Evolution of Mitochondrial Gene Orders in Echinoderms

Marleen Perseke [c], Guido Fritzsch [a,c], Kai Ramsch [b], Matthias Bernt [b], Daniel Merkle [b], Martin Middendorf [b], Detlef Bernhard [c], Peter F. Stadler [d,a,e,f,g,*], Martin Schlegel [c]

[a] *Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*

[b] *Parallel Computing and Complex Systems Group, Faculty of Mathematics and Computer Science, University of Leipzig, Johannisgasse 26, D-04103 Leipzig, Germany*

[c] *Institute of Biology II: Zoologie, Molekulare Evolution und Systematik der Tiere, University of Leipzig, Talstrasse 33, D-04103 Leipzig, Germany*

[d] *Bioinformatics Group, Department of Computer Science, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*
`studla@bioinf.uni-leipzig.de`

[e] *Department of Theoretical Chemistry University of Vienna, Währingerstraße 17, A-1090 Wien, Austria*

[f] *Fraunhofer Institut für Zelltherapie und Immunologie – IZI Deutscher Platz 5e, 04103 Leipzig, Germany*

[g] *Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA*

Abstract

A comprehensive analysis of the mitochondrial gene orders of all previously published and two novel *Antedon mediterranea* (Crinoidea) and *Ophiura albida* (Ophiuroidea) complete echinoderm mitochondrial genomes shows that all major types of rearrangement operations are necessary to explain the evolution of mitochondrial genomes. In addition to protein coding genes we include all tRNA genes as well as the control region in our analysis. Surprisingly, 7 of the 16 genomes published in the GenBank database contain misannotations, mostly unannotated tRNAs and/or mistakes in the orientation of tRNAs, which we have corrected here. Although the gene orders of mt genomes appear very different, only 8 events are necessary to explain the evolutionary history of echinoderms with the exception of the ophiuroids. Only two of these rearrangements are inversions, while we identify three tandem-duplication-random-loss events and three transpositions.

Key words: Mitochondrial genomes, Echinodermata, rearrangement, tandem-duplication-random-loss, inversion, transposition

1 Introduction

Mitochondrial genomes have been a particularly fruitful data set for phylogenetic reconstructions due to their limited size and the availability of a large number of informative data sets. In addition to the sequence data of its proteins, rRNAs, and tRNAs, the order of the genes on the circular mitogenome of animals has received extensive attention as a phylogenetic marker since the seminal work of Watterson *et al.* (1982); Sankoff *et al.* (1992). So far, most *computational* approaches utilizing mitochondrial gene orders for phylogenetic reconstruction use either inversions and transpositions as the sole edit operation (e.g. Sankoff (1993); Blanchette *et al.* (1999)), or employ an easy to compute distance measure such as break point distance (Blanchette *et al.*, 1999). The disadvantage of distance-based methods is that they do not necessarily reconstruct a sequence of evolutionary events.

In the last years it has become evident that inversions and transpositions are not the only mechanism that reshapes mitogenomes. Lavrov *et al.* (2002) suggest the evolution of gene orders in millipeds by duplication of the entire mitogenome followed by blocks of gene loss. A similar event explains e.g. the gene order in eels (Inoue *et al.*, 2003). More recently, Mueller and Boore (2005) showed that extensive gene rearrangement in lung-less salamanders (Plethodontidae) can be understood in terms of a duplication of part of the mitochondrial genome and subsequent differential gene loss. In this case, presumably functional copies of mitochondrial genes as well as multiple identify able pseudogenes provide direct support for the duplication-mediated mechanism. Both tandem and non-tandem duplications are present in these genomes, suggesting different duplication mechanisms. Another intriguing example are the frequent rearrangements of the "WANCY" tRNA cluster in certain vertebrate lineages (San Mauro *et al.*, 2006).

While a *"tandem-duplication-random-loss"* mechanism may explain many or most of the observed rearrangements, inversions are better explained by *"intermitochondrial recombination"* (Shao and Barker, 2003; Miller *et al.*, 2004). Furthermore, one effectively observes inversions (often called "reversals" in the computer science literature), transpositions, and reverse transpositions (Figure 1) so frequently that for practical purposes they constitute elementary operations, even though mechanistically they may be the results of duplications with subsequent *non-random* loss.

In contrast to inversions, transpositions, and reverse transpositions, which are "reversible" operations in the sense that they do not imply a direction, tandem-duplication-random-loss (TDRL) moves are in general "irreversible" (Chaudhuri *et al.*, 2006) and hence imply a direction of the corresponding edge in the phylogenetic tree. This property makes them particularly valuable

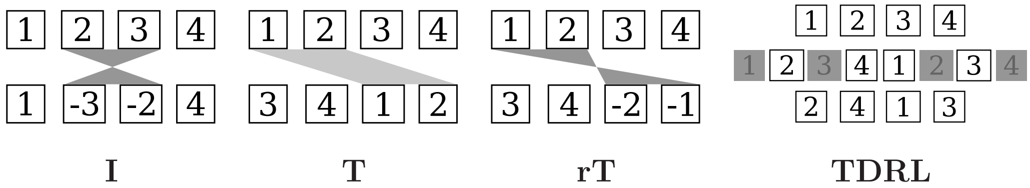

Fig. 1. Rearrangement operations acting on mitogenomes. From left to rigth: inversion (**I**), transposition (**T**), reverse transposition (**rT**), tandem-duplication-random-loss (**TDRL**).

for phylogenetic studies and suggests that a detailed reconstruction of the rearrangement history of mitogenes can lead to more detailed and more certain phylogenetic conclusions.

In this contribution, we use Echinodermata as a case study. The purpose of our contribution is three-fold. First, we describe the two newly sequenced complete mitogenomes of the crinoid *Antedon mediterranea* and the ophiuroid *Ophiura albida*. Second, we re-annotate all available mitogenomes from echinoderms to test for new and improved tools for mitogenome annotation. To our surprise we identified a number of minor mistakes, inaccuracies, and omissions in the GenBank annotation and to lesser extent also in the published literature. Most of these problems affect tRNAs and rRNAs, and may influence phylogenetic analysis based on these data. Third, we reconstruct – to the extent possible – the sequence of genome rearrangement events within echinoderms.

2 Materials and Methods

2.1 Specimens, DNA Extraction, Amplification, and Sequencing

Specimens of *Antedon mediterranea* and *Ophiura albida* were kindly supplied from the Natural History Museum Senckenberg (Frankfurt). *Ophiura albida* was collected from Helgoland (Germany) and *Antedon mediterranea* from Mediterranean sea (Croatia). The DNA was prepared from tissue of ethanol-preserved specimens by phenol-chloroform extraction following proteinase K digestion. Parts of *Cox1*, *Cox3*, *CytB*, and *ND4-ND5* were amplified using degenerate oligonucleotide primers, which were designed from consensus sequences of echinoderms, and used to design specific "Long PCR primers" for both species. The amplification of the complete mt genome from both species was done with three overlapping fragments using the Long PCR Enzym Mix Kit (Fermentas). The fragments were sequenced directly by primer walking using an ABI Prism 3100 automated sequencer and BigDye Termination v3.1 Cycle Sequencing Kit (Applied Biosystems) in both directions. A complete list of the primer sequences and their positions may be obtained from the supplement material. Protein coding genes and the rRNA genes were identified by alignments with homologous echinoderm genes. The boundaries of the rRNA genes were determined by the flanked genes. All tRNA genes were identified by the program tRNAscan-SE version 1.23 (Lowe and Eddy, 1997). The mt genomes of *A. mediterranea* and *O. albida* have been submitted to GenBank under the accession numbers ***AM404181*** and ***AM404180***.

2.2 Analysis of the unassigned sequence fragments

Following the annotation of the two novel mitogenomes, further sequence analysis focused on all echinoderm mitochondrial sequences which were not identified as tRNAs, rRNAs, mitochondrial protein-coding genes, or putative control region in the corresponding GenBank entries. Unassigned sequences of echinoderm species were aligned using NCBI `blast` (version 2.2.14, nucleotide match score 2, nucleotide mismatch cost 1, gap open cost 2 and gap extension cost 1) against all echinoderm mitochondrial sequences. Local alignments with $E < 10^{-5}$ were analyzed further by retrieving and comparing the annotation of the matching database sequences. The entire procedure was incorporated into an automatic pipeline that summarizes the search results and displays them on a web browser. The weblink for an automatic tool may be obtained from the authors on request. In addition, we again used `tRNAscanSE`.

2.3 Genome rearrangements

Genome rearrangements were studied using the new `CREx` tool developed by this group of authors. The software, which is available at `http://pacosy.informatik.uni-leipzig.de/crex`, is described in more detail in Bernt *et al.* (2007). Thus we only sketch the methods here. The basic idea is to employ an efficient tree data-structure that represents the gene-orders of pairs of sister taxa in such a way that the different rearrangement operations are easy to identify in most cases.

Since gene groups are often preserved during evolution, it is important that methods for reconstructing rearrangements respect constraints. For the computation of optimal *sorting scenarios* (more precisely, sequences of a minimal number of inversions respecting such constraints that transforms two given gene orders into each other), Bérard *et al.* (2007) recently proposed so-called strong interval trees, a variant of so-called PQ-trees (Booth and Leuker, 1976; Parida, 2006). It turns out that this data structure is particularly suitable when gene groups have to be conserved throughout a sorting scenario. Mathematically, this conservation constraint is reflected by *common intervals*: Given

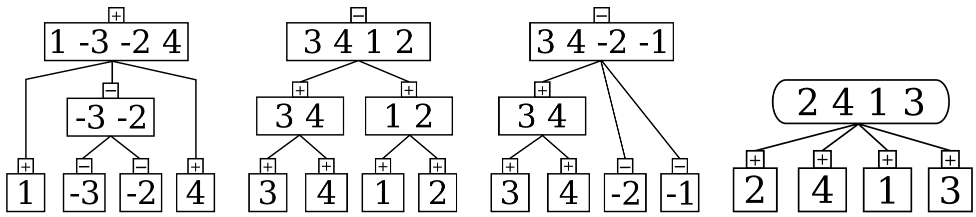

Fig. 2. Patterns arising in strong interval trees through the application of one rearrangement. One of the two gene orders in the example is 1234 and the other gene order has been obtained by one of the following operations (the corresponding order can be found in the root of the tree) a) inversion, b) transposition, c) reverse transposition, d) tandem duplication random loss (a prime node where the children can be sorted with one TDRL). Prime nodes are depicted by ellipses, and linear nodes by rectangles, where the sign in the square on top of a node indicates if the node is increasing or decreasing.

two gene orders, a set of genes is a common interval (Bérard *et al.*, 2007) if the genes in that set appear consecutively in both gene orders. A certain subset of all common intervals, the so-called strong common intervals can be computed very efficiently and determines all common intervals. The strong common intervals have the property that they do not overlap each other, i.e., they are either disjoint or one is completely contained in the other. Therefore, the set of strong common intervals can be represented as the nodes of a special type of tree. The children of a node (strong common interval) are simply the strong common intervals that it includes entirely. The nodes are given a sign. If the children of a node appear in the same order in both input gene orders, the node is called "linear increasing" (+); if the children of a node appear in exactly the opposite order, it is "linear decreasing" (-); otherwise the node is called prime. The importance of this common interval tree is that it greatly facilitates the identification of the genome rearrangement operations, see figure 2 for an example.

Currently, we consider four operations: inversions (reversals), transpositions, reverse transpositions, and tandem duplication random losses (TDRLs). One can show that TDRLs always lead to prime nodes. Hence, prime nodes in strong interval trees are a good indicator for TDRLs.

In order to resolve a rearrangement problem between two sister taxa the strong interval tree of the corresponding gene orders is searched for patterns that belong to one of the four types of rearrangement operations. When the entire rearrangement can be resolved by these operations, we next try to determine for each operation on which of the two branches to the sister taxa it occurred. For inversions, transpositions, or reverse transpositions this can not be determined without additional information such as the gene order of an outgroup. The distance measure between gene orders that is based on the minimum number of TDRLs is not symmetric (Chaudhuri *et al.*, 2006). In particular, in many cases a rearrangement can be explained by a single TDRL in one direction, while reversing the rearrangement would require more than a single operation. In such cases TDRLs yield directional information, allowing the inference of the ancestral state from the comparison of only two taxa. This inference can then be further verified by comparison with a gene order of an outgroup.

2.4 Phylogenetic Analysis

To determine the phylogenetic position of *A. mediterranea* and *O. albida* their mitochondrial sequences were aligned together with all available mitochondrial sequences of echinoderms (accession numbers in Table 1) and two hemichordates (NC_001887 and NC_007438). All sequences of the 13 protein coding genes were aligned separately. The two complete mt genomes of Hemichordata were used as outgroup due to their close relationship to the Echinodermata. The sequences were aligned using the `ClustalW` algorithm implemented in Mega 3.1 (Kumar *et al.*, 2004) with default parameters and the Gonnet protein weight matrix.

The different evolution models and their parameters were tested with the program `ProtTest` (version 1.3) (Abascal *et al.*, 2005) and then used for the phylogenetic reconstruction with neighbor joining, maximum likelihood and bayesian analysis. MP and NJ analyses were performed with `PAUP` (Swofford, 2002), maximum likelihood with `PHYML` v.2.4.4 (Guindon and Gascuel, 2003) and bayesian analysis with `MrBayes` v.3.1.2 (Ronquist and Huelsenbeck, 2003). To test the robustness of bifurcations, bootstrap analysis were calculated with 10.000 replicates for NJ, 2.000 replicates for MP and 100 replicates for ML. Bayesian analysis was run with the best-fit model as inferred by ProtTest for 1.000.000 generations, with a sampling frequency of 10 generations. From the 100.000 trees found, we determined a subset of trees for building the consensus tree by inspecting likelihood values of trees saved by MrBayes. The burn-in was set to 25.000 trees to ensure that stable likelihood values were achieved.

3 Results and Discussion

3.1 Two Novel Mitochondrial Genomes

The complete mt genomes have been determined for the crinoid *Antedon mediterranea* and the ophiuroid *Ophiura albida*. The mitochondrial genomes of *A. mediterranea* (16,169 bp) and *O. albida* (16,580 bp) are circular DNA molecules and contain the complete set of 13 protein coding genes, 22 tRNA genes and 2 rRNA genes, as found in other mitochondrial genomes of Metazoa.

Most of the protein coding genes initiate with ATG except of ND5 (GTG) and ATP8 (GTG) in *A. mediterranea*. All protein coding genes of both echinoderm species possess full termination codons (TAA or TAG).

Within echinoderms, a strand-specific nucleotide bias of the third codon position of protein-encoding genes is reported (Scouras *et al.*, 2004) and the references therein). Hence, mitochondrial protein genes of the major sense strand demonstrate a low T/C ratio in the third codon position whereas mitochondrial protein genes of the minor sense strand possess a higher T/C ratio (Scouras *et al.*, 2004). Only the nucleotide composition of crinoids demonstrates an exactly opposite orientated nucleotide bias (Scouras *et al.*, 2004). This noteworthy heterogeneity is also found in this study: the nucleotide composition of *O. albida* reveals the main nucleotide bias, whereas *A. mediterranea* shows the exactly opposite orientated nucleotide bias (Suppl. Mat.).

The mitochondrial gene order of both, *A. mediterranea* and *O. albida*, (Figure 3) corresponds to other arrangements in echinoderms (Boore, 1999). The gene order of *O. albida* is identical to the gene order of *O. lutkeni* (Scouras *et al.*, 2004). In *O. albida* three UAS regions are found, which also occur at the same locations compared to *O. lutkeni*. Similar to *O. lutkeni*, the UASI region (686 bp) is flanked by genes ND6 and tRNA^{Gly} and may be involved in regulation of the mitochondrial replication because it contains a guanine stretch enclosed by TA-rich sequences. UAS II (492 bp) region is flanked by the genes tRNA^{Gly} and 16S rRNA and UAS III (62 bp) region is located between the tRNA^{Thr} and tRNA^{Trp} genes.

Both, UAS I and UAS II regions of *O. albida* possess repetitive elements: UAS I contains five identical tandem repeats (31 bp) followed by the first 22 bp of a sixth repeat; UAS II includes two identical tandem repeats (212 bp) followed by the first 46 bp of a third repeat. In *O. lutkeni* only the UAS I region reveals six complete identical tandem repeats with a length of 167 bp followed by one incomplete repeat with 32 bp (Scouras *et al.*, 2004).

In contrast, the gene order of *A. mediterranea* reveals some unique rearrangements. There are two distinct "hot spots" where some rearrangements occurred. The first "hot spot" is located between the genes tRNA^{His} and ND5. Between these genes, only the $\text{tRNA}^{\text{Ser1}}$ is located in all known mt genomes of echinoderms. In *A. mediterranea*, the $\text{tRNA}^{\text{Ser1}}$ gene and a UAS region (UAS IV, 170bp) flank tRNA^{Arg} and ND4L which is normally located between Cox1 and Cox2. One further exception to this rule has been described previously: in the crinoid *Gymnocrinus richeri* where the gene ND4L is located between the

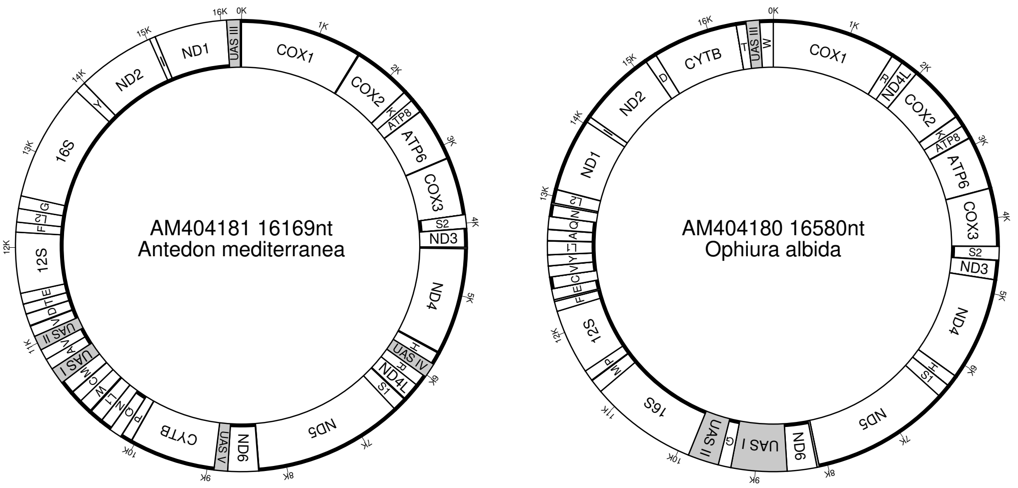

Fig. 3. Maps of the mitogenomes of the crinoid *Antedon mediterranea* and the ophiuroid *Ophiura albida*. The images were generated from the GenBank files with the mitochondrial visualization tool 'mtviz' (application note in preparation). It can be found at `http://pacosy.informatik.uni-leipzig.de/mtviz`.

genes ATP6 and CoxIII. Interestingly, a part of the UAS IV shows high similarity with tRNA$^{\mathrm{Ser1}}$ (see Suppl. Mat.) and thus most likely arose by means of a duplication event.

The second "hot spot" surrounds the putative control region. In the three known mt genomes of crinoids, the putative control region is located within a 12-tRNA-gene cluster that is flanked by CytB and the 12S rRNA. In comparison, few single tRNA rearrangements (tRNA$^{\mathrm{Ala}}$, tRNA$^{\mathrm{Val}}$, tRNA$^{\mathrm{Asp}}$) occurred in the mitochondrial genome of *A. mediterranea*. These caused a shift in the putative control region (UAS I, 200 bp). Furthermore, two variants of the tRNA$^{\mathrm{Val}}$ gene were found. These are separated by the UAS II (175 bp) which exists only in *A. mediterranea* and exhibits a high similarity with the region containing the genes tRNA$^{\mathrm{Asp}}$, tRNA$^{\mathrm{Thr}}$ and tRNA$^{\mathrm{Glu}}$ (Suppl. Mat.). We therefore conclude that the UAS II region and the tRNA$^{\mathrm{Val}}$ gene in this region originated from a tandem duplication of the mitochondrial fragment containing the genes tRNA$^{\mathrm{Val}}$, tRNA$^{\mathrm{Asp}}$, tRNA$^{\mathrm{Thr}}$ and tRNA$^{\mathrm{Glu}}$.

The UAS III (76 bp) and UAS V (69 bp) regions of *A. mediterranea* are identical in location compared to other mitochondrial genomes of crinoids. UAS III region is flanked by the genes ND1 and CoxI and UAS V region is located between the genes ND6 and CytB.

To our surprise, not all database entries of previously published echinoderm mitogenomes feature the complete set of 22 tRNAs and 2 rRNAs characteristic for metazoan mitochondria. We have therefore re-annotated all published echinoderm mitogenomes using the new tool described in Sect. 2.2 above. The results are summarized in Table 1, showing that features missing from the annotation are indeed present in the respective genomes. The only exception is the mitogenome of *Pisaster ochraceus* ***NC_004610*** (Smith *et al.*, 1990) which is not totally complete; tRNA^{Glu} and tRNA^{Thr} appear to be located in the missing data.

While the `blast` support alone is quite weak for some tRNAs, the results agree with `tRNAscanSE` and have been corroborated by manual inspection. In addition to missing annotation features in *Luidia quinalia* ***NC_006664***, *Asterias amurensis* ***NC_006665***, *Astropecten polyacanthus* ***NC_006666***, there are many cases in which the orientation of tRNAs is incorrectly annotated.

Figure 4 summarizes the distribution of UAS in echinoderm mitogenomes. The majority of the UAS is shorter than 20 nt, at least roughly following an exponential length distribution. UAS with a length of 60-100nt are often tRNAs missed in the annotation or remnants of duplication events in which tRNA pseudogenes can be found. Most of the very long UAS are rRNAs. The most interesting cases for further study are those with a length $100 < n < 500$, such as UASII in *Antedon*, see above.

3.3 Mitochondrial gene rearrangements

As mentioned above, *O. albida* possesses the same gene order as *O. lutkeni*, which differs substantially from *Ophiopholis aculeata* (Scouras *et al.*, 2004). For ophiuroids thus only two distinct gene orders, which are very different, are known. Unfortunately, `CREx` is not able to resolve a plausible rearrangement sequence because too little is conserved between the two orders. Thus, as in Scouras *et al.* (2004), the ancestral state of ophiuroids remains unresolved.

The gene order of *A. mediterranea* is most similar to the consensus gene order of crinoids, which is represented by *F. serratissima* and *P. gracilis* (Scouras and Smith, 2001). It deviates, however, from this consensus in three areas. The most unusual features are two variants of tRNA^{Val} gene. They are more similar to each other than to any other echinoderm tRNA^{Val} gene (Suppl. Mat.). The duplicated tRNA^{Val} is absent in *Antedon bifida* (M.P., unpublished data) and hence probably originated in a very recent duplication.

Table 1
Missing or wrong annotations in the GenBank files, changes are highlighted; acc: the accession number; gene: the name of the gene; codon: the codon of tRNAs; ±, start, end: orientation, start and end positions, support: t: tRNAscan support value, b: number of species, where blast hits occurred in the gene of the missing annotation

species	acc.	gene	codon	±	start	end	support
Strongylocentrotus purpuratus	***NC_001453***						
Paracentrotus lividus	***NC_001572***						
Patiria pectinifera	***NC_001627***	Phe	GAA	**+**	5930	6001	t: 56.62
		Asn	GTT	**-**	11910	11981	t: 46.36
Arbacia lixula	***NC_001770***						
Florometra serratissima	***NC_001878***						
Pisaster ochraceus	***NC_004610***	Glu					
		Thr			16S -T	-E -12S	
Ophiopholis aculeata	***NC_005334***						
Cucumaria miniata	***NC_005929***						
Ophiura lutkeni	***NC_005930***						
Luidia quinalia	***NC_006664***	**Asp**	**GTC**	**+**	**3583**	**3652**	t: 14.7 b: 3
		Trp	**TCA**	**-**	**3869**	**3938**	t: 50.48 b: 13
		Ser	???	**+**	**10801**	**10869**	b: 7
		12S		**+**	**>14529**	**<15412**	b: 15
		16S		**-**	**>15556**	**<1212**	b: 16
Asterias amurensis	***NC_006665***	Ile	GAT	**-**	1087	1158	t: 33.81
		Trp	**TCA**	**-**	**2706**	**2773**	t: 39.17 b: 11
		Leu	TAG	**-**	2857	2928	t: 53.94
		Lys	**CTT**	**+**	**5783**	**5857**	t: 8.77 b: 7
		Ser	**GCT**	**+**	**9555**	**9620**	t: 9.15 b: 8
		12S		**+**	**>13285**	**<14177**	b: 17
		16S		**-**	**>14325**	**<24**	b: 17
Astropecten polyacanthus	***NC_006666***	Cys	GCA	**-**	4261	4329	t: 26.4
		Trp	**TCA**	**-**	**4330**	**4398**	t: 39.68 b: 14
		Leu	TAG	**-**	4470	4542	t: 52.81
		Lys	**CTT**	**+**	**7403**	**7474**	t: 9.78 b: 6
		Ser	**GCT**	**+**	**11208**	**11274**	t: 5.78 b: 8
		12S		**+**	**>14912**	**<15812**	b: 16
		16S		**-**	**>15956**	**<1696**	b: 17
Gymnocrinus richeri	***NC_007689***						
Phanogenia gracilis	***NC_007690***						
Acanthaster planci	***NC_007788***	Gly	TCC	-	3773	3841	t: 22
Acanthaster brevispinus	***NC_007789***	Gly	TCC	-	3769	3837	t: 27.75

In crinoids, all rearrangements can be completely resolved by means of `CREx`: starting from the consensus gene order of crinoids we observe one transposition (ND4L + tRNA$^{\mathrm{Arg}}$ = `T2`), and one TDRL event of 6 tRNA genes and the

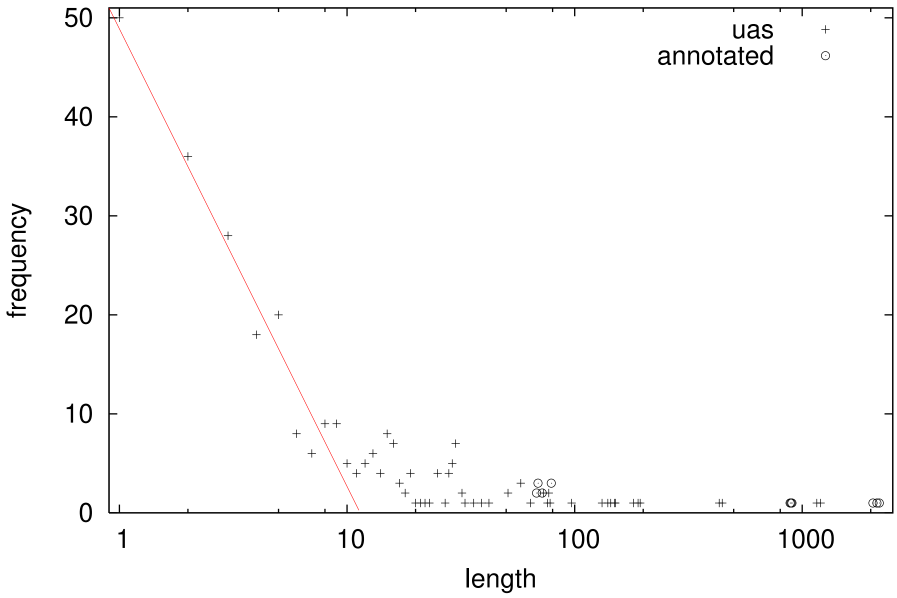

Fig. 4. Length distribution of unassigned sequences (UAS) in published echinoderm mitochondrial genomes. Many of the longer UAS can be annotated (from left to right) marked by circles: tRNAs: 68 ***NC_006666*** tRNA^{Ser}; 69 ***NC_006664*** tRNA^{Ser}, ***NC_006665*** tRNA^{Ser}, and ***NC_006666*** tRNA^{Trp}; 72 ***NC_006664*** tRNA^{Asp}, and ***NC_006666*** tRNA^{Lys}; 79 ***NC_006664*** tRNA^{Trp}, ***NC_006665*** tRNA^{Trp}, and ***NC_006665*** tRNA^{Lys}. Ribosomal RNAs: 884 ***NC_006664*** 12S rRNA; 893 ***NC_006665*** 12S rRNA; 901 ***NC_006666*** 12S rRNA; 2045 ***NC_006666*** 16S rRNA; 2127 ***NC_006665*** 16S rRNA; 2181 ***NC_006664*** 16S rRNA. The thin line emphasizes the exponential distribution of very short unassigned sequences.

control region (TDRL2), see figure 5. Following this event, UAS II was created by duplication of the four tRNAs (tRNA^{Val}-tRNA^{Asp}-tRNA^{Thr}-tRNA^{Glu}) and subsequent loss of the copies of tRNA^{Asp}, tRNA^{Thr}, and tRNA^{Glu}.

The reconstructed ancestral state of crinoids is identical to the gene orders of *F. serratissima* and *P. gracilis*, which is in agreement with Scouras and Smith (2006).

The known gene orders of the other echinoderm groups (Asteroidea, Echinoidea and Holothuroidea) demonstrate no rearrangements within each class and only few rearrangements are necessary to transform the gene orders from one group into each other (see Suppl. Mat.). One inversion of 16 genes (I1) is common to all asteroids and echinoids, see e.g. (Asakawa *et al.*, 1995; Smith *et al.*, 1990) and the references therein. At present, only a single holothuroid gene order is known which can be deduced from the echinoid gene order by a single TDRL event of a tRNA cluster (TDRL1) (Arndt and Smith, 1998). The

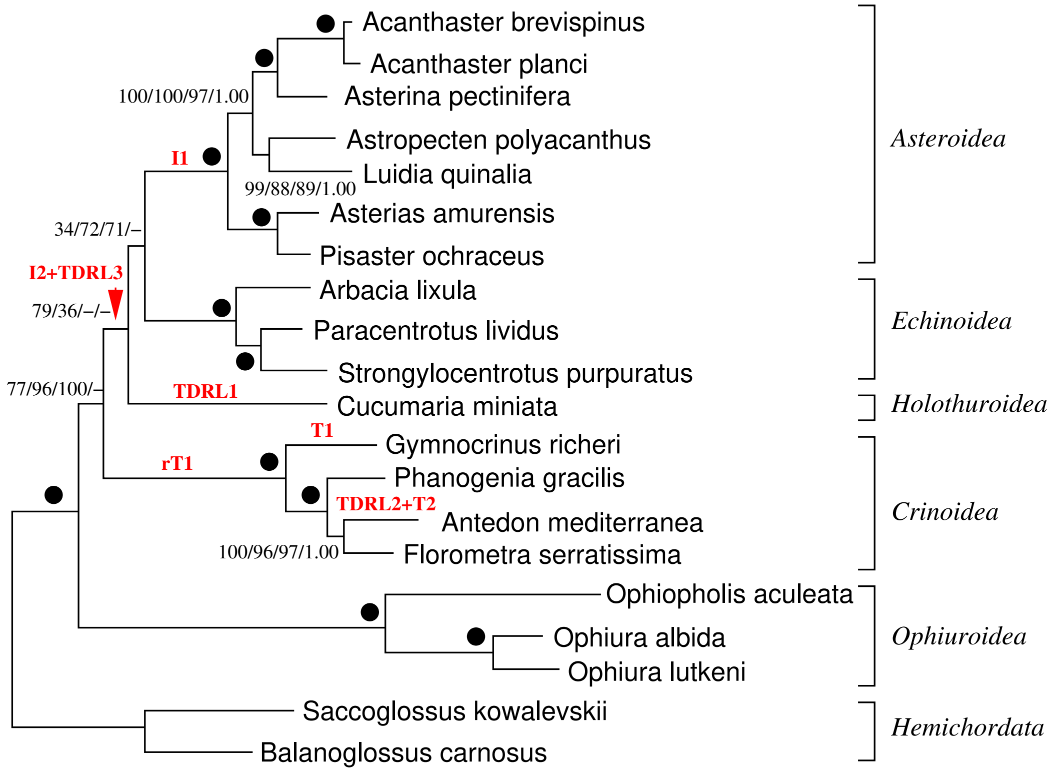

Fig. 5. Maximum likelihood analysis of thirteen protein coding genes of the echinoderm mt genomes. The numbers show the bootstrap values for ML/NJ/MP and the posterior probability of bayesian analysis. If this values are 100 resp. 1.00 they are represented by solid points to save space. The tree has been rooted with two hemichordates. Branch lengths are proportional to evolutionary distance.

`CREx` analysis implies that Echinoidea represents the ancestral state within these three groups because all events occurred on the branches leading to the Asteroidea or Holothuroidea. However, the phylogenetic relationships between these three classes cannot be unambiguously defined based on both genome rearrangements and on the sequence analysis.

The differences between the echinoid gene order and ancestral crinoid gene order can be explained by one inversion (I2), see (Scouras *et al.*, 2004), followed by one tandem duplication random loss (TDRL3), and one reverse transposition (rT1). The I2+TDRL3 event affects the 16S rRNA, the genes ND1, ND2, as well as tRNA^{Ile}, $\text{tRNA}^{\text{Leu2}}$, tRNA^{Gly}, and tRNA^{Tyr} and consists of an inversion together with duplication and subsequent reduction to the normal gene set. We suggest that the I2+TDRL3 event are mechanistically coupled, i.e., constitute a single rearrangement event. Alternatively, the putative I2+TDRL3 event can be explained by an inversion with two additional single tRNA rearrangements (see Suppl. Mat.), or as a decoupled sequence of an inversion and TDRL3. The direction of the TDRL3 implies that it occurred on

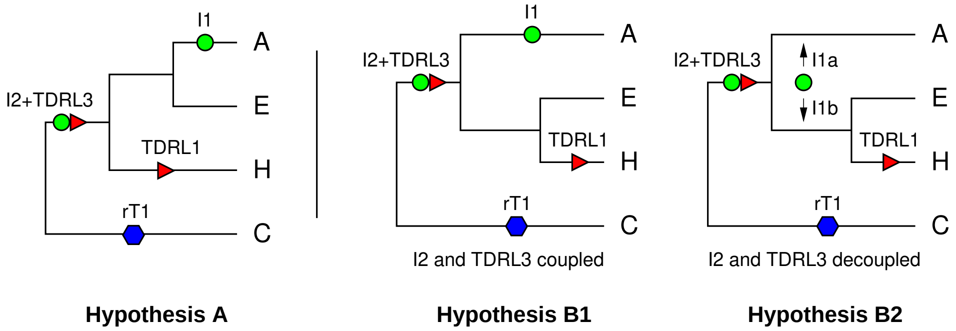

Fig. 6. Two phylogenetic hypotheses are consistent with the most parsimonious rearrangement scenario. Hypothsis A: based on phylogenetic analyses of the amino acid sequences. This scenario implies that the events I2 and TDRL3 must be coupled. Hypothsis B: based on the gene order analysis and in this scenario the events I2 and TDRL3 can be either coupled or not. Both variants are shown. A = Asteroidea, E = Echinoidea, H = Holothuroidea, C = Crinoidea

the branch leading to echinoids, asteroids and holothuroids. In contrast, the reverse transposition rT1 of the fragment containing the 12S rRNA and three tRNAs provides no direct information about its location on the two branches. Scouras and Smith (2001); Scouras *et al.* (2004); Scouras and Smith (2006) suggested, based mostly on the nucleotide bias in crinoids and the putative reverse orientation of the control region relative to the protein-coding genes, that the reverse transposition occurred in the crinoid lineage. This is consistent with our data. Accepting that rT1 is crinoid specific and that I2+TDRL3 are indeed part of the same rearrangement event, we can derive the uniquely determined gene order for the ancestor of asterioids, echinoids, holothuroids, and crinoids. It is the crinoid arrangement without the reverse transposition rT1. Furthermore, the ancestral arrangement of the echinoid, holothuroid, and asteroid ancestor coincides with the extant echinoid gene order. This scenario is consistent with two phylogenetic hypotheses (A and B1), figure. 6

If we reject the coupling of the I2 and the TDRL3 events (Hypothesis B2 in figure 6), an alternative scenario becomes plausible, which places a inversion I1 on the Echinozoa (echinoid+holothuroid) branch and makes the Asteroidea gene order the ancestral state of the echinoid+holothuroid+asteroid group. Note that in this scenario the inversion I1 implies two different gene orders (see Suppl. Mat.) depending on which group (Asteroidea, I1a, or Echinozoa, I1b) represents the ancestral state.

In the analysis reported here we have included the (putative) control regions (as annotated in GenBank and/or the corresponding literature). Inclusion of the control region resulted in better support for TDRL2 (reversing the direction of the rearrangement now requires two TDRLs instead of only two transpositions). Interestingly, most of the eight rearrangements contain or are located close to the control region. A frequent involvement of the control region in genome rearrangements was also noted in chordates (Boore and Brown, 1998).

In figure 5 we map the rearrangement operations determined by `CREx` to the consensus phylogenetic tree obtained from a careful analysis of the mitochondrial protein sequences. First, we observe that the results of `CREx` are consistent with the molecular phylogeny. The `CREx` data are, however, not sufficient to completely resolve the phylogenic relationships. We note that the gene order analysis fails to provide unambiguous information exactly for those nodes that contradict the preferred phylogenetic hypothesis, in particular the position of the ophiuroids, see below.

3.4 Phylogenetic analysis

Phylogenetic analyses of the amino-acid sequences of all 13 protein-coding genes resulted in mostly congruent tree topologies but also yielded some differences dependent from the analysis method used (Figure 5). All five echinoderm classes could be recovered as monophyletic groups in each case.

Furthermore, rooting the trees with the two hemichordate species always opposed the Ophiuroidea to a clade containing crinoids, asteroids, echinoids and holothuroids. This branching is highly supported by all analyses methods. There is, however, some evidence for long branch attraction presumably caused by the rapid evolution of Ophiuroidea, which is consistent with their heavily rearranged gene orders.

Within the remaining groups the crinoids branch off next well supported by ML, MP, and NJ analyses. However, bayesian analysis resulted in a clearly different branching order in which the Holothuroidea branch off first followed by the Asteroidea and the Echinoidea and Crinoidea [O(H(A(C+E)))]. If this scenario is true, it becomes necessary, however, to postulate additional genome rearrangement events at the base of the crinoid lineage. The grouping composed of Holothuroidea, Echinoidea and Asteroidea as well as the clade consisting of Asteroidea and Echinoidea are only weakly supported by the aminoacid sequence data. On the other hand, it is strongly supported by the gene orders.

The molecular phylogeny showing the Ophiuroidea most basal within the Echinodermata is in conflict with traditional views of echinoderm phylogeny (see e.g. (Littlewood *et al.*, 1997)), but agrees in part with other molecular (Scouras *et al.*, 2004; Janies, 2001) and morphological (Gudo, 2005) analyses. Since the ancestral gene order of Ophiuroidea could not be resolved, its phylogenetic position as suggested by these characters remains inconclusive.

4 Conclusion

We present here a comprehensive analysis of 16 complete echinoderm mitogenomes from the literature together with novel data for *Antedon mediterranea* and *Ophiura albida*. In contrast to previous studies, we also consider all tRNAs and the control region in our automated analysis. Consistent to the literature, a re-annotation of the GenBank data showed that all Asteroidea share the same gene order. The same is true for echinoids. These two classes differ by a single inversion of 16 genes (Smith *et al.*, 1990; Asakawa *et al.*, 1995). In contrast, several rearrangement events can be observed within crinoids and ophiuroids.

In order to study the rearrangements scenarios in more detail we used here the new web-based tool `CREx` (Bernt *et al.*, 2007). This program utilizes so-called strong interval trees, a data structure that highlights the differences between two input gene orders, as a basis for a heuristic reconstruction of rearrangement scenarios. At present, `CREx` can identify inversions, transpositions, reverse transpositions, and tandem-duplication-random-loss events. Duplications of at least two tRNAs in the genome of *Antedon mediterranea* provide further evidence for a dominating duplication-based mechanism for mitochondrial genome rearrangements. Interestingly, the reconstructed events affect predominantly tRNAs and most of them contain the control region. Using these full data, we are able to resolve the evolution of echinoderm gene orders with the exception of the ophiuroids. The two known gene orders within this class are so different that it has been impossible to reconstruct a credible ancestral state. In order to resolve this issue, additional data will be indispensable.

The evolution of echinoderm gene orders (without ophiuroids) can be explained with only 8 events: two inversions, three transpositions (two within crinoids), and three TDRLs. The latter events are phylogenetically of particular interest since they provide directional information. One of the TDRLs separates *Cucumaria* from the ancestral gene order of echinoids, another one occurred within crinoids. The third TDRL groups holothuroids, echinoids, and asteroids together. The same tree topology was obtained from the aminoacid sequence data using multiple analysis methods.

Acknowledgements. We thank Michael Gudo for collecting specimens of *Antedon mediterranea* and *Ophiura albida*. This work was partially supported by DFG grants SCHL 229/14-1 and STA 850/3-1 as part of the SPP-1174 (Metazoan Deep Phylogeny).

References

Abascal F, Zardoya R, Posada D, 2005. Prottest: Selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105.

Arndt A, Smith M, 1998. Mitochondrial gene rearrangement in the sea cucumber genus, cucumaria. Mol Biol Evol 15:1009–1016.

Asakawa S, Himeno H, Miura K, Watanabe K, 1995. Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. Genetics 140:1047–1060.

Bérard S, Bergeron A, Chauve C, Paul C, 2007. Perfect sorting by reversals is not always difficult. IEEE ACM Transaction on Computational Biology and Bioinformatics 4:4–16.

Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, Bernhard D, Schlegel M, Stadler PF, Middendorf M, 2007. `CREx`: Infering genomic rearrangement based on common intervals. Bioinformatics Submitted.

Blanchette M, Kunisawa T, Sankoff D, 1999. Gene order breakpoint evidence in animal mitochondrial phylogeny. J Mol Evol 49:193–203.

Boore JL, 1999. Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780.

Boore JL, Brown WM, 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opinion Gen Devel 8:668–674.

Booth K, Leuker G, 1976. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. J Comp System Sci 13:335–379.

Chaudhuri K, Chen K, Mihaescu R, Rao S, 2006. On the tandem duplication-random loss model of genome rearrangement. In: Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms, (pp. 564–570). ACM.

Gudo M, 2005. Hydromechanical considerations on the evolution and diversification of the echinoderm bauplans. URL `http://arxiv.org/abs/q-bio.PE/0505036(q-bio.PE/050503)`.

Guindon S, Gascuel O, 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst Biol 52:696–704.

Inoue JG, Miya M, Tsukamoto K, Nishida M, 2003. Evolution of the deep-sea gulper eel mitochondrial genomes: large-scale gene rearrangements originated within the eels. Mol Biol Evol 20:1917–1924.

Janies D, 2001. Phylogenetic relationships of extant echinoderm classes. Canadian Journal of Zoology 79:1232–1250.

Kumar S, Tamura K, Nei M, 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5:150–163.

Lavrov DV, Boore JL, Brown WM, 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. Mol Biol Evol 19:163–169.

Littlewood DT, Smith AB, Clough KA, Emson RH, 1997. The interrelationships of the echinoderm classes: morphological and molecular evidence. Biol J Linn Soc Lond 61:409–438.

Lowe TM, Eddy SR, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25:955–964.

Miller AD, Nguyen TTT, Burridge CP, Austin CM, 2004. Complete mitochondrial DNA sequence of the australian freshwater crayfish, *Cherax destructor* (Crustacea: Decapoda: Parastacidae): a novel gene order revealed. Gene 331:65–72.

Mueller RL, Boore JL, 2005. Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. Mol Biol Evol 22:2104–2112.

Parida L, 2006. A PQ framework for reconstructions of common ancestors and phylogeny. In: Comparative Genomics, vol. 4205, (pp. 141–155). Berlin: Springer. RECOMB 2006 International Workshop, RCG 2006 Montreal, Canada, September 24-26, 2006 Proceedings.

Ronquist F, Huelsenbeck JP, 2003. Mrbayes 3: Bayesian phylogenetic inference under mixed models. Bioinf 19:1572–1574.

San Mauro D, Gower DJ, Zardoya R, Wilkinson M, 2006. A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. Mol Biol Evol 23:227–234.

Sankoff D, 1993. Analytical approaches to genomic evolution. Biochimie 75:409–413.

Sankoff D, Leduc G, Antoine N, Paquin B, Lang BF, Cedergren R, 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc Natl Acad Sci USA 89:6575–6579.

Scouras A, Beckenbach K, Arndt A, Smith JM, 2004. Complete mitochondrial genome dna sequence for two ophiuroids and a holothuroid: the utility of protein gene sequence and gene maps in the analyses of deep deuterostome phylogeny. Mol Phylogenet Evol 31:50–65.

Scouras A, Smith JM, 2001. A novel mitochondrial gene order in the crinoid echinoderm *Florometra serratissima*. Mol Biol Evol 18:61–73.

Scouras A, Smith JM, 2006. The complete mitochondrial genomes of the sea lily *Gymnocrinus richeri* and the feather star *Phanogenia gracilis*: signature nucleotide bias and unique nad4L gene rearrangement within crinoids. Mol Phylogenet Evol 39:323–34.

Shao R, Barker SC, 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. Mol Biol Evol 20:362–370.

Smith JM, Banfield DK, Doteval K, Gorski S, Kowbel DJ, 1990. Nucleotide sequence of nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star *Pisaster ochraceus*. J Mol Evol 31:195–204.

Swofford DL, 2002. `PAUP*`: Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4.0b10. Sunderland, MA: Sinauer Associates.

Handbook and Software.
Watterson GA, Ewens WJ, Hall TE, Morgan A, 1982. The chromosome inversion problem. J Theor Biol 99:1–7.