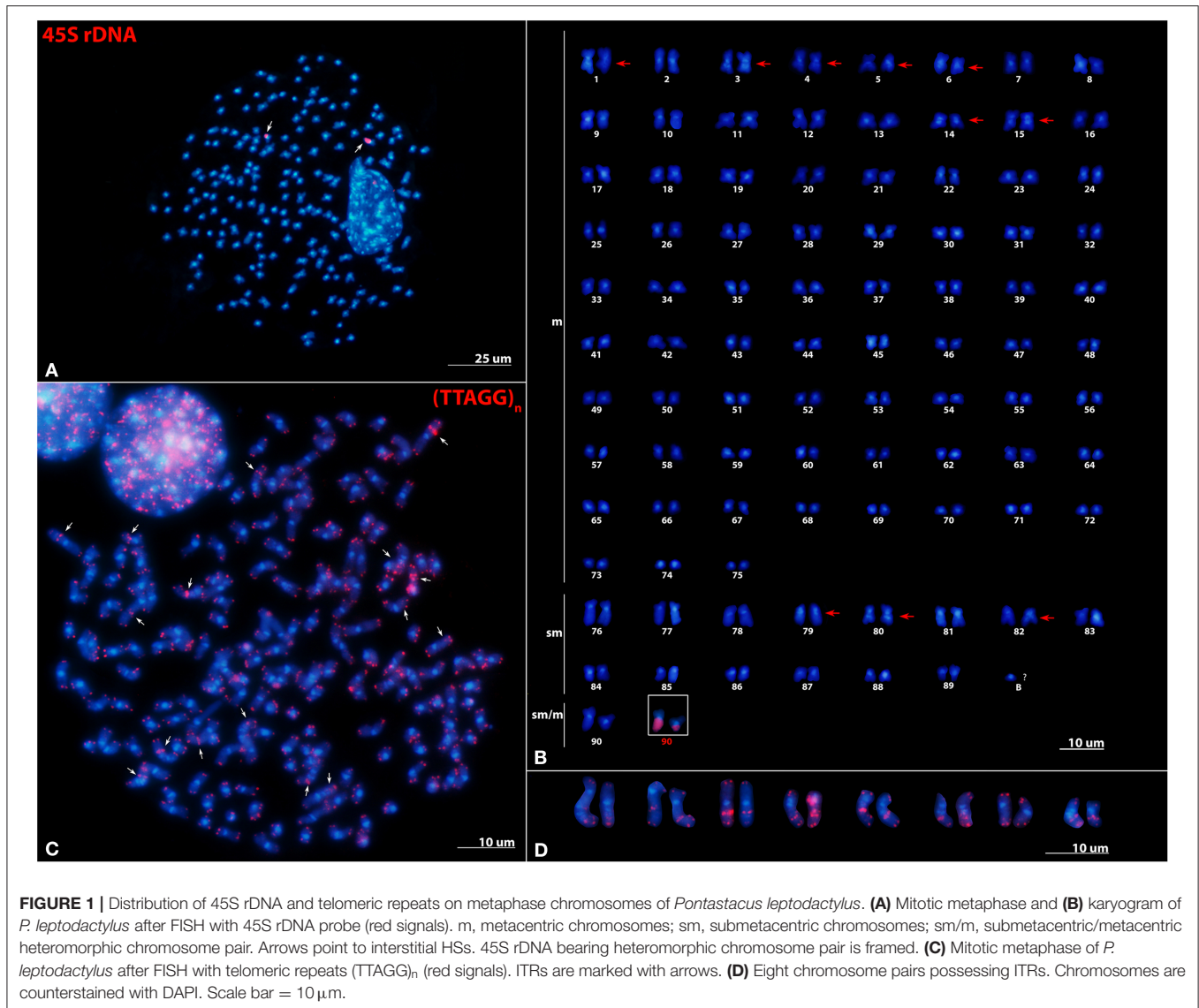
The genome size of *P. leptodactylus* was measured in three individuals from a single population. Results showed that the average 1C DNA value was 18.7 Gbp (Figure 2). Clustering of $2 \times 125,000$ paired-end reads resulted in 19,092 clusters. The nuclear repetitive DNA constituted 54.85% of the genome (Table 3). Of all the repetitive elements, 84.1% were classified to the known repetitive element groups (belonging to 37 major categories), while 4.48% remained unclassified as “other.” Satellite repeats were the most abundant elements, representing 27.52% of the genome, of which minisatellites (10–100 bp) comprised 24.7%, while conventional satellites (>100 bp) comprised 2.87% of the genome. Transposable elements (TEs) contributed 22.67% to the *P. leptodactylus* nuclear genome. Repeats classified as LTR retrotransposons represented the major fraction of the TEs of *P. leptodactylus*, comprising 15.32% (71 clusters) of nuclear DNA, followed by DIRS, LINE, and Penelope elements that comprised 3.57% (4 clusters), 2.23% (33 clusters), 1.00% (2

clusters) of nuclear DNA, respectively. LTR retrotransposons were mostly represented by Ty3/*gypsy* elements (14.95%, 55 clusters), followed by Ty1/*copia* (0.1%, 4 clusters), BEL (0.05%, 2 clusters), and ERV (0.03%, 1 cluster). DNA transposons constituted 0.51% (23 clusters) of the nuclear genome, with Helitrons as the most abundant (0.15%, 6 clusters). Ribosomal RNA genes (45S rDNA) represented 0.01% (1 cluster) of the genome (Table 3).

Based on the RepeatExplorer pipeline, 258 satellite DNA families have been identified. Satellite DNA families have been designated as PISAT1-21, through PISAT258-57 (stands for *Pontastacus leptodactylus* satellite 1 through to 258 in decreasing genomic abundance, with the respective monomer length separated by a dash; Supplementary Table 2). Their unit lengths ranged from 14 to 664 bp (median value 59 bp; Supplementary Table 2). The distribution of the lengths was biased due to the predominance of short satellite repeats, with more than half (240) being classified as minisatellites. The A+T content of the consensus satDNA sequences varied between 29.17 and 73.14% among the families, with a median value of 54.34%, which indicated a slight bias toward A+T rich satellites. Spearman’s rank correlation test showed no significant correlation between satellite length and A+T content (p -value: 0.368, correlation coefficient: 0.056) (Figure 3). Only one monomer of the perfect telomeric sequence motif (TTAGG/CCTAA) was present within the consensus sequence of 13 satellite elements, while monomers of other satDNAs contained no telomeric sequence motifs. Based on BLAST searches the satDNA sequences showed no similarity with any other DNA sequence deposited in non-redundant databases. Supplementary Table 2 shows the reconstruction of representative monomer sequences for each satDNA family. Genomic abundance of satellite DNAs ranged from 0.01% up to 10.91% of the genome (Supplementary Table 2). SatDNA family PISAT1-21 showed the highest abundance (10.91%), followed by PISAT2-21 (3.79%) and PISAT3-411 (1.29%).

Detailed Characterization and Chromosomal Localization of PISAT3-411, PISAT6-70, PISAT7-134, and PISAT14-79 Satellite DNA Families

Five satellite DNA families, PISAT1-21, PISAT3-411, PISAT6-70, PISAT7-134, and PISAT14-79 were selected for further analysis (Table 2, Supplementary Figures 1, 4–6). Firstly, to confirm their tandem arrangement the predicted monomer sequences of selected satellite DNA families have been validated by performing PCR with *P. leptodactylus* genomic DNA as a template using primers designed to face outwards from the reconstructed monomer consensus (Table 2, Supplementary Figure 1). In this arrangement, the amplification can occur only between the primer pairs located in adjacent tandemly repeated arrays. All five putative repeats tested using this assay produced the expected amplification products, and their cloned sequences (from one to four per satellite) matched the predicted consensus with 82–100% similarity. The lowest similarity (82%) was observed between



cloned and predicted consensus PISAT7-134 repeat, while the other four satellite families exhibited 95–100% similarity between cloned and predicted consensus sequence. We selected the one with the highest identity to the reference monomer as the probe for subsequent hybridizations.

Dot plot analysis of PISAT1-21, PISAT3-411, PISAT6-70, PISAT7-134, and PISAT14-79 did not reveal any consecutive tandem sub-repeats, although multiple poly-A and poly-T repetitions were observed in GC poor satellite repeat families PLSAT3-411 and PLSAT7-134 (Supplementary Figure 6).

Chromosome mapping of the PISAT1-21, PISAT3-411, PISAT6-70, PISAT7-134, and PISAT14-79 satellites revealed distinct hybridization sites, with reproducible and unambiguous markings for all analyzed mitotic metaphases (Figure 4). PISAT1-21 satellite family hybridized to the interstitial positions in the vicinity to the probable (peri)centromeric HSS on some chromosomes (Figure 4A). The PISAT3-411

satellite hybridized in the (peri)centromeric regions, labeling all probable (peri)centromeric HSS on all *P. leptodactylus* chromosomes (Figure 4B). The PISAT7-134 and PISAT6-70 satellite families hybridized to the interstitial positions of some chromosomes (Figures 4C,D). The PISAT14-79 satellite family co-localized with the AT-rich DAPI-positive probable (peri)centromeric heterochromatin on some chromosomes and is also located subterminally and intercalary on some chromosomes (Figure 4D). Besides, the PISAT14-79 probe marked the whole shorter arm of one chromosome pair. The PISAT6-70, PISAT7-134, and PISAT14-79 signals co-localized with some interstitial probable HSS.

Similarity Between satDNA Families of *P. leptodactylus*

Some longer satDNA families showed similarity to other shorter families. Of 258 satellite repeats characterized in

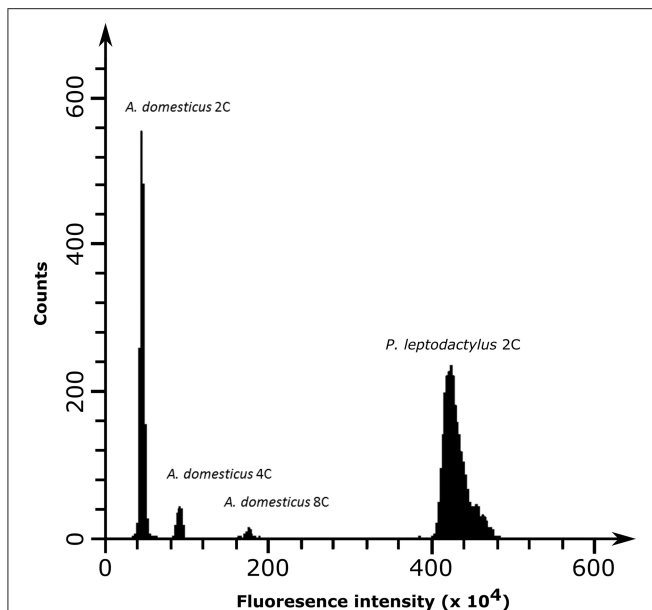


FIGURE 2 | Flow cytometry histograms of neural tissue from house cricket *Acheta domesticus* 2C (first peak), *A. domesticus* 4C (second peak), *A. domesticus* 8C (third peak), and vascular tissue from *P. leptodactylus* (fourth peak) obtained by PI fluorescence dye excitation and counts representing the cell population.

P. leptodactylus, 39 repeats showed similarities, forming 18 groups. Each group consisted of two or three satellite repeats. Similarity within each group ranged from 55 to 78%, average similarity is 63%. Only one satDNA family, PISAT75-664, showed complex units including sub-repeats with high percentages of similarity to other shorter family, PISAT3-411 (Figure 5). Detailed analysis showed that PISAT75-664 unit includes the complete PISAT3-411 unit and four direct sub-repeats, each ~70 bp long, each showing high similarity (79.9, 80.6, 70.21, and 56.82%) to 3' end of the core of the PLSAT3-411 unit (Figure 5).

DISCUSSION

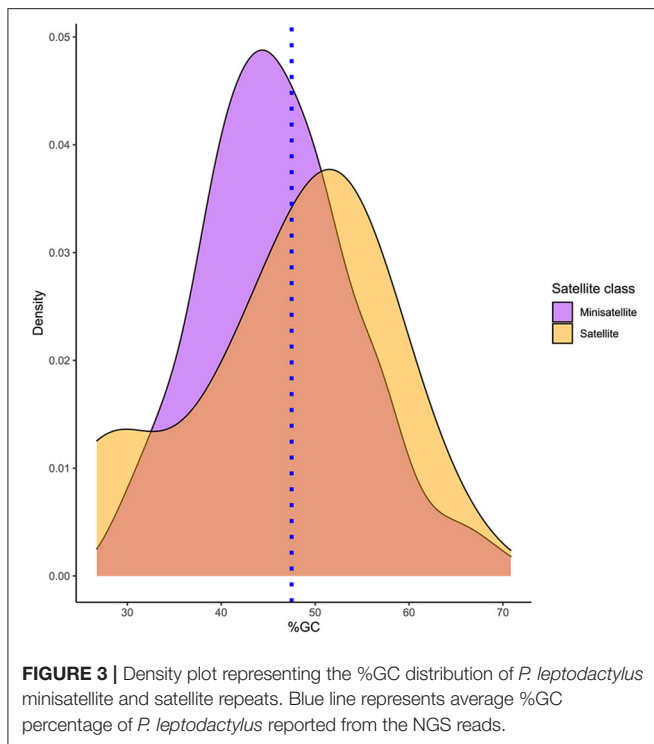
Phylogenetic Placement of *P. leptodactylus* Individuals Used in This Study

Although, *Pontastacus leptodactylus* is naturally distributed across Europe, previous study by Njegovan et al. (2017) indicated that its presence in Lake Maksimir is the consequence of human mediated translocation. Phylogenetic reconstruction indicated that the individuals of *P. leptodactylus* belong to the Asian lineage *sensu* Maguire et al. (2014), specifically they are closely related to haplotypes originating from Armenia. Although, we observed two haplotypes within the Lake Maksimir population, they differed only in one base (site 378: Hap1-C, Hap2-A), collapsed to a single haplotype in the network reconstruction. This supports a theory by Njegovan et al. (2017) that the crayfish were introduced into lakes from the local market, supplied from the Armenian breeders. Further sampling and population studies,

TABLE 3 | Major types of repetitive DNA in *P. leptodactylus* (classification according to Wicker et al., 2007).

Categories/Superfamilies	Abundance (%)	Clusters (n)
Satellites	27.57	258
Minisatellites	24.70	240
Satellites	2.87	18
Class I (retrotransposons)		
LTR elements	15.32	71
<i>Ty3/gypsy</i>	14.95	55
<i>Ty1/copia</i>	0.10	4
<i>BEL</i>	0.05	2
<i>ERV</i>	0.03	1
<i>Integrated virus DNA</i>	0.19	9
DIRS	3.57	4
Penelope	1.00	2
LINE	2.23	33
<i>R1</i>	0.04	2
<i>R2</i>	0.03	2
<i>RTE</i>	0.01	1
<i>RTEX</i>	0.03	1
<i>Jockey</i>	0.10	7
<i>I</i>	0.21	1
<i>L1</i>	0.02	1
<i>Ingi</i>	0.05	1
<i>CRE</i>	0.01	1
<i>CR1</i>	1.32	6
<i>CR2</i>	0.03	2
<i>Nimb</i>	0.03	2
<i>Kiri</i>	0.02	1
<i>Daphne</i>	0.01	1
<i>Unclassified LINE</i>	0.32	4
SINE	0.07	1
Class II (DNA transposons)—Subclass 1		
TIR	0.27	14
<i>Mariner</i>	0.05	4
<i>hAT</i>	0.05	2
<i>CACTA</i>	0.03	1
<i>Harbinger</i>	0.03	1
<i>piggyBAC</i>	0.03	1
<i>Dada</i>	0.02	1
<i>Ginger2</i>	0.02	1
<i>Sola</i>	0.02	1
<i>Ginger3</i>	0.01	1
<i>Transisb</i>	0.01	1
Class II (DNA transposons)-Subclass 2		
Helitron	0.15	6
Polinton	0.09	3
rDNA	0.01	1
Unclassified repetitive	4.48	97
Total repetitive DNA	54.85	490

coupled with a multigene approach may help in resolving the taxonomic status of the three lineages within the *P. leptodactylus* species complex.



P. leptodactylus Karyotype and Genomic Organization of 45S rDNA and Telomeric (TTAGG)_n Repeats

In this study, FISH results showed one 45S rDNA locus and ten probable interstitial HSs in the studied *P. leptodactylus*, which is different from the previous work on *P. leptodactylus* that reported two 45S rDNA loci and six interstitial HSs (Mlinarec et al., 2011). The observed discrepancy suggests the presence of intraspecific variability within *P. leptodactylus*, and we could speculate that differences in rDNA loci number as well as in the number of interstitial HSs could possibly be lineage specific. In particular, samples analyzed in Mlinarec et al. (2011) belonged to European lineage *sensu* Maguire et al. (2014), while samples used in the present study belong to the Asian lineage *sensu* Maguire et al. (2014). Intraspecific variability has been reported in other groups of organisms such as two fish species from genus *Schistura* (Sember et al., 2015), as well as in plants *Phaseolus vulgaris* and *Tanacetum cinerariifolium* (Pedrosa-Harand et al., 2006; Mlinarec et al., 2019). Different mechanisms can lead to intrachromosomal variability such as unequal crossing-over, non-homologous recombination and movement mediated by transposons (Liu et al., 2003; Nguyen et al., 2010; Pereira et al., 2013; Vershinina et al., 2015; Mlinarec et al., 2019).

Large AT-rich probable HSs positioned in the (peri)centromeric position on all chromosomes and interstitially on some chromosomes suggest a high amount of repetitive DNA in the genome of *P. leptodactylus* (this study; Mlinarec et al., 2011). Large (peri)centromeric HSs have been found in different crustacean families such as Astacidae (Mlinarec

et al., 2011, 2016), Nephropidae (Deiana et al., 1996; Coluccia et al., 2001; Salvadori et al., 2002), Scyllaridae (Deiana et al., 2007), Palinuridae (Coluccia et al., 1999, 2005; Cannas et al., 2004), Cambaridae (Salvadori et al., 2014), and Palaemonidae (González-Tizón et al., 2013; Torrecilla et al., 2017; Molina et al., 2020).

In this study it was observed that telomeres of *P. leptodactylus* consist of (TTAGG)_n pentameric repeats, same as in all decapod crustaceans studied until now and in most arthropods (Vítková et al., 2005; Salvadori et al., 2012, 2014). However, this study showed that a significant part of telomeric repeats is located interstitially in the chromosomes of *P. leptodactylus*. ITRs were also observed in other crustaceans such as *Jasus lalandii* and *Procambarus clarkii* (Salvadori et al., 2012, 2014). In *J. lalandii*, ITRs are associated with rDNA (Salvadori et al., 2012), while in *P. leptodactylus* and *P. clarkii* co-localization of ITRs with rDNA loci has not been observed (Salvadori et al., 2014). The occurrence of ITRs outside of the chromosomal termini is not fully understood. ITRs in (peri)centromeric regions could represent remnants of structural chromosome fusions (Ruiz-Herrera et al., 2008; Bolzán, 2012). This is unlikely in *P. leptodactylus* as there were no ITRs in (peri)centromeric positions. ITRs might have originated from the transposition of telomeric repeats by transposable elements or during repair of double stranded breaks (Aksenova and Mirkin, 2019) or might simply reflect the fact that telomeric sequences are present within repetitive DNA components like in some plants (Tek and Jiang, 2004; Mlinarec et al., 2009; Emadzade et al., 2014). The last case is unlikely as in *P. leptodactylus*, satellite repeats do not contain stretches of telomeric repeats.

Pontastacus leptodactylus Repeatome

This work represents the most comprehensive characterization of the repetitive elements in any species belonging to the family Astacidae. In this study, we showed that *P. leptodactylus* harbors a large variety of repetitive elements, accounting for about 54.85% of its genome. As repeats may escape their detection by degradation, we consider this value as an underrepresentation. Degraded repeats arise from point mutations, indels and rearrangements, and they may be so substantial that they contribute repeats into tracks of unique or low-copy sequences. This is supported by recent studies on 101 species showing that in the large genomes, such as the genome of *P. leptodactylus*, the proportion of single and low-copy (up to 20 copies) sequences significantly increases with genome size, which is accompanied by a significant decrease in the genome proportion of medium-copy repeats (Novák et al., 2020a).

The analyses of draft genomes of *C. quadricarinatus* and *P. virginalis* showed that they have a significantly lower amount of repetitive DNA, 33.73 and 27.52%, respectively (Gutekunst et al., 2018; Tan et al., 2020), in comparison with *P. leptodactylus*. Furthermore, in *P. leptodactylus* satellite repeats and Ty3/gypsy elements are the most abundant, while in *C. quadricarinatus* and *P. virginalis*, LINE elements are the most abundant repetitive elements in the genome (Tan et al., 2020). However, comparison of the results of this study with those of Tan et al. (2020)

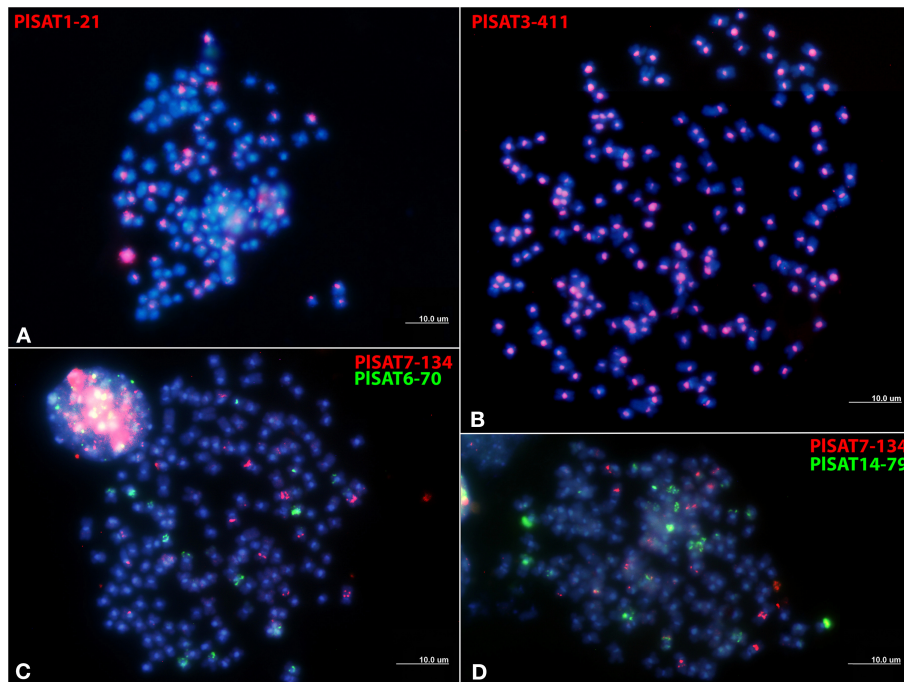
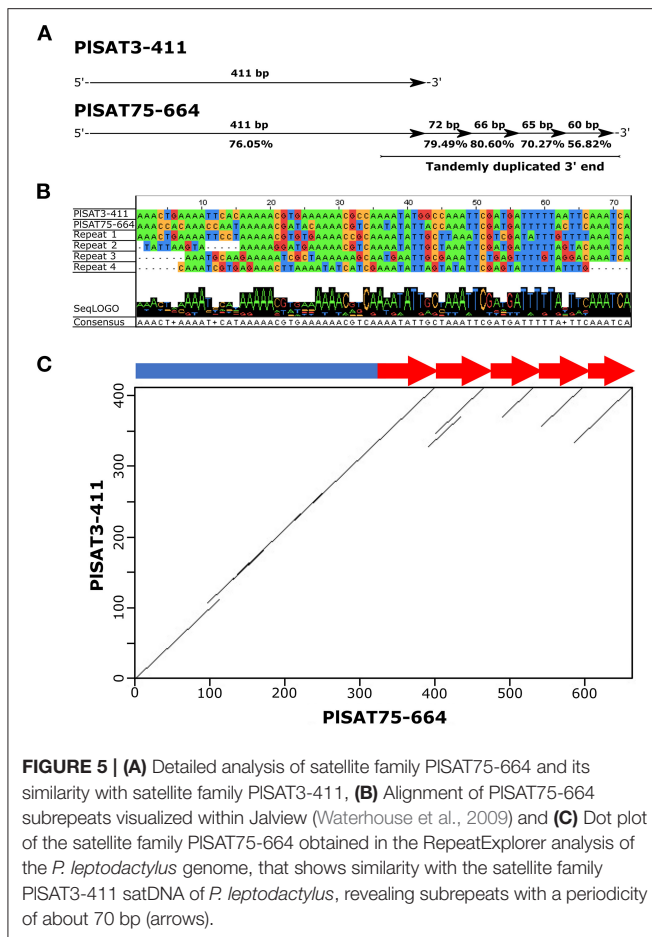


FIGURE 4 | Distribution of satellite repeat families on metaphase chromosomes of *Pontastacus leptodactylus*. **(A)** PISAT1-21 (in red), **(B)** PISAT3-411 (in red), **(C)** PISAT7-134 (in red) and PISAT6-70 (in green), and **(D)** PISAT7-134 (in red) and PISAT6-70 (in green) as probes. Chromosomes are counterstained with DAPI. Scale bar = 10 μ m.

should be taken with caution since different methods have been applied for repeat identification. Estimation of the repeat abundance from the *de novo* genome assemblies generated by short-read sequencing as in Tan et al. (2020), can lead to the underrepresentation of the highly repetitive elements. These elements are often clustered into a single contig or fragmented across multiple short contigs due to the inherited characteristics of the *de novo* genome assembly tools, therefore misrepresenting the abundance of the repetitive elements in the genome (Chu et al., 2016). The flow cytometry method estimated 1C = 18.7 Gbp size for the *P. leptodactylus* genome, providing the first report on genome size for any species within the family Astacidae. However, there is still a general lack of genome sizes for the infraorder Astacidea. As far as we are aware, genome size is available for several members of the family Cambaridae (5 species) and Parastacidae (1 species) ranging from 3.82 to 6.06 Gbp (Gregory, 2020; Tan et al., 2020). This makes *P. leptodactylus* (1C = 18.7 Gbp) species with the highest genome size of all known members of the infraorder Astacidea. In *P. leptodactylus*, genome expansion can be a result of the accumulation of short tandem repeats and retroelements as it is shown in this study that the genome of this species is rich in satellite DNA and retroelements. A large genome size as well as a highly repetitive genome explains difficulties generated during the genome assembly process, which limit the generation of available genomic resources from crustacean species (Tan et al., 2020; Van Quyen et al., 2020).

In *P. leptodactylus*, satellite repeats are the most abundant group of repetitive elements, accounting for 27.52% of its genome. Although, the knowledge about repetitive DNA composition in the genomes of decapod crustaceans is scarce, it is likely that a great expansion of satellites occurred in the genome of *P. leptodactylus*. The large amount of satellite repeats has been reported in other organisms such as insects *Drosophila virilis* and *Triatoma infestans* (Wei et al., 2014; Pita et al., 2018). In *D. virilis* nearly 50% of the genome is composed of satDNA, while in *T. infestans* satellite repeats make up 25-33% of the genome and are arranged into at least 42 satellite DNA families (Wei et al., 2014; Pita et al., 2018). Furthermore, we found great diversity of satDNA repeats in the genome of *P. leptodactylus* with a total of 258 satellite families which is by far the most satellite-rich species discovered to date. A large number of different satDNA elements is found in other organisms such as the fish *Megaleporinus macrocephalus* (Teleostei, Anostomidae) where 164 satellite repeats have been described (Utsunomia et al., 2019). Similar to *P. leptodactylus*, in *M. macrocephalus*, short satellites dominate in the genome. Among plants, *Luzula elegans* (Poaceae) has the highest number of satellites, 37, constituting 9.9% of the genome (Heckmann et al., 2013). The species *Vicia faba* (Fabaceae) is another example of the plant species with a high number of satellites, over 30, that together constitutes 935 Mbp (7%) of its genome (Ávila Robledillo et al., 2018). Large satDNA abundance and diversity is not a common characteristic for all animal and plant genomes, as there are, as far as we know, many more reports on the organisms poor in satellite DNA using



similar approaches. In *Tanacetum cinerariifolium* (Asteraceae), only three among the 58,204 clusters obtained were classified as satellites, representing 1.04% of the genome (Mlinarec et al., 2019). Similarly, after the investigation of *Passiflora edulis* by RepeatExplorer, only two of the 233 repetitive elements were satellites, representing less than 0.1% of the genome (Pamponé et al., 2019).

It is tempting to speculate where the diversity of *P. leptodactylus* satellites originate from. Novel satellite DNA families may arise from the independent duplication of different genomic sequences, such as intergenic spacers, or even from those derived from other satellite DNAs (Garrido-Ramos, 2017). The satDNA sequences can interact with transposable elements to create new repetitive DNA (Pita et al., 2018). It is suggested that transposable elements provide the mechanism by which satDNA repeats could propagate in the genome through dispersed short repeat arrays (Macas et al., 2011; Bardella et al., 2014). The *P. leptodactylus* genome is rich in the LTR retrotransposons.

Minisatellites (monomer size 10–100 bp) were found to be surprisingly numerous in the *P. leptodactylus* genome, accounting for about 24.7% of the genome. High content of minisatellites in the *P. leptodactylus* genome might indicate a high level of DNA polymerase slippage as it is generally

considered that short tandem repeats (<100 bp) expand through DNA polymerase slippage (Garrido-Ramos, 2017). The most abundant satDNA in the genome of *P. leptodactylus* is a minisatellite PISAT1-21. Its short monomer size of 21 bp is unusual for a tandem repeat of high abundances, which generally consist of 160–180 or 320–360 bp monomers (Garrido-Ramos, 2017). This underpins that satellites with short monomer lengths can form very large arrays as observed here for PISAT1-21. In the hermit crab *Pagurus pollicaris*, a minisatellite AGTGCAG(CTG)_n constitutes a large fraction of its genome (Chambers et al., 1978). An exceptional abundance of microsatellite and SSR sequences has also been found in the genome of freshwater prawns of the genus *Macrobrachium* (Palaemonidae) as well as in the penaeid shrimp *Litopenaeus vannamei* (Zhang et al., 2019; Molina et al., 2020), suggesting that short tandem repeats are a significant component of decapod crustaceans genomes.

In *P. leptodactylus*, 171 (66.27%) satellite DNA families showed A+T content higher than 50%, and could be classified as AT-rich (Figure 3). Furthermore, there is no correlation between A+T content and satellite length. The high A+T content could be a consequence of satDNA being subject to epigenetic modifications such as the methylation of cytosines, consequently deamination of 5-methylcytosines forming more AT base pairs in *P. leptodactylus* satDNAs. In the fish *Megaleporinus microcephalus* short (<100 bp) and long (>100 bp) satellites had a similar amount of A+T content (Utsunomia et al., 2019). In the fern *V. speciosa* satDNAs longer unit length showed a higher A+T content (Ruiz-Ruano et al., 2019). In *V. faba*, most of the satellite sequences had an elevated A+T content (65–80%) (Ávila Robledillo et al., 2018).

In *P. leptodactylus*, the satellites are abundant in the (peri)centromeric region, on both ends of the chromosomes and some of them are distributed on the interstitial regions of the chromosomes. This is in line with previous results which show that subtelomere and centromere regions contain large parts of satellite repeats (Melters et al., 2013; Garrido-Ramos, 2017). Conventional satellites (monomer size >100 bp) and minisatellites (monomer size 10–100 bp) are conventionally differentiated by their location (Garrido-Ramos, 2017). While classic satDNAs are usually located as long arrays at the heterochromatin segments, minisatellites are generally proper of euchromatic regions (Garrido-Ramos, 2017). In *P. leptodactylus*, the classic satellite family PISAT3-411 constitutes (peri)centromeric HSs, while minisatellites PISAT6-70 and PISAT14-79 as a part of euchromatic regions are located along the chromosome arms.

(Peri)centromeric Satellite Family PLSAT3-411

Centromeres are often packaged into heterochromatin, containing large amounts of repetitive DNA (Wang et al., 2009; Mehta et al., 2010). Here we showed that the probable (peri)centromeric heterochromatic segments located on all *P. leptodactylus* chromosomes are formed by a specific highly amplified satellite family PISAT3-411. The arrangement of the (peri)centromeric satDNA family PISAT3-411 can be

explained by the principle of equilocality, according to which, heterochromatin accumulates at equivalent positions in each chromosome within a genome (Garrido-Ramos, 2017). The most consistent form of equilocality relates to the heterochromatin in the vicinity of centromeres (John et al., 1985), which is true for PISAT3-411 being present in the (peri)centromeric regions of all chromosomes. Following the survey of tandem satellite repeats in 282 species from various kingdoms (Melters et al., 2013), PISAT3-411 is an ideal candidate for centromeric repeat sequences. It is one of the most abundant satellite repeats accounting for 1.29% of the genome and it is A+T-rich. It has been found that centromeric satDNAs are generally A+T rich (Garrido-Ramos, 2015; Yuan et al., 2018). Most animal species investigated so far have a single or only a few centromeric satellites with monomers hundreds of nucleotides long that are shared by all chromosomes, an observation that is explained by their coevolution with kinetochore proteins (Garrido-Ramos, 2015). The (peri)centromeric satellite family PISAT3-411 is common in that respect. (Peri)centromere composition of *P. leptodactylus* calls for the investigation of additional species from different genera to get a more representative insight into the evolution of the (peri)centromeric satellite family. The frequent accumulation of satDNA in centromeric regions is explained by its role in centromere functions, such as kinetochore assembly and chromosome segregation during mitosis or meiosis, or even some epigenetic regulations, or simply by passive accumulation due to the absence of recombination-based elimination mechanisms (McFarlane and Humphrey, 2010; Plohl et al., 2014; Catania et al., 2015). To fully confirm that PISAT3-411 is a true centromeric satellite family, underlying the functional kinetochore CENH3-ChIP followed by sequencing is needed.

Similarity Between satDNA Families

Most of the satDNA families described in this study did not show any conserved features or sequence similarities between each other suggesting their independent origin. Only 39 of the 258 satDNA repeats described in *P. leptodactylus*, showed similarities, however, their similarity is not high, ranging from 55 to 78%, average similarity is 63%. Two satellites, PISAT3-411 and PISAT75-664, were among the most interesting. The longer unit PISAT75-664 is organized into HOR (higher order repeat) structures that consist of PISAT3-411 basic monomer and four times directly repeated ~70 bp long sequence that shows high similarity to PISAT3-411 (Figure 5). The similarity between PISAT3-411 and PISAT75-664 indicates the existence of a satDNA superfamily (SF), derived from a common ancestor satDNA. In the most parsimonious scenario, HOR structure might have formed after a ~70 bp fragment was four times amplified within the satDNA, resulting in a new repeat unit of 664. It is known that the simultaneous amplification and homogenization of two or more adjacent monomers leads to the formation of HORs (Garrido-Ramos, 2017). Furthermore, it is generally considered that shorter repeats originate by replication slippage, while longer units originate by unequal crossing over (Garrido-Ramos, 2017). Therefore in *P. leptodactylus*, replication slippage might be the mechanism for the origin of the four times tandemly repeated ~70 bp subunits within PISAT75-664. The combination of short

repeat units into longer units constituting HORs is a common trend in satDNA evolution (Plohl et al., 2008; Garrido-Ramos, 2017). Regular HORs, usually dimeric, have been found in several species of beetles (Palomeque and Lorite, 2008; Vlahović et al., 2017). Complex HORs, shaped from interspersed and/or inversely oriented monomers and frequently with extraneous sequence elements, have been found in non-human mammals, such as mouse, pig, bovid, horse, dog, elephant, insect, and fish (Palomeque and Lorite, 2008; Vlahović et al., 2017; Utsunomia et al., 2019).

The present study is the first one focusing on the repeatome of *P. leptodactylus* and enables a new perspective into the evolution of this complex species. *P. leptodactylus* repeatome serves as an important and valuable resource to support ongoing comparative genomic, cytogenomic, fundamental, and applied biology studies. To gain a more comprehensive understanding of chromosome evolution and genomic compositions of freshwater crustaceans, chromosome and genome resources are much needed for more species across taxonomic groups.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ebi.ac.uk/ena/browser/view/SRR7698976>, SRR7698976; <https://www.ncbi.nlm.nih.gov/nucleotide/KX279350.1>, KX279350.

AUTHOR CONTRIBUTIONS

The study was conceived by IM, VB, JM, LLB, and LB, while LLB, LB, and JM designed the experimental part of the study. Field work was carried out by LLB, LA, LB, and LL, while lab work (DNA isolation, cytogenetic experiments) by LLB, LA, LB, LL, and IM. CG and AH conducted flow cytometry analysis. LLB carried out the bioinformatic analyses and JM designed the primers. JM, LLB, LB, IM, VB, FG, and CA discussed and interpreted the results. JM wrote the paper with the help of LLB and LB. The initial version of the manuscript was drafted by IM, VB, LLB, and LB. All authors read, edited, enhanced original version of the manuscript, and approved its final version.

FUNDING

This work was supported by the institutional project financed by the University of Zagreb, by the University of Zagreb Student union grant for the project development of new karyotypization techniques of freshwater crayfish (family: Astacidae), and LL by ESF-DOK-2018-01-9589.

ACKNOWLEDGMENTS

We would like to thank Lana Jelić and JU Park Maksimir for providing us with their support in conducting this research and raising awareness of the presence of native freshwater crayfish species in the City of Zagreb.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.611745/full#supplementary-material>

Supplementary Table 1 | *COI* gene sequences from *P. leptodactylus* used to construct a phylogenetic tree.

Supplementary Table 2 | *Pontastacus leptodactylus* satellitome. Repeat unit lengths, G+C content (%), abundances (%), and consensus sequence.

Supplementary Table 3 | Morphometric measurements of *P. leptodactylus* chromosomes ($n = x = 90$).

Supplementary Figure 1 | Consensus sequences of PISAT1-21, PISAT3-411, PISAT6-70, PISAT7-134, and PISAT14-79 repeats in fasta format. Primer sequences for amplification of tandem repeat specific probes are underlined and bolded.

Supplementary Figure 2 | Median-joining network of the *COI* barcode region haplotypes. Haplotype from the lake Maksimir is marked with an arrow. Haplotype size reflects relative frequency. Each branch represents one mutational step, unless otherwise noted (numbers in red above branches). Black circles represent missing intermediate haplotypes. Different colored circles denote the share of distinct haplotypes within countries (legend is shown in the upper-left corner).

Supplementary Figure 3 | An idiogram of *P. leptodactylus* chromosomes with marked localization of 45S rDNA and DAPI positive AT-rich heterochromatin bands. This idiogram was generated based on the FISH information from **Figures 1A,B**.

Supplementary Figure 4 | The graph layout corresponding to read clusters of **(A)** PISAT1-21, **(B)** PISAT3-411, **(C)** PISAT6-70, **(D)** PISAT7-134, and **(E)** PISAT14-79. The percentage indicates the genome proportion of each cluster.

Supplementary Figure 5 | Sequence logos showing the level of sequence divergence.

Supplementary Figure 6 | Dot plot analysis of **(A)** PISAT1-21, **(B)** PISAT3-411, **(C)** PISAT6-70, **(D)** PISAT7-134, and **(E)** PISAT14-79 obtained in TAREAN analysis of *P. leptodactylus*.

REFERENCES

- Akhan, S., Bektas, Y., Berber, S., and Kalayci, G. (2014). Population structure and genetic analysis of narrow-clawed crayfish (*Astacus leptodactylus*) populations in Turkey. *Genetica* 142, 381–395. doi: 10.1007/s10709-014-9782-5
- Aksenova, A. Y., and Mirkin, S. M. (2019). At the beginning of the end and in the middle of the beginning: structure and maintenance of telomeric DNA repeats and interstitial telomeric sequences. *Genes* 10:118. doi: 10.3390/genes10020118
- Ávila Robledillo, L., Koblížková, A., Novák, P., Böttinger, K., Vrbová, I., Neumann, P., et al. (2018). Satellite DNA in *Vicia faba* is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. *Sci. Rep.* 8:5838. doi: 10.1038/s41598-018-24196-3
- Bandelt, H. J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036
- Bardella, V. B., Aristeu, J. A., and Vanzela, A. L. L. (2014). Origin and distribution of AT-rich repetitive DNA families in *Triatoma infestans* (Heteroptera). *Infect. Genet. Evol.* 23, 106–114. doi: 10.1016/j.meegid.2014.01.035
- Blackburn, E. H. (2005). Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS Lett.* 579, 859–862. doi: 10.1016/j.febslet.2004.11.036
- Bolzán, A. D. (2012). Chromosomal aberrations involving telomeres and interstitial telomeric sequences. *Mutagenesis* 27, 1–15. doi: 10.1093/mutage/ger052
- Bracken-Grissom, H. D., Ah Yong, S., Wilkinson, R. D., Feldmann, R. M., Schweizer, C. E., Breinholt, J. W., et al. (2014). The emergence of lobsters: phylogenetic relationships, morphological evolution and divergence time comparisons of an ancient group (Decapoda: Achelata, Astacidea, Glypheidea, Polychelida). *Syst. Biol.* 63, 457–479. doi: 10.1093/sysbio/syu008
- Cannas, R., Deiana, A. M., Salvadori, S., and Coluccia, E. (2004). Preliminary data on the karyology of *Panulirus regius* De Brito Capello, 1864 (Crustacea Decapoda). *Biol. Mar. Mediterr.* 11, 721–723.
- Catania, S., Pidoux, A. L., and Allshire, R. C. (2015). Sequence features and transcriptional stalling within centromere DNA promote establishment of CENP-A chromatin. *PLoS Genet.* 11:e1004986. doi: 10.1371/journal.pgen.1004986
- Chambers, C. A., Schell, M. S., and Skinner, D. M. (1978). The primary sequence of a crustacean satellite containing a family of repeats. *Cell* 13, 97–110. doi: 10.1016/0092-8674(78)90141-1
- Chu, C., Nielsen, R., and Wu, Y. (2016). REPdenovo: inferring *de novo* repeat motifs from short sequence reads. *PLoS One* 11:e0150719. doi: 10.1371/journal.pone.0150719
- Coluccia, E., Cannas, R., Deiana, A. M., Milia, A., Salvadori, S., and Libertini, A. (1999). Genome size and A-T base content in five Palinuridae species (Crustacea Decapoda). *Biol. Mar. Mediterr.* 6, 688–691.
- Coluccia, E., Cau, R. A., Cannas, R., Milia, A., Salvadori, S., and Deiana, A. M. (2001). Mitotic and meiotic chromosomes of the American lobster *Homarus americanus* (Nephropidae, Decapoda). *Hydrobiologia* 449, 149–152. doi: 10.1023/A:1017557523022
- Coluccia, E., Salvadori, S., Cannas, R., Milia, A., and Deiana, A. M. (2005). First data on the karyology of *Palinurus gilchristi* (Crustacea Decapoda). *Biol. Mar. Mediterr.* 12, 661–663.
- Crandall, K. A., and Buhay, J. E. (2008). Global diversity of crayfish (Astacidae, Cambaridae, and Parastacidae-Decapoda) in freshwater. *Hydrobiologia* 595, 295–301. doi: 10.1007/s10750-007-9120-3
- Crandall, K. A., and De Grave, S. (2017). An updated classification of the freshwater crayfishes (Decapoda: Astacidea) of the world, with a complete species list. *J. Crust. Biol.* 37, 615–653. doi: 10.1093/jcbl/rux070
- Deiana, A. M., Cau, A., Cannas, R., Coluccia, E., and Salvadori, S. (2007). “Genetics of Slipper Lobsters,” in *The Biology and Fisheries of the Slipper Lobster*, eds K. L. Lavalli and E. Spanier (Boca Raton, FL: CRC Press), 53–67.
- Deiana, A. M., Coluccia, E., Milia, A., and Salvadori, S. (1996). Supernumerary chromosomes in *Nephrops norvegicus* L. (Crustacea, Decapoda). *Heredity* 76, 92–99. doi: 10.1038/hdy.1996.12
- Diupotex Chong, M. E., Foster, N. R., and Zarate, L. A. (1997). A cytogenetic study of the crayfish *Procambarus digueti* (Bouvier, 1897) (Decapoda Cambaridae) from Lake Camecuaro, Michoacan, Mexico. *Crustaceana* 70, 875–885. doi: 10.1163/156854097X00492
- Emadzade, K., Jang, T. S., Macas, J., Kovarik, A., Novák, P., Parker, J., et al. (2014). Differential amplification of satellite PaB6 in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). *Ann. Bot.* 114, 1597–1608. doi: 10.1093/aob/mcu178
- Eschscholtz, F. F. (1823). Descriptio novae *Astacorum speciei* Rossicae. *Mem Soc Imper Nat Moscu* 6, 109–110, 1 plate.
- Fasten, N. (1914). Spermatogenesis of the American crayfish, *Cambarus virilis* and *C. immunitis*, with special reference to synopsis and the chromatoid bodies. *J. Morphol.* 25, 587–649. doi: 10.1002/jmor.1050250403
- Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Garrido-Ramos, M. A. (2015). Satellite DNA in plants: more than just rubbish. *Cytogenet. Genome Res.* 146, 153–170. doi: 10.1159/000437008
- Garrido-Ramos, M. A. (2017). Satellite DNA: an evolving topic. *Genes* 8:230. doi: 10.3390/genes8090230

- González-Tizón, A. M., Rojo, V., Menini, E., Torrecilla, Z., and Martínez-Lage, A. (2013). Karyological analysis of the shrimp *Palaemon serratus* (Decapoda: Palaemonidae). *J. Crust. Biol.* 33, 843–848. doi: 10.1163/1937240X-00002185
- Gregory, T. R. (2020). *Animal Genome Size Database*. Available online at: <http://www.genomesize.com>.
- Gross, R., Kóiv, K., Pukk, L., and Kaldre, K. (2017). Development and characterization of novel tetranucleotide microsatellite markers in the noble crayfish (*Astacus astacus*) suitable for highly multiplexing and for detecting hybrids between the noble crayfish and narrow-clawed crayfish (*A. leptodactylus*). *Aquaculture* 472, 50–56. doi: 10.1016/j.aquaculture.2016.04.015
- Gutekunst, J., Andriantsoa, R., Falckenhayn, C., Hanna, K., Stein, W., Rasamy, J., et al. (2018). Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nat. Ecol. Evol.* 2, 567–573. doi: 10.1038/s41559-018-0467-9
- Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., and Chen, J. (2020). RIdiogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.* 6:e251. doi: 10.7717/peerj-cs.251
- Hare, E. E., and Johnston, J. S. (2011). “Genome size determination using flow cytometry of propidium iodide-stained nuclei,” in *Molecular Methods for Evolutionary Genetics, Methods in Molecular Biology*, eds V. Orgogozo and M. V. Rockman (Totowa, NJ: Humana Press), 3–12.
- Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., et al. (2013). The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. *Plant J.* 73, 555–565. doi: 10.1111/tpj.12054
- Holdich, D. M., Reynolds, J. D., Souty-Grosset, C., and Sibley, P. J. (2009). A review of the ever increasing threat to European crayfish from non-indigenous crayfish species. *Knowl. Manag. Aquat. Ecosyst.* 11, 394–395. doi: 10.1051/kmae/2009025
- Ijdo, J. W., Wells, R. A., Baldini, A., and Reeders, S. T. (1991). Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. *Nucleic Acids Res.* 19:4780. doi: 10.1093/nar/19.17.4780
- Indy, J. R., Arias-Rodríguez, L., Páramo-Delgado, S., Hernández-Vidal, U., Álvarez-González, C. A., and Contreras-Sánchez, W. M. (2010). Mitotic karyotype of the tropical freshwater crayfish *Procambarus* (*Austrocambarus*) *llamasi* (Decapoda, Cambaridae). *Rev. Biol. Trop.* 58, 655–662. doi: 10.15517/rbt.v58i2.5236
- Inkscape Project. (2020). *Inkscape*. Retrieved from: <https://inkscape.org>.
- John, B., King, M., Schweizer, D., and Mendelak, M. (1985). Equilocality of heterochromatin distribution and heterochromatin heterogeneity in acridid grasshoppers. *Chromosoma* 91, 185–200. doi: 10.1007/BF00328216
- Jurka, J., Klonowski, P., Dagman, V., and Pelton, P. (1996). CENSOR - a program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem.* 20, 119–121. doi: 10.1016/S0097-8485(96)80013-1
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Khoshkholgh, M., and Nazari, S. (2019). The genetic diversity and differentiation of narrow-clawed crayfish *Pontastacus leptodactylus* (Eschscholtz, 1823) (Decapoda: Astacidea: Astacidae) in the Caspian Sea Basin, Iran as determined with mitochondrial and microsatellite DNA markers. *J. Crust. Biol.* 2, 112–120. doi: 10.1093/jcblol/ruy113
- Komagata, N., and Komagata, K. (1992). Chromosome numbers in three species of Japanese fresh-water crayfishes. *Chromosome Inf. Serv.* 53, 15–16.
- Kostyuk, V. S., Garbar, A. V., and Mezhzherin, S. V. (2013). Karyotypes and morphological variability of crayfish *Pontastacus leptodactylus* and *P. angulosus* (Malacostraca, Decapoda). *Vestn. Zool.* 47, 11–16. doi: 10.2478/vzoo-2013-0020
- Kouba, A., Petrussek, A., and Kozák, P. (2014). Continental-wide distribution of crayfish species in Europe: update and maps. *Knowl. Manag. Aquat. Ecosyst.* 413, 5. doi: 10.1051/kmae/2014007
- Kuo, H. F., Olsen, K. M., and Richards, E. J. (2006). Natural variation in a subtelomeric region of Arabidopsis: implications for the genomic dynamics of a chromosome end. *Genetics* 173, 401–417. doi: 10.1534/genetics.105.055202
- Liu, Z. L., Zhang, D., Hong, D. Y., and Wang, X. R. (2003). Chromosomal localization of 5S and 18S-5.8S-25S ribosomal DNA sites in five Asian pines using fluorescence in situ hybridization. *Theor. Appl. Genet.* 106, 198–204. doi: 10.1007/s00122-002-1024-z
- Louis, E. J., and Vershinin, A. V. (2005). Chromosome ends: different sequences may provide conserved functions. *Bioessays* 27, 685–697. doi: 10.1002/bies.20259
- Lukhtanov, V. A., and Iashenkova, Y. (2019). Linking karyotypes with DNA barcodes: proposal for a new standard in chromosomal analysis with an example based on the study of Neotropical Nymphalidae (Lepidoptera). *Comp. Cytogenet.* 13, 435–449. doi: 10.3897/CompCytogen.v13i4.48368
- Macas, J., Kejnovský, E., Neumann, P., Novák, P., Koblížková, A., and Vyskot, B. (2011). Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. *PLoS One* 6:e27335. doi: 10.1371/annotation/4ccaacb2-92d7-445a-87da-313cedf18feb
- Maguire, I., and Dakić, L. (2011). Comparative analyses of *Astacus leptodactylus* morphological characteristics from Croatia and Armenia. *Biol. Sect. Zool.* 66, 491–498. doi: 10.2478/s11756-011-0045-y
- Maguire, I., Podnar, M., Jelić, M., Štambuk, A., Schrimpf, A., Schulz, H., and Klobučar, G. (2014). Two distinct evolutionary lineages of the *Astacus leptodactylus* species-complex (Decapoda: Astacidae) inferred by phylogenetic analyses. *Invertebr. Syst.* 28, 117–123. doi: 10.1071/IS13030
- Martin, P., Thonagel, S., and Scholtz, G. (2015). The parthenogenetic Marmorkrebs (Malacostraca: Decapoda: Cambaridae) is a triploid organism. *J. Zool. Syst. Evol. Res.* 54, 13–21. doi: 10.1111/jzs.12114
- McFarlane, R. J., and Humphrey, T. C. (2010). A role for recombination in centromere function. *Trends Genet.* 26, 209–213. doi: 10.1016/j.tig.2010.02.005
- Mehta, G. D., Agarwal, M. P., and Ghosh, S. K. (2010). Centromere identity: a challenge to be faced. *Mol. Genet. Genomics* 284, 75–94. doi: 10.1007/s00438-010-0553-4
- Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J. G., et al. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* 14, 1–20. doi: 10.1186/gb-2013-14-1-r10
- Mlinarec, J., Chester, M., Siljak-Yakovlev, S., Papes, D., Leitch, A. R., and Besendorfer, V. (2009). Molecular structure and chromosome distribution of three repetitive DNA families in *Anemone hortensis* L. (Ranunculaceae). *Chromosome Res.* 17, 331–346. doi: 10.1007/s10577-009-9025-2
- Mlinarec, J., Mužić, M., Pavlica, M., Šrut, M., Klobučar, G., and Maguire, I. (2011). Comparative karyotype investigations in the European crayfish *Astacus astacus* and *A. leptodactylus* (Decapoda, Astacidae). *Crustaceana* 84, 1497–1510. doi: 10.1163/156854011X607015
- Mlinarec, J., Porupski, I., Maguire, I., and Klobučar, G. (2016). Comparative karyotype investigations in the white-clawed crayfish *Austropotamobius pallipes* (Lereboullet, 1858) species complex and stone crayfish *A. torrentium* (Schrank, 1803) (Decapoda: Astacidae). *J. Crustacean Biol.* 36, 87–93. doi: 10.1163/1937240X-00002390
- Mlinarec, J., Skuhala, A., Jurković, A., Malenica, N., McCann, J., Weiss-Schneeweiss, H., et al. (2019). The repetitive DNA composition in the natural pesticide producer *Tanacetum cinerariifolium*: interindividual variation of subtelomeric tandem repeats. *Front. Plant Sci.* 10:613. doi: 10.3389/fpls.2019.00613
- Molina, W. F., Costa, G., Cunha, I., Bertollo, L., Ezaz, T., Liehr, T., et al. (2020). Molecular cytogenetic analysis in freshwater prawns of the genus *Macrobrachium* (Crustacea: Decapoda: Palaemonidae). *Int. J. Mol. Sci.* 21:2599. doi: 10.3390/ijms21072599
- Monat, C., Tando, N., Tranchant-Dubreuil, C., and Sabot, F. (2016). LTRclassifier: a website for fast structural LTR retrotransposons classification in plants. *Mob. Genet. Elements* 6:6. doi: 10.1080/2159256X.2016.1241050
- Nguyen, P., Sahara, K., Yoshido, A., and Marec, F. (2010). Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica* 138, 343–354. doi: 10.1007/s10709-009-9424-5
- Niiyama, H. (1962). On the unprecedentedly large number of chromosomes of the crayfish *Astacus trowbridgii* Stimpson. *Annot. Zool. Japon.* 35, 229–233.
- Njegovan, V., Orešković, M., Hudina, S., and Maguire, I. (2017). Crayfish fauna of the Maksimir lakes in the urban area of Zagreb. *Freshwater Crayfish* 23, 1–11. doi: 10.5869/fc.2017.v23-1.1
- Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., and Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* 45:e111. doi: 10.1093/nar/gkx257

- Novák, P., Guignard, M. S., Neumann, P., Kelly, L. J., Mlinarec, J., Koblížková, A., et al. (2020a). Repeat sequence turnover shifts fundamentally in species with large genomes. *Nature Plants* 6, 1325–1329. doi: 10.1038/s41477-020-00785-x
- Novák, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform.* 11:378. doi: 10.1186/1471-2105-11-378
- Novák, P., Neumann, P., and Macas, J. (2020b). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* 15, 3745–3776. doi: 10.1038/s41596-020-0400-y
- Novák, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793. doi: 10.1093/bioinformatics/btt054
- Palomeque, T., and Lorite, P. (2008). Satellite DNA in insects: a review. *Heredity* 100, 564–573. doi: 10.1038/hdy.2008.24
- Pamponét, V. C. C., Souza, M. M., Gonçalo, S. S., Micheli, F., Ferreira de Melo, C. A., Gomes de Oliveira, S., et al. (2019). Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: cytogenomic characterization of transposable elements and satellite DNA. *BMC Genomics* 20:262. doi: 10.1186/s12864-019-5576-6
- Pedrosa-Harand, A., de Almeida, C. C., Mosiolek, M., Blair, M. W., Schweizer, D., and Guerra, M. (2006). Extensive ribosomal DNA amplification during Andean common bean (*Phaseolus vulgaris* L.) evolution. *Theor. Appl. Genet.* 112, 924–933. doi: 10.1007/s00122-005-0196-8
- Pereira, C. S. A., Aboim, M. A., Ráb, P., and Collares-Pereira, M. J. (2013). Introgressive hybridization as a promoter of genome reshuffling in natural homoploid fish hybrids (Cyprinidae, Leuciscinae). *Heredity* 112, 343–350. doi: 10.1038/hdy.2013.110
- Pita, S., Mora, P., Vela, J., Palomeque, T., Sánchez, A., Panzera, F., et al. (2018). Comparative analysis of repetitive DNA between the main vectors of chagas disease: *Triatoma infestans* and *Rhodnius prolixus*. *Int. J. Mol. Sci.* 19:1277. doi: 10.3390/ijms19051277
- Plohl, M., Luchetti, A., Mestrovic, N., and Mantovani, B. (2008). Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409, 72–82. doi: 10.1016/j.gene.2007.11.013
- Plohl, M., Meštrović, N., and Mravinac, B. (2014). Centromere identity from the DNA point of view. *Chromosoma* 123, 313–325. doi: 10.1007/s00412-014-0462-0
- R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna. Retrieved from: <https://www.R-project.org/>.
- Riethman, H., Ambrosini, A., and Paul, S. (2005). Human subtelomere structure and variation. *Chromosome Res.* 13, 505–515. doi: 10.1007/s10577-005-0998-1
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Ruiz-Herrera, A., Nergadze, S. G., Santagostino, M., and Giulotto, E. (2008). Telomeric repeats far from the ends: mechanisms of origin and role in evolution. *Cytogenet. Genome Res.* 122, 219–228. doi: 10.1159/000167807
- Ruiz-Ruano, F. J., Navarro-Domínguez, B., Camacho, J., and Garrido-Ramos, M. A. (2019). Characterization of the satellitome in lower vascular plants: the case of the endangered fern *Vandenboschia speciosa*. *Ann. Bot.* 123, 587–599. doi: 10.1093/aob/mcy192
- Sakamoto, Y., and Zacaro, A. A. (2009). *LEVAN, An ImajeJ Plugin for Morphological Cytogenetic Analysis of Mitotic and Meiotic Chromosomes*. Initial version. An Open Source Java Plugin Distributed Over the Internet from: <http://rsbweb.nih.gov/ij/>.
- Salvadori, S., Coluccia, E., Deidda, F., Cau, A., Cannas, R., and Deiana, A. M. (2012). Comparative cytogenetics in four species of Palinuridae: B chromosomes, ribosomal genes and telomeric sequences. *Genetica* 140, 429–437. doi: 10.1007/s10709-012-9691-4
- Salvadori, S., Coluccia, E., Deidda, F., Cau, A., Cannas, R., Lobina, C., et al. (2014). Karyotype, ribosomal genes, and telomeric sequences in the crayfish *Procambarus clarkii* (Decapoda: Cambaridae). *J. Crust. Biol.* 34, 525–531. doi: 10.1163/1937240X-00002247
- Salvadori, S., Coluccia, E., Milia, A., Cannas, R., and Deiana, A. M. (2002). Study of mitotic and meiotic chromosomes of *Homarus gammarus* (Crustacea Decapoda). *Biol. Mar. Mediterr.* 9, 889–891.
- Scalici, M., Solano, E., and Gibertini, G. (2010). Karyological analyses on the Australian crayfish *Cherax destructor* (Decapoda: Parastacidae). *J. Crust. Biol.* 30, 528–530. doi: 10.1651/09-3200.1
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Sember, A., Bohlen, J., Šlechtová, V., Altmanová, M., Symonová, R., and Ráb, P. (2015). Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. *BMC Evol. Biol.* 15:251. doi: 10.1186/s12862-015-0532-9
- Tan, M. H., Gan, H. M., Lee, Y. P., Bracken-Grissom, H., Chan, T. Y., Miller, A. D., et al. (2019). Comparative mitogenomics of the Decapoda reveals evolutionary heterogeneity in architecture and composition. *Sci. Rep.* 9:3617. doi: 10.1038/s41598-019-47145-0
- Tan, M. H., Gan, H. M., Lee, Y. P., Grandjean, F., Croft, L. J., and Austin, C. M. (2020). A giant genome for a giant crayfish (*Cherax quadricarinatus*) with insights into *cox1 Pseudogenes* in decapod genomes. *Front. Genet.* 11:201. doi: 10.3389/fgene.2020.00201
- Tan, X., Qin, J. G., Chen, B., Chen, L., and Li, X. (2004). Karyological analyses on red-claw crayfish *Cherax quadricarinatus* (Decapoda: Parastacidae). *Aquaculture* 234, 65–76. doi: 10.1016/j.aquaculture.2003.12.020
- Tek, A. L., and Jiang, J. (2004). The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. *Chromosoma* 113, 77–83. doi: 10.1007/s00412-004-0297-1
- Torreccilla, Z., Martínez-Lage, A., Perina, A., González-Ortegón, E., and González-Tizón, A. M. (2017). Comparative cytogenetic analysis of marine *Palaemon* species reveals a X1X1X2X2/X1X2Y sex chromosome system in *Palaemon elegans*. *Front. Zool.* 14:47. doi: 10.1186/s12983-017-0233-x
- Torres-Ruiz, R. A., and Hemleben, V. (1994). Pattern and degree of methylation in ribosomal RNA genes of *Cucurbita pepo* L. *Plant Mol. Biol.* 26, 1167–1179. doi: 10.1007/BF00040697
- Utsunomia, R., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Goes, C. A. G., Melo, S., Ramos, L. P., et al. (2019). Satellitome landscape analysis of *Megaloporus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* 9:5856. doi: 10.1038/s41598-019-42383-8
- Van Quyen, D., Gan, H. M., Lee, Y. P., Nguyen, D. D., Nguyen, T. H., Tran, X. T., et al. (2020). Improved genomic resources for the black tiger prawn (*Penaeus monodon*). *Marine Genomics* 52:100751. doi: 10.1016/j.margen.2020.100751
- Vershina, A., Anokhin, B., and Lukhtanov, V. (2015). Ribosomal DNA clusters and telomeric (TTAGG)_n repeats in blue butterflies (Lepidoptera, Lycaenidae) with low and high chromosome numbers. *Comp. Cytogenet.* 9, 161–171. doi: 10.3897/CompCytogen.v9i2.4715
- Vítková, M., Král, J., Traut, W., Zrzavý, J., and Marec, F. (2005). The evolutionary origin of insect telomeric repeats, (TTAGG)_n. *Chromosome Res.* 13, 145–156. doi: 10.1007/s10577-005-7721-0
- Vlahović, I., Glunčić, M., Rosandić, M., Ugarković, Đ*, and Paar, V. (2017). Regular higher order repeat structures in beetle *Tribolium castaneum* genome. *Genome Biol. Evol.* 9, 2668–2680. doi: 10.1093/gbe/evw174
- Wang, G., Zhang, X., and Jin, W. (2009). An overview of plant centromeres. *J. Genet. Genomics* 36, 529–537. doi: 10.1016/S1673-8527(08)60144-7
- Wang, S. Y., Biesiot, P. M., and Skinner, D. M. (1999). Toward an understanding of satellite DNA function in Crustacea. *Amer. Zool.* 39, 471–486.
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wei, K. H., Grenier, J. K., Barbash, D. A., and Clark, A. G. (2014). Correlated variation and population differentiation in satellite DNA abundance among lines of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18793–18798. doi: 10.1073/pnas.1421951112
- Weiss-Schneeweiss, H., Leitch, A. R., McCann, J., Jang, T. S., and Macas, J. (2015). “Chapter 5: Employing next generation sequencing to explore the repeat landscape of the plant genome,” in *Next-Generation Sequencing in Plant Systematics*, eds E. Hörandl and M. S. Appelhans (Washington, DC: IAPT), 1–25.

- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165
- Yuan, Z., Zhou, T., Bao, L., Liu, S., Shi, H., Yang, Y., et al. (2018). The annotation of repetitive elements in the genome of channel catfish (*Ictalurus punctatus*). *PLoS One* 13:e0197371. doi: 10.1371/journal.pone.0197371
- Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y., Yu, Y., et al. (2019). Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nat. Commun.* 10:356. doi: 10.1038/s41467-018-08197-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Boštjančić, Bonassin, Anušić, Lovrenčić, Besendorfer, Maguire, Grandjean, Austin, Greve, Hamadou and Mlinarec. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.