SCIENTIFIC REPORTS

OPEN

Received: 15 November 2016
Accepted: 11 May 2017
Published online: 22 June 2017

The highly rearranged mitochondrial genomes of the crabs *Maja crispata* and *Maja squinado* (Majidae) and gene order evolution in Brachyura

Andrea Basso[1], Massimiliano Babbucci[1], Marianna Pauletto[1], Emilio Riginella[2], Tomaso Patarnello[1] & Enrico Negrisolo[1]

We sequenced the mitochondrial genomes of the spider crabs *Maja crispata* and *Maja squinado* (Majidae, Brachyura). Both genomes contain the whole set of 37 genes characteristic of Bilaterian genomes, encoded on both α- and β-strands. Both species exhibit the same gene order, which is unique among known animal genomes. In particular, all the genes located on the β-strand form a single block. This gene order was analysed together with the other nine gene orders known for the Brachyura. Our study confirms that the most widespread gene order (BraGO) represents the plesiomorphic condition for Brachyura and was established at the onset of this clade. All other gene orders are the result of transformational pathways originating from BraGO. The different gene orders exhibit variable levels of genes rearrangements, which involve only tRNAs or all types of genes. Local homoplastic arrangements were identified, while complete gene orders remain unique and represent signatures that can have a diagnostic value. Brachyura appear to be a hot-spot of gene order diversity within the phylum Arthropoda. Our analysis, allowed to track, for the first time, the fully evolutionary pathways producing the Brachyuran gene orders. This goal was achieved by coupling sophisticated bioinformatic tools with phylogenetic analysis.

The true crabs belong to Brachyura, the largest clade (an infraorder) of the crustacean Decapoda order (Crustacea, Malacostraca)[1]. Shrimps, prawns, crayfishes and lobsters, some of the most popular crustaceans, are also contained in Decapoda[1]. Currently, more than 7,250 species belong to the Brachyura[2]. Crabs form a big taxonomic group and exhibit a broad array of forms and adaptations, what make them one of the key group to study important biological and evolutionary issues[3]. Several Brachyuran species play an important role as food source for humans and have a relevant commercial value in the fish markets worldwide[4].

Currently, the Brachyura are divided in the five major clades Dromiodea, Homoloidea, Cyclodorippoidea, Raninoidea and Eubrachyura (Fig. 1)[2,5,6]. The first two taxa form a monophyletic group, as well as Cyclodorippoidea and Raninoidea (Fig. 1). This latter clade is sister taxon of Eubrachyura, the biggest and most differentiated lineage of crabs, encompassing the vast majority of the species. The Eubrachyura are split in two major groups named Heterotremata and Thoracotremata (Fig. 1). Within the Eubrachyura, the phylogenetic position of the primary-freshwater crabs belonging to the families Gecarcinucidae, Potamidae, Potamonautidae, Pseudothelphusidae and Tricodactylidae, is particularly debated[6]. According to the applied phylogenetic method and the type of analysed characters they have been placed within Heterotremata or Thoracotremata[5,6]. Notably, in the most recent phylogeny available for Brachyura, the freshwater crabs are included within Heterotremata[5,6], thus this arrangement is initially followed in the present paper (Fig. 1).

[1]University of Padova, Department of Comparative Biomedicine and Food Science (BCA), 35020, Agripolis, Legnaro (PD), Italy. [2]University of Padova, Department of Biology, 35131, Padova, Italy. Andrea Basso and Massimiliano Babbucci contributed equally to this work. Correspondence and requests for materials should be addressed to E.N. (email: enrico.negrisolo@unipd.it)

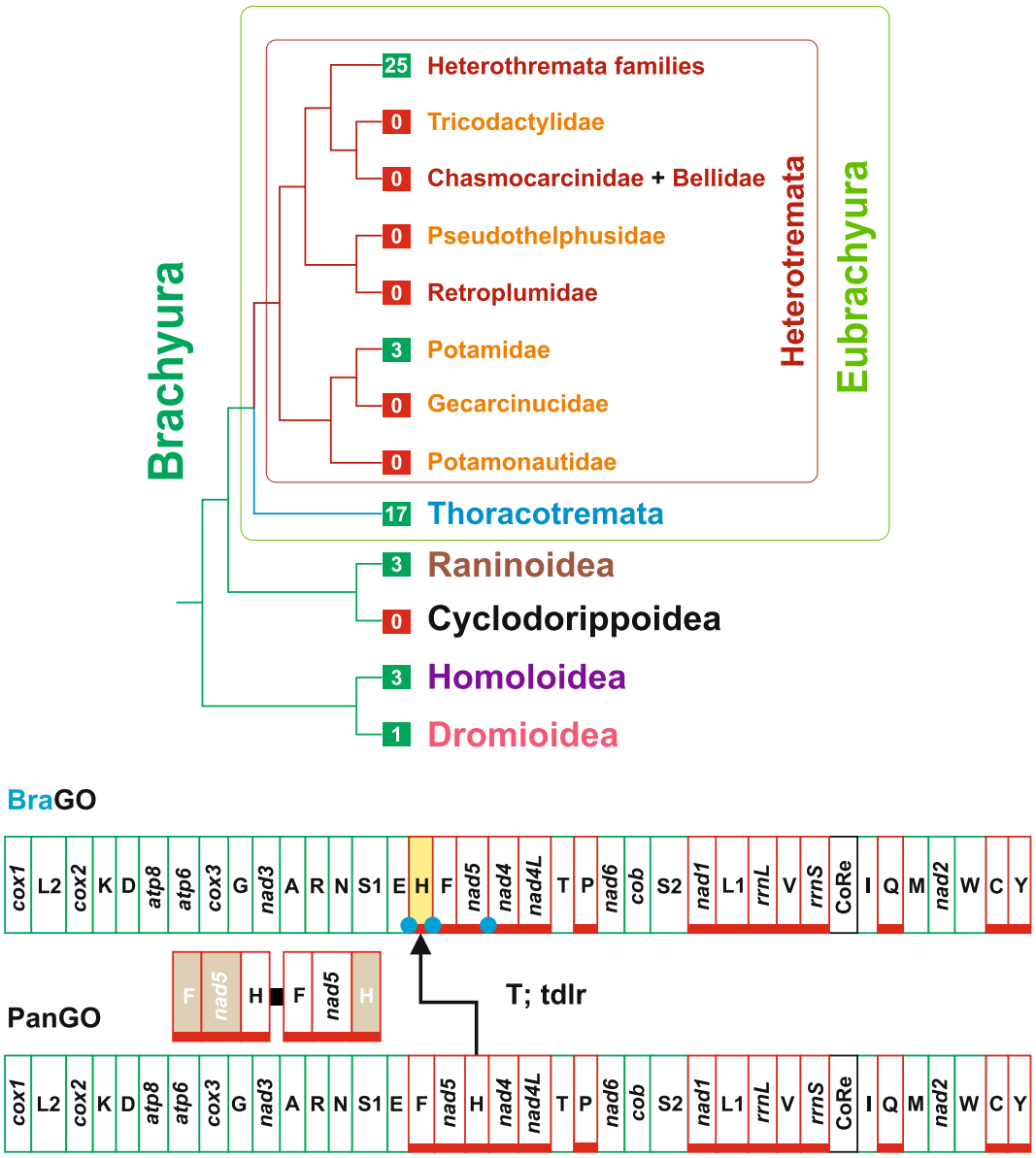


Figure 1. Phylogenetic relationships among major Brachyuran clades and BraGO vs PanGO. The tree depicts the phylogenetic relationships among the major Brachyura lineages cited in the main text. The values in the green/red box refer to the full-length mtDNAs available for that lineage. The taxa names are coloured according to the current placement of taxa themselves within the major Brachyuran groups (*e.g.* all families belonging to fresh water crabs are in orange). The genomic transformation from PanGO to BraGO is provided below the tree. PanGO is linearized starting from *cox1*. The genes encoded on the α-strand (orientation from right to left in Fig. 1) are green-boxed, while those encoded on the β-strand (orientation from left to right in Fig. 1) are underlined and red-boxed. Nomenclature: *atp6* and *atp8*: ATP synthase subunits 6 and 8; *cob*: apocytochrome b; *cox1-3*: cytochrome c oxidase subunits 1–3; *nad1-6* and *nad4 L*: NADH dehydrogenase subunits 1–6 and 4 L; *rrnS* and *rrnL*: small and large subunit ribosomal RNA (rRNA) genes; X: transfer RNA (tRNA) genes, where X is the one-letter abbreviation of the corresponding amino acid, in particular L1 (CTN codon family) L2 (TTR codon family), S1 (AGN codon family) S2 (TCN codon family); CoRe: Control Region. T: transposition event. Tdrl: tandem duplication random loss mechanism producing the observed rearrangement. *TrnH*, that changed its position relative to PanGO, through a transposition event, is shown with a yellow background. Conversely, the passively-shifted genes are figured with their original background (see Methods section). The extra copy of every gene that is lost in the genomic rearrangement is figured with a light brown background. A blue circle marks an intergenic spacer present in a position associated to genomic rearrangement.

The mitochondrial genome (mtDNA) of Crustacea is usually a double helical and circular molecule spanning 15–18 kb[7]. The most noftable exception is represented by several Isopoda mtDNAs consisting in a combination of a linear molecule approximately 14 kb long, and a circular molecule, made by two linear molecules connected in a head-to-head arrangement[8,9].

The Crustacean mtDNA usually contains 37 genes including 13 protein-encoding genes, 22 tRNAs and the small and large ribosomal RNAs (Fig. 1)[7]. Exceptions exist to this largely prevalent scheme and some genes can be absent[10], or duplicated[11].

The mitochondrial genes are encoded on both strands of DNA (hereinafter referred to as the α- and β-strands). Genes can overlap, be adjacent or separated by a variable number of nucleotides (*i.e.*, intergenic spacers). The major intergenic spacer that is always present is the Control Region (CoRe) harbouring the mtDNA origin of replication[12]. The gene order (GO) is not always conserved and Crustacean mtDNAs exhibit different GOs[10]. With respect to a reference GO, genes can be transposed (*i.e.*, moved to a different placement on the same strand), inverted (*i.e.*, moved to the opposite strand), or both inverted and transposed (a combination of the first two events). Mitochondrial genes rearrangements are not completely unveiled, although various models have been proposed[13–17]. A transposition can be explained by a tandem duplication and random loss model (tdrl)[13,14]. A gene inversion is modelled through an intra-mitochondrial recombination[15], while the inverted transposition can be described through the combination of these two mechanisms. A tandem duplication and random loss (TDRL) event can be applied to analyse the global rearrangement pattern[16,17]. According to Bernt and Middendorf[17], TDRL involves a tandem duplication of a continuous segment of genes such that the original segment and its copy are placed consecutively and followed by the loss of one copy of each redundant genes. Multiple genes simultaneously change their position in a TDRL event.

Different GOs have proven to be highly diagnostic in defining animal groups at various taxonomic ranks[18]. In particular, the sister-taxon relationship between Crustacea and Hexapoda (*i.e.*, the clade Pancrustacea) is strongly supported by their exclusively shared GO (hereafter named PanGO) (Fig. 1)[19].

The crab mtDNA usually contains the whole set of genes mentioned above. A peculiar condition is found in the potamid crab *Geothelphusa dehani*, which exhibits a case of tRNA remoulding (also named recruitment)[20]. In this case, a point mutation in the anticodon (TAG → TAA) transformed an extra-copy of *trnL1* in a functional *trnL2*, while the true *trnL2* was lost. The remoulding of tRNA is a general process, occurring sparsely in both Eukaryota and Prokaryota, which can involve multiple tRNAs[21–24]. Within Decapoda mtDNAs, the tRNA remoulding has been recorded in species of hermit crabs (Anomura)[22], as well as in mud shrimps (Gebiidea and Axiidea)[25]. However, in these taxa the point mutation (TAA → TAG) occurred in an extra copy of *trnL2*, which became a functional *trnL1*, while the true *trnL1* is no longer present.

Crabs exhibit different mtDNA GOs, none identical to PanGO[20,26–33] (Supplementary Table S1). The most common, the Brachyuran basic GO (hereafter named BraGO)[26,34], is depicted in Fig. 1. BraGO differs from PanGO for the transposition of *trnH*, which is located between *trnE* and *trnF*, instead of its placement downstream to *nad5* in PanGO (Fig. 1). Currently, full-length mtDNAs are available for representatives of all major Brachyuran clades[20,27,28,32,33,35–41] (Fig. 1) (see also Supplementary Table S1). However, the taxon coverage is very sparse and the sequencing of new genomes is a high priority. To improve our knowledge on Brachyuran mtDNAs we sequenced the complete genomes of the two spider crabs *Maja crispata* and *Maja squinado* (Majidae). The mtDNAs of both species exhibit the same GO (hereafter MajGO), which is different from any other known animal GO and very re-arranged with respect to PanGO and BraGO. The MajGO is described in details in the Results and Discussion section. After describing the MajGO, we downloaded all complete, or near complete, Brachyuran mtDNAs available in GenBank (release 30.09.2016) and analysed them, in combination with the newly sequenced *Maja* genomes. The goals of present paper were (a) to establish the transformational pathways that led to the diverse GOs observed in Brachyura; (b) to identify the plesiomorphic condition among Brachyuran GOs; (c) to trace the evolutionary steps that produced each unique GO; (d) to test the value of GOs as molecular signatures for the Brachyuran clades.

Methods

Ethics statement. No specific permits were required for the work described here. Individuals included in the present study were bought in a fish market or directly collected by one of the authors and they were not subjected to any experimental manipulation. The study was performed in accordance with the EU directive 2010/63/EU and Italian DL 2014/26. The experiments, as well as the euthanasia procedure, were monitored and carried out by authorized staff to minimise animals' suffering.

Sampling of Maja crispata and Maja squinado. The specimen of *M. crispata* used in the present study was collected by Emilio Riginella in the Venice Lagoon (Italy). The specimen of *M. squinado*, caught in the North Adriatic Sea, was acquired in the fish marked of Chioggia (Italy) by Enrico Negrisolo. The samples were preserved in pure ethanol at 4 °C until DNA extraction.

Total DNA was extracted using the ZR Genomic DNA-Tissue Midiprep (Zymo Research corp.) Kit. DNA quality was assessed through electrophoresis. The DNA concentration was determined using the (high sensitivity) Qubit DNA quantification kit (Invitrogen, USA).

Mitochondrial genome sequencing. The total DNAs, at a concentration of at least 100 ng/μl, were sent to the IGA Technology facility (http://www.igatechnology.com/) (Udine, Italy) to be sequenced using Next-Generation Sequencing (NGS) Illumina HiSeq 2000 and following a 100PE strategy (See the IGA Technology Services for further details on the sequencing strategy). After the sequencing process, 25,946,982 and 32,836,146 paired sequences were obtained for *M. crispata* and *M. squinado*, respectively.

Genome assembly and identification of the full length mitochondrial genome. Global assembly of the Illumina reads obtained for *M. crispata* and *M. squinado* was accomplished with the software CLC Genomics Workbench v8.5 (http://www.clcbio.com). After a BLAST search against the non-redundant database available at the NCBI web site[42], the sequences that had a high score match with mitochondrial genes (E 10^{-20})

were fully annotated using the strategy described in the next section. Afterwards, a single sequence for both *M. crispata* and *M. squinado* covering at least 95% of the final full length mtDNA (see below), was selected as the template for successive assembly performed using the MITObim program[43]. This second analysis provided a final assembly encompassing the full length mitochondrial genome (mtDNA) for both *M. crispata* and *M. squinado* Statistics on the final assemblies were calculated with CLC Genomics Workbench v8.5.

The full length sequences of both mtDNAs can be accessed from the EBI/GenBank (*M. crispata*, KY650651; *M. squinado*, KY650652).

Mitochondrial genome annotation. The nomenclature of genes and strands are according to Negrisolo *et al.*[44]. The names used to indicate strands are very variable in mtDNA literature[7,10,20,26,27,30,34,36,44]. In this paper, the strand encoding the majority of genes is listed as α-strand[20,34,44]. First/majority/plus/Heavy (H) -strand are alternative names for the α-strand[7,10,26,27,30,36]. Conversely, the strand encoding the minority of genes is listed here as β-strand. Second/minority/minus/Light (L) -strand are alternative names for the β-strand[7,10,26,27,30,36]. Initially, the mtDNA sequence was translated into putative proteins using the Transeq program available on the EBI website (https://www.ebi.ac.uk/Tools/st/emboss_transeq/). The identity of these polypeptides was verified using the BLAST program[42] available at the NCBI website. The boundaries of genes were determined as follows: the 5′ ends of protein-coding genes (PCGs) were defined as the first legitimate in-frame start codon (ATN, GTG, TTG, GTT) in the open reading frame (ORF) that was not located within an upstream gene encoded on the same strand. The only exceptions were *atp6* and *nad4*, which overlap with their upstream genes (*atp8* and *nad4L*, respectively) in many mtDNAs[45]. The PCG terminus was defined as the first in-frame stop codon that was encountered. When the stop codon was located within the sequence of a downstream gene encoded on the same strand, a truncated stop codon (T or TA) adjacent to the beginning of the downstream gene was designated as the termination codon. This codon was thought to be completed by polyadenylation, thereby producing a complete TAA stop codon after transcript processing. Finally, pairwise comparisons with orthologous proteins were performed using ClustalW[46] to better define the limits of the PCGs.

Regardless of the real initiation codon, a formyl-Met was assumed to be the starting amino acid for all proteins as previously reported in other mitochondrial genomes[47,48].

Transfer RNA genes were identified using the tRNAscan-SE program[49] or recognised manually as sequences having the appropriate anticodon and capable of folding into the typical cloverleaf secondary structure of tRNAs[45]. The validity of these predictions was further enhanced by comparison, based on multiple alignment and structural information, to published orthologous counterparts.

The boundaries of the ribosomal *rrnL* and *rrnS* genes were determined by comparison to the orthologous counterparts present in the mtDNAs of the Brachyura species already sequenced, as well as structural information implied by direct modelling (data not presented here).

Data set construction. All partial or complete mtDNAs published or publicly available, used in the present paper, were downloaded from GenBank and re-annotated following the approach described above to produce very high-quality annotations. This approach led us to correct the genes boundaries for several taxa. A more drastic change was done with the re-placement of the CoRe in the mtDNA of the potamid crab *Sinopotamon xiushuiense* (KU042041). According to the annotation provided in GenBank (unpublished, see Supplementary Table S1), the CoRe is located between *trnY* and *rrnL* genes. However, in the re-annotation process of *S. xiushuiense* mtDNA we discovered, inside of the intergenic spacer (1,221 bp long) located between *rrnS* and *trnI*, the unique signature AACTTATATTACCTA$(AT)_{27}$, which is shared by the CoRe of *G. dehani*, the other potamid crab of our data set. Thus, in the present study, this spacer is considered as the true CoRe of *S. xiushuiense* mtDNA. Two additional evidences support our choice. First, the CoRe is located between *rrnS* and *trnI* in most of the GOs observed in Brachyura (see below) and more in general in Arthropoda. Secondly, peculiar signatures similar to that presented here are known for other groups of Pancrustacea. For example, the ATAGA$(T)_n$ ($n > 10$) motif characterizes the vast majority of CoRes in the mtDNAs of Lepidoptera[50].

The availability of new mtDNA sequences in GenBank is a continuous evolving process, occurring at an unpredictable pace. More than fifty partial or complete mtDNAs of species belonging to the major lineages of Brachyura were available in GenBank at the first September 2016. For 50 mtDNAs (partial or complete) it was possible to unambiguously determine their complete GOs, and use them in the full array of analyses presented in this paper. Six outgroups belonging to the infraorders Anomura, Axiidea and Gebiidea (Decapoda, Crustacea) were added to the final set that contains 56 Taxa (T56) (Supplementary Table S1). The mtDNAs of *Huananpotamon lichuanense*, and *Sesarma neglectum* became available too late in GenBank to be fully considered (Supplementary Table S1). Therefore, they were included only in some analyses.

Multiple alignments of orthologous genes and proteins. Initially each set of the 13 orthologous protein-coding genes, derived from the 56 mtDNAs (Supplementary Table S1), was aligned using the ClustalW program implemented in the MEGA 5.2.2. program[51]. Each alignment was performed with the option "Codons" activated, which ensures that the alignment of DNA sequences is obtained using as backbone the multiple alignment derived from the amino acid counterparts. Following recent findings provided by Tan *et al.*[52] we did not filter alignments to select blocks of conserved positions, because this process can produce incorrect, statistically supported, trees[52]. Successively, the 13 alignments were concatenated in two data sets (56 T.DNA and its translated counterpart 56 T.PRO) that were used in the phylogenomic analysis. The 56 T.DNA and 56 T.PRO sets spanned respectively 11,208 and 3,736 positions.

Statistics of DNA/amino acid sequences. The AT-skew = (A − T)/(A + T) and the GC-skew = (G − C)/(G + C) were computed for the α strand of the Brachyuran mtDNAs in order to evaluate the compositional biases[53]. The base compositions were determined with the EditSeq program from the Lasergene software package (DNAStar, Madison, WI).

The total number of codons present in the mitochondrial protein-coding genes was calculated with the MEGA program. Stop codons were excluded from the calculation, because they are not linked to a tRNA family. Analogously, the start codons were omitted, because different codons determine the same formyl-Met as starting amino acid[47,48]. The abundance of each codon family was expressed as number of codons per thousand codons (CDSpT). The skews computations as well as other statistical calculations were performed with the spreadsheet Microsoft Excel (Microsoft™).

Mitochondrial phylogenomics of Brachyura. A preliminary analysis on the phylogenetic information present in 56 T.DNA and 56 T.PRO sets was performed according to the likelihood mapping approach[54] implemented in the IQ-TREE 1.5.2 program[55]. This analysis revealed that the maximum phylogenetic information was present in the 56 T.PRO set (data not shown). Thus, this set was used in the tree searches described below.

Phylogenetic analyses were performed according to the maximum likelihood (ML) method on the 56 T.PRO data set[56]. The ML trees were computed with the program IQ-TREE 1.5.2[55]. In the tree search analysis 100 independent runs were performed in order to avoid/minimize the possibility to be entrapped in sub-optimal trees. The optimal partitioning scheme as well as best fitting evolutionary models were selected with the program IQ-TREE 1.5.2[55,57]. The best partitioning/evolutionary models were the following: partition 1 (COX1, COX2, COX3, ATP6, CYTB), model mtZOA + I + G4[58]; partition 2 (ATP8, NAD2, NAD3, NAD6), model mtMAM + I + G4[59]; partition 3 (NAD1, NAD4, NAD4L, NAD5), model mtZOA + F + I + G4[58]. In order to minimize the possibility of long-branch attraction phenomena, 56 T.PRO data set was analysed also according to the empirical profile mixture models (C10-C60)[60] implemented in the IQ-TREE program. The C10-C60 models are the maximum likelihood counterparts of the CAT model developed for Bayesian analysis[61]. The C10-C60 models were applied alone or in combination with the mtZOA[58] and mtMAM[59] substitution matrices. All these analyses provided topologies fully congruent with that obtained from the gamma-based models listed above. However, the C10-C60 approaches required much higher computational times than the gamma-based analyses, making unfeasible to use them in the calculation of bootstrap values. Thus, the gamma-based approach was applied to complete the phylogenetic analyses.

The ultrafast bootstrap test (UFBoot) was performed to assess the robustness of ML tree topology (10,000 replicates)[62]. Alternative topologies were evaluated using the Weighted Shimodaira and Hasegawa and the Almost Unbiased tests[55,63].

Gene order analysis of Brachyuran mitochondrial genomes. *The pairwise approach using the CREx program.* Pairwise-comparisons between different GOs were performed with the CREx program[16]. This software analyses genomic rearrangement pathways using common intervals[16,17,64]. A common interval is a subset of genes that appear consecutively in two (or more) GOs being investigated[16].

The CREx program models rearrangements involving transpositions, inversions, inverse transpositions as well as TDRLs[13–17,64]. CREx produces transformational pathways in which the common intervals, shared by the pairs of GOs, are preserved in all intermediate steps. Once the whole set of common intervals has been determined for a pair of GOs (*e.g.*, GO1 and GO2) CREx heuristically identifies the most parsimonious transformational pathways that connect GO1 to GO2 and vice versa. For the reader interested on this topic, a detailed description on the functioning of CREx is provided in an open access paper recently authored by our group[18].

The number of shared common intervals (NSCI) is a measure that can be used to compare the level of similarity of two GOs. Identical GOs share the maximum NSCI value while highly divergent GOs have low NSCIs. Pairwise NSCI-based similarity values were calculated for the Brachyuran GOs. Given the fundamental role played by the control region CR, this latter was considered also in the computation of NSCI values (see results). In the CREx analyses, the software was allowed to compute up to ten alternative scenarios (option max. alternatives = 10) in every search[65]. The output for the different GO reconstructions was always a single transformational pathway. However, the current version of CREx program, which has a heuristic strategy of search, does not explore all possible alternatives, due to an overwhelming computational complexity that would be required for performing this type of analysis. Thus, CREx preferentially provides a single unique transformational scenario, and computes alternative scenarios only in specific cases. Therefore, the transformational pathway reconstructed by CREx is not the only possible and not necessarily the most parsimonious.

Current knowledge on the molecular mechanisms generating the GO rearrangements is very limited and largely insufficient. Thus, it is necessary to rely on mathematical models, implemented in bioinformatic programs, to identify the more probable transformational pathways generating the GOs. Currently, the CREx program is the most flexible and sophisticated software, available to perform this task. The combinatorial mathematics which is used by CREx is rapidly evolving and a natural lag exists between the formulation of new algorithms and their implementation in the software[66]. What is emerging is that, when the reconstructed pathway implies multiple TDRLs, there is not always the certainty that it is the only plausible scenario[66]. The presence of intergenic spacers, located in the genomic positions involved in TDRLs, is regarded as a first, even if weak, independent evidence supporting the most complex pathways[66]. A more conclusive evidence is supposed to be the presence of remnants of the copies of the genes located in these spacers, which were lost in genomic rearrangements, especially TDRLs[66,67].

Intergenic spacers, not linked to rearrangements, are common in animal mtDNAs and exhibit a random genomic distribution. A DNA slippage, during the genome replication, is supposed to be the most common mechanism generating these genomic elements[50]. The spacers produced by DNA slippage have usually, but not

always, a small size (20 bases ≤). The spacers linked to genomic rearrangements are very variable in size, but often they span from some tens to several hundreds of bases (*e.g.* Supplementary Fig. S1). Thus in the most favourable situation, it is possible to identify within these spacers the remnants of extra genes copies[67]. Unfortunately, this expectation is often highly diminished by the fact that the size of the spacers, even the largest ones, is much smaller than that of the initial genomic portions involved in the TDRLs. This empirical evidence implies that, once generated, the spacers are subject to a very rapid shrinking. Even worse, local phenomena of slippage and/or a fast substitution rate can further modify these spacers. Additionally, some rearrangements may have occurred very far in the past, leaving small or no spacers at all. Finally, if the reconstructed evolutionary scenario implies multiple TDRLs, the probability to find large size spacers linked to the earlier events should be low. Thus, identify the remnants of lost genes can be a daunting task, impossible to obtain even if very desirable. Conversely, the co-occurrence of multiple intergenic spacers, with genomic positions congruent with the inferred rearrangement pathway, should generate a distributional pattern difficult to explain in terms of pure chance. If this hypothesis holds, the spacers distributional pattern, easily identifiable, becomes a reasonable support (even if not conclusive) to the transformational pathway inferred by CREx.

The occurrence of intergenic spacers associated to rearrangements was checked for every GO to corroborate the obtained evolutionary scenarios identified by CREx. The presence of the remnants of the genes copies located in the intergenic spacers was tested by pair-wise alignments performed with ClustalW[46].

The phylogenetic approach using the TreeREx program. When multiple and highly variable GOs are analysed, it is necessary to apply the phylogenetic approach, implemented in the program TreeREx for inferring the evolutionary pathways leading to the observed diversity of GOs[64]. A fully bifurcating rooted reference tree is necessary. On this tree the pairwise scenarios computed by CREx are mapped along the branches using TreeREx software that can also infer the putative GOs at the internal nodes. Every node is successively labelled, according to a reliability scale implemented in TreeREx, as (a) consistent node, (b) 1-consistent node, and (c) fallback node. In the TreeREx analysis, the consistent nodes are considered to be the most reliable, the 1-consistent nodes exhibit an intermediate level of certainty, and the fallback nodes have the highest level of uncertainty for what concerns the reconstructed GO. More details on the functioning of TreeREx are provided by[18].

The TreeREx analysis was performed with default settings, as suggested at the website: -s, *i.e.* strong consistency method applied; -w, *i.e.* weak consistency method applied; -W, *i.e.* parsimonious weak consistency method applied; -o, *i.e.* get alternative bp scenario for prime nodes; -m = 0, *i.e.* maximum number of inversions + TDRL scenarios considered (http://pacosy.informatik.uni-leipzig.de/185-0-TreeREx.html)[64]. The settings above represent a global strategy to search for alternative rearrangements scenarios. In doing so, every node of the reference phylogenetic tree was defined by a GO, regardless of the certainty level for that node.

Both CREx and TreeRex require that analysed GOs include an identical set of genes. Thus if a gene is lacking in a GO it must be removed also from other genomic arrangements. Similarly, if multiple copies of the same gene are present, only one is retained in GO comparisons.

Tracking the changes in gene orders. It is well established that genomic rearrangements are modelled through transformational pathways that minimize the number of genes involved in active movements[18]. In describing the above rearrangement leading to BraGO, we attributed it to the transposition of *trnH* between *trnE* and *trnF* (Fig. 1). We did not consider the alternative scenario implying the repositioning of the *trnF* + *nad5* block between *trnH* and *nad4*, because this alternative hypothesis, involving multiple genes, is a less parsimonious explanation for the appearance of BraGO. Of course, the *trnF* + *nad5* block changed its placement downstream (right) to its original position in PanGO, but this movement is viewed as a passive shift, determined by the upstream (left) transposition of *trnH*.

When a transposition occurs between two adjacent genes, the ambiguity on who is the gene moving is resolved by assuming that the repositioning upstream (left) to the original placement is the transposition event, while the shift downstream (to the right) is the passive effect. This is an arbitrary choice because also the alternative scenario is equally parsimonious. The genes, that changed their placement through an obvious transposition event, are figured with a yellow background in this paper whereas the passively-shifted genes are identified with their original background.

However, it is not always possible to determine if the positions exhibited by the genes after a rearrangement are due to active or passive movements. This situation occurs when several genes are involved in complex patterns of movements, which are modelled by single/multiple TDRLs covering large part of the genome. In these cases, the placement of a gene, upstream or downstream to its position in the reference GO, can be the effect of the active re-positioning of the gene itself or a passive-shift due to the movements of the surrounding genes, or a combination of both. Thus, we label with a light-blue background all the genes involved in a repositioning that cannot be identified unambiguously as the result of an active or passive movement. Within this framework, the common intervals, encompassing two or more genes, shared by the re-arranged genomic portion and the reference GOs, are marked with an upper light-blue bar.

Results and Discussion

The mtDNAs of *Maja crispata* and *Maja squinado*. In this study, the complete mtDNAs of the spider crabs *M. crispata* and *M. squinado* were sequenced and annotated. The final assembly of *M. crispata* was 16,592 bp long and contained 24,734 reads. Additional statistics for the *M. crispata* mtDNA were: base coverage = 100%; mismatch = 0%; average coverage depth = 147.25; maximum coverage depth = 409. The final assembly of *M. squinado* was 16,598 bp long and contained 20,432 reads. The other statistics for the *M. squinado* assembly were: base coverage = 100%; mismatch = 0%; average coverage depth = 121.50; and the maximum coverage depth was 227. The mtDNAs of *M. crispata* and *M. squinado* contain the full set of 37 genes found in metazoan mtDNAs

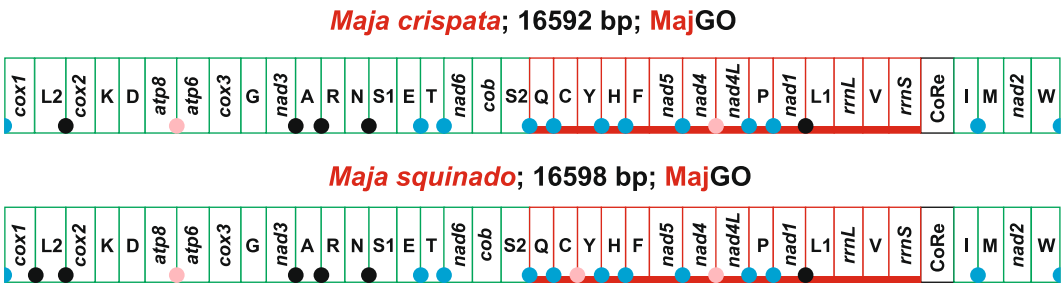


Figure 2. The mitochondrial genomes of *Maja crispata* and *Maja squinado*. The MajGO gene order is depicted and linearized starting from *cox1*. Graphical representation of the mtDNAs and nomenclature of genes as in Fig. 1. A blue circle marks an intergenic spacer assumed to be linked to the genomic rearrangement. A black circle marks an intergenic spacer supposed to be the result of a DNA slippage, during the genome replication. A pink circle marks an overlap between adjacent genes. The α-strand genes are green-boxed, while those of β-strand are underlined and red-boxed.

(Fig. 2). Both mtDNAs present intergenic spacers of variable size (Fig. 2, Supplementary Fig. S1). The newly determined mtDNAs share the same gene order MajGO, which is different from any other animal GO so far sequenced (Fig. 2). In MajGO, all the genes located on the β-strand form a single block, placed between *trnE* and CoRe. For more details on *Maja* mtDNAs, the reader should refer to the Supplementary Figs S1–S6.

The phylogeny of Brachyura. The mitochondrial phylogenomics of Brachyura obtained from 56 T.PRO set is shown in Fig. 3. The majority (39 up to 50) of the nodes of the tree receive good statistical support (UFBoot values > 90%). Only three basal nodes of the Thoracotremata clade does not receive any statistical support (UFBoot values < 50%). The overall phylogenetic outputs are in agreement with those presented in Fig. 1. *Dynomene pilumnoides* (Dromioidea) and the Homolidae species (Homoloidea) cluster together. The Raninidae taxa (Raninoidea) are sister group of Eubrachyura. Within this clade, Heterotremata and Thoracotremata are monophyletic groups receiving very strong statistical support. Taking into account the different taxon sampling, the relationships of the species included within Heterotremata and Thoracotremata are in agreement with the most complete phylogeny of Brachyura, which is based on nuclear genes[5]. Likewise, the superfamilies Ocypodoidea and Grapsoidea do not form monophyletic groups, in perfect agreement with Tsang *et al.*[5]. The only point of strong disagreement is the placement of Potamidae, which is sister group of Thoracotremata in Fig. 3, while is nested within the Heterotremata in the nuclear genes-based phylogeny (Fig. 1). The alternative placement of Potamidae as sister group of Hetereotremata was tested (Fig. 3). However, this relationship is rejected (p-value < 0.01) by both Weighted Shimodaira and Hasegawa and Almost Unbiased tests[63]. The placement of Potamidae shown in Fig. 3 perfectly agrees with the results obtained by other authors working with mitochondrial sequences (DNA and/or proteins)[33,68]. The placement of primary freshwater crabs, Potamidae and other families, remains a contentious issue[5,6]. The sparse taxon sampling of mtDNAs does not allow to fully disentangle this problem and a broadening of the taxonomic coverage is a high priority task of future research. A wider taxon and gene sampling will help to ascertain if the discrepancies are due to the nature of the markers, to the different taxon coverage, to the inadequacy of the phylogeny reconstruction algorithms currently available, or a combination of these factors.

The TreeREx program requires a fully resolved tree to reconstruct the evolutionary pathways producing the Brachyuran GO diversity. Initially both the alternative topologies depicted in Fig. 3 were considered. Given that the global evolutionary scenarios do not change for the largest majority of the nodes, the results presented below are referred exclusively to the tree obtained from the analysis of 56 T.PRO set.

The evolution of mitochondrial gene order in Brachyura. *The global pattern.* Currently ten different mtDNA GO are known for Brachyura (Fig. 4, Supplementary Fig. S7). The most widespread is BraGO (Figs 1 and 4). Six of the other GOs are restricted to single species (*i.e.*, *Dynomene pilumnoides* DynGO; *Damithrax spinosissimus*, DamGO; *Geothelphusa dehaani*, GeoGO; *Huananpotamon lichuanense*, HuaGO; *Sinopotamon xiushuiense* SinGO; *Xenograpsus testudinatus* XenGO). MajGO is shared by both species of *Maja* sequenced within the present study. Finally, SesGO is shared by the crabs belonging to Sesarmidae, while MaVaGO is found in Macrophthalmidae and Varunidae (Fig. 4). The mapping of GOs shows that BraGO occurred at the onset of Brachyura clade. Taking into account that the oldest fossil crabs are known from early Jurassic[69], BraGO appeared 200 MYA, and since then it has remained unchanged for many Brachyuran taxa (Fig. 4). BraGO shares 1,258 out of 1,400 common intervals with PanGO (Fig. 4). The other Brachyuran GOs, which have evolved from BraGO, can be roughly divided in three groups: (a) very re-arranged GOs (*i.e.* MajGO, MaVaGO, and XenGO), which share 312 or less common intervals with BraGO; (b) medium re-arranged GOs (SinGO) (NSCI = 732); (c) low re-arranged GOs (DamGO, DynGO, HuaGO and SesGO), which share 1058 or more common intervals with BraGO.

In the low re-arranged GOs only some tRNAs have changed their placement. Conversely, in both medium and highly re-arranged GOs all types of genes have been involved in the movements. However, none gene was inverted (Fig. 4). The level of rearrangement in various GOs does not appear to be linked, at least in a clearly detectable pattern, to A + T and G + C contents, AT- and GC-skews, as well as codon usage (Supplementary Figs S2–S4). The transformational pathways producing the Brachyuran GOs diversity are described in the following paragraphs.

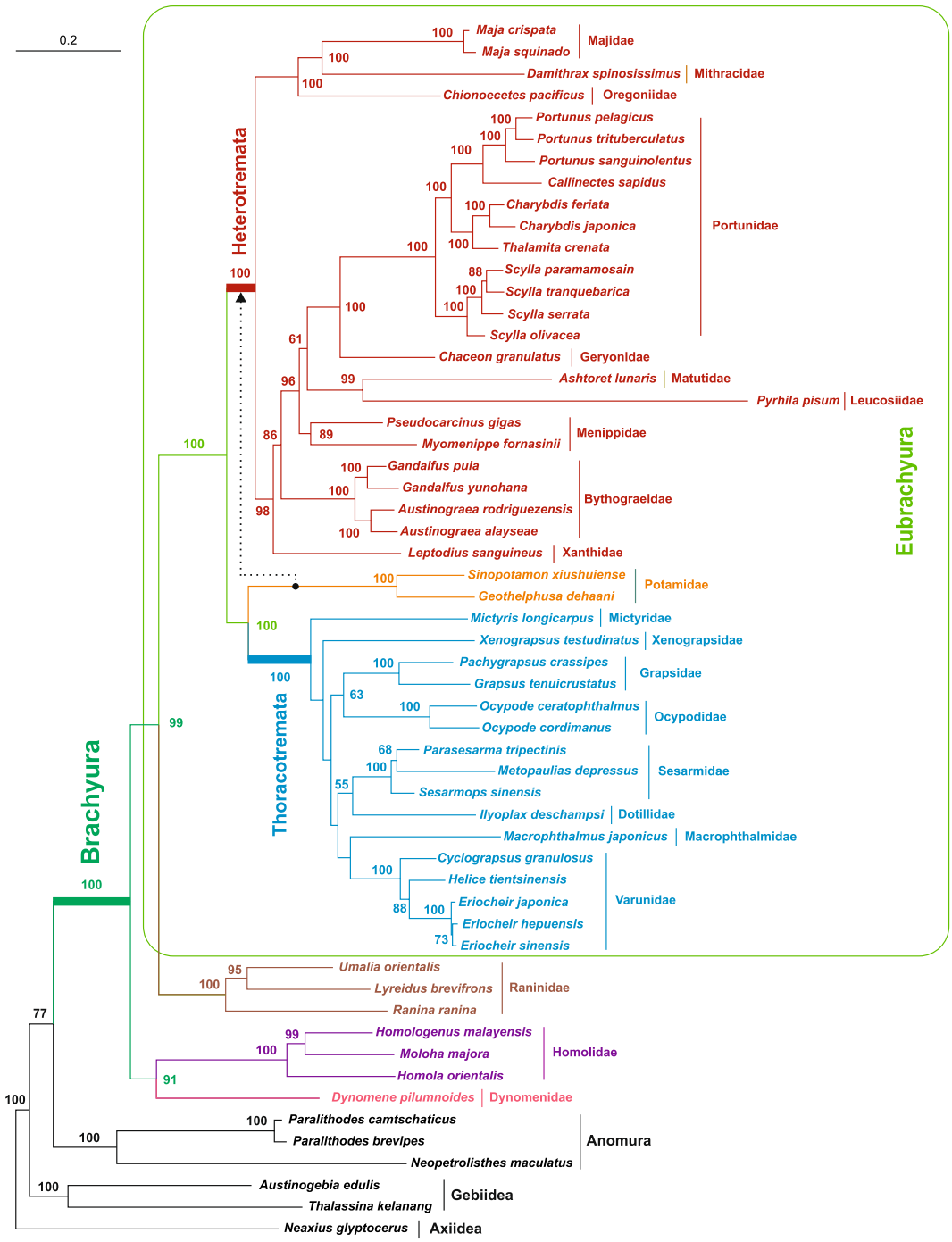
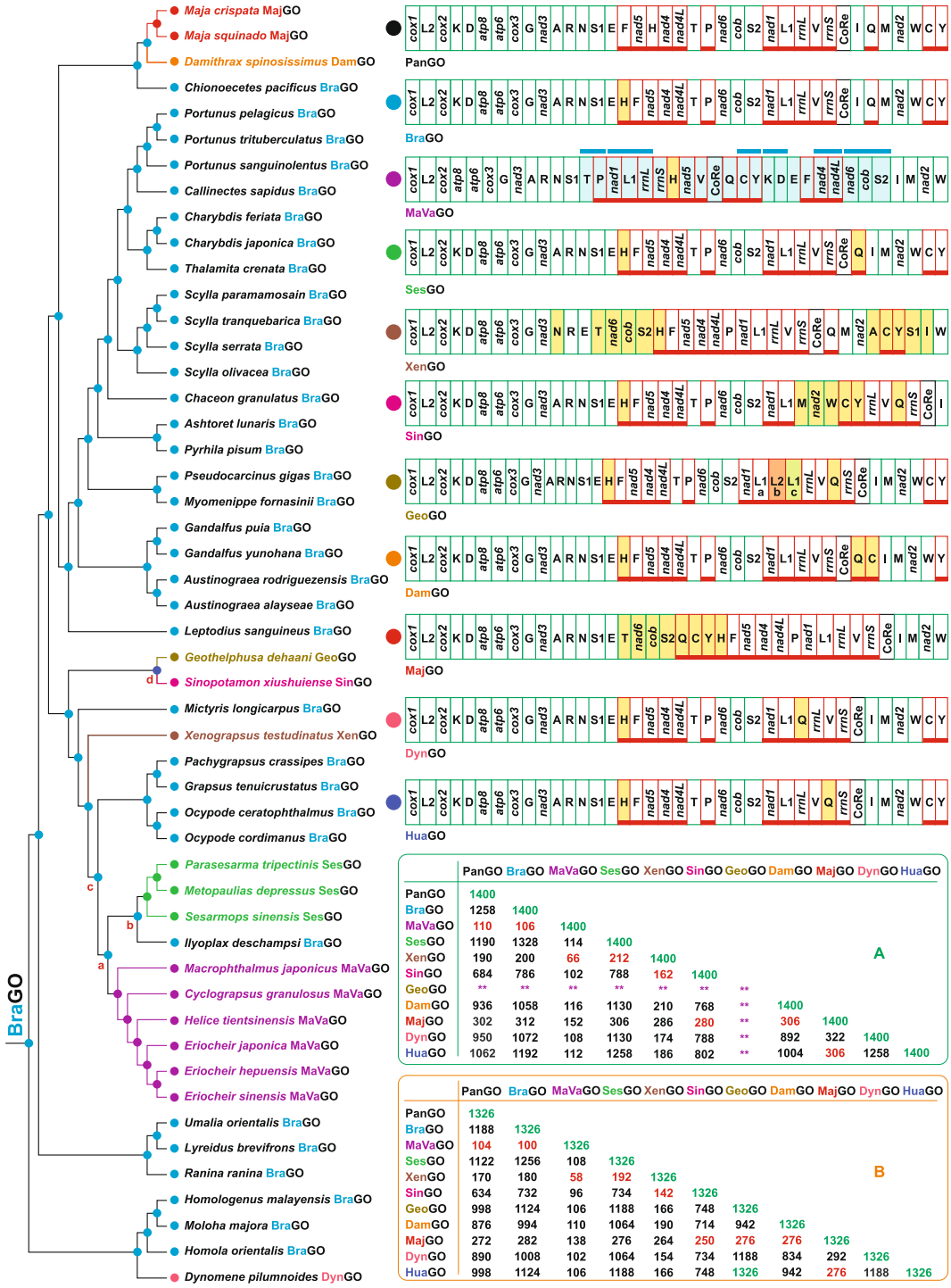


Figure 3. The mitochondrial phylogenomic tree of Brachyura. ML tree (-ln = 112327.8522) obtained from the analysis performed on the 56 T.PRO multiple alignment. Ultrafast bootstrap values ≥ 50% are provided for each node. The alternative phylogenetic placement of Potamidae (see main text) is depicted as a dotted line. The scale bar represents 0.2 substitution/site. Taxa names are coloured as in Fig. 1.

Initially, the reconstruction of GO evolution was inferred with the TreeREx program (Fig. 4, Supplementary Fig. S7). In particular, TreeREx assigned a GO identical to SesGO to nodes a-c, which exhibited lower consistency values (Supplementary Fig. S7). All species, except *Ilyoplax deschampsi*, deriving from nodes a-b exhibit a CoReQ local arrangement in their mtDNA (Fig. 4), determined by the transposition of *trnQ* immediately downstream to the control region, which is a hotspot of genomic rearrangements[18]. However, the CoReQ arrangement is shared also with DamGO reported in the mtDNA of the unrelated crab *D. spinosissimus* (Fig. 4). These findings imply that the transpositions of the mobile *trnQ* generate homoplastic rearrangements in crab GOs. Furthermore, Sesarmidae and Macrophthalmidae + Varunidae, which in Fig. 4 are closely linked, do not result so closely related when a broad taxon sampling exists[5]. Finally, the occurrence of SesGO at nodes a-c implies that two secondary



	PanGO	BraGO	MaVaGO	SesGO	XenGO	SinGO	GeoGO	DamGO	MajGO	DynGO	HuaGO
PanGO	1400										
BraGO	1258	1400									
MaVaGO	110	106	1400								
SesGO	1190	1328	114	1400							
XenGO	190	200	66	212	1400						
SinGO	684	786	102	788	162	1400					
GeoGO	**	**	**	**	**	**	**				
DamGO	936	1058	116	1130	210	768	**	1400			
MajGO	302	312	152	306	286	280	**	306	1400		
DynGO	950	1072	108	1130	174	788	**	892	322	1400	
HuaGO	1062	1192	112	1258	186	802	**	1004	306	1258	1400

A

	PanGO	BraGO	MaVaGO	SesGO	XenGO	SinGO	GeoGO	DamGO	MajGO	DynGO	HuaGO
PanGO	1326										
BraGO	1188	1326									
MaVaGO	104	100	1326								
SesGO	1122	1256	108	1326							
XenGO	170	180	58	192	1326						
SinGO	634	732	96	734	142	1326					
GeoGO	998	1124	106	1188	166	748	1326				
DamGO	876	994	110	1064	190	714	942	1326			
MajGO	272	282	138	276	264	250	276	276	1326		
DynGO	890	1008	102	1064	154	734	1188	834	292	1326	
HuaGO	998	1124	106	1188	166	748	1326	942	276	1188	1326

B

Figure 4. The evolution of mitochondrial gene orders in Brachyura. For most of the nodes the GO was inferred directly with the program TreeREx. For nodes a–d, the GO assignment was performed manually (see main text). L2b, derived from a tRNA remoulding process, not orthologous to true *trnL2s*. L1c, extra copy of *trnL1* not considered in the TreeREx analysis. Table A, NSCI values computed with CREx program through pairwise-comparisons of complete GOs. Table B, NSCI values computed through pairwise-comparisons of GO deprived of L2s. **NSCI values not computed (see main text). The genomic and genetic nomenclature, as well as the colour scheme, are the same as in Fig. 1. The genes that changed their position relative to PanGO, through a transposition event, are shown with a yellow background. The passively-shifted genes are figured with their original background. The genes involved in a repositioning, which cannot be identified unambiguously as the result of a transposition or a passive shift, are figured with a light blue background. In this latter case, the common intervals, encompassing two or more genes, shared by the re-arranged GO with PanGO, are highlighted with a light blue bar.

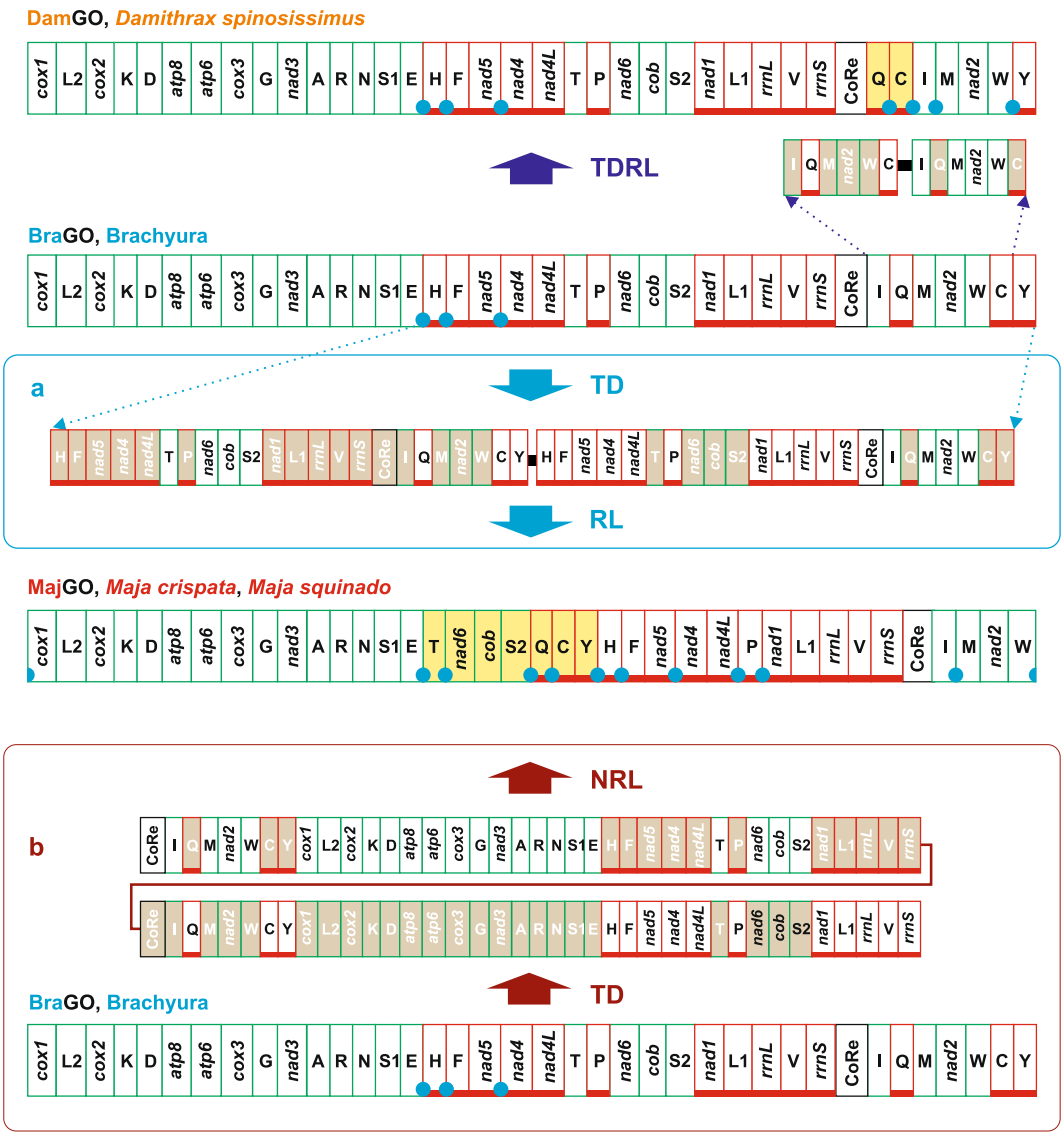


Figure 5. The evolutionary pathways generating DamGO and MajGO arrangements. The rearrangements in the GOs of *D. spinosissimus* and *Maja* species are investigated and depicted with respect to BraGO. TDRL, TD/RL, tandem duplication/random loss event. TD/NRL tandem duplication/non random loss event. The genomic and genetic nomenclature, as well as the colour scheme, are the same as in Fig. 1. The genes that changed their position relative to PanGO, through a transposition event, are shown with a yellow background. The passively-shifted genes are figured with their original background. A blue circle marks an intergenic spacer present in a position associated to genomic rearrangement.

independent reversals to the plesiomorphic condition represented by BraGO occurred in *I. deschampsi* mtDNA and in the common ancestor of Grapsidae + Ocypodidae taxa (Figs 3 and 4). Such reversions are extremely improbable events, provided that they happen. Indeed, to our knowledge, they have never been documented in the Bilaterian animals. To complete our reasoning, it must be added here that the accuracy of TreeREx reconstructions is influenced by the coverage of taxon sampling[64]. We regard the output provided by TreeREx program for nodes a-c as an implausible scenario, determined by the sparse taxon sampling coupled with the homoplastic CoReQ arrangement, and influenced by the fact that TreeREx works locally on small subtrees. Thus, in Fig. 4 we manually assign BraGO to nodes a-c. Finally, we manually attributed HuaGO to the node d (Fig. 4). The reasons leading to this assignment are described in details in the paragraph dealing with the evolution of GOs in Potamid crabs.

All analysed GOs present intergenic spacers located in the genomic positions congruent with the inferred transformational pathways (Figs 5–8)[66]. For most of the GOs, it was possible to identify some possible remnants of genes located within these intergenic spacers (see Methods) (Supplementary Fig. S8). However, in agreement with the expectation expressed in Methods section, the evidence that they are true remnants is often weak. The occurrence of these spacers supports, at minimum, the view that the positions harbouring them played a pivotal role in the transformation pathways, which generated the new GOs. However, provided that alternative

transformational routes may involve the same positions and that the current version of CREx reconstructs heuristically preferentially only one scenario, it is not possible to conclude that pathways depicted in Figs 5–8, particularly the most complex ones, represent the only explanation for describing the GO rearrangements. More realistically, each pathway must be regarded as one of the plausible evolutionary scenarios for the studied GO.

The evolutionary pathways generating DamGO and MajGO arrangements. The DamGO has been produced through a TDRL event that involved the genomic portion included between *trnI* and *trnC* (Fig. 5). In DamGO, *trnQ* and *trnC* are transposed with respect to BraGO. In the *Maja* species a single TDRL event is necessary to explain the final rearrangement. However, in this latter case the genomic portion involved in the process contains 22 genes plus the CoRe (Fig. 5a). Consequently, the MajGO is the third most re-arranged Brachyuran GO. It shares 312 out of 1,400 common intervals with BraGO (Fig. 4). To our knowledge, MajGO is unique among the animal GOs so far determined, not only for the global placement of the genes, but also for the arrangement in a single unbroken block of the β-strand genes. Alternatively, the MajGO arrangement could be the result of a tandem duplication non-random loss event (TDNRL) (Fig. 5b and Supplementary Fig. S9)[70,71]. In a TDNRL event, after the duplication of the complete mtDNA, the genes located only on one strand of each duplicated copy are lost. The final GO is strand-biased (further details in Supplementary Fig. S9). Both *Maja* mtDNAs exhibit intergenic spacers (Figs 1 and 5, and Supplementary Fig. S2) at all the positions involved in the TDRL event (Fig. 5a). The remnants of some of genes seem to occur within these spacers (Supplementary Fig. S8), except for *trnE-trnT*, which is very short. Both TDRL and TDNRL models implicitly/explicitly allow the presence of intergenic spacers associated to GO rearrangements. Thus, the occurrence of intergenic spacers does not allow to decide what model (TDRL or TDNRL) describes better the evolutionary pathway that generated the *Maja* GO.

The evolution of GOs in Potamid crabs. Currently, three full length mtDNAs are available for the Potamid crabs. Their GOs (GeoGO in *G. dehaani*; HuaGO in *H. lichuanense*; SinGO in *S. xiushuiense*; Supplementary Table S1) are distinct and different from BraGO (Figs 4 and 6). GeoGO and SinGO have been included in all our analyses. Due to its late availability, HuaGO has been considered only in the comparisons presented in Figs 4 and 6 and Supplementary Fig. S10. The transformational pathway producing SinGO implies two transpositions (T1-T2) (Fig. 6). T1 generates a GO that is identical to HuaGO of *H. lichuanense*. In this case, the CREx exhibits a good predictive capability, because it identified a GO that has a counterpart (HuaGO) in an existing crab. The T2 event produced the transposition of a block of contiguous genes leading to SinGO. BraGO and SinGO shares 786 common intervals (Fig. 4). A partial mtDNA is available in GenBank for the Potamid crab *Sinopotamon yangtsekiense*[29], which does not cover some tRNAs, included *trnQ*. In this mtDNA the block *nad2-trnY* retains the standard placement observed in BraGO (Fig. 6). However, all genes in the block present multiple frameshift (*nad2*) or mismatches, which severely jeopardize the secondary structure of tRNAs (data not shown). These findings support the hypothesis that this partial mtDNA sequence is not good. Thus new full-length mtDNAs of *S. yangtsekiense* and other species of *Sinopotamon* are necessary to study the GO evolution in these Potamid crabs.

A multi-steps strategy was necessary to define the changes leading to GeoGO, due to the peculiar condition exhibited by this genome (Figs 4 and 6, and Supplementary Fig. S10). As mentioned in the introduction, the *G. dehaani* mtDNA presents a tRNA remoulding, which involves the *trnL1* and *trnL2*. The true orthologous *trnL2* is lost. Furthermore, two copies of *trnL1* surrounding the functional L2 are found in GeoGO (Figs 4 and 6). The current version of TreeREx program is capable to analyse only GOs that contain an identical set of genes (see Methods). Thus, a first TreeREx search was performed, with a version of GeoGO obtained from the original one by removing *trnL1* (c in Fig. 6). In this analysis, *trnL2* (b) was treated as the true orthologous of the L2 genes present in other GOs (Figs 4 and 6). The TreeREx search allowed to identify the overall scenario presented in Fig. 4, except for nodes a-d. Successively, after removing *trnL2s* (orthologous or functional) from the GOs set, a second analysis was performed with TreeREX (Fig. 4). This reconstruction identified for the node d a GO that is identical to HuaGO deprived of L2 (HuaGO-L2) (Fig. 4). Furthermore, the GeoGO-reduced, a second version of GeoGO deprived of L2 (b) and L1 (c), exhibits an arrangement identical to HuaGO-L2. GeoGO-reduced shares 1,124 out of 1,326 common intervals with BraGO (Fig. 4). As shown above, the first step leading to SinGO was a transposition generating a GO identical to HuaGO (Fig. 6). *G. dehaani* and *S. xiushuiense* are members of the same phyletic lineage. Given the identical arrangement shared by GeoGO-reduced and HuaGO-L2, and the T1 event observed in the transformational pathway of SinGO, the most parsimonious scenario is to consider the transposition of *trnQ* as the first step shared by the evolutionary changes that produced GeoGO and SinGO (Fig. 6). Thus, combining all the findings presented above, we identified HuaGO as the common first event of the transformational pathways, that generated the GeoGO and SinGO (Fig. 6). HuaGO is considered also the more plausible GO reconstruction for the node d of Fig. 4.

In the evolution of GeoGO, two successive (D1-D2) duplications of *trnL1* followed the transposition T1. These multiplicative steps were the fundament prerequisite for the gene remoulding (GR) process, which generated two functional *trnL2*. The last event was the loss of the true *trnL2*. The deletion of the true *trnL2* could not predate other events. Indeed, *trnL2* is associated to the most numerous codon families not only in Brachyura (see Supplementary Fig. S5) but in animal taxa[50,67,72]. It is implausible that the true *trnL2* was lost before the remoulding process occurred.

In an alternative pathway, the D2 duplication could be placed after the remoulding event or after the deletion of the true *trnL2* (Supplementary Fig. S10). In this case a duplication of the tandem genes L1(a)-L2(b) followed by the loss of only the extra copy of L2 (b) is the more plausible scenario. Alternatively, a complicated and improbable event would be necessary, *i.e.* the duplication of *trnL2* (b) followed by a back-mutation process leading, through a reverse remoulding, to *trnL1* (c).

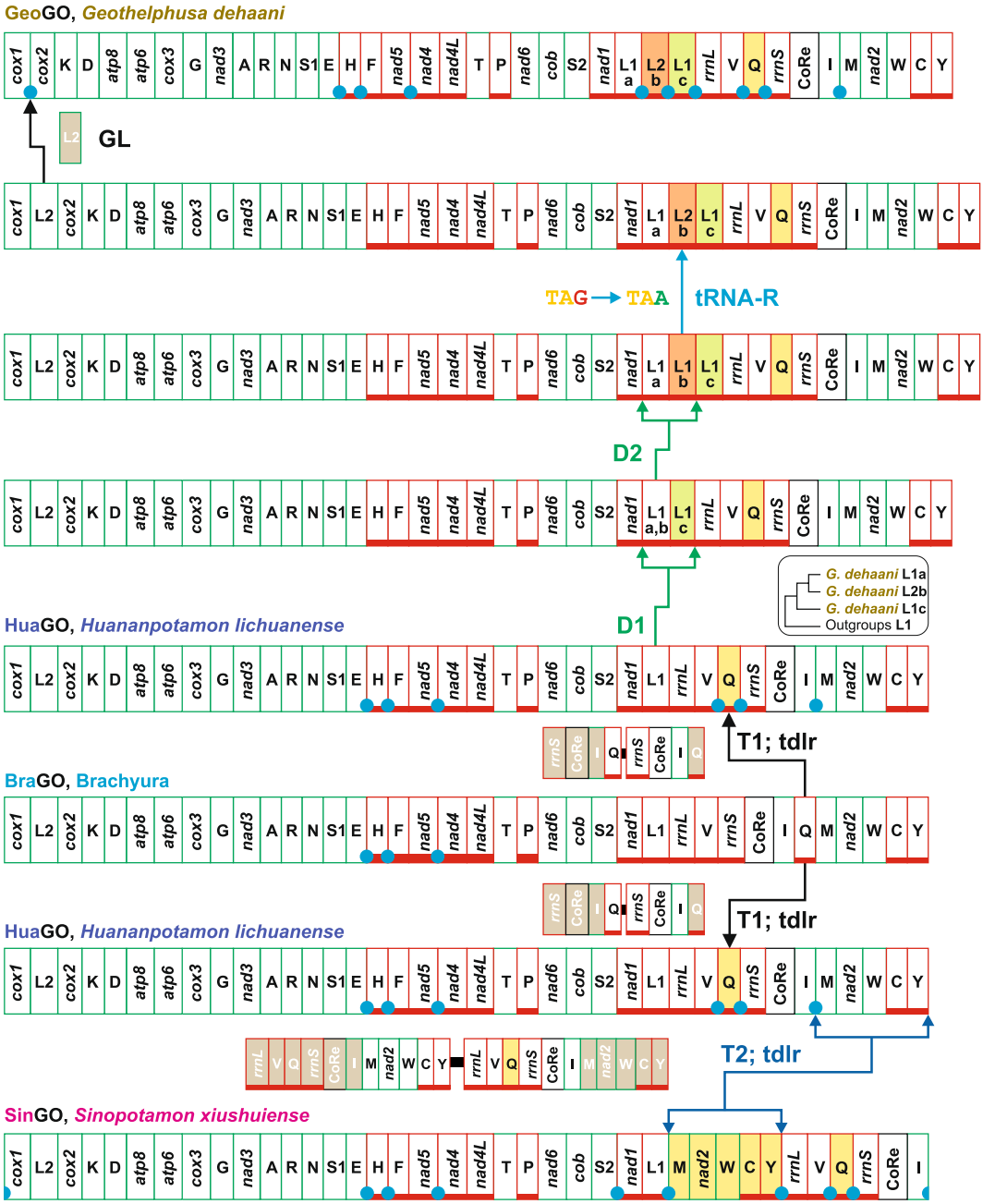


Figure 6. The evolution of GOs in Potamid crabs. Transformational pathways generating GeoGO, HuaGO and SinGO. The rearrangements in the GOs of Potamid species are investigated and depicted with respect to BraGO. T1-T2, transposition events; tdrl, tandem duplication random loss mechanism producing the observed rearrangement; D1-D2, gene duplication events; tRNA-R, tRNA remoulding event; GL, gene loss event. The genomic and genetic nomenclature, as well as the colour scheme, are the same as in Fig. 1. The genes that changed their position relative to PanGO, through a transposition event, are shown with a yellow background. The passively-shifted genes are figured with their original background. A blue circle marks an intergenic spacer present in a position associated to genomic rearrangement.

The alternative scenario described above, implies that L1 (a) and L1 (c) are sister sequences (Supplementary Fig. S10). However, the phylogenetic analysis[24] of Brachyuran *trnL1s* reveals that L1 (a) and L1 (c) are not sister sequences (Fig. 6, Supplementary Fig. S10). Conversely, the pathway presented in Fig. 6 is fully consistent with the available data, and represent in our view the most parsimonious and plausible explanation of the GeoGO evolution.

The evolution of GOs in Sesarmidae and Xenograpsus testudinatus. SesGO is the least re-arranged among crab GOs as proved by the highest number of shared common intervals (1,328 on 1,400) with BraGO (Fig. 4). The transposition of *trnQ* downstream to CoRe is the molecular signature characterizing all the Sesarmid mtDNAs

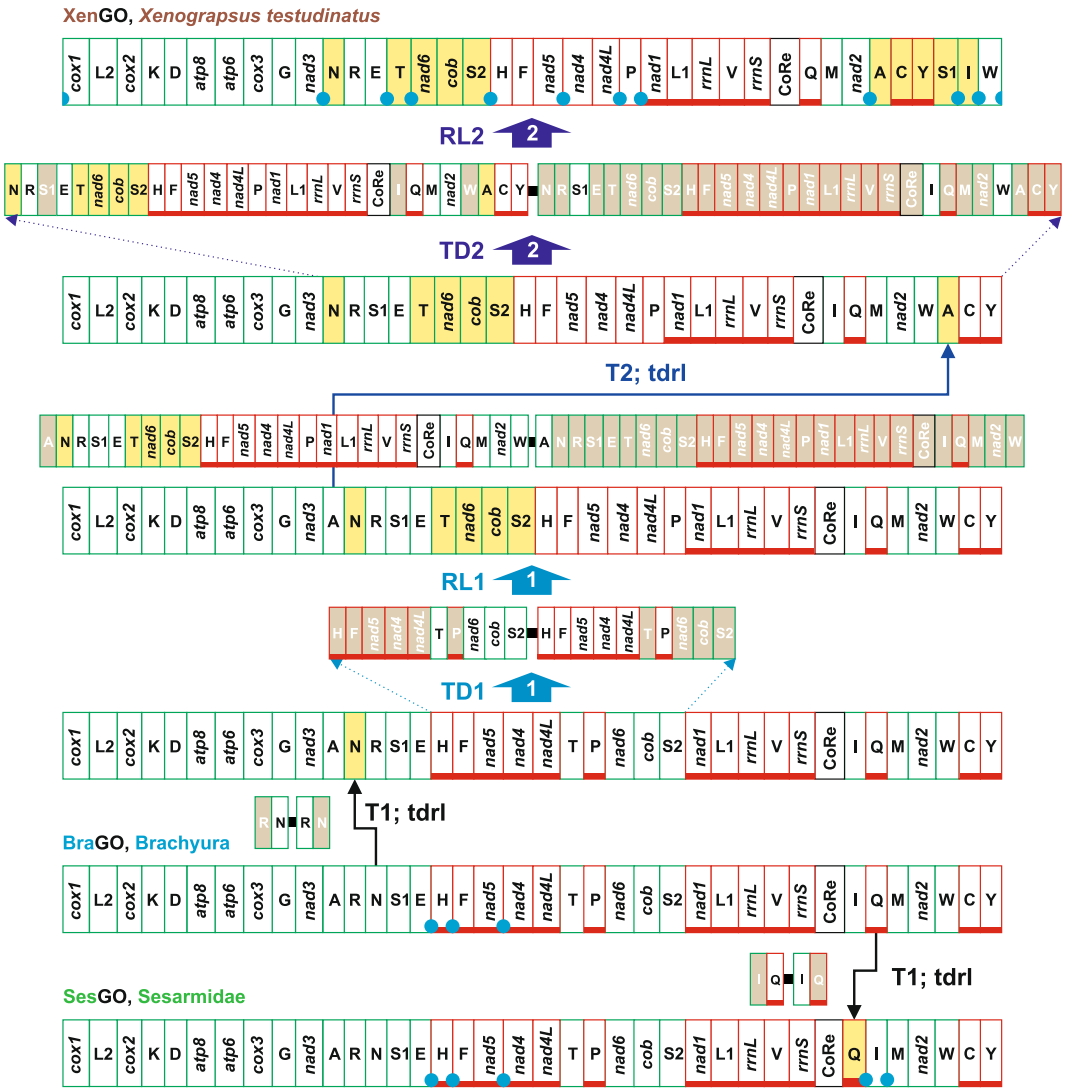


Figure 7. The evolution of GOs in Sesarmidae and *X. testudinatus*. The rearrangements in the GOs of Sesarmidae and *X. testudinatus* are investigated and depicted with respect to BraGO. T1-T2, transpositions event; tdrl, tandem duplication random loss mechanism producing the observed rearrangement. TD/RL1-2, tandem duplication/random loss events. The genomic and genetic nomenclature, as well as the colour scheme, are the same as in Fig. 1. The genes that changed their position relative to PanGO, through a transposition event, are shown with a yellow background. The passively-shifted genes are figured with their original background. A blue circle marks an intergenic spacer present in a position associated to genomic rearrangement.

(Fig. 7). In fact, in addition to the taxa analysed here, SesGO has been described very recently[73] also for *Sesarma neglectum*, a species whose sequence became available in GenBank too late (2016-10-31) to be analysed here. At opposite, XenGO, is a very modified GO, ranking second among the most re-arranged GOs (Figs 4 and 7). XenGO derived from BraGO through a complex pathway implying two transpositions (*trnN* and *trnA*) and two TDRL events (Fig. 7), computed as the most parsimonious scenario by CREx. These latter involved large blocks of mtDNA in the process of genomic rearrangement (Fig. 7). Intergenic spacers, occurring in several points involved in TDRLs, and possible remnants of genes further favour the XenGO pathway (Fig. 7, Supplementary Fig. S8).

The evolutionary pathways producing the DynGO and MaVaGO arrangements. We include in this paragraph the evolutionary reconstructions relative to GOs obtained from distantly related crabs. The choice is dictated simply by practical reasons of presentation of our analysis.

The transformational processes determining the appearance of DynGO are reported in Fig. 8. The transposition of *trnQ* downstream to *trnL1* generated DynGO, which shares 1,008 common intervals with BraGO (Fig. 4). Conversely, MaVaGO is the result of a complex mechanism of rearrangements. Two successive TDRL events involving all 37 genes plus CoRe, with the exclusion of *cox1*, *trnL2* and *cox2*, generated MaVaGO, which characterizes the Macrophthalmidae + Varunidae mtDNAs. The high level of rearrangement is corroborated by the number of common intervals (106) shared by MaVaGO and BraGO. This value is the smallest among the known

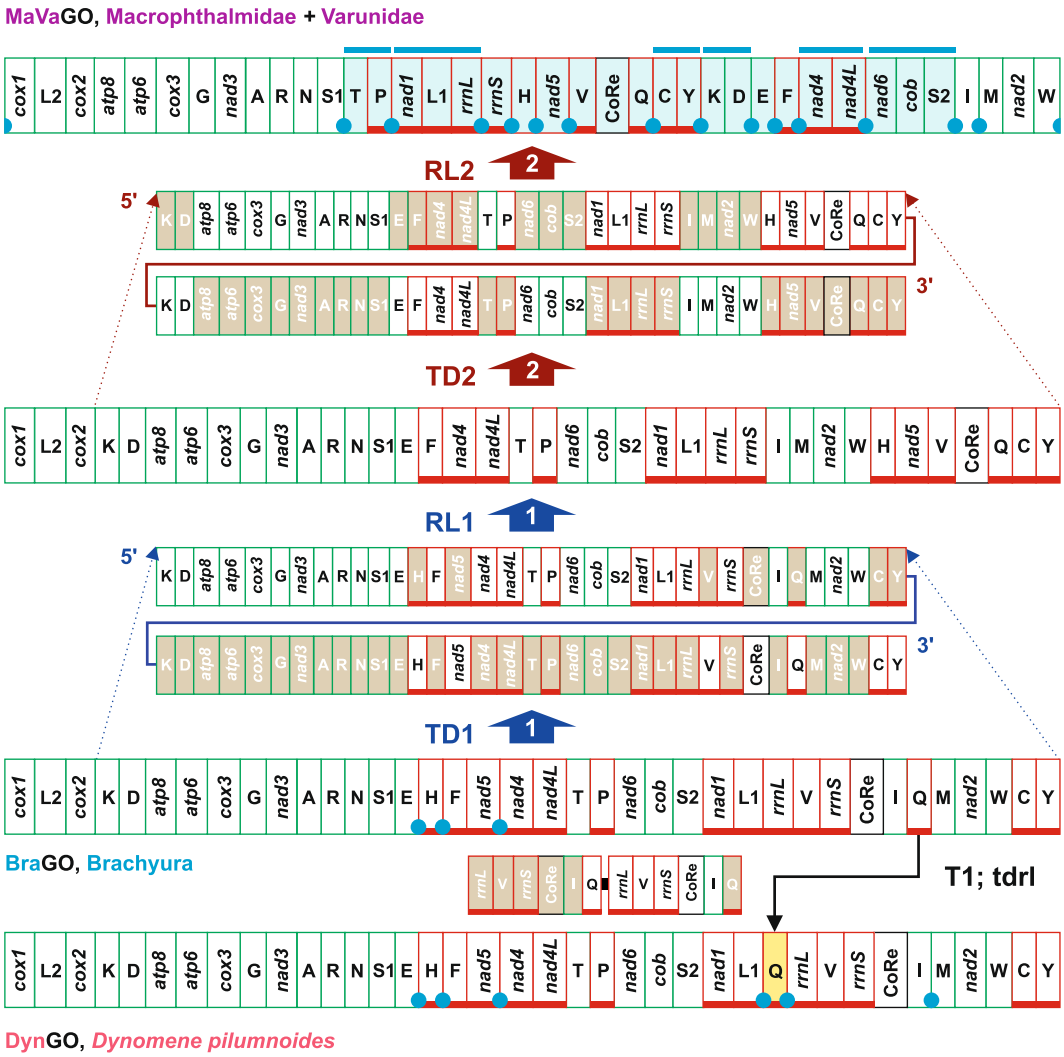


Figure 8. The evolutionary pathways producing the DynGO and MaVaGO arrangements. The rearrangements in the GOs of *D. pilumnoides* and Macrophthalmidae + Varunidae are investigated and depicted with respect to BraGO. T1, transposition events; tdrl, tandem duplication random loss mechanism producing the observed rearrangement. TD/RL1-2, tandem duplication/random loss events. The genomic and genetic nomenclature, as well as the colour scheme, are the same as in Fig. 1. The genes that changed their position relative to PanGO, through a transposition event, are shown with a yellow background. The passively-shifted genes are figured with their original background. The genes involved in a repositioning, which cannot be identified unambiguously as the result of a transposition or a passive shift, are figured with a light blue background. In this latter case, the common intervals, encompassing two or more genes, shared by MaVaGO with PanGO, are highlighted with a blue bar. A blue circle marks an intergenic spacer present in a position associated to genomic rearrangement.

crab GOs (Fig. 4). Notably, the transformational pathway going from BraGO to MaVaGO, computed as the most parsimonious scenario by CREx, implies only two TDRL events, without an early transposition of *trnQ*, an event hypothesised in the TreeREx reconstruction (see above). The CREx reconstruction further corroborates the manual assignment of BraGO to nodes a-c of Fig. 4. Intergenic spacers, occurring in positions involved in TDLRs, and possible remnants of genes support the MaVaGO scenario (Fig. 8, Supplementary Fig. S8).

The Brachyuran GOs as molecular signatures. Even restricting the analysis of GOs to the placement of the genes in a linearized genome and ignoring their orientation, the 37 standard animal mitochondrial genes can be arranged in an astonishing number of GOs (*i.e.*, $37! = 1.367 \times 10^{43}$ or 38! if the CoRe is also included), provided that the movement of every gene is equally probable[74]. However, an always increasing amount of evidence shows that the equally probable movement scenario is totally unrealistic and that some genes are much more mobile than others[18]. In reality, the gene movements in mtDNAs occur preferentially along specific pathways. This characteristic reduces the number of possible arrangements that are likely to be observed, and drastically increases the probability of convergent evolution in GOs. Convergence can be limited to the sharing of local homoplastic rearrangements or involve the full rearrangement of a GO[18]. The tRNAs are the most mobile genes[13,18,74].

Furthermore, the genes most prone to homoplastic rearrangements are contiguous in the genome or located around the origin of replication of the mtDNA[18,75,76].

The transposition of *trnH* upstream to *trnF* was the first event of Brachyuran mtDNA. It characterizes still most of the GOs except MaVaGO, where successive rearrangements have disrupted this pairing. *TrnH* is not close to CoRe in PanGO (Figs 1 and 4). Furthermore, the transposition of *trnH* in BraGO, can be described as a long range movement, because it implies the repositioning upstream to the largest coding gene of mtDNA, *i.e. nad5* (Fig. 1). BraGO is known only for true crabs and our expectation is that this GO represents a strong molecular signature for the Brachyuran clade, in agreement with earlier suggestions[34].

The analysis of the other Brachyuran GOs reveals that with the exception of the highly re-arranged MajGO, MaVaGO, and XenGO, and the peculiar situation of GeoGO, the remaining GOs have been generated by the transpositions of *trnQ* and *trnC* (Figs 4–8). This hypermobility has generated local homoplastic arrangements. The most widespread is CoReQ, which is shared by DamGO, MaVAGO, SesGO, and XenGO, even if the mechanisms generating this arrangement are different (Figs 5–8). Similarly, a CoReQC homoplastic arrangement occurs in DamGO and MaVaGO (Figs 4–5 and 8).

These findings show that in Brachyura only complete GOs must be considered as molecular signatures, a result mirroring a general behaviour of animal GOs[18].

MajGO, MaVaGO, and XenGO exhibit high level of rearrangements involving multiple genes (Figs 4–8). The probability of multiple independent appearances of these GOs seems very low. MajGO is shared by *M. crispata* and *M. squinado*, that are sister species[77]. Thus, our expectation is that MajGO is a synapomorphy characterizing at minimum this subclade of *Maja*, and possibly the whole genus. The range of occurrence of MajGO requires further sequencing efforts. Also, MaVaGO is expected to be a true synapomorphy defining a clade containing all crabs sharing this GO. Finally, XenGO is an apomorphy currently known only for *X. testudinatus*.

DynGO, DamGO, HuaGO and SinGO are molecular signatures for taxa possessing them. The low level of rearrangement coupled with the type of genes transposed suggest caution in assigning them the status of mitochondrial genomic apomorphies. The invitation to cautiousness is supported by the increasing evidence that same homoplastic GO can be shared by unrelated animal taxa as recently demonstrated for butterflies, some ants and crickets[18,78]. Even in these cases the homoplastic GOs represent molecular signatures. However, they cannot be used alone as diagnostic feature exclusively characterizing the taxa possessing them. The homoplastic GOs must be evaluated in a phylogenetic context[18].

The present study, which is the first to be conducted coupling sophisticated bioinformatic tools with phylogenetic analysis, confirms that the Brachyura are a hot-spot of GOs diversity among Arthropoda. Currently a full length mtDNA is available for less than 1% (52 out of 7250) of the crab species[2] (Supplementary Table S1). The high number of crab GOs (10) so far determined lead us to suggest that new GOs will be discovered with the increase of the taxon coverage.

References

1. De Grave, S. *et al*. A classification of living and fossil genera of decapod Crustaceans. *Raffles Bull. Zool.* **21**(Suppl), 1–109 (2009).
2. Davie, P. J. F., Guinot, D. & Ng, P. K. L. Systematics and classification of Brachyura. In *The Crustacea*, *complementary to the Traité de Zoologie*, *Volume 9* Part C-*II* (eds Castro, P., Davie, P. J. F., Guinot, D., Schram, F. R. & von Vaupe Klei, J. C.) 1049–1130 (Koninklijke NV, Brill, 2015).
3. Castro, P., Davie, P. J. F., Guinot, D., Schram, F. R. & von Vaupe Klei, J. C. (eds) *The Crustacea*, *complementary to the Traité de Zoologie*, *Volume 9 Part C-I and Part C-II*, 1–1221 (Koninklijke NV Brill, 2015).
4. Carpenter, K. E. & Niem, V. H. (eds) FAO species identification guide for fishery purposes. The living marine resources of the Western Central Pacific. Volume 2. Cephalopods, crustaceans, holothurians and sharks. 687–1396 (FAO, 1998).
5. Tsang, L. M. *et al*. Evolutionary history of true crabs (Crustacea: Decapoda: Brachyura) and the origin of freshwater crabs. *Mol. Biol. Evol.* **31**, 1173–1187 (2014).
6. Davie, P. J. F., Guinot, D. & Ng, P. K. L. Phylogeny of Brachyura in *The Crustacea*, *complementary to the Traité de Zoologie*, *Volume 9* Part C-*II* (eds Castro, P., Davie, P. J. F., Guinot, D., Schram, F. R. & von Vaupe Klei, J. C.) 921–979 (Koninklijke NV Brill, 2015).
7. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res* **27**, 1767–1780 (1999).
8. Raimond, R. *et al*. Organization of the large mitochondrial genome in the isopod *Armadillidium vulgare*. *Genetics* **151**, 203–210 (1999).
9. Marcadé, I. *et al*. Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda, Crustacea). *J. Mol. Evol.* **65**, 651–659 (2007).
10. Kilpert, F. & Podsiadlowski, L. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. *BMC Genomics* **7**, 241, doi:10.1186/1471-2164-7-241 (2006).
11. Tang, G. *et al*. Complete mitochondrial genome of *Rhynchocinetes durbanensis* (Rhynchocinetidae: *Rhynchocinetes*). *Mitochondrial DNA Part B* **1**, 245–246 (2016).
12. Saito, S., Tamura, K. & Aotsuka, T. Replication origin of mitochondrial DNA in Insects. *Genetics* **171**, 1695–1705 (2005).
13. Moritz, C., Dowling, T. E. & Brown, W. M. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* **18**, 269–292 (1987).
14. Boore, J. L. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In *Comparative genomics* (eds Sankoff, D. & Nadeau, J. H.), 133–147 (Kluwer Academic Publishers, 2000).
15. Dowton, M. & Campbell, N. J. H. Intramitochondrial recombination – is it why some mitochondrial genes sleep around? *Trends Ecol. Evol* **16**, 269–271 (2001).
16. Bernt, M. *et al*. CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics* **23**, 2957–2958 (2007).
17. Bernt, M. & Middendorf, M. A method for computing an inventory of metazoan mitochondrial gene order rearrangements. *BMC Bioinformatics* **12**(Suppl 9), S6, doi:10.1186/1471-2105-12-S9-S6 (2011).
18. Babbucci, M., Basso, A., Patarnello, T. & Negrisolo, E. Is it an ant or a butterfly? Convergent evolution in the mitochondrial gene order of Hymenoptera and Lepidoptera. *Genome Biol. Evol* **6**, 326–3343, doi:10.1093/gbe/evu265 (2014).
19. Boore, J. L., Lavrov, D. V. & Brown, W. M. Gene translocation links insects and crustaceans. *Nature* **392**, 667–668 (1998).
20. Segawa, R. D. & Aotsuka, T. The mitochondrial genome of the Japanese freshwater crab, *Geothelphusa dehaani* (Crustacea: Brachyura): evidence for its evolution via gene duplication. *Gene* **355**, 28–39 (2005).
21. Cantatore, P., Gadaleta, M. N., Roberti, M., Saccone, C. & Wilson, A. C. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. *Nature* **329**, 853–855 (1987).

22. Rawlings, T. A., Collins, T. M. & Bieler, R. Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. *Proc. Natl. Acad. Sci. USA* **100**, 15700–15705 (2003).
23. Wang, X. & Lavrov, D. V. Gene recruitment – a common mechanism in the evolution of transfer RNA gene families. *Gene* **475**, 22–29 (2011).
24. Sahyoun, A. H. *et al*. Towards a comprehensive picture of alloacceptor tRNA remolding in metazoan mitochondrial genomes. *Nucleic Acids Res.* **43**, 8044–8056 (2015).
25. Lin, F. J. *et al*. Evolution and phylogeny of the mud shrimps (Crustacea: Decapoda) revealed from complete mitochondrial genomes. *BMC Genomics* **13**, 631, doi:10.1186/1471-2164-13-631 (2012).
26. Yamauchi, M. M., Miya, M. U. & Nishida, M. Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (Crustacea: Decapoda: Brachyura). *Gene* **311**, 129–135 (2003).
27. Sun, H., Zhou, K. & Song, D. Mitochondrial genome of the Chinese mitten crab *Eriocheir japonica sinenesis* (Brachyura: Thoracotremata: Grapsoidea) reveals a novel gene order and two target regions of gene rearrangements. *Gene* **349**, 207–217 (2005).
28. Ki, J. S., Dahms, H. U., Hwang, J. S. & Lee, J. S. The complete mitogenome of the hydrothermal vent crab *Xenograpsus testudinatus* (Decapoda, Brachyura) and comparison with brachyuran crabs. *Comp. Biochem. Physiol. Part D* **4**, 290–299 (2009).
29. Ji, Y. K. *et al*. Mitochondrial genomes of two brachyuran crabs (Crustacea: Decapoda) and phylogenetic analysis. *J. Crustacean Biol* **34**, 494–503 (2014).
30. Hui, M., Liu, Y. & Cui, Z. First complete mitochondrial genome of primitive crab *Homologenus malayensis* (Decapoda: Brachyura: Podotremata: Homolidae). *Mitochondrial DNA Part A* **27**, 859–860 (2016).
31. Kim, S. J., Moon, J. W. & Ju, S. J. Complete mitochondrial genome of the blind vent crab *Gandalfus puia* (Crustacea: Bythograeidae) from the Tonga Arc. *Mitochondrial DNA Part A* **27**, 2719–2720 (2016).
32. Márquez, E. J., Hurtado-Alarcón, J. C., Isaza, J. P., Alzate, J. F. & Campos, N. H. Mitochondrial genome of the Caribbean king crab *Damithrax spinosissimus* (Lamarck, 1818) (Decapoda: Majidae). *Mitochondrial DNA* (in press), doi:10.3109/19401736.2014.961140 (2017).
33. Shi, G. *et al*. Unusual sequence features and gene rearrangements of primitive crabs revealed by three complete mitochondrial genomes of Dromiacea. *Comp. Biochem. Physiol. Part D: Gen. Prot* **20**, 65–73 (2016).
34. Miller, A. D., Murphy, N. P., Burridge, C. P. & Austin, C. M. Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (Menippidae) and *Macrobrachium rosenbergii* (Palaemonidae). *Mar. Biotech.* **7**, 339–349 (2005).
35. Yu, Y. Q., Ma, W. M., Yang, W. J. & Yang, J. S. The complete mitogenome of the lined shore crab *Pachygrapsus crassipes* Randall 1840 (Crustacea: Decapoda: Grapsidae). *Mitochondrial. DNA* **25**, 263–264 (2014).
36. Ma, H. *et al*. First mitochondrial genome for the red crab (*Charybdis feriata*) with implication of phylogenomics and population genetics. *Sci. Rep* **5**, 11524, doi:10.1038/srep11524 (2015).
37. Shi, G. *et al*. The complete mitochondrial genomes of *Umalia orientalis* and *Lyreidus brevifrons*: The phylogenetic position of the family Raninidae within Brachyuran crabs. *Mar. Genom* **21**, 53–61 (2015).
38. Sung, J. M., Lee, J. H., Kim, S. G., Zafer Karagozlu, M. & Kim, C. B. Analysis of complete mitochondrial genome of *Ocypode cordimanus* (Latreille, 1818) (Decapoda, Ocypodidae). *Mitochondrial DNA Part B* **1**, 363–364 (2016).
39. Sung, J. M., Lee, J. H., Kim, S. G., Zafer Karagozlu, M. & Kim, C. B. Complete mitochondrial genome of *Leptodius sanguineus* (Decapoda, Xanthidae). *Mitochondrial DNA Part B* **1**, 500–501 (2016).
40. Tan, M. H., Gan, H. M., Lee, Y. P. & Austin, C. M. The complete mitogenome of the soldier crab *Mictyris longicarpus* (Latreille, 1806) (Crustacea: Decapoda: Mictyridae). *Mitochondrial DNA Part A* **27**, 2121–2122 (2016).
41. Tan, M. H., Gan, H. M., Le, Y. P. & Astin, C. M. Te complete mitogenome of the moon crab *Ashtoret lunaris* (Forskal, 1775), (Crustacea; Decapoda; Matutidae). *Mitochondrial DNA Part A* **27**, 1313–1314 (2016).
42. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
43. Hahn, C., Bachmann, L. & Cevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - a baiting and iterative mapping approach. *Nucleic Acids Res.* **41**, e129 (2013).
44. Negrisolo, E., Babbucci, M. & Patarnello, T. The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondriomics of neuropterid insects. *BMC Genomics* **12**, 221, doi:10.1186/1471-2164-12-221 (2011).
45. Wolstenholme, D. R. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol* **141**, 173–216 (1992).
46. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680 (1994).
47. Smith, A. E. & Marcker, K. A. N-formylmethionyl transfer RNA in mitochondria from yeast and rat liver. *J. Mol. Biol.* **38**, 241–243 (1968).
48. Fearnley, I. M. & Walker, J. E. Initiation codons in mammalian mitochondria: differences in genetic code in the organelle. *Biochemistry* **26**, 8247–8251 (1987).
49. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955–964 (1997).
50. Salvato, P., Simonato, M., Battisti, A. & Negrisolo, E. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* **9**, 331, doi:10.1186/1471-2164-9-331 (2008).
51. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol* **28**, 2731–2739 (2011).
52. Tan, G. *et al*. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* **64**, 778–791 (2015).
53. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41**, 353–358 (1995).
54. Strimmer, K. & von Haeseler, A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Nat. Acad. Sci. USA* **94**, 6815–6819 (1997).
55. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol* **32**, 268–274 (2015).
56. Felsenstein, J. *Inferring phylogenies* 1–580 (Sinauer Associates, 2004).
57. Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* **14**, 82, doi:10.1186/1471-2148-14-82 (2014).
58. Rota-Stabelli, O., Yang, Z. & Telford, M. J. MtZoa: A general mitochondrial amino acid substitutions model for animal evolutionary studies. *Mol. Phylogenet. Evol.* **52**, 268–272 (2009).
59. Yang, Z., Nielsen, R. & Hasegawa, M. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.* **15**, 1600–1611 (1998).
60. Quang, L. S., Gascuel, O. & Lartillot, N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317–2323 (2008).
61. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol* **21**, 1095–1109 (2004).
62. Minh, B. Q., Nguyen, L. T. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
63. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).

64. Bernt, M., Merkle, D. & Middendorf, M. An algorithm for inferring mitogenome rearrangements in a phylogenetic tree in *Comparative Genomics*, RECOMB-CG *2008*, *LNB 5267* (eds Nelson, C. E. & Vialette, S.) 143–157 (Springer, 2008).
65. Bernt, M. *et al*. Finding all sorting tandem duplication random loss operations. *J. Disc. Algorithms* **9**, 32–48 (2011).
66. Hartmann, T., Chu, A. C., Middendorf, M. & Bernt, M. Combinatorics of tandem duplication random loss mutations on circular genomes. *Trans. Comput. Biol. Bioinformatics* (in press), 1–14, doi:10.1109/TCBB.2016.2613522 (2017).
67. Jühling, F. *et al*. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acid. Res* **40**, 2833–2845 (2012).
68. Shen, H., Braband, A. & Scholtz, G. Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. *Mol. Phylogen. Evol* **66**, 776–789 (2013).
69. Jagt, J. W. M., Van Bakel, B. W. M., Guinot, D., Fraaije, R. H. B. & Artal, P. Fossil Brachyura. In *The Crustacea*, *complementary to the Traité de Zoologie*, *Volume 9* Part C-II (eds Castro, P., Davie, P. J. F., Guinot, D., Schram, F. R. & von Vaupe Klei, J. C.) 847–920 (Koninklijke NV Brill, 2015).
70. Lavrov, D. V., Boore, J. L. & Brown, W. M. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. *Mol. Biol. Evol.* **19**, 163–169 (2002).
71. Beckenbach, A. T. Mitochondrial genome sequences of Nematocera (Lower Diptera): evidence of rearrangement following a complete genome duplication in a Winter Crane Fly. *Genome Biol. Evol* **4**, 89–101, doi:10.1093/gbe/evr131 (2012).
72. Montelli, S., Peruffo, A., Patarnello, T., Cozzi, B. & Negrisolo, E. Back to water: signature of adaptive evolution in Cetacean mitochondrial tRNAs. *PLoS One* **6**, e0158129, doi:10.1371/journal.pone.0158129 (2016).
73. Xing, Y. *et al*. The complete mitochondrial genome of the semiterrestrial crab, *Chiromantes neglectum* (Eubrachyura: Grapsoidea: Sesarmidae). *Mitochondr. DNA B: Resources* **1**, 461–463 (2016).
74. Cameron, S. L. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59**, 95–117 (2014).
75. Boore, J. L. & Brown, W. M. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev* **8**, 668–674 (1998).
76. Boore, J. L. The use of genome-level characters for phylogenetic reconstruction. *Trends Ecol. Evol.* **21**, 439–446 (2006).
77. Sotelo, G., Morán, P. & Posada, D. Molecular phylogeny and biogeographic history of the European *Maja* spider crabs (Decapoda, Majidae). *Mol. Phylogen. Evol* **53**, 314–319 (2009).
78. Yang, J., Ye, F. & Huang, Y. Mitochondrial genomes of four katydids (Orthoptera: Phaneropteridae): new gene rearrangements and their phylogenetic implications. *Gene* **575**, 702–711 (2016).

Acknowledgements

This work was supported by two grants (PRA-2010, CPDA105877; ex60% 2014) provided by the University of Padua to Enrico Negrisolo. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank prof. Carlotta Mazzoldi (University of Padova) for providing help in the collection of crab specimens.

Author Contributions

E.N. conceived the studies. E.R. collected and identified specimens of *M. crispata*. A.B., M.B. and E.N. performed the laboratory experiments. A.B., M.B. and E.N. performed the bioinformatic analyses. M.P. and E.N. wrote the manuscript. A.B., M.B., E.R., T.P., M.P. and E.N. revised the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-04168-9

Competing Interests: The authors declare that they have no competing interests.

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

Supplementary Information

The highly rearranged mitochondrial genomes of the crabs *Maja crispata* and *Maja squinado* (Majidae) and gene order evolution in Brachyura

Andrea Basso[1*], Massimiliano Babbucci[1*], Marianna Pauletto[1], Emilio Riginella[2], Tomaso Patarnello[1] & Enrico Negrisolo[1]

[1] University of Padova, Department of Comparative Biomedicine and Food Science (BCA), 35020, Agripolis - Legnaro (PD), Italy.

[2] University of Padova, Department of Biology, 35131, Padova, Italy.

[*] These authors contributed equally to this work.

Correspondence and requests for materials should be addressed to E.N.

enrico.negrisolo@unipd.it.

Table S1. List of taxa, taxonomy, accession numbers, references and crab gene orders

	Infraorder		Superfamily	Family	Species	Accession number	GenBank availability	Reference	crab GOs
outgroups	Axiidea			Strahlaxiidae	*Neaxius glyptocercus* (von Martens, 1868)	JN897379	03.12.2012	P;09	***
	Gebiidea			Upogebiidae	*Austinogebia edulis* (Ngoc-Ho & Chan, 1992)	JN897376	03.12.2012	P;09	***
				Thalassinidae	*Thalassina kelanang* Moh & Chong, 2009	JN897378	03.12.2012	P;09	***
	Anomura			Lithodidae	*Paralithodes brevipes* (H. Milne Edwards & Lucas, 1841)	AB735677	28.05.2013	U;38	***
				Lithodidae	*Paralithodes camtschaticus* (Tilesius, 1815)	JX944381	22.07.2013	P;06	***
				Porcellanidae	*Neopetrolisthes maculatus* (H. Milne Edwards, 1837)	KC107816	20.03.2013	P;22	***
1	Brachyura	Eubrachyura	Dromioidea	Dynomenidae	*Dynomene pilumnoides* Alcock, 1900	KT182070	20.08.2016	P;24	DynGO
2			Homoloidea	Homolidae	*Homola orientalis* Henderson, 1888	KT182071	30.06.2016	P;24	BraGO
3				Homolidae	*Homologenus malayensis* Ihle, 1912	KJ612407	31.12.2014	P;03	BraGO
4				Homolidae	*Moloha majora* (Kubo, 1936)	KT182069	25.01.2016	P;24	BraGO
5			Raninoidea	Raninidae	*Lyreidus brevifrons* Sakai, 1937	KM983394	15.03.2015	P;23	BraGO
6				Raninidae	*Ranina ranina* (Linnaeus, 1758)	KM189817	26.08.2014	P;02	BraGO
7				Raninidae	*Umalia orientalis* (Sakai, 1963)	KM365084	15.03.2015	P;23	BraGO
8			Potamoidea	Potamidae	*Geothelphusa dehaani* (White, 1847)	AB187570	08.08.2012	P;21	GeoGO
9			Potamoidea	Potamidae	*Sinopotamon xiushuiense* Dai, Zhou & Peng, 1995	KU042041	16.12.2015	U;35	SinGO
10			Grapsoidea	Grapsidae	*Grapsus tenuicrustatus* (Herbst, 1783)	KT878721	08.03.2016	P;27	BraGO
11			Grapsoidea	Grapsidae	*Pachygrapsus crassipes* Randall, 1840	KC878511	11.09.2014	P;43	BraGO
12			Grapsoidea	Sesarmidae	*Metopaulias depressus* Rathbun, 1896	KX118277	11.06.2016	U;08	SesGO
13			Grapsoidea	Sesarmidae	*Parasesarma tripectinis* (Shen, 1940)	KU343209	23.04.2016	U;18	SesGO
14			Grapsoidea	Sesarmidae	*Sesarmops sinensis* (H. Milne Edwards, 1834)	KR336554	01.05.2016	U;11	SesGO
15			Grapsoidea	Varunidae	*Cyclograpsus granulosus* H. Milne Edwards, 1853	LN624373	26.06.2015	P;29	MaVaGO
16			Grapsoidea	Varunidae	*Eriocheir hepuensis* Dai, 1991	FJ455506	22.11.2008	P;36	MaVaGO
17			Grapsoidea	Varunidae	*Eriocheir japonica* (De Haan, 1835)	FJ455505	22.11.2008	P;36	MaVaGO
18			Grapsoidea	Varunidae	*Eriocheir sinensis* H. Milne Edwards, 1853	AY274302	18.04.2005	P;25	MaVaGO
19			Grapsoidea	Varunidae	*Helice tientsinensis* Rathbun, 1931	KR336555	01.05.2016	U;11	MaVaGO
20			Grapsoidea	Xenograpsidae	*Xenograpsus testudinatus* N. K. Ng, Huang & Ho, 2000	EU727203	03.11.2009	P;05	XenGO
21			Ocypodoidea	Dotillidae	*Ilyoplax deschampsi* (Rathbun, 1913)	JF909979	22.07.2014	P;04	BraGO
22			Ocypodoidea	Macrophthalmidae	*Macrophthalmus japonicus* (De Haan, 1835)	KU343211	23.04.2016	U;18	MaVaGO
23			Ocypodoidea	Mictyridae	*Mictyris longicarpus* Latreille, 1806	LN611670	02.10.2014	P;31	BraGO
24			Ocypodoidea	Ocypodidae	*Ocypode ceratophthalmus* (Pallas, 1772)	LN611669	02.10.2014	P;32	BraGO
25			Ocypodoidea	Ocypodidae	*Ocypode cordimanus* Latreille, 1818	KT896743	08.03.2016	P;26	BraGO
26		Heterotremata	Majoidea	Majidae	*Maja crispata* Risso, 1827	**KY650651**	**NEW**	P;01	MajGO
27			Majoidea	Majidae	*Maja squinado* (Herbst, 1788)	**KY650652**	**NEW**	P;01	MajGO
28			Majoidea	Mithracidae	*Damithrax spinosissimus* (Lamarck, 1818)	KM405516	25.10.2014	P;14	DamGO
29			Majoidea	Oregoniidae	*Chionoecetes pacificus* Sakai, 1978	AB735678	04.07.2013	U;38	BraGO
30			Bythograeoidea	Bythograeidae	*Austinograea alayseae* Hessler & Martin, 1989	JQ035660	11.02.2013	P;40	BraGO
31			Bythograeoidea	Bythograeidae	*Austinograea rodriguezensis* Tsuchida & Hashimoto, 2002	JQ035658	11.02.2013	P;40	BraGO
32			Bythograeoidea	Bythograeidae	*Gandalfus puia* McLay, 2007	KR002727	21.06.2015	P;07	BraGO
33			Bythograeoidea	Bythograeidae	*Gandalfus yunohana* (Takeda, Hashimoto & Ohta, 2000)	EU647222	27.10.2010	P;41	BraGO
34			Calappoidea	Matutidae	*Ashtoret lunaris* (Forskål, 1775)	LK391941	20.06.2014	P;33	BraGO
35			Eriphioidea	Menippidae	*Myomenippe fornasinii* (Bianconi, 1851)	LK391943	20.06.2014	P;34	BraGO
36			Eriphioidea	Menippidae	*Pseudocarcinus gigas* (Lamarck, 1818)	AY562127	15.02.2006	P;17	BraGO
37			Portunoidea	Geryonidae	*Chaceon granulatus* (Sakai, 1978)	AB769383	31.01.2014	U;39	BraGO
38			Portunoidea	Portunidae	*Callinectes sapidus* Rathbun, 1896	AY363392	19.05.2005	P;19	BraGO
39			Portunoidea	Portunidae	*Charybdis feriata* (Linnaeus, 1758)	KF386147	13.08.2015	P;12	BraGO
40			Portunoidea	Portunidae	*Charybdis japonica* (A. Milne-Edwards, 1861)	FJ460517	01.07.2010	P;10	BraGO
41			Portunoidea	Portunidae	*Portunus pelagicus* (Linnaeus, 1758)	KM977882	14.01.2015	P;15	BraGO
42			Portunoidea	Portunidae	*Portunus sanguinolentus* (Herbst, 1783)	KT438509	19.10.2015	P;16	BraGO
43			Portunoidea	Portunidae	*Portunus trituberculatus* (Miers, 1876)	AB093006	12.07.2003	P;37	BraGO
44			Portunoidea	Portunidae	*Scylla olivacea* (Herbst, 1796)	FJ827760	31.03.2009	U;20	BraGO
45			Portunoidea	Portunidae	*Scylla paramamosain* Estampador, 1949	FJ827761	31.03.2009	P;13	BraGO
46			Portunoidea	Portunidae	*Scylla serrata* (Forskål, 1775)	FJ827758	31.03.2009	U;20	BraGO
47			Portunoidea	Portunidae	*Scylla tranquebarica* (Fabricius, 1798)	FJ827759	31.03.2009	U;20	BraGO
48			Portunoidea	Portunidae	*Thalamita crenata* Rüppell, 1830	LK391945	20.07.2014	P;30	BraGO
49			Xanthoidea	Xanthidae	*Leptodius sanguineus* (H. Milne Edwards, 1834)	KT896744	07.04.2016	P;28	BraGO
50			Leucosioidea	Leucosiidae	*Pyrhila pisum* (De Haan, 1841)	KU343210	23.04.2016	U;18	BraGO
!		Thor	Grapsoidea	Sesarmidae	*Sesarma neglectum* de Man, 1887	KX156954	**31.10.2016**	P; 43	SesGO
!			Potamoidea	Potamidae	*Huananpotamon lichuanense* Dai, Zhou & Peng, 1995	KX639824	**02.10.2016**	U; 44	HuaGO

P, published reference; U, Unpublished; GO, Gene order

Species taxonomy based on various sources. In particular, for marine taxa the nomenclature is that found in the WoRMS data base (World Register of Marine Species)
WoRMS Editorial Board (2016). World Register of Marine Species. Available from **http://www.marinespecies.org** at VLIZ. Accessed 2016-09-23. doi:10.14284/170

***, GO not presented here; **BraGO,** Brachyuran basic GO**;** **DamGO**, *Damithrax spinosissimus* GO*;* **DynGO**, *Dynomene pilumnoides* GO*;* **GeoGO**, *Geothelphusa dehaani* GO; **HuaGO,** *Huananpotamon lichuanense* GO, **MajGO**, *Maja* genus GO; **MaVaGO**, Macrophthalmidae + Varunidae GO; **SesGO**, Sesarmidae GO; **SinGO**, *Sinopotamon xiushuiense* GO; **XenGO**, *Xenograpsus testudinatus* GO.

!, this mtDNA became available too late to be incorporated in all the analyses. It was considered only in some analyses (see main text). Thor, Thoracothremata.

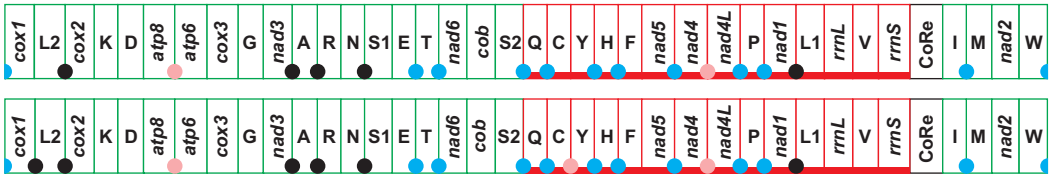
References

1. Basso et al. **this paper**
2. Cheng, J., Jiang W., Shi, H., Sha, Z. The complete mitochondrial genome of red frog crab *Ranina ranina* (Crustacea: Decapoda: Brachyura: Raninidae). *Mitochondrial DNA*. **27**, 1368–1369, doi: 10.3109/19401736.2014.947584 (2016).
3. Hui, M, Liu, Y, Cui, Z. First complete mitochondrial genome of primitive crab *Homologenus malayensis* (Decapoda: Brachyura: Podotremata: Homolidae). *Mitochondrial DNA* **27**, 859–860, doi: 10.3109/19401736.2014.919476 (2016).
4. Ji, Y.K., Wang, A., Lu, X.L., Song, D.H., Jin, Y.H., Lu, J.J., Sun, HY. Mitochondrial genomes of two brachyuran crabs (Crustacea: Decapoda) and phylogenetic analysis. *Journal of Crustacean Biology*. **34**, 494–503, doi: 10.1163/1937240X-00002252 (2014).
5. Ki, J.S., Dahms, H.U., Hwang, J.S., Lee, J.S. The complete mitogenome of the hydrothermal vent crab *Xenograpsus testudinatus* (Decapoda, Brachyura) and comparison with brachyuran crabs. *Comparative Biochemistry and Physiology, Part D.* **4**, 290–299, doi: 10.1016/j.cbd.2009.07.002 (2009).
6. Kim, S., Choi, H.G., Park, J.K., Min, G.S. The complete mitochondrial genome of the subarctic red king crab, *Paralithodes camtschaticus* (Decapoda, Anomura). *Mitochondrial DNA* **24**, 350–352, doi: 10.3109/19401736.2012.760555 (2013).
7. Kim, S.J., Moon, J.W., Ju, SJ. Complete mitochondrial genome of the blind vent crab *Gandalfus puia* (Crustacea: Bythograeidae) from the Tonga Arc. *Mitochondrial DNA Part A* **27**, 2719–2720, doi: 10.3109/19401736.2015.1046162 (2016).
8. Lesny, P., Schubart, C.D., Podsiadlowski, L. **Unpublished**. *Metopaulias depressus*
9. Lin, F.J., Liu, Y., Sha, Z., Tsang, L.M., Chu, K.H., Chan, T.Y., Liu, R., Cui, Z. Evolution and phylogeny of the mud shrimps (Crustacea: Decapoda) revealed from complete mitochondrial genomes. *BMC Genomics* **13**, 631, doi: 10.1186/1471-2164-13-631 (2012).
10. Liu, Y., Cui, Z. Complete mitochondrial genome of the Asian paddle crab *Charybdis japonica* (Crustacea: Decapoda: Portunidae): gene rearrangement of the marine brachyurans and phylogenetic considerations of the decapods. *Mol. Biol. Rep*. **37**, 2559–2569, doi: 10.1007/s11033-009-9773-2 (2010).
11. Liu, Q.N., Tang, B.P. **Unpublished**. *Sesarmops sinensis, Helice tientsinensis*
12. Ma, H., Ma, C., Li, C., Lu, J., Zou, X., Gong, Y., Wang, W., Chen, W., Ma, L., Xia L. First mitochondrial genome for the red crab (*Charybdis feriata*) with implication of phylogenomics and population genetics. *Scientific Reports* **5**, 11524, doi: 10.1038/srep11524 (2015).
13. Ma, H., Ma, C., Li, X., Xu, X., Feng, N., Ma, L. The complete mitochondrial genome sequence and gene organization of the mud crab (*Scylla paramamosain*) with phylogenetic consideration. *Gene* **519**, 120–127, doi: 10.1016/j.gene.2013.01.028 (2013).
14. Márquez, E.J., Hurtado-Alarcón, J.C., Isaza, J.P., Alzate, J.F., Campos, N.H. Mitochondrial genome of the Caribbean king crab *Damithrax spinosissimus* (Lamarck, 1818) (Decapoda: Majidae). *Mitochondrial DNA* (in press); doi:10.3109/19401736.2014.961140 (2016).
15. Meng, X., Jia, F., Liu, P., Li, J. The complete mitogenome of blue swimming crab *Portunus pelagicus* Linnaeus, 1766 (Crustacea: Decapoda: Portunidae), *Mitochondrial DNA Part A* **27**, 2789–2790, doi: 10.3109/19401736.2015.1053068, (2016).
16. Meng, X., Jia, F., Zhang, X., Liu, P., Li, J. Complete sequence and characterization of mitochondrial genome in the swimming crab *Portunus sanguinolentus* (Herbst, 1783) (Decapoda, Brachyura, Portunidae). *Mitochondrial DNA Part A* **27**, 3052–3053, doi: 10.3109/19401736.2015.1063130 (2016).
17. Miller, A.D., Murphy, N.P., Burridge, C.P., Austin, C.M. Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (Menippidae) and *Macrobrachium rosenbergii* (Palaemonidae). *Marine Biotechnology*. **7**, 339–349, doi: 10.1007/s10126-004-4077-8 (2005).
18. Park,Y.J., Park, C.E., Lee, S.H., Ko, H.S., Shin, J.H. **Unpublished**. *Parasesarma tripectinis, Macrophthalmus japonicus, Pyrhila pisum*
19. Place, A.R., Feng, X., Steven, C.R., Fourcade, H.M., Boore, J.L. Genetic markers in blue crabs (*Callinectes sapidus*) II: complete mitochondrial genome sequence and characterization of genetic variation. *Journal of Experimental Marine Biology and Ecology*. **319**, 15–27, doi:10.1016/j.jembe.2004.03.024 (2005).
20. Sangthong, P. **Unpublished**, *Scylla olivacea, Scylla serrata*, *Scylla tranquebarica*
21. Segawa, R.D., Aotsuka, T. The mitochondrial genome of the Japanese freshwater crab, *Geothelphusa dehaani* (Crustacea: Brachyura): evidence for its evolution via gene duplication. *Gene* **355**, 28–39, doi: 10.1016/j.gene.2005.05.020 (2005).
22. Shen, H., Braband, A., Scholtz, G. Mitogenomic analysis of decapod crustacean phylogeny corroborates traditional views on their relationships. *Molecular Phylogenetics and Evolution.* **66**, 776–789, doi: 10.1016/j.ympev.2012.11.002 (2013).
23. Shi, G., Cui, Z., Hui, M., Liu, Y., Chan, T.Y., Song, C. The complete mitochondrial genomes of *Umalia orientalis* and *Lyreidus brevifrons*: The phylogenetic position of the family Raninidae within Brachyuran crabs. *Marine Genomics*. **21**, 53–61, doi: 10.1016/j.margen.2015.02.002 (2015).
24. Shi, G., Cui, Z., Hui, M., Liu, Y., Chan, T.Y., Song, C. Unusual sequence features and gene rearrangements of primitive crabs revealed by three complete mitochondrial genomes of Dromiacea. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics* **20**, 65–73, doi: 10.1016/j.cbd.2016.07.004 (2016).
25. Sun, H., Zhou, K., Song, D. Mitochondrial genome of the Chinese mitten crab *Eriocheir japonica sinenesis (*Brachyura: Thoracotremata: Grapsoidea) reveals a novel gene order and two target regions of gene rearrangements. *Gene* **349**, 207–217, doi: 10.1016/j.gene.2004.12.036 (2005).
26. Sung, J.M., Lee, J.H., Kim, S.G., Zafer Karagozlu, M., Kim, C.B. Analysis of complete mitochondrial genome of *Ocypode cordimanus* (Latreille, 1818) (Decapoda, Ocypodidae), *Mitochondrial DNA Part B* **1**, 363–364, doi: 10.1080/23802359.2016.1168718 (2016).
27. Sung, J.M., Lee, J.H., Kim, S.K., Zafer Karagozlu, M., Kim, C.B. The complete mitochondrial genome of *Grapsus tenuicrustatus* (Herbst, 1783) (Decapoda, Grapsidae). *Mitochondrial DNA Part B* **1**, 441–442, doi: 10.1080/23802359.2016.1180559 (2016).
28. Sung, J.M., Lee, J.H., Kim, S.G., Zafer Karagozlu, M., Kim, C.B. Complete mitochondrial genome of *Leptodius sanguineus* (Decapoda, Xanthidae). *Mitochondrial DNA Part B* **1**, 500–501, doi: 10.1080/23802359.2016.1192505 (2016).
29. Tan, M.H., Gan, H.M., Lee, Y.P., Austin, C.M. The complete mitogenome of purple mottled shore crab *Cyclograpsus granulosus* H. Milne-Edwards, 1853 (Crustacea: Decapoda: Grapsoidea). *Mitochondrial DNA* (in press), doi:10.3109/19401736.2014.989514 (2016).
30. Tan, M.H, Gan, H.M., Lee, Y.P., Austin, C.M. The complete mitogenome of the swimming crab *Thalamita crenata* (Rüppell, 1830) (Crustacea; Decapoda; Portunidae). *Mitochondrial DNA Part A* **27**, 1275–1276, doi: 10.3109/19401736.2014.945553 (2016).
31. Tan, M.H., Gan, H.M., Lee, Y.P., Austin, C.M. The complete mitogenome of the soldier crab *Mictyris longicarpus* (Latreille, 1806) (Crustacea: Decapoda: Mictyridae). *Mitochondrial DNA Part A* **27**, 2121–2122, doi:10.3109/19401736.2014.982585 (2016).
32. Tan, M.H., Gan, H.M., Lee, Y.P., Austin, C.M. The complete mitogenome of the ghost crab *Ocypode ceratophthalmus* (Pallas, 1772) (Crustacea: Decapoda: Ocypodidae). *Mitochondrial DNA Part A* **27**, 2123–2124, doi:10.3109/19401736.2014.982587 (2016).
33. Tan, M.H., Gan, H.M., Lee, Y.P., Austin, C.M. The complete mitogenome of the moon crab *Ashtoret lunaris* (Forskal, 1775), (Crustacea; Decapoda; Matutidae). *Mitochondrial DNA Part A* **27**, 1313–1314, doi: 10.3109/19401736.2014.945572 (2016).
34. Tan, M.H., Gan, H.M., Lee, Y.P., Austin, C.M. The complete mitogenome of the stone crab *Myomenippe fornasinii* (Bianconi, 1851) (Crustacea: Decapoda: Menippidae). *Mitochondrial DNA Part A* **27**, 1374–1375, doi: 10.3109/19401736.2014.947587 (2016).
35. Wang, Y., Zhou, X., Xu, W., Bai, J., Zeng, Q. **Unpublished**. *Sinopotamon xiushuiense*
36. Wang, J., Huang, L., Cheng, Q., Lu, G., Wang. C. Complete mitochondrial genomes of three mitten crabs, *Eriocheir sinensis*, *E. hepuensis*, and *E. japonica*. *Mitochondrial DNA Part A* **27**, 1175–1176, doi: 10.3109/19401736.2014.936425 (2016).
37. Yamauchi, M.M., Miya, M.U., Nishida, M. Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (Crustacea: Decapoda: Brachyura). *Gene* **311**, 129–135, doi: 10.1016/S0378-1119(03)00582-1 (2003).
38. Yanagimoto,T., Kobayashi, T. **Unpublished**. *Paralithodes brevipes, Chionoecetes pacificus*
39. Yanagimoto,T. **Unpublished**. *Chaceon granulatus*
40. Yang, J.S., Lu, B., Chen, D.F., Yu, Y.Q., Yang, F., Nagasawa, H., Tsuchida, S., Fujiwara, Y., Yang, W.J. When did decapods invade hydrothermal vents? Clues from the Western pacific and Indian oceans. *Molecular Biology and Evolution.* **30**, 305–309, doi: 10.1093/molbev/mss224 (2013).
41. Yang, J.S., Nagasawa, H., Fujiwara, Y., Tsuchida, S., Yang, W.J. The complete mitogenome of the hydrothermal vent crab *Gandalfus yunohana* (Crustacea: Decapoda: Brachyura): a link between the Bythograeoidea and Xanthoidea. *Zoologica Scripta*. **39**, 621 – 630, doi: 10.1111/j.1463-6409.2010.00442.x (2010).
42. Yu, Y.Q., Ma, W.M., Yang, W.J., Yang, J.S. The complete mitogenome of the lined shore crab *Pachygrapsus crassipes* Randall 1840 (Crustacea: Decapoda: Grapsidae). *Mitochondrial DNA* **25**, 263–264, doi: 10.3109/19401736.2013.800497 (2014).
43. [!], Xing,Y., Ma, X., Wei, Y., Pan, D., Liu, W. Sun,H. The complete mitochondrial genome of the semiterrestrial crab, *Chiromantes neglectum* (Eubrachyura: Grapsoidea: Sesarmidae). *Mitochondrial DNA B Resour* **1**, 461-463, doi: 10.1080/23802359.2016.1186509 (2016).
44. [!], Wang, Y., Bai, J., Zhu, C.C., Zou, J.X. Zhou, X.M. **Unpublished**. *Huananpotamon lichuanense*

Maja crispata; 16592 bp; MajGO

	start	end	size		
cox1	1	1534	1534	ATG	T(aa)
trnL2	1535	1598	64	taa	
isp (*trnL2-cox2*)	1599	1604	6		
cox2	1605	2292	688	ATG	T(aa)
trnK	2293	2359	67	ttt	
trnD	2360	2425	66	gtc	
atp8	2426	2584	159	GTG	TAA
atp6	2578	3251	674	ATT	TA(a)
cox3	3252	4041	790	ATG	T(aa)
trnG	4042	4105	64	tcc	
nad3	4106	4459	354	ATT	TAA
isp (*nad3-trnA*)	4460	4463	4		
trnA	4464	4528	65	tgc	
isp (*trnA-trnR*)	4529	4531	3		
trnR	4532	4594	63	tcg	
trnN	4595	4661	67	gtt	
isp (*trnN-trnS1*)	4662	4663	2		
trnS1	4664	4729	66	tct	
trnE	4730	4794	65	ttc	
isp (*trnE-trnT*)	4795	4829	35		
trnT	4830	4893	64	tgt	
isp (*trnT-nad6*)	4894	4933	40		
nad6	4934	5442	509	ATC	TA(a)
cob	5443	6577	1135	ATG	T(aa)
trnS2	6578	6642	65	tga	
isp (*trnS2-trnQ*)	6643	7568	926		
trnQ	7569	7637	69	ttg	

	start	end	size		
isp (*trnQ-trnC*)	7638	7829	192		
trnC	7830	7891	62	gca	
trnY	7892	7955	65	gta	
isp (*trnY-trnH*)	7956	8062	107		
trnH	8063	8130	68	gtg	
isp (*trnH-trnF*)	8131	8132	2		
trnF	8133	8197	65	gaa	
nad5	8198	9931	1734	ATG	TAA
isp (*nad5-nad4*)	9932	9935	4		
nad4	9936	11309	1374	ATG	TAG
nad4L	11303	11602	300	ATG	TAA
isp (*nad4L-trnP*)	11603	11635	33		
trnP	11636	11702	67	tgg	
isp (*trnP-nad1*)	11703	11809	107		
nad1	11810	12763	954	ATT	TAG
isp (*nad1-trnL1*)	12764	12768	5		
trnL1	12769	12834	66	tag	
rrnL	12835	14113	1279		
trnV	14114	14185	72	tac	
rrnS	14186	14989	804		
CoRe	14990	15321	332		
trnI	15322	15386	65	gat	
isp (*trnI-trnM*)	15387	15405	19		
trnM	15406	15471	67	cat	
nad2	15472	16478	1007	ATG	TA(a)
trnW	16479	16546	68	tca	
isp (*trnW-cox1*)	16547	16592	46		



Maja squinado; 16598 bp; MajGO

	start	end	size		
cox1	1	1536	1536	ATG	TAA
isp (*cox1-trnL2*)	1537	1538	2		
trnL2	1539	1602	64	taa	
isp (*trnL2-cox2*)	1603	1608	6		
cox2	1609	2296	688	ATG	T(aa)
trnK	2297	2363	67	ttt	
trnD	2364	2428	65	gtc	
atp8	2439	2587	159	GTG	TAA
atp6	2581	3254	674	ATT	TA(a)
cox3	3255	4044	790	ATG	T(aa)
trnG	4045	4108	64	tcc	
nad3	4109	4462	354	ATT	TAA
isp (*nad3-trnA*)	4463	4466	4		
trnA	4467	4530	64	tgc	
isp (*trnA-trnR*)	4531	4533	3		
trnR	4534	4597	64	tcg	
trnN	4598	4664	67	gtt	
isp (*trnN-trnS1*)	4665	4666	2		
trnS1	4667	4732	66	tct	
isp (*trnS1-trnE*)	4733	4733	1		
trnE	4734	4798	65	ttc	
isp (*trnE-trnT*)	4799	4833	35		
trnT	4834	4896	63	tgt	
isp (*trnT-nad6*)	4897	4936	40		
nad6	4937	5445	509	ATC	TA(a)
cob	5446	6580	1135	ATG	T(aa)
trnS2	6581	6645	65	tga	
isp (*trnS2-trnQ*)	6646	7562	917		

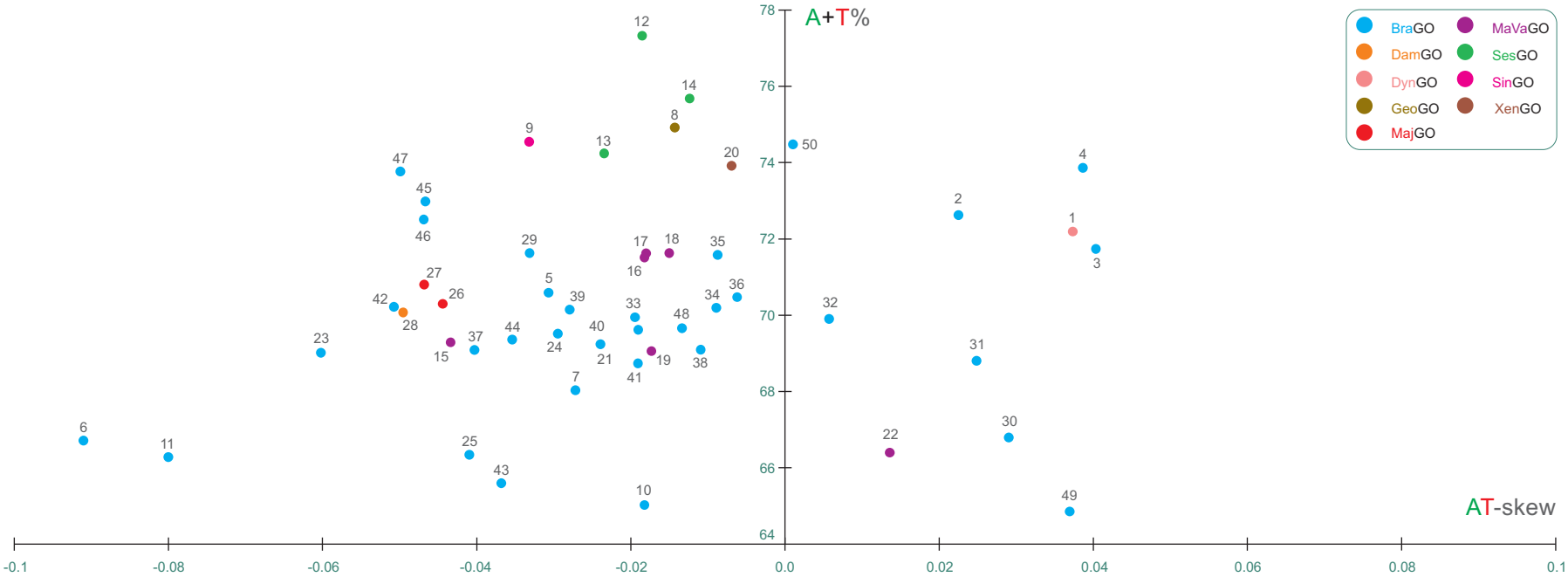
	start	end	size		
trnQ	7563	7631	69	ttg	
isp (*trnQ-trnC*)	7632	7822	191		
trnC	7823	7883	61	gca	
trnY	7883	7948	66	gta	
isp (*trnY-trnH*)	7949	8072	124		
trnH	8073	8140	68	gtg	
isp (*trnH-trnF*)	8141	8142	2		
trnF	8143	8207	65	gaa	
nad5	8208	9941	1734	ATG	TAA
isp (*nad5-nad4*)	9942	9951	10		
nad4	9952	11319	1368	ATG	TAG
nad4L	11313	11612	300	ATG	TAA
isp (*nad4L-trnP*)	11613	11644	32		
trnP	11645	11712	68	tgg	
isp (*trnP-nad1*)	11713	11820	108		
nad1	11821	12777	957	ATG	TAG
isp (*nad1-trnL1*)	12778	12779	2		
trnL1	12780	12845	66	tag	
rrnL	12846	14127	1282		
trnV	14128	14199	72	tac	
rrnS	14200	15002	803		
CR	15003	15339	337		
trnI	15340	15404	65	gat	
isp (*trnI-trnM*)	15405	15435	31		
trnM	15436	15503	68	cat	
nad2	15504	16510	1007	ATG	TA(a)
trnW	16511	16578	68	tca	
isp (*trnW-cox1*)	16579	16598	20		

Figure S1. The mitochondrial genomes of *Maja crispata* and *Maja squinado*.
The MajGO gene order is depicted and linearized starting from cox1. Graphical representation of the mtDNAs and nomenclature of genes as in Figure 1. Isp, intergenic spacer; start, start of the genomic element; end, end of the genomic element; size, size of the genomic element. Start and stop are referred to the α-strand placement. For protein-coding genes, the start codon is provided in **blue** and the stop codon in **red** (with incomplete stop codons written in parentheses). The anticodon is provided (**purple**) for every tRNA (e.g., tga for *trnS2*). A **blue** circle marks an intergenic spacer associated to a genomic rearrangement (see main text). A **black** circle marks an intergenic spacer supposed to be the result of a DNA slippage, during the genome replication[1]. A **pink** circle marks an overlap between adjacent genes.

The mtDNAs of *Maja crispata* and *Maja squinado*

In this study, the complete mtDNAs of the spider crabs *M. crispata* and *M. squinado* were sequenced and annotated. The final assembly of *M. crispata* was 16,592 bp long and contained 24,734 reads. Additional statistics for *M. crispata* mtDNA were: base coverage = 100%; mismatch = 0%; average coverage depth = 147.25; maximum coverage depth = 409. The final assembly of *M. squinado* was 16,598 bp long and contained 20,432 reads. The other statistics for *M. squinado* assembly were: base coverage = 100%; mismatch = 0%; average coverage depth = 121.50; and the maximum coverage depth was 227. The mtDNAs of M. crispata and M. squinado contain the full set of 37 genes found in metazoan mtDNAs (Fig. S1).
Both mtDNAs present intergenic spacers (isp in Fig. S1). These latter range from 1 base (*trnS1-trnE*, *M. squinado*) to 926 bases (*trnS2-trnQ*, *M. squinado*). The *trnS2-trnQ* intergenic spaces exhibit the largest size (Fig. S1). The newly determined mtDNAs share the same gene order MajGO, which is different from any other animal GO so far sequenced (Fig. S1). In MajGO, all the genes located on the β-strand form a single block, placed between *trnE* and CoRe. The A+T contents, the G+C contents, the AT-skews and GC-skews, as well as the codon usages of *M. crispata* and *M. squinado* mtDNAs fall within the ranges computed for the Brachyura mtDNAs (Supplementary Figs S2-S4). The 22 tRNAs are capable to produce the typical cloverleaf secondary structures (Supplementary Figs S5-S6). The *rrnL* and *rrnS* genes are also capable to fold in the secondary structures characterizing these genes. The *rrnS* and *rrnL* structures were determined through a homology modelling process, using as templates the several structures already available for crustacean and more in general for Arthropoda (e.g. Salvato *et al.*[1]; Babbucci *et al.*[2]). The detailed description of these structures will be presented in a paper dealing with the evolution of *rrnSs* and *rrnLs* of crustacean Decapoda, which is in preparation in our laboratory.

1. Salvato, P., Simonato, M., Battisti, A., & Negrisolo, E. The complete mitochondrial genome of the bag-shelter moth Ochrogaster lunifer (Lepidoptera, Notodontidae). *BMC Genomics* **9**, 331; 10.1186/1471-2164-9-331 (2008).
2. Babbucci, M., Basso, A., Patarnello, T., & Negrisolo E. Is It an ant or a butterfly? Convergent evolution in the mitochondrial gene order of Hymenoptera and Lepidoptera. *Genome Biol. Evol.* **6**, 326–3343; doi:10.1093/gbe/evu265 (2014).



1 *Dynomene pilumnoides* Alcock, 1900
2 *Homola orientalis* Henderson, 1888
3 *Homologenus malayensis* Ihle, 1912
4 *Moloha majora* (Kubo, 1936)
5 *Lyreidus brevifrons* Sakai, 1937
6 *Ranina ranina* (Linnaeus, 1758)
7 *Umalia orientalis* (Sakai, 1963)
8 *Geothelphusa dehaani* (White, 1847)
9 *Sinopotamon xiushuiense* Dai, Zhou & Peng, 1995
10 *Grapsus tenuicrustatus* (Herbst, 1783)
11 *Pachygrapsus crassipes* Randall, 1840
12 *Metopaulias depressus* Rathbun, 1896
13 *Parasesarma tripectinis* (Shen, 1940)
14 *Sesarmops sinensis* (H. Milne Edwards, 1834)
15 *Cyclograpsus granulosus* H. Milne Edwards, 1853
16 *Eriocheir hepuensis* Dai, 1991
17 *Eriocheir japonica* (De Haan, 1835)
18 *Eriocheir sinensis* H. Milne Edwards, 1853
19 *Helice tientsinensis* Rathbun, 1931
20 *Xenograpsus testudinatus* N. K. Ng, Huang & Ho, 2000
21 *Ilyoplax deschampsi* (Rathbun, 1913)
22 *Macrophthalmus japonicus* (De Haan, 1835)
23 *Mictyris longicarpus* Latreille, 1806
24 *Ocypode ceratophthalmus* (Pallas, 1772)
25 *Ocypode cordimanus* Latreille, 1818
26 *Maja crispata* Risso, 1827
27 *Maja squinado* (Herbst, 1788)
28 *Damithrax spinosissimus* (Lamarck, 1818)
29 *Chionoecetes pacificus* Sakai, 1978
30 *Austinograea alayseae* Hessler & Martin, 1989
31 *Austinograea rodriguezensis* Tsuchida & Hashimoto, 2002
32 *Gandalfus puia* McLay, 2007
33 *Gandalfus yunohana* (Takeda, Hashimoto & Ohta, 2000)
34 *Ashtoret lunaris* (Forskål, 1775)
35 *Myomenippe fornasinii* (Bianconi, 1851)
36 *Pseudocarcinus gigas* (Lamarck, 1818)
37 *Chaceon granulatus* (Sakai, 1978)
38 *Callinectes sapidus* Rathbun, 1896
39 *Charybdis feriata* (Linnaeus, 1758)
40 *Charybdis japonica* (A. Milne-Edwards, 1861)
41 *Portunus pelagicus* (Linnaeus, 1758)
42 *Portunus trituberculatus* (Miers, 1876)
43 *Portunus sanguinolentus* (Herbst, 1783)
44 *Scylla olivacea (Herbst, 1796)*
45 *Scylla paramamosain* Estampador, 1949
46 *Scylla serrata* (Forskål, 1775)
47 *Scylla tranquebarica* (Fabricius, 1798)
48 *Thalamita crenata* Rüppell, 1830
49 *Leptodius sanguineus* (H. Milne Edwards, 1834)
50 *Pyrhila pisum* (De Haan, 1841)

Figure S2. AT-skew vs. A+T% in the 50 Brachyuram mtDNAs.
The values were calculated on the **α-strand** of the mtDNA genomes. The X axis provides the **AT-skew** values, while the Y axis provides the **A+T%** values.
BraGO, Brachyuran basic GO; **DamGO**, *Damithrax spinosissimus* GO; **DynGO**, *Dynomene pilumnoides* GO; **GeoGO**, *Geothelphusa dehaani* GO; **MajGO**, *Maja* genus GO; **MaVaGO**, Macrophthalmidae + Varunidae GO; **SesGO**, Sesarmidae GO; **SinGO**, *Sinopotamon xiushuiense* GO; **XenGO**, *Xenograpsus testudinatus* GO.

1 *Dynomene pilumnoides* Alcock, 1900
2 *Homola orientalis* Henderson, 1888
3 *Homologenus malayensis* Ihle, 1912
4 *Moloha majora* (Kubo, 1936)
5 *Lyreidus brevifrons* Sakai, 1937
6 *Ranina ranina* (Linnaeus, 1758)
7 *Umalia orientalis* (Sakai, 1963)
8 *Geothelphusa dehaani* (White, 1847)
9 *Sinopotamon xiushuiense* Dai, Zhou & Peng, 1995
10 *Grapsus tenuicrustatus* (Herbst, 1783)
11 *Pachygrapsus crassipes* Randall, 1840
12 *Metopaulias depressus* Rathbun, 1896
13 *Parasesarma tripectinis* (Shen, 1940)
14 *Sesarmops sinensis* (H. Milne Edwards, 1834)
15 *Cyclograpsus granulosus* H. Milne Edwards, 1853
16 *Eriocheir hepuensis* Dai, 1991
17 *Eriocheir japonica* (De Haan, 1835)
18 *Eriocheir sinensis* H. Milne Edwards, 1853
19 *Helice tientsinensis* Rathbun, 1931
20 *Xenograpsus testudinatus* N. K. Ng, Huang & Ho, 2000
21 *Ilyoplax deschampsi* (Rathbun, 1913)
22 *Macrophthalmus japonicus* (De Haan, 1835)
23 *Mictyris longicarpus* Latreille, 1806
24 *Ocypode ceratophthalmus* (Pallas, 1772)
25 *Ocypode cordimanus* Latreille, 1818
26 *Maja crispata* Risso, 1827
27 *Maja squinado* (Herbst, 1788)
28 *Damithrax spinosissimus* (Lamarck, 1818)
29 *Chionoecetes pacificus* Sakai, 1978
30 *Austinograea alayseae* Hessler & Martin, 1989
31 *Austinograea rodriguezensis* Tsuchida & Hashimoto, 2002
32 *Gandalfus puia* McLay, 2007
33 *Gandalfus yunohana* (Takeda, Hashimoto & Ohta, 2000)
34 *Ashtoret lunaris* (Forskål, 1775)
35 *Myomenippe fornasinii* (Bianconi, 1851)
36 *Pseudocarcinus gigas* (Lamarck, 1818)
37 *Chaceon granulatus* (Sakai, 1978)
38 *Callinectes sapidus* Rathbun, 1896
39 *Charybdis feriata* (Linnaeus, 1758)
40 *Charybdis japonica* (A. Milne-Edwards, 1861)
41 *Portunus pelagicus* (Linnaeus, 1758)
42 *Portunus trituberculatus* (Miers, 1876)
43 *Portunus sanguinolentus* (Herbst, 1783)
44 *Scylla olivacea (Herbst, 1796)*
45 *Scylla paramamosain* Estampador, 1949
46 *Scylla serrata* (Forskål, 1775)
47 *Scylla tranquebarica* (Fabricius, 1798)
48 *Thalamita crenata* Rüppell, 1830
49 *Leptodius sanguineus* (H. Milne Edwards, 1834)
50 *Pyrhila pisum* (De Haan, 1841)

Figure S3. GC-skew vs. G+C% in the 50 Brachyuram mtDNAs.
The values were calculated on the **α-strand** of the mtDNA genomes. The X axis provides the **GC-skew** values, while the Y axis provides the **G+C%** values.
BraGO, Brachyuran basic GO; **DamGO**, *Damithrax spinosissimus* GO; **DynGO**, *Dynomene pilumnoides* GO; **GeoGO**, *Geothelphusa dehaani* GO; **MajGO**, *Maja* genus GO; **MaVaGO**, Macrophthalmidae + Varunidae GO; **SesGO**, Sesarmidae GO; **SinGO**, *Sinopotamon xiushuiense* GO; **XenGO**, *Xenograpsus testudinatus* GO.

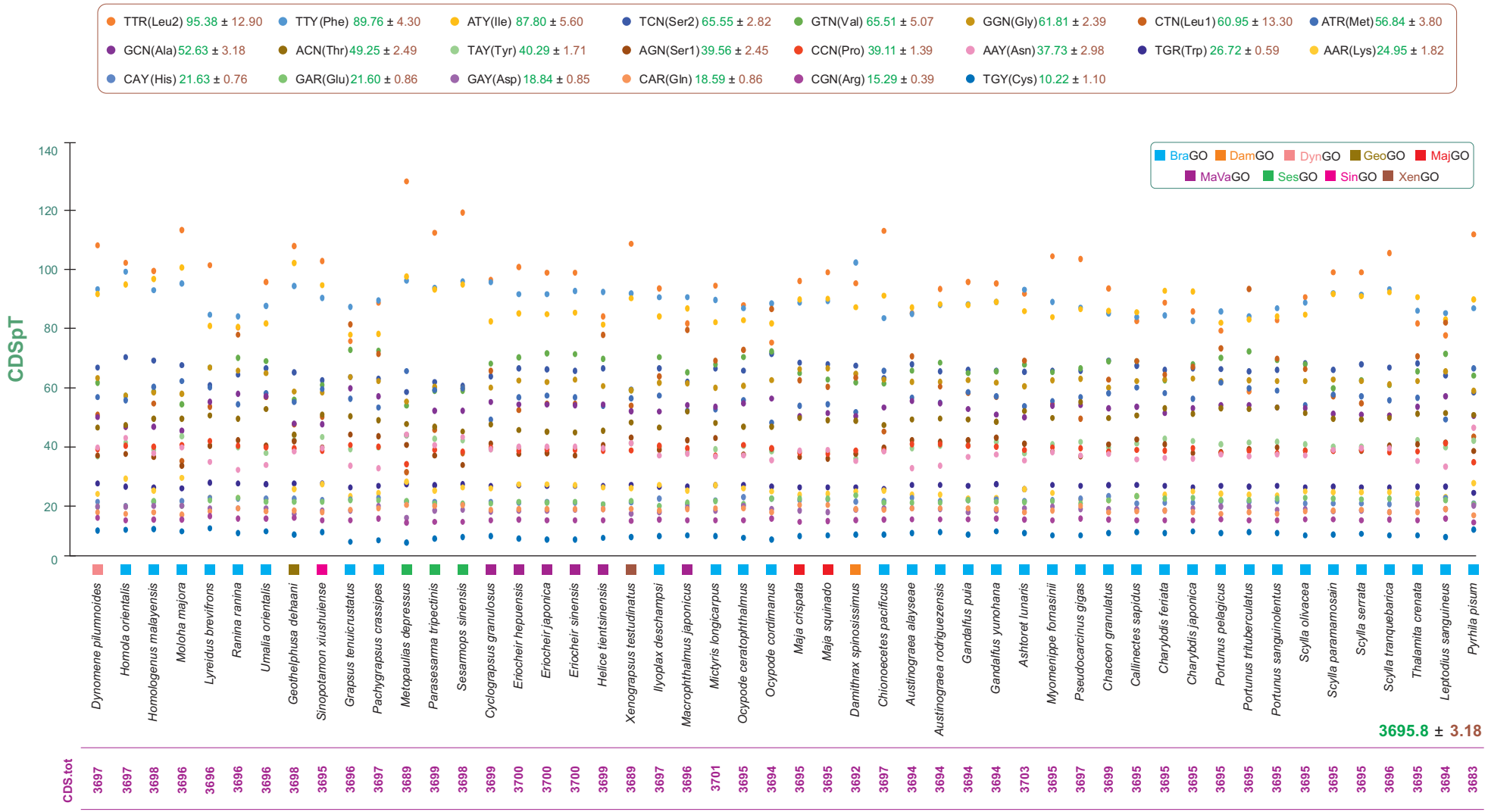


Figure S4. Codon distribution in Brachyuran mtDNAs.
CDS.tot, the total number of codons. **CDspT**, codons per thousand codons. Codon families are provided in the inset above the graphic. The **average value** and the **standard deviation**, computed for the 50 Brachyuran mtDNAs, are provided for each codon family.
BraGO, Brachyuran basic GO; **DamGO**, *Damithrax spinosissimus* GO; **DynGO**, *Dynomene pilumnoides* GO; **GeoGO**, *Geothelphusa dehaani* GO; **MajGO**, *Maja* genus GO; **MaVaGO**, Macrophthalmidae + Varunidae GO; **SesGO**, Sesarmidae GO; **SinGO**, *Sinopotamon xiushuiense* GO; **XenGO**, *Xenograpsus testudinatus* GO.

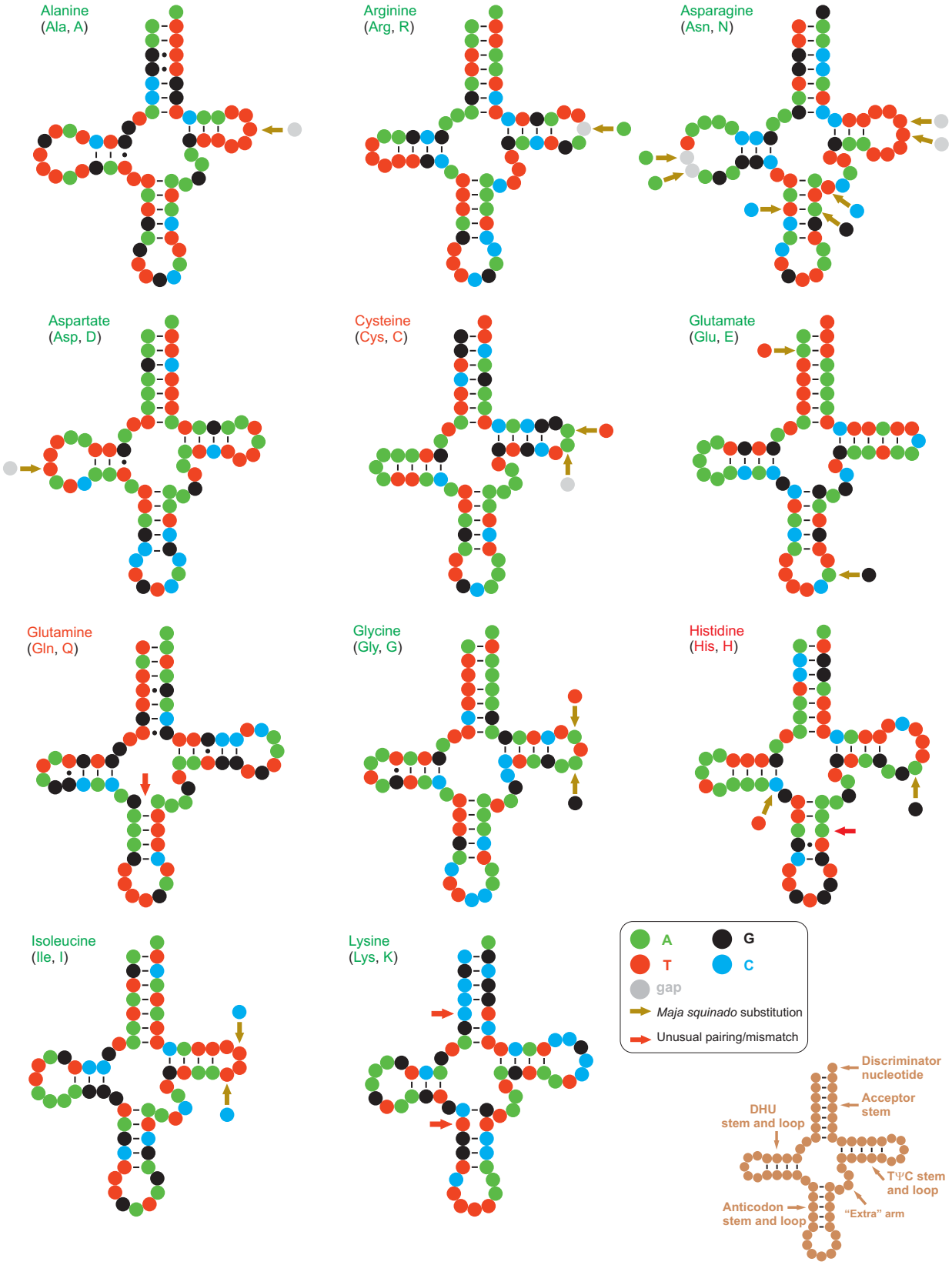


Figure S5. Secondary structure of *Maja* tRNAs and level of conservation (*trnA*-*trnK*).
Secondary structures are based on the tRNA sequences of *Maja crispata*. A substitution occurring in the same position for the tRNAs of *Maja squinado* is marked with a **light-brown** arrow. An unusual pairing/mismatch in the secondary structure is marked with a **red** arrow. A standard pairing is figured with a dash (-), while the G T pairing is figured wit a dot.

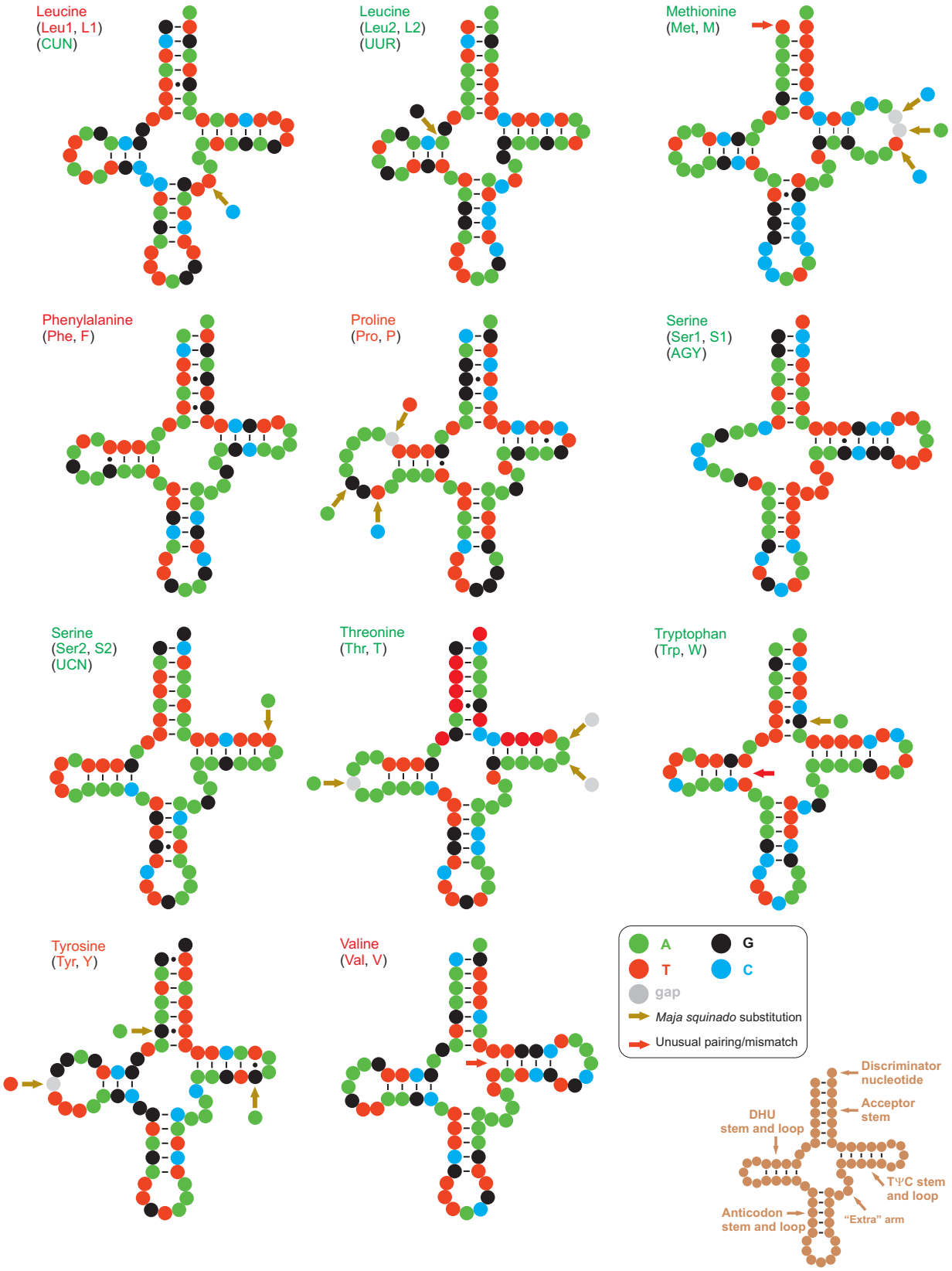


Figure S6. Secondary structure of *Maja* tRNAs and level of conservation (*trnL1*-*trnV*).
Secondary structures are based on the tRNA sequences of *Maja crispata*. A substitution occurring in the same position for the tRNAs of *Maja squinado* is marked with a **light-brown** arrow. An unusual pairing/mismatch in the secondary structure is marked with a **red** arrow. A standard pairing is figured with a dash (-), while the G T pairing is figured wit a dot.

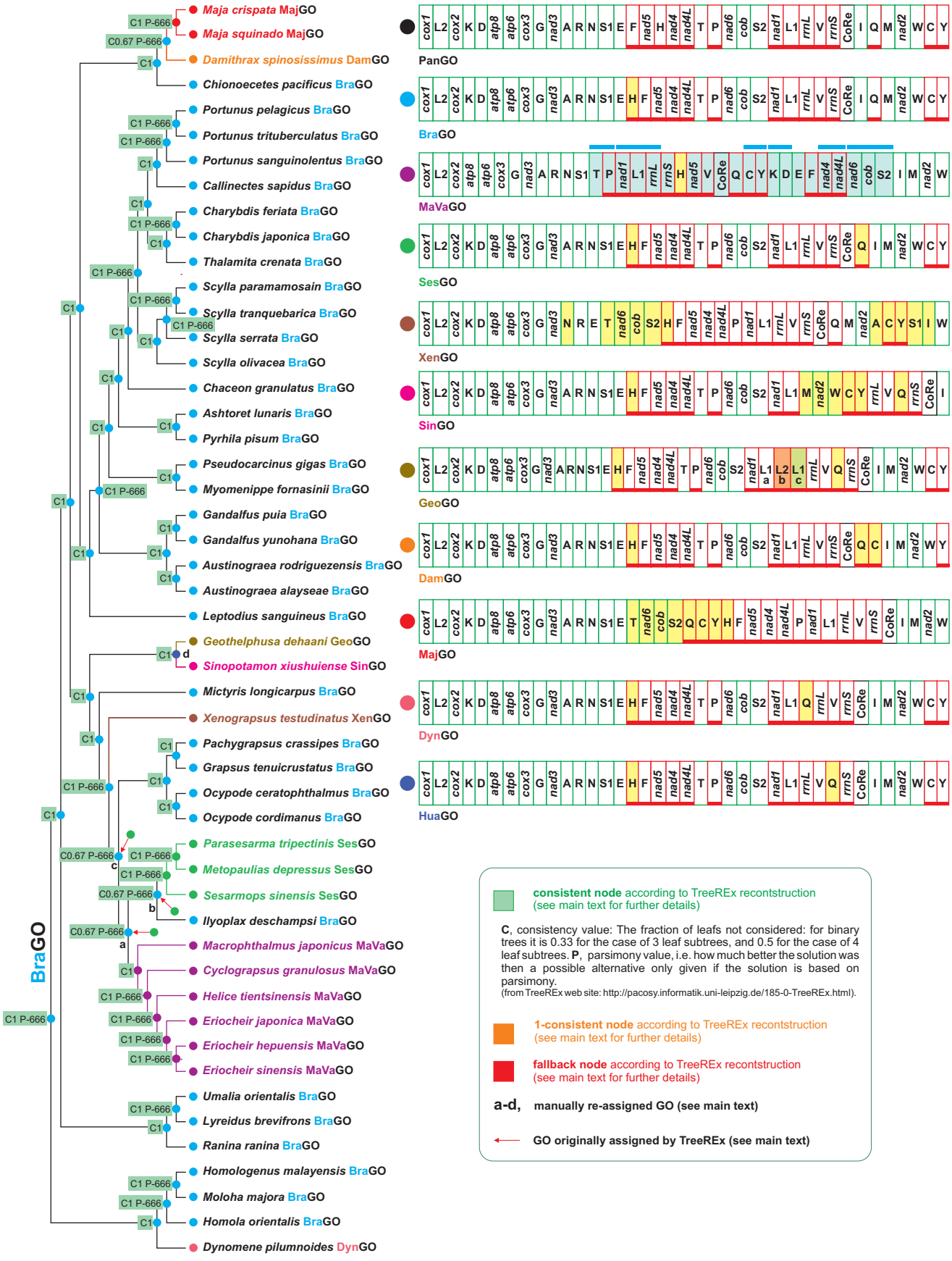


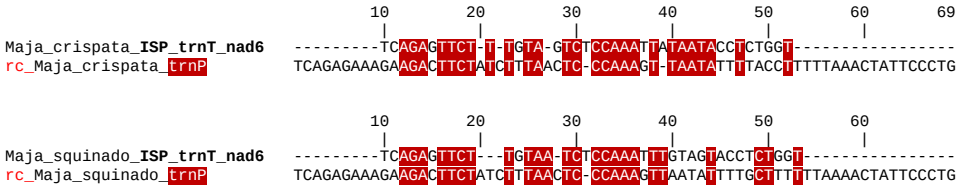
Figure S7. Evolution of GOs in Brachyura: the TreeREx outputs.
The GO assigned at each node by TreeREx is presented together with the associated score (**consistent**, **1-consistent**, **fallback** node). **Green square**, consistent node. **Orange square**, 1-consistent node. **Red square**, fallback node. The different colours reflect the level of uncertainty that characterises the reconstruction of the GO. The red squares indicate the highest level of uncertainty with respect to alternative GOs, while orange and particularly green squares point to more reliable GO reconstructions. (See main text for details).
The genes that changed their position relative to PanGO, through a transposition event, are shown with a yellow background. The passively-shifted genes are figured with their original background. The genes involved in a repositioning, which cannot be identified unambiguously as the result of a transposition or a passive shift, are figured with a light blue background. In this latter case, the common intervals, encompassing two or more genes, shared by the re-arranged GO with PanGO, are highlighted with an upper light blue bar.

Figure S8. Pairwise-alignments of genes involved in genomic rearrangements and the associated intergenic spacers.
The pairwise-alignments were performed with the ClustalW program, available at the PRABI/Rhone-Alpes Bioinformatics Center (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html). Successively, the alignments were manually improved through visual inspection. rc_sequence_name = reverse complement sequence of a gene encoded in the β-strand

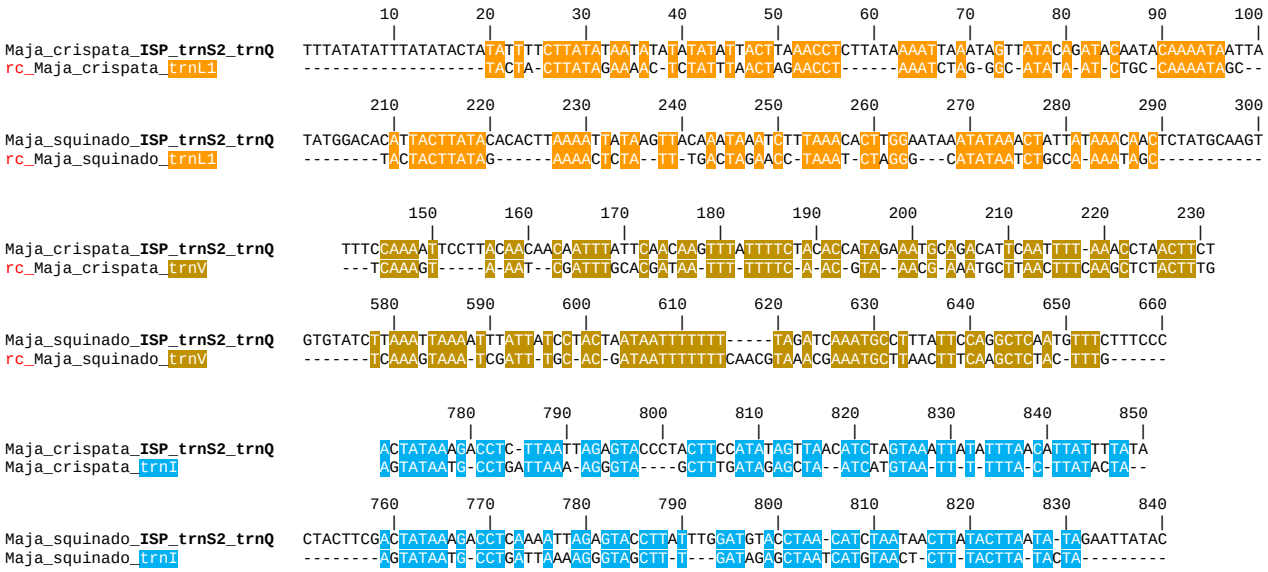
Figure S8.a

MajGO

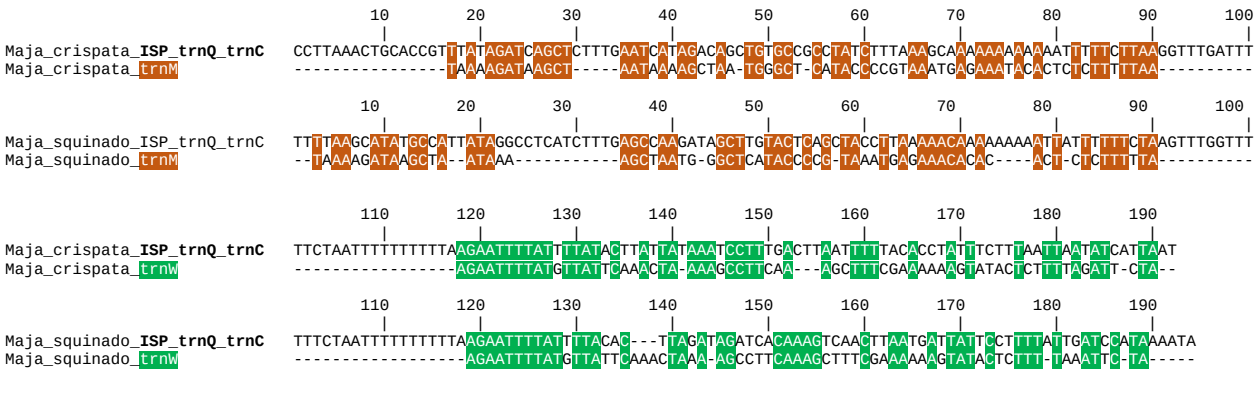
Intergenic spacer: ISP_trnT_nad6 (see Figure 5, main text)



Intergenic spacer: ISP_trnS2_trnQ (see Figure 5, main text)



Intergenic spacer: ISP_trnQ_trnC (see Figure 5, main text)



Intergenic spacer: ISP_nad4L_trnP (see Figure 5, main text)

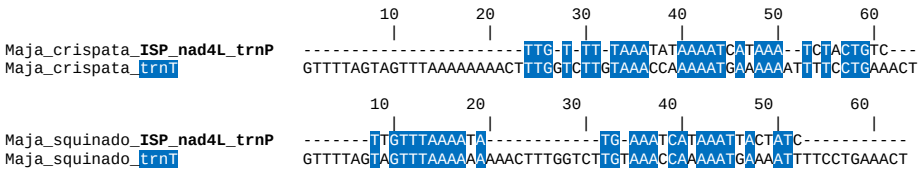
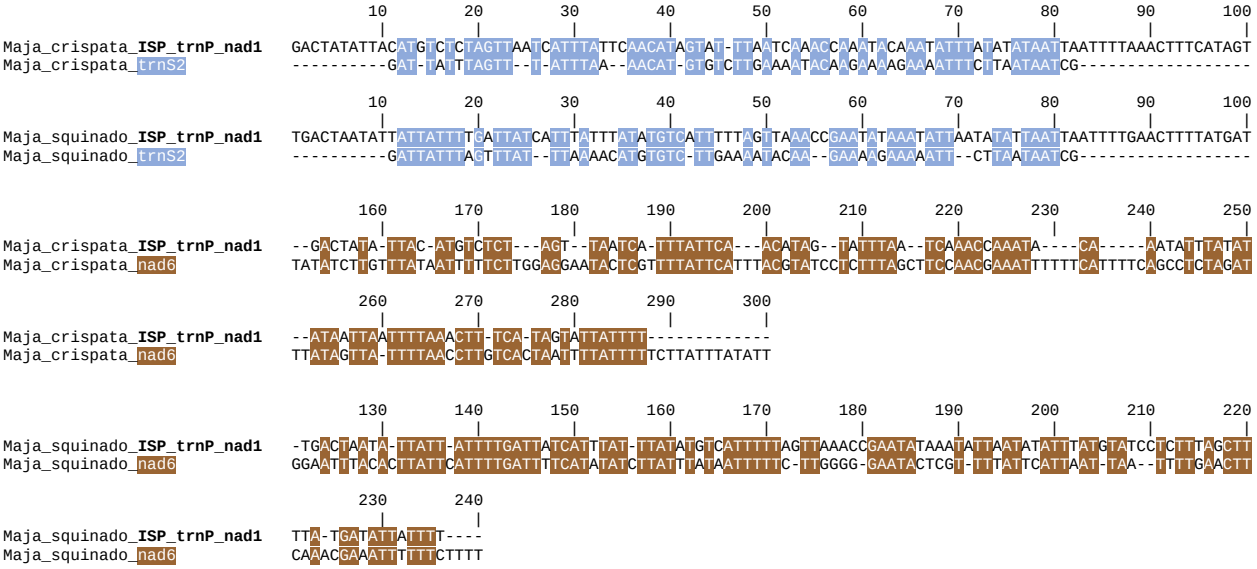


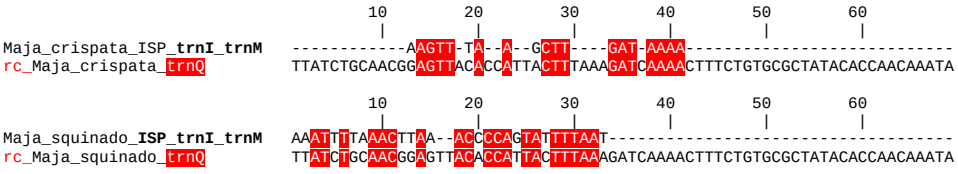
Figure S8.b

MajGO

Intergenic spacer: ISP_trnP_nad1 (see Figure 5, main text)

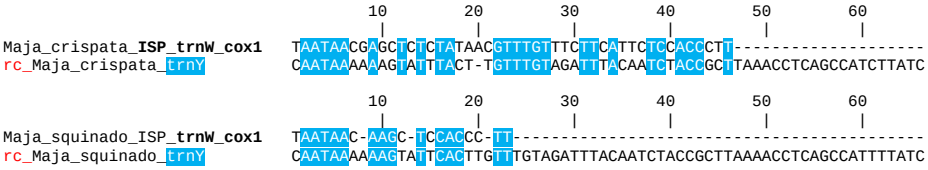


Intergenic spacer: ISP_trnI_trnM (see Figure 5, main text)



rc = reverse complement sequence of a gene encoded in the β-strand

Intergenic spacer: ISP_trnW_cox1 (see Figure 5, main text)



DamGO

Intergenic spacer: ISP_trnQ_ trnC (see Figure 5, main text)

Intergenic spacer: ISP_trnI_ trnM (see Figure 5, main text)

HuaGO

Intergenic spacer: ISP_trnV_trnQ (see Figure 6, main text)

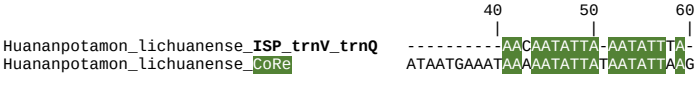
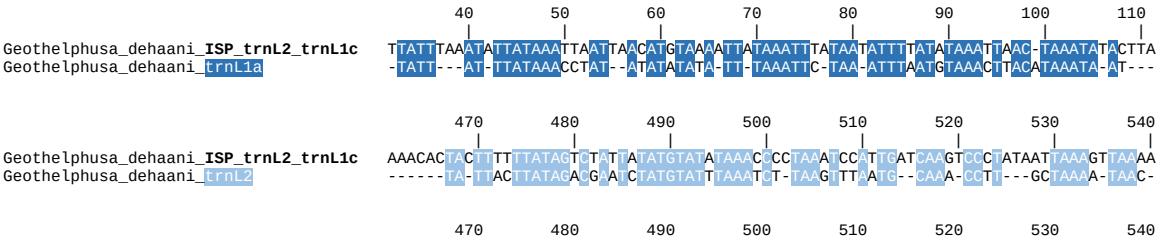


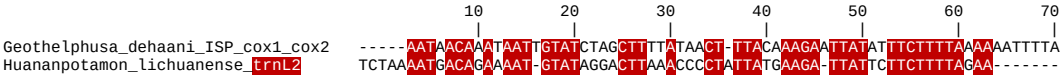
Figure S8.c

GeoGO

Intergenic spacer: ISP_trnL2_trnL1 (see Figure 6, main text)

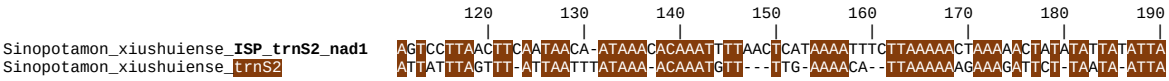


Intergenic spacer: ISP_cox1_cox2 (see Figure 6, main text)

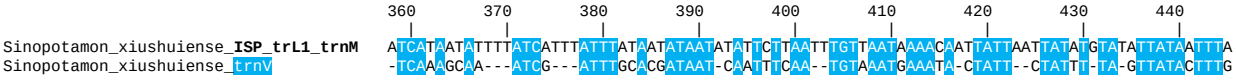


SinGO

Intergenic spacer: ISP_trnS2_nad1 (see Figure 6, main text)

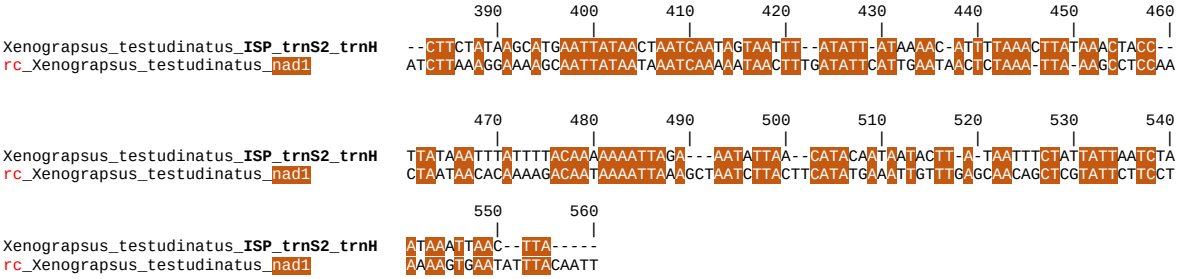


Intergenic spacer: ISP_trnL1_trnM (see Figure 6, main text)

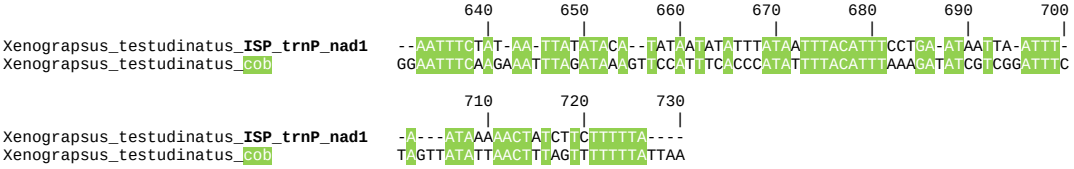


XenGO

Intergenic spacer: ISP_trnS2_trnH (see Figure 7, main text)



Intergenic spacer: ISP_trnP_nad1 (see Figure 7, main text)



Intergenic spacer: ISP_trnE_trnT (see Figure 7, main text)

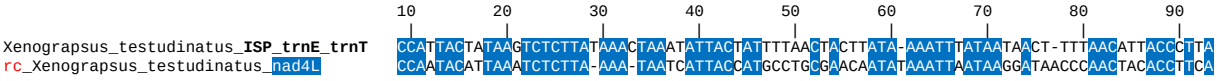
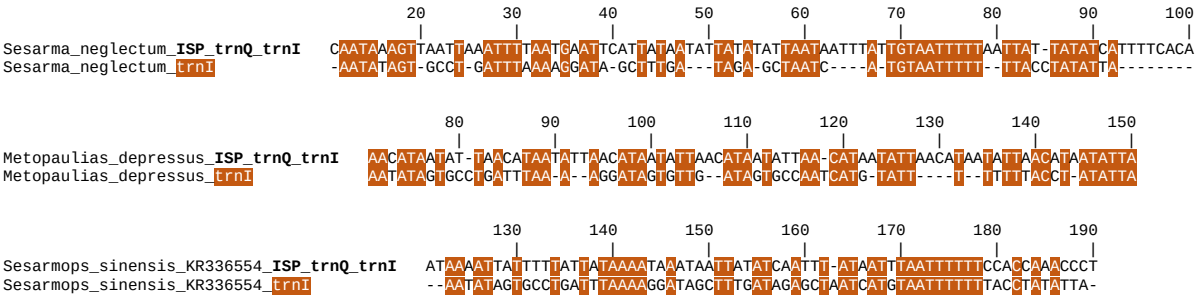


Figure S8.d

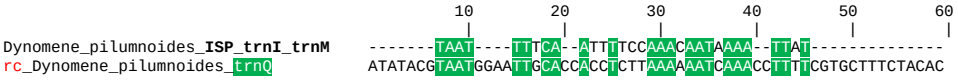
SesGO

Intergenic spacer: ISP_trnQ_trnI (see Figure 7, main text)

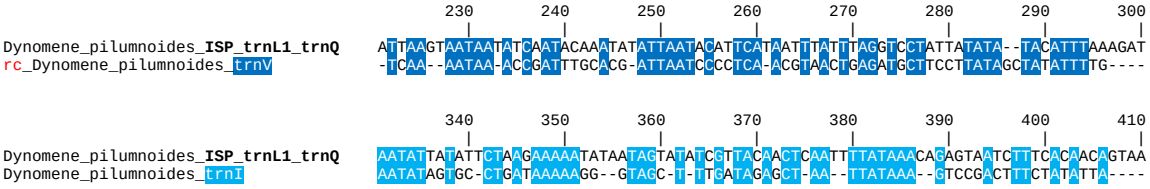


DynGO

Intergenic spacer: ISP_trnI_trnM (see Figure 8, main text)

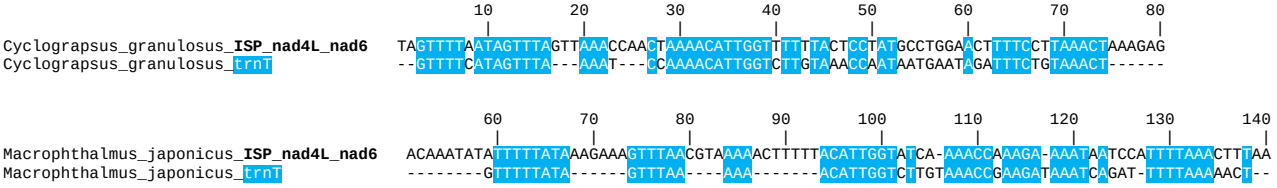


Intergenic spacer: ISP_trnL1_trnQ (see Figure 8, main text)

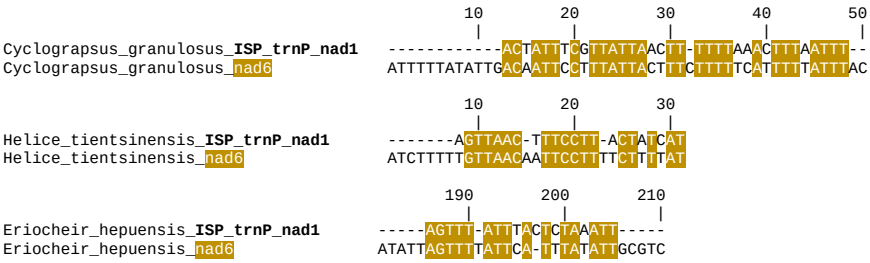


MaVaGO

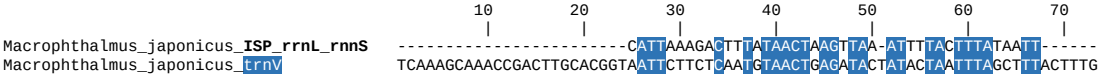
Intergenic spacer: ISP_nad4L_nad6 (see Figure, 8 main text)



Intergenic spacer: ISP_trnP_nad1 (see Figure 8, main text)



Intergenic spacer: ISP_rrnL_rrnS (see Figure 8, main text)



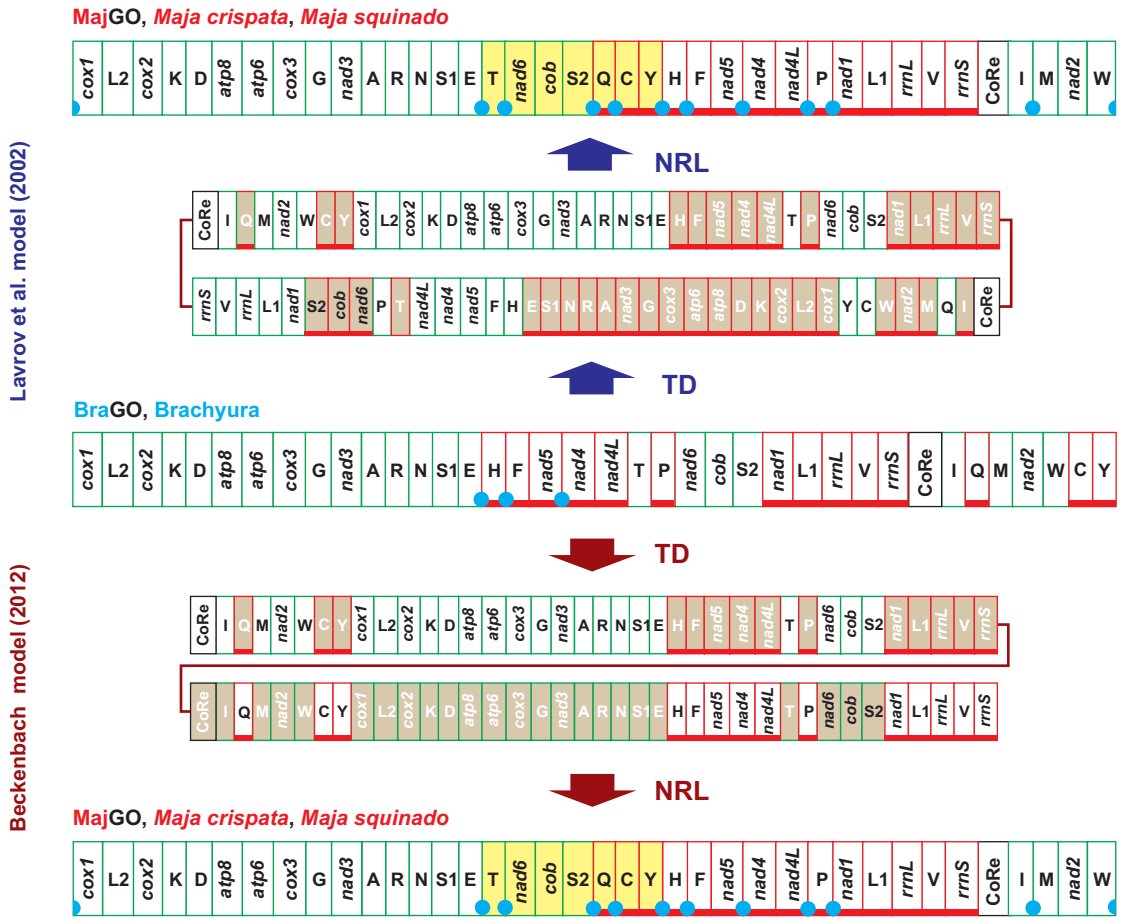


Figure S9. The evolutionary pathways generating the MajGO modelled by tandem duplication nonrandom loss (TDNRL) models.
The rearrangement in the GO of *Maja* species is investigated, and depicted with respect to **BraGO**. TD/NRL tandem duplication/ non random loss event. The genomic and genetic nomenclature, as well as the colour scheme, are the same as in Figure 1. The genes that changed their position relative to BraGO, through a transposition event, are shown with a yellow background. The passively-shifted genes are figured with their original background. A blue circle marks an intergenic spacer present in a position associated to genomic rearrangement (see Figure 5, main text).

Text associated to Figure S9.

The MajGO arrangement can be explained by the tandem duplication nonrandom loss (TDNRL) models (Fig. S9). Two TDNRLs models exhist[1,2]. The model first was developed by Lavrov et al.[1] to explain the arrangement of two Diplopoda. In these millipedes most of the genes located in the same strand are contiguous, thus generating a strand-biased arrangement1. More recently, a second TDNRL model was presented by Beckenbach[2] to explain the genomic arrangement of the winter crane fly *Paracladura trichoptera*, which exhibits a GO with a stand-biased distribution of the genes, different however than that described for millipedes. Both models implied the duplication of the complete mtDNA. This process generates a dimeric molecule where the monomers are covalently linked head to tail (Lavrov et al.[1]) or head to head (Beckenbach[2]). For reason of space only the TDNRL model of Beckenbach[2] is presented in Figure 5 of main text.
The model of Lavrov et al.[1] implies the occurrence of a single, bidirectional, transcription promotor located in the CoRe, while the model of Beckenbach[2] assumes the presence of multiple transcription promotors located on both strands and associated with block of genes.
Both TDNRLs imply that, after the complete duplication of the mtDNA, the transcriptional promotor/s, which are located in the same strand, are lost/inactivated in the first monomer and the same process occurs in the opposite strand of the second monomer.
Thus, the genes, located in the strand deprived of their promotors, cannot be transcribed. This effect rapidly transforms the genes in pseudogenes, which are lost at the end. The final rearrangement is thus strand-biased and non random. However, no-biased transposition/inversion can also occur. The model of Lavrov et al.[1] explicitly predicts the presence of a second intergenic large spacer (indeed the duplication of the CoRe), while the model of Beckenbach[2] does not make any assumption on the presence/absence of intergenic spacers. Finally, none of the two TDNRL models denies the possibility that intergenic spacers are present in positions different than CoRe as the effect of the TDNRL process.
In *Maja* species, intergenic spacers of variable size (Supplementary Figs. S1, S9) are present in all the positions associated with genomic rearrangements. However, the presence of these spacers does not allow to decide what model (TDRL or TDNRL) describes better the evolutionary pathway that generated the *Maja* GO. Indeed the presence of these spacers is consistent with both models.

References

1. Lavrov, D.V. Boore, J.L., & Brown, W.M. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. *Mol. Biol. Evol*. **19**, 163–169 (2002).
2. Beckenbach, A.T. Mitochondrial genome sequences of Nematocera (Lower Diptera): evidence of rearrangement following a complete genome duplication in a Winter Crane Fly. *Genome Biol. Evol*. **4**, 89–101; doi:10.1093/gbe/evr131 (2012).

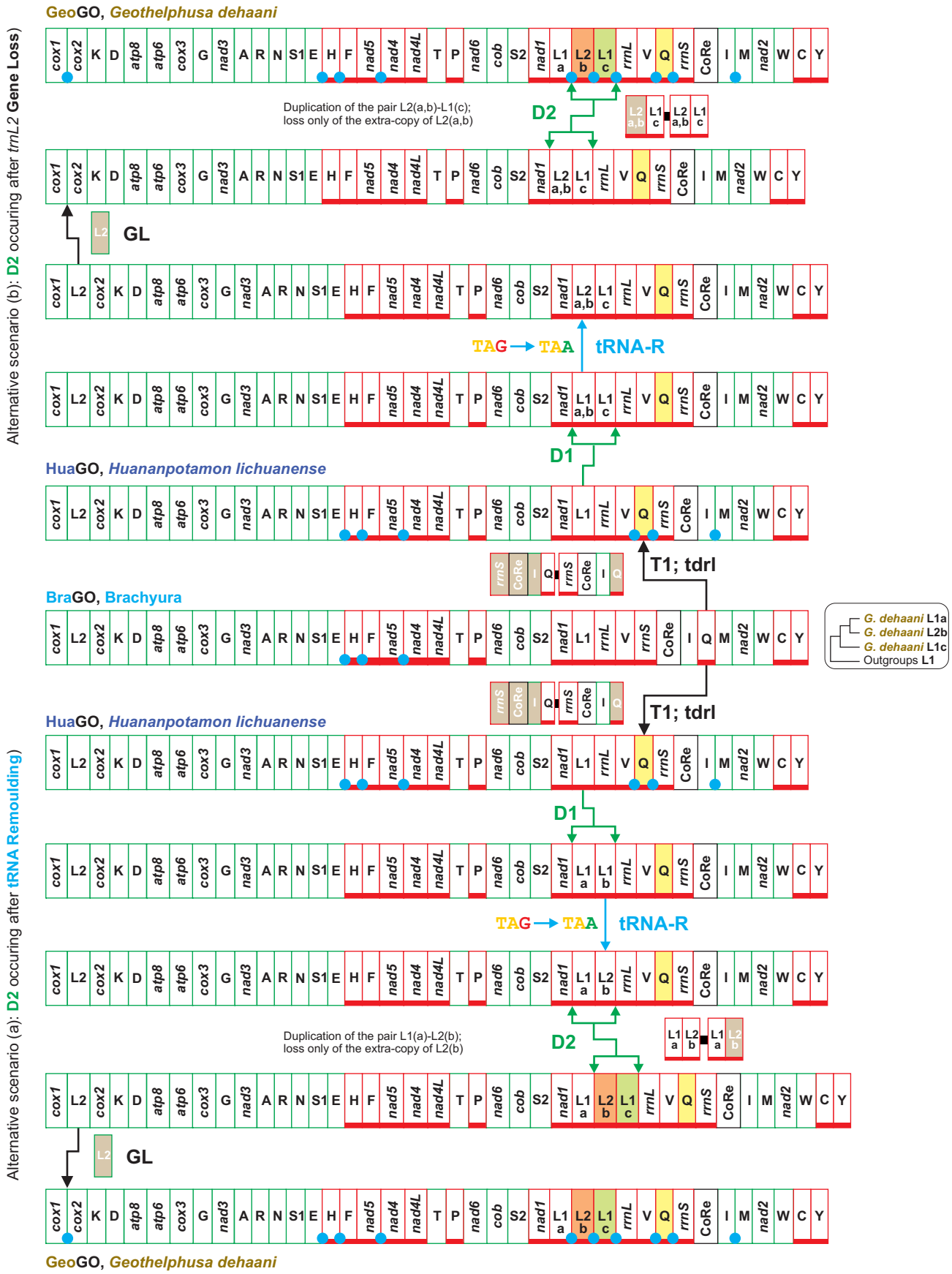


Figure S10. Alternative transformational pathways for GeoGO.
The rearrangements in the GOs of Potamid species are investigated and depicted with respect to BraGO. **T1**, transposition event; tdlr, duplication random loss, mechanism producing the observed re-arrangement; **D1-D2**, gene duplication events; **tRNA-R**, tRNA remoulding event; **GL**, gene loss event. The genomic and genetic nomenclature, as well as the colour scheme, are the same as in Figure 1 of main text.