Clonal genome evolution and rapid invasive spread of the marbled crayfish

Julian Gutekunst[1], Ranja Andriantsoa[1], Cassandra Falckenhayn[1], Katharina Hanna[1], Wolfgang Stein [2], Jeanne Rasamy[3] and Frank Lyko [1]*

The marbled crayfish *Procambarus virginalis* is a unique freshwater crayfish characterized by very recent speciation and parthenogenetic reproduction. Marbled crayfish also represent an emerging invasive species and have formed wild populations in diverse freshwater habitats. However, our understanding of marbled crayfish biology, evolution and invasive spread has been hampered by the lack of freshwater crayfish genome sequences. We have now established a de novo draft assembly of the marbled crayfish genome. We determined the genome size at approximately 3.5 gigabase pairs and identified >21,000 genes. Further analysis confirmed the close relationship to the genome of the slough crayfish, *Procambarus fallax*, and also established a triploid AA'B genotype with a high level of heterozygosity. Systematic fieldwork and genotyping demonstrated the rapid expansion of marbled crayfish on Madagascar and established the marbled crayfish as a potent invader of freshwater ecosystems. Furthermore, comparative whole-genome sequencing demonstrated the clonality of the population and their genetic identity with the oldest known stock from the German aquarium trade. Our study closes an important gap in the phylogenetic analysis of animal genomes and uncovers the unique evolutionary history of an emerging invasive species.

Freshwater crayfish are important keystone species that play critical roles in the maintenance of their ecosystems[1,2]. Taxonomically, they belong to the order of decapod crustaceans, which includes crabs, lobsters, prawns and shrimps. Surprisingly, however, complete genome sequences from these ecologically and economically important groups remain to be established. Currently, the only crustacean genomes available are those of the water flea (*Daphnia pulex*) and the sand flea (*Parhyale hawaiensis*)[3,4], leaving decapods as a major gap in the phylogenetic analysis of genomes.

The marbled crayfish *Procambarus virginalis* (Fig. 1a) is a freshwater crayfish species[5] that holds a unique position among decapod crustaceans due to its parthenogenetic mode of reproduction[6]. Marbled crayfish are descendants of the sexually reproducing slough crayfish *Procambarus fallax* and reproduce by apomictic parthenogenesis[7,8]. We have previously suggested that marbled crayfish originated through an evolutionarily very recent macromutation in *P. fallax*, consistent with the first known appearance of marbled crayfish in the German aquarium trade in 1995 (ref. [8]). Subsequent distribution via the pet trade and anthropogenic releases resulted in increasing numbers of wild populations in several countries[9–16]. The propagation of marbled crayfish is facilitated by their parthenogenetic mode of reproduction and their high fecundity[17], which allows the establishment of large populations from single animals[18], and may serve as a model for the spread of invasive species. However, our understanding of marbled crayfish distribution, origins, diversification and ability to adapt to new environments is severely limited by the lack of genetic information.

Available genome sequences of parthenogenetic animals are currently limited to certain nematodes[19–22] and the bdelloid rotifer *Adineta vaga*[23]. Their analysis revealed several interesting features that are likely to reflect important strategies for the evolutionary robustness of these parthenogenetic animals, including the presence of allelic regions on the same chromosome in *A. vaga*[23] and substantial

heterozygosity, combined with the loss of key meiosis genes in *Diploscapter pachys* and *Diploscapter coronatus*[21,22]. However, these features were identified in genomes that have been shaped by asexual reproduction for millions of years. Due to its very young evolutionary age, the marbled crayfish provides a unique opportunity to analyse the structure of a recently formed parthenogenetic genome and to track its evolution at a very early stage.

Here, we provide a draft genome assembly of the marbled crayfish to investigate the genome structure, evolutionary history, population structure and invasive spread of this unique animal.

Results

Previous studies based on microsatellite analyses and karyotyping have shown that the marbled crayfish is a triploid organism with 276 chromosomes, which corresponds to the exact triplicate number of the haploid set of chromosomes in *P. fallax*[24]. Furthermore, marbled crayfish represent an evolutionarily very young species[8,24], which contrasts with other known parthenogenetic animals, such as bdelloid rotifers and asexually reproducing nematodes, and suggests that the three genome copies are still highly similar. We therefore assumed that the marbled crayfish genome represents a triplicate version of the original 1 N genotype from *P. fallax*. To quantitatively determine the marbled crayfish genome size, we analysed the DNA content of haemocytes by flow cytometry (Fig. 1b). Haploid genome size estimates using human and mouse blood cells as internal references suggested genome sizes of 3.9 and 3.5 gigabase pairs (Gbp), respectively (Fig. 1c and Supplementary Fig. 1). An in silico genome size estimate based on k-mer frequencies provided a slightly lower, but overall consistent value (3.3 Gb; Supplementary Fig. 1). Taken together, these findings suggest that the 1 N-equivalent genome size of the marbled crayfish is approximately 3.5 Gbp.

To establish the complete genome sequence of the marbled crayfish, we used genomic DNA from a single animal of the 'Petshop'

[1]Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center (DKFZ), Heidelberg, Germany. [2]School of Biological Sciences, Illinois State University, Normal, IL, USA. [3]Mention Zoologie et Biodiversité Animale, Université d'Antananarivo, Antananarivo, Madagascar.
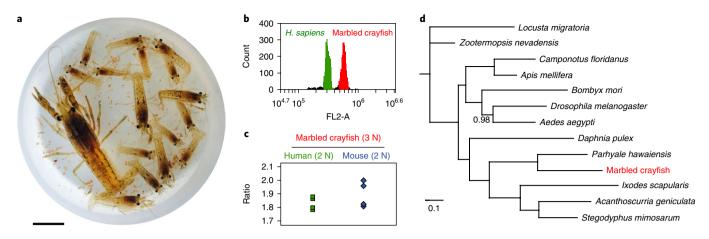*e-mail: f.lyko@dkfz.de

Fig. 1 | De novo assembly of the marbled crayfish genome. a, Three-month-old marbled crayfish from the same clutch, showing substantial phenotypic variation. Scale bar, 1 cm. **b**, Flow cytometry of haemocytes from marbled crayfish and peripheral blood from *Homo sapiens*. The FL2-A axis represents DNA content. **c**, Ratios of the DNA signals from marbled crayfish haemocytes to internal controls from human (left) or mouse blood (right). Quantitative genome size estimates are based on four independent measurements and haploid (1 N) genome sizes of 3.2 Gbp (human) and 2.8 Gbp (mouse), respectively. N indicates ploidy. **d**, Phylogenetic clustering of 138 orthologues from recently published arthropod genomes. Shimodaira–Hasegawa-like branch support <1 is indicated by numbers.

laboratory strain[8] to prepare various libraries for Illumina sequencing and obtained 350 Gbp of DNA sequence (Supplementary Table 1). After contig assembly, scaffolding and gap filling, we generated a draft genome sequence with a total length of 3.3 Gb and a weighted mean sequence length (N50) of 39.4 kilobases (kb). Benchmarking with universal single-copy orthologues[25] suggested that the quality of the marbled crayfish genome assembly was comparable to other, recently published arthropod genomes[26–28] (Supplementary Fig. 2). Phylogenetic placement among various published arthropod genomes confirmed that the marbled crayfish is most closely related to *D. pulex* and *P. hawaiensis* (Fig. 1d).

The gene length in marbled crayfish averaged 6.7 kb. Average exon and intron sizes were 0.3 kb and 2 kb, respectively, thus placing marbled crayfish gene lengths between those of *P. hawaiensis* and *D. pulex*. Gene annotation was performed using the MAKER genome annotation pipeline[29], which provided important starting points for further analysis. For example, we detected multiple genomic locations for cellulase genes of the GH9 family (Fig. 2a). Genomically encoded cellulase genes are relatively rare in higher animals, but are
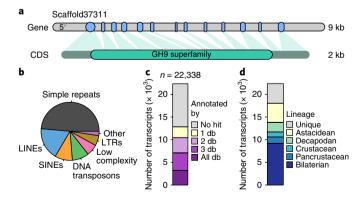
generally assumed to play a key role for omnivorousness in freshwater crayfish[30]. Furthermore, repeat annotation detected 484,313 repeats that were subclassified into 7 major categories and covered 8.8% of the annotated genome assembly (Fig. 2b). Repeat coverage is likely to increase substantially in future versions of the marbled crayfish genome assembly, as the fragmentation of the genome assembly currently represents a major bottleneck for algorithmic repeat detection. The most recent version of the genome assembly and annotation can be accessed through a dedicated internet portal (http://marmorkrebs.dkfz.de).

In parallel, we also established the marbled crayfish transcriptome from a normalized sequencing library that was generated from several distinct tissues. Benchmarking again confirmed that the quality was comparable to other, recently published arthropod transcriptomes (Supplementary Fig. 2). The transcriptome consists of 22,338 transcripts (Fig. 2c), which corresponds closely to the numbers of predicted genes (21,772) and messenger RNAs (22,205) in the genome assembly. Comparisons with other publicly available transcriptomes revealed homologues for the majority (81%) of predicted proteins, while 19% (*n* = 4,306) of the predicted proteins were classified as unique (Fig. 2d).

The analysis of two other parthenogenetic genomes had revealed the presence of homologous but diverged blocks that reflect genome rearrangements typically associated with asexual reproduction[19,23]. However, the average copy number of the 1,066 universal single-copy orthologues was 1.01, which argues against the existence of divergent homologues. Similarly, the coverage distribution of genes showed only a single peak (Fig. 3a). Finally, we could only detect a very low number (*n* = 66) of collinear genes on different scaffolds, all of which had rather high E-values (Supplementary Table 2), indicating that they probably represent artefacts. Together, these findings suggest that the genome rearrangements described in long-standing parthenogens are not detectable in the marbled crayfish genome, which is consistent with its very young evolutionary age.

Additional features of the marbled crayfish genome were revealed by the analysis of heterozygous sequence variants. The global rate of heterozygosity was 0.53%, which is relatively high compared with other sequenced genomes, including *P. fallax* from the pet trade (0.03%; Fig. 3b). Furthermore, the allelic frequency of marbled crayfish sequence variants peaked at 0.33 (Fig. 3c), in agreement with heterozygous positions in a triploid genome. In contrast, the frequencies of *P. fallax* and *P. alleni* sequence variants and polymorphisms
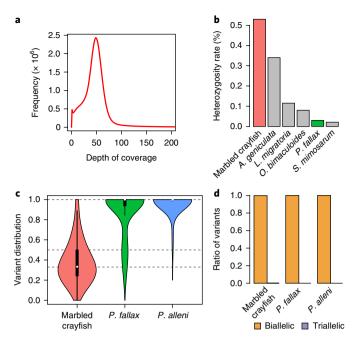

Fig. 2 | Annotation of the marbled crayfish genome. a, Gene structure of the marbled crayfish GH9 cellulase gene and coding sequence (CDS). **b**, Automatically annotated repeats (*n* = 520,403) distributed into seven major categories. LINEs/SINEs, long/short interspersed nuclear elements; LTRs, long terminal repeats. **c**, Numbers of transcripts found within one or multiple different functional annotation databases (db). **d**, Transcripts classified by lineage based on sequence homology.

Fig. 3 | Characterization of the marbled crayfish genome. a, Read coverage depth in genes extracted from the marbled crayfish genome sequence. Only a single peak can be observed, indicating the absence of divergent homologues. **b**, Genome-wide heterozygosity in marbled crayfish compared with other organisms. **c**, Violin plots illustrating the combined frequency of alternative nucleotides in all types of biallelic variants among the three different *Procambarus* species. **d**, Combined distribution of biallelic (orange) and triallelic (purple) variants in marbled crayfish, *P. fallax* and *P. alleni*.
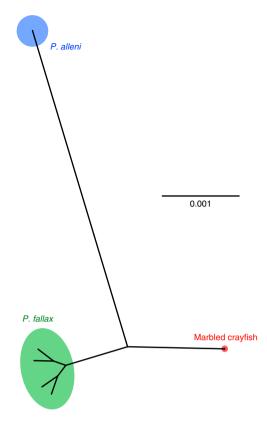

Fig. 4 | Phylogenetic relationships among marbled crayfish and closely related freshwater crayfish species. Phylogenetic tree based on the SNVs of five marbled crayfish, four *P. fallax* and one *P. alleni* genome sequences. The five marbled crayfish genomes are too similar to become resolved into visible individual branches and are therefore shown as a single branch.

peaked at 0.5 and 1.0 (Fig. 3c), reflecting their diploid nature and polymorphisms towards the marbled crayfish genome, respectively. Finally, the triploid marbled crayfish genome showed a negligible fraction (0.15%) of triallelic sequence polymorphisms (Fig. 3d). Together, these findings provide strong support for an AA′B genotype that may have originated from an autopolyploid gamete[8].

Previous studies have suggested that marbled crayfish reproduce by apomictic parthenogenesis[31,32], which should result in the establishment of a genetically homogeneous population. We therefore sequenced the genomes of four additional marbled crayfish from diverse sources: (1) an animal from the longest-known stock ('Heidelberg', founded in 1995); (2) an animal from a German wild population caught in 2013 ('Moosweiher'); (3) an animal from a market purchase in Madagascar ('Madagascar 1'); and (4) an animal from an American laboratory stock, which originated from another pet shop purchase in Germany[17] ('Petshop 2'). In addition, we generated genome sequences from two closely related species, *P. fallax* (four animals from an aquarium supplier) and *P. alleni* (one animal from an aquarium supplier). Sequencing and mapping to the marbled crayfish reference genome resulted in genome coverages ranging from 16–72× (Supplementary Table 3). We then used single-nucleotide polymorphisms to analyse phylogenetic relationships among the nine sequenced animals. The results confirmed the clonality of the four analysed marbled crayfish genomes and their separation from *P. fallax* and *P. alleni* (Fig. 4). Taken together, our findings illustrate a unique path of animal genome evolution that involves genome duplication, triploidy and clonal expansion.

Despite their clonality and very recent emergence, it has been suggested that marbled crayfish are successful invaders of new territories and environments[9,11,18]. For example, in 2007, a novel crayfish was found in the capital of Madagascar. The animals were characterized

as marbled crayfish based on their morphology and DNA sequencing of a 16 S mitochondrial DNA fragment[9,10]. However, there were no further reports on their genetic characteristics and potential spread. The availability of genome sequences for the closely related and morphologically similar marbled crayfish *P. fallax* and *P. alleni* allowed us to identify genomic sites with a high degree of sequence diversity among the three species, to confirm the identity and track the spread of the animals on Madagascar. Fieldwork was conducted in two phases, with a first series of collections in the central highland, followed by a more comprehensive study covering large parts of the country (Supplementary Table 4). In a pilot analysis, we sequenced polymerase chain reaction amplicons for a mitochondrial (cytochrome b; 214 bp) and nuclear (Dnmt1; 220 bp) locus from 24 independent animals that were collected in four regions from the central highland (Fig. 5a). The results showed 100% sequence identity with the marbled crayfish reference sequence for all analysed samples and substantial sequence differences towards *P. fallax* and *P. alleni* (Fig. 5a). These findings unambiguously classify the collected animals as marbled crayfish. Additionally, our systematic field collections and morphological analyses detected large populations of marbled crayfish (Fig. 5b) in diverse freshwater habitats, such as lakes and rice fields on the central highland, as well as in swamps close to the coastline (Supplementary Table 4 and Fig. 5c). We therefore analysed an additional 25 animals from 8 diverse regions by DNA sequencing of cytochrome b and Dnmt1 and identified only 6 mismatches among >20,000 bases analysed in total (Supplementary Table 5), which is commensurate to the normal level of polymerase chain reaction and/or sequencing errors. We estimate that between 2007 and 2017, the size of the marbled crayfish distribution area increased about 100-fold from 10^3 km² to more than 10^5 km²
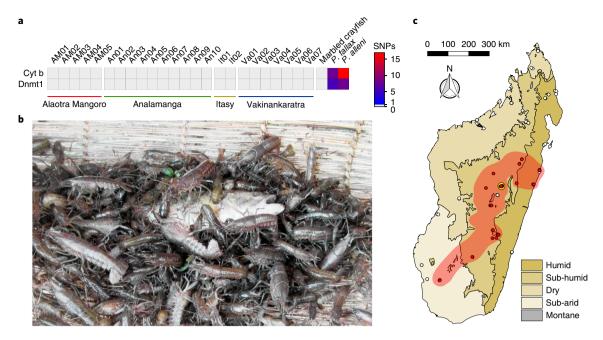
Fig. 5 | Invasive spread of marbled crayfish on Madagascar. a, Representative genotyping results of 24 animals collected on four distinct collection sites. The heatmap indicates sequence similarities with the marbled crayfish reference sequence. *P. fallax* and *P. alleni* were included as controls. Cyt b, cytochrome b; SNPs, single-nucleotide polymorphisms. **b**, Marbled crayfish by-catch in a traditional fishing tool or 'tandroho'. **c**, Distribution of marbled crayfish on Madagascar (as of March 2017) in major biogeographical zones[55], as indicated. Red dots indicate discovery sites where the presence of marbled crayfish was confirmed by DNA sequencing. White dots indicate sites where no marbled crayfish were found. The small central yellow circle indicates the distribution range reported for the year 2007 (ref. [9]).

(Fig. 5c) and that the current population on Madagascar comprises millions of animals.

To further characterize the marbled crayfish population on Madagascar, we used whole-genome sequencing. The sequencing of 5 animals from diverse collection sites and mapping to the marbled crayfish reference genome resulted in genome coverages ranging from 17× to 36× (Supplementary Table 3). Sequence comparisons revealed extremely low numbers of polymorphisms in the analysed marbled crayfish genomes (Fig. 6a). In marked contrast, the *P. fallax* genome showed a substantial number of polymorphisms towards the marbled crayfish reference genome sequence (Fig. 6a). These results provide additional, strong support for the clonality of the marbled crayfish population.

To further explore the relationship between the animals found on Madagascar and the German stocks of marbled crayfish, we obtained two additional whole-genome sequences of animals from Germany (Supplementary Table 3). Our final dataset thus consisted of 11 genome sequences from diverse sources (Supplementary Table 6). Genetic variants were extracted and filtered for single base substitutions and mapping artefacts were eliminated by remapping of sequencing reads from the genome reference individual (see Methods for details). This identified a strikingly low number of only 416 single-nucleotide variants (SNVs) in a highly diverse group of animals (Supplementary Table 7). The maximum number of non-synonymous SNVs per animal was four (Supplementary Table 7), which further illustrates the extremely low genetic complexity of the marbled crayfish population. Finally, the comparison of SNVs also provided interesting insight into the relationships of the sequenced animals. The results showed an overlapping distribution of animals from Germany and Madagascar (Fig. 6b), indicating that the Malagasy population originates from a German stock. In addition, a separate cluster was formed by two aquarium stocks that were independently founded by animals from different stores of the same German pet shop chain more than ten years ago (Fig. 6b).
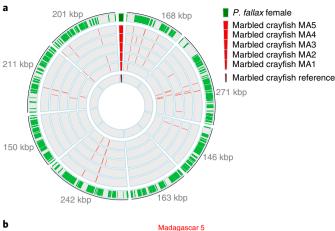
In summary, our findings thus establish the marbled crayfish as a potent invader of freshwater ecosystems and demonstrate a unique genetic structure of the invasive population.

Discussion

Our study establishes the genome assembly of a decapod crustacean, thus providing an important resource for further research for an economically and scientifically important group of animals. The marbled crayfish is a particularly important example due to its recent emergence, parthenogenetic mode of reproduction and invasive potential. Our results show that the marbled crayfish genome consists of two almost identical copies of one genotype and a third copy of a comparably divergent, but still homologous genotype. These findings are consistent with the model that the marbled crayfish genome originated from an autopolyploid *P. fallax* gamete and the mating of two distantly related *P. fallax* individuals (Supplementary Fig. 3), possibly from distant populations and in captivity. Alternative hypotheses involving allopolyploid formation with *P. alleni* appear unlikely due to the lack of hybrid morphological features[7] and the considerable genetic differences. Interestingly, triploidy and heterozygosity might provide a significant evolutionary advantage for marbled crayfish as they could buffer the effects of deleterious genetic mutations (Muller's ratchet[33]) and also increase the capacity for rapid adaptation[34]. Evolution of the marbled crayfish genome towards effective haploidy, as predicted by the Meselson effect[35,36], was not detectable. This is probably explained by the very young evolutionary age of marbled crayfish and represents an important difference from the genomes of other asexually reproducing animals, such as *Meloidogyne incognita*[19] and *A. vaga*[23].

Our results unambiguously demonstrate the clonality of the marbled crayfish genome, consistent with the proposed mode of reproduction by apomictic parthenogenesis[31,32]. The generation of genetic diversity will be shaped by a complex set of factors, including the intrinsic mutability of the genome, environmental mutagens, genetic drift and selective pressure. All these factors are known to
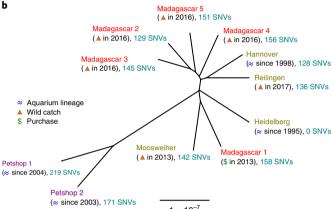
Fig. 6 | Clonality of the marbled crayfish population. a, Schematic overview of sequence polymorphisms of marbled crayfish from Madagascar. The plot consists of eight segments representing eight arbitrarily chosen genomic scaffolds (scaffold lengths are indicated) in concentric rings, representing different animals. Vertical lines represent polymorphic positions to the reference genome. The genome sequences of *P. fallax* (green) and the marbled crayfish reference genome sequence (dark red) are shown for comparison. kbp, kilobase pairs. **b,** Phylogenetic tree of 11 marbled crayfish from diverse sources, as determined by the distribution of the 416 SNVs detected in the population. Names shown in red indicate animals from Madagascar, while mustard yellow indicates animals from Germany. Animals originating from a German pet shop chain are shown in purple.

play an important role in the evolution of tumour genomes[37,38]. The analysis of mutations in marbled crayfish populations provides an opportunity to detect the generation, fixation and elimination of genetic changes with particularly high sensitivity and robustness and could therefore disentangle the specific contributions of individual factors. As such, it will be interesting to further explore marbled crayfish as a model system for clonal genome evolution in cancer[39,40].

Our results also provide detailed genetic information about the marbled crayfish population on Madagascar. While increasing numbers of marbled crayfish introductions have been reported in northern countries[11–16], there are currently no scientific records of active range expansions. This contrasts with the situation on Madagascar, where rapid invasive spread is not only facilitated by anthropogenic distribution, but also supported by favourable environmental conditions, such as relatively high temperatures and a dense network of freshwater habitats[41,42] (Supplementary Fig. 4). The marbled crayfish population on Madagascar will have to be monitored closely to prevent any adverse impact on the unique local freshwater communities[43,44].

This includes seven endemic crayfish species[45,46] in habitats that are adjacent to or overlapping with the current distribution range of marbled crayfish.

Finally, our results also demonstrate that the Madagascar population is genetically homogeneous and extremely similar to the oldest known stock of marbled crayfish founded in Germany in 1995. These findings support the notion that the global marbled crayfish population represents a single clone. The rapid invasion of diverse habitats is particularly noteworthy, as it appears to be independent of genetic variants, which are generally considered to be the major determinants of ecological adaptation[47]. This suggests that alternative mechanisms, such as stochastic epigenetic variation and/or epigenetic plasticity[48,49], play a prominent role in the rapid adaptation of marbled crayfish.

Methods

Genome size estimation by flow cytometry. A detailed protocol for cell preparation can be found in the Supplementary Materials. Briefly, 100 μl aliquots of cells treated with 2 μl RNase A stock solution (50 mg ml⁻¹) were mixed with 5 μl propidium iodide stock solution (1 mg ml⁻¹). Samples were diluted with 100 μl 1× phosphate buffered saline and briefly mixed. Propidium-iodide-stained cells were counted and the fluorescence intensity per cell was measured. After determining the cell density of each sample (cell counts μl⁻¹), equal volumes of stained cells from different organisms were mixed together and analysed again with the flow cytometer. The genome size was obtained by calculating the median fluorescence signal of stained cells per haploid genome multiplied by the known genome size of the reference samples.

Library generation and genome sequencing. A detailed protocol for DNA isolation and quality control can be found in the Supplementary Materials. Library generation, sequencing and pre-processing of read data was performed by Eurofins MWG. DNA from one individual female from a laboratory strain (Petshop 1) was used for sequencing. Fragmentation of shotgun libraries was performed on a Covaris E210 according to the manufacturer's instructions. Ligation products were size selected by agarose gels, with a targeted insert size of 500 bp. Library generation for long jumping distance (LJD) sequencing was performed following a mate-pair library protocol provided by Illumina. The protocol was modified using adaptor-guided ligation of genomic fragments, which achieves higher accuracies. Targeted insert sizes were 3, 8, 20 and 40 kb.

In total, six shotgun libraries and six LJD libraries were produced. Clusters were generated using an Illumina cBot, followed by sequencing which differed between the shotgun library and LJD runs. Shotgun libraries were sequenced using a HiSeq 2500 platform (HiSeq Control Software 2.0.12.0) with 2× 150 bp paired-end sequencing. LJD libraries were sequenced using Illumina HiSeq 2000 with 2× 100 bp paired-end sequencing. Additionally, one MiSeq library was generated in the German Cancer Research Center Genomics and Proteomics Core Facility following a standard MiSeq protocol with a library with 900-bp insert sizes and a 300-bp read length.

Genome assembly and annotation. A detailed description can be found in the Supplementary Materials.

Phylogenetic relationships. A set of 138 single-copy orthologues were extracted from recently published arthropod genomes. To estimate genetic divergence, a multiple sequence alignment was calculated with ClustalW (2.1)[50]. Furthermore, alignments were reduced to focus conserved blocks using Gblocks (0.91b)[51]. Phylogenetic tree construction was based on maximum likelihood estimation from PhyML (20120412)[52] with default parameters. Branch support was calculated using a Shimodaira–Hasegawa-like procedure..

Analysis of genetic variants. Sequencing data of all individuals (marbled crayfish, *P. fallax* and *P. alleni*) were mapped to the marbled crayfish reference genome using Bowtie2 (2.2.6)[53]. Subsequently, a subset of sequences larger than 10 kb was extracted, resulting in a total of 6.7×10^8 callable sites comprising about 19% of the genome assembly. Variant calling was performed using Freebayes (0.9.21-g7dd41db). Remapping shotgun reads and eliminating sites with no reference allele observations enabled filtering of potential mapping error sites. Variants were categorized as biallelic and triallelic, describing their occurrence on variant alleles. Additionally, polymorphisms were denoted as SNVs in one allele (marbled crayfish) or in up to two alleles (*P. fallax and P. alleni*) to the reference genome. Before polymorphism calling, sites were restricted to non-heterozygous loci, as determined by variant calling in the genome individual.

Population genetic mutations were quantified by mapping read data to the reference genome, as described above. Polymorphic sites were further restricted by data from the reference strain to positions with a minimum reference coverage of 10 and a maximum reference coverage of 200. Furthermore, each site was required

to have a valid genotype (as determined by Freebayes) in all samples. Samples were filtered for at most one heterozygous substitution to the reference genome. For each site, a minimum coverage of eight was required for each sample to account for differences in the sequencing yield. Phylogenetic trees were generated based on pairwise polymorphism information between different *Procambarus* species and between different marbled crayfish animals.

Determination of heterozygosity levels. High-coverage-sequence information (~72×) of the marbled crayfish individual used for genome assembly was used for estimating heterozygosity levels. Non-normalized read-mapping data, as described in variation analysis, were extracted from sequences of ≥10 kb. Variant calling was performed using Freebayes with the parameter settings described in the Supplementary Materials. Heterozygous positions were extracted after filtering the output quality of at least 30 and coverage of at least 15. Heterozygosity information for other genomes, estimated by a similar approach, was extracted from publications[26,27,54]. The heterozygosity level of *P. fallax* was estimated by assembling high-coverage reads into a raw contig assembly. Due to a lower total read coverage than in marbled crayfish, the coverage cutoff was 10. Heterozygosity levels were calculated as the total number of heterozygous single-nucleotide polymorphisms divided by the actual number of nucleotides (without ambiguous bases).

Fieldwork. Fieldwork was conducted on Madagascar between March 2016 and March 2017 and covered 15 regions, 33 sites and 88 sampling stations. Sites were selected based on biogeographical parameters with a wide range of different habitats, as well as feedback from awareness campaigns and interactions with local residents. Two to five sampling stations were chosen per site, according to the availability of water access and local collaborators, such as fishermen or crayfish collectors. Animals were caught in lakes, streams, ponds, rice field channels and swamps for one to three hours during the morning using traditional fishing tools, such as creels (50 cm × 30 cm × 30 cm) called 'tandroho' and/or nets called 'harato'. Ecological characteristics of the habitats and physicochemical parameters of the water were noted. Collected animals (up to 100 per catch) were morphologically analysed to preliminarily confirm their identity. Tissue samples from three to five animals per sampling station were preserved in ethanol for genotyping.

Genotyping. Crayfish genomic DNA was isolated and purified from 100 mg abdominal musculature using a Tissue Ruptor and the DNeasy Blood and Cell Culture Kit (both Qiagen), following the manufacturer's instructions. For genotyping, two fragments were amplified: a 274-bp fragment from the mitochondrial cytochrome b gene was amplified using primers: 5′-CAG GAC GTG CTC CGA TTC ATG-3′ and 3′-GAC CCA GAT AAC TTC ATC CCA G-5′. In addition, a 334-bp fragment of the Dnmt1 nuclear gene was amplified using primers 5′-GCT TTC TGG TCT CGT ATG GTG-3′ and 3′-CTG CAC ACA GCC TAA GAT GC-5′. A polymerase chain reaction was performed with thin-wall tubes using a Bio-Rad Peltier Thermal Cycler. Some 5 µl of genomic DNA was added to a reaction mixture (25 µl final volume) containing 2 µl (10 µmol) of reverse and forward primers, 1 µl of deoxynucleoside triphosphates (10 mmol), 0.5 µl of FireTaq blue polymerase (1 U µl⁻¹; Steinbrenner), 2.5 µl of 10× Reaction Buffer and 14 µl of water. Samples were preheated at 96 °C for 3 min followed by amplification under the following conditions: denaturation at 96 °C for 30 s, annealing at 57 °C for 30 s and elongation at 72 °C for 30 s. A total of 30 cycles were performed and then followed by a final elongation step at 72 °C for 3 min. The resulting polymerase chain reaction amplicons were analysed and purified by agarose gel electrophoresis and cloned using the TOPO TA Cloning Kit (Invitrogen) according to the manufacturer's instructions. Purified plasmids were sequenced by GATC Biotech and the sequences were aligned in FinchTV version 1.4.0.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. All sequencing data have been deposited as a National Center for Biotechnology Information BioProject (accession number PRJNA356499).

References

1. Creed, R. P. Jr Direct and indirect effects of crayfish grazing in a stream community. *Ecology* **75**, 2091–2103 (1994).
2. Crandall, K. A. & Buhay, J. E. Global diversity of crayfish (Astacidae, Cambaridae, and Parastacidae–Decapoda) in freshwater. *Hydrobiologia* **595**, 295–301 (2008).
3. Colbourne, J. K. et al. The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555–561 (2011).
4. Kao, D. et al. The genome of the crustacean *Parhyale hawaiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *eLife* **5**, e20062 (2016).
5. Lyko, F. The marbled crayfish (Decapoda: Cambaridae) represents an independent new species. *Zootaxa* **4363**, 544–552 (2017).
6. Scholtz, G. et al. Ecology: parthenogenesis in an outsider crayfish. *Nature* **421**, 806 (2003).
7. Martin, P., Dorn, N. J., Kawai, T., van der Heiden, C. & Scholtz, G. The enigmatic Marmorkrebs (marbled crayfish) is the parthenogenetic form of *Procambarus fallax* (Hagen, 1870). *Contrib. Zool.* **79**, 107–118 (2010).
8. Vogt, G. et al. The marbled crayfish as a paradigm for saltational speciation by autopolyploidy and parthenogenesis in animals. *Biol. Open.* **4**, 1583–1594 (2015).
9. Jones, J. P. G. et al. The perfect invader: a parthenogenic crayfish poses a new threat to Madagascar's freshwater biodiversity. *Biol. Invasions* **11**, 1475–1482 (2009).
10. Kawai, T. et al. Parthenogenetic alien crayfish (Decapoda: Cambaridae) spreading in Madagascar. *J. Crust. Biol.* **29**, 562–567 (2009).
11. Chucholl, C., Morawetz, K. & Groß, H. The clones are coming—strong increase in Marmorkrebs [*Procambarus fallax* (Hagen, 1870) f. *virginalis*] records from Europe. *Aquat. Invasions* **7**, 511–519 (2012).
12. Bohman, P., Edsman, L., Martin, P. & Scholtz, G. The first Marmorkrebs (Decapoda: Astacida: Cambaridae) in Scandinavia. *Bioinvasions Rec.* **2**, 227–232 (2013).
13. Patoka, J. et al. Predictions of marbled crayfish establishment in conurbations fulfilled: evidences from the Czech Republic. *Biologia* **71**, 1380–1385 (2016).
14. Lokkos, A. et al. The alien, parthenogenetic marbled crayfish (Decapoda: Cambaridae) is entering Kis-Balaton (Hungary), one of Europe's most important wetland biotopes. *Knowl. Manag. Aquat. Ecosyst.* **417**, 16 (2016).
15. Liptak, B. et al. Expansion of the marbled crayfish in Slovakia: beginning of an invasion in the Danube catchment? *J. Limnol.* **75**, 305–312 (2016).
16. Novitsky, R. A. & Son, M. O. The first records of Marmorkrebs [*Procambarus fallax* (Hagen, 1870) f. *virginalis*] (Crustacea, Decapoda, Cambaridae) in Ukraine. *Ecol. Monte* **5**, 44–46 (2016).
17. Seitz, R., Vilpoux, K., Hopp, U., Harzsch, S. & Maier, G. Ontogeny of the Marmorkrebs (marbled crayfish): a parthenogenetic crayfish with unknown origin and phylogenetic position. *J. Exp. Zool. A Comp. Exp. Biol.* **303**, 393–405 (2005).
18. Feria, T. P. & Faulkes, Z. Forecasting the distribution of Marmorkrebs, a parthenogenetic crayfish with high invasive potential, in Madagascar, Europe, and North America. *Aquat. Invasions* **6**, 55–67 (2011).
19. Abad, P. et al. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* **26**, 909–915 (2008).
20. Blanc-Mathieu, R. et al. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS. Genet.* **13**, e1006777 (2017).
21. Fradin, H. et al. Genome architecture and evolution of a unichromosomal asexual nematode. *Curr. Biol.* **27**, 2928–2939 (2017).
22. Hiraki, H. et al. Genome analysis of *Diploscapter coronatus*: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. *BMC Genom.* **18**, 478 (2017).
23. Flot, J. F. et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* **500**, 453–457 (2013).
24. Martin, P., Thonagel, S. & Scholtz, G. The parthenogenetic Marmorkrebs (Malacostraca: Decapoda: Cambaridae) is a triploid organism. *J. Zool. Syst. Evol. Res.* **54**, 13–21 (2016).
25. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
26. Sanggaard, K. W. et al. Spider genomes provide insight into composition and evolution of venom and silk. *Nat. Commun.* **5**, 3765 (2014).
27. Wang, X. et al. The locust genome provides insight into swarm formation and long-distance flight. *Nat. Commun.* **5**, 2957 (2014).
28. Gulia-Nuss, M. et al. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat. Commun.* **7**, 10507 (2016).
29. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinforma.* **48**, 1–39 (2014).
30. Byrne, K. A., Lehnert, S. A., Johnson, S. E. & Moore, S. S. Isolation of a cDNA encoding a putative cellulase in the red claw crayfish *Cherax quadricarinatus*. *Gene* **239**, 317–324 (1999).
31. Vogt, G., Tolley, L. & Scholtz, G. Life stages and reproductive components of the Marmorkrebs (marbled crayfish), the first parthenogenetic decapod crustacean. *J. Morphol.* **261**, 286–311 (2004).
32. Martin, P., Kohlmann, K. & Scholtz, G. The parthenogenetic Marmorkrebs (marbled crayfish) produces genetically uniform offspring. *Naturwissenschaften* **94**, 843–846 (2007).
33. Muller, H. J. The relation of recombination to mutational advance. *Mutat. Res.* **106**, 2–9 (1964).
34. Selmecki, A. M. et al. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–352 (2015).
35. Birky, C. W. Jr Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* **144**, 427–437 (1996).

36. Mark Welch, D. & Meselson, M. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**, 1211–1215 (2000).

37. Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nat. Rev. Genet.* **13**, 795–806 (2012).

38. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483–1489 (2015).

39. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).

40. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).

41. IUCN/UNEP/WWF in *Madagascar, An Environmental Profile* (ed. Jenkins, M. D.) 1–10 (IUCN, Gland and Cambridge, 1987).

42. Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen–Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**, 1633–1644 (2007).

43. Benstead, J. P. et al. Conserving Madagascar's freshwater diversity. *Biosciences* **53**, 1101–1111 (2003).

44. Brown, J. L. et al. Spatial biodiversity patterns of Madagascar's amphibians and reptiles. *PLoS. ONE* **11**, e0144076 (2016).

45. Rabearisoa, B., Elouard, J. M. & Ramanankasina, E. in *Biogeography of Madagascar* (ed. Lourenco, W. R.) 559–562 (ORSTOM, Paris, 1996).

46. Jones, J. P. G., Andriahajaina, F. B., Hockley, N. J., Crandall, K. A. & Ravoahangimalala, O. R. The ecology and conservation status of Madagascar's endemic freshwater crayfish (Parastacidae; Astacoides). *Freshw. Biol.* **52**, 1820–1833 (2007).

47. Savolainen, O., Lascoux, M. & Merila, J. Ecological genomics of local adaptation. *Nat. Rev. Genet.* **14**, 807–820 (2013).

48. Vogt, G. et al. Production of different phenotypes from the same genotype in the same environment by developmental variation. *J. Exp. Biol.* **211**, 510–523 2008).

49. Pujadas, E. & Feinberg, A. P. Regulated noise in the epigenetic landscape of development and disease. *Cell* **148**, 1123–1131 (2012).

50. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

51. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).

52. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).

53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

54. Albertin, C. B. et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220–224 (2015).

55. Schatz G. E. in *Diversity and Endemism in Madagascar* (eds Lourenco, W. R. & Goodman, S. M.) 1–9 (ORSTOM, Paris, 2000).

Acknowledgements

Author contributions

J.G. and C.F. carried out the computational analyses. R.A. and J.R. carried out the fieldwork. R.A., C.F. and K.H. carried out the experiments. W.S. contributed sequencing data. F.L. conceived the study and wrote the manuscript with contributions from J.G., R.A. and W.S. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41559-018-0467-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to F.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

natureresearch

Corresponding author(s): Frank Lyko

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▶ Experimental design

1. **Sample size**

   Describe how sample size was determined.

   > We analyzed a clonal population of >150 animals (Tab. S5), >50 of which were analyzed by PCR and sequencing (Fig. 5). Furthermore, 11 animals were analyzed by whole-genome sequencing. The results confirmed the clonality of the animals.

2. **Data exclusions**

   Describe any data exclusions.

   > No data were excluded.

3. **Replication**

   Describe whether the experimental findings were reliably reproduced.

   > The main finding of our paper is the clonality of the marbled crayfish population, which was confirmed repeatedly and by highly sensitive methodology (e.g. whole-genome sequencing of 11 animals from diverse sources).

4. **Randomization**

   Describe how samples/organisms/participants were allocated into experimental groups.

   > Not applicable, due to the clonality of the population analyzed.

5. **Blinding**

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   > Blinding was not relevant, as we describe a clonal population.

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. **Statistical parameters**

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	A statement indicating how many times each experiment was replicated
☒	☐	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
☒	☐	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

   See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

> All software packages (including details, such as version numbers) are provided in the Methods section.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

> All materials are freely available from the authors.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

> Not applicable.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

> Not applicable, no cell lines were used.

b. Describe the method of cell line authentication used.

> Not applicable, no cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination.

> Not applicable, no cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

> Not applicable, no cell lines were used.

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

> Marbled crayfish, female, 1-4 years old. Research Permit N° 262/16/MEEF/SG/DGF/DSAP/SCB.Re (Ministry of Ecology, Environment and Forest of Antananarivo, Madagascar) for field work on Madagascar.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

> The study did not involve human research participants.