



# Monophyly of Anthozoa (Cnidaria): why do nuclear and mitochondrial phylogenies disagree?

Marine Pratlong, Corinne Rancurel, Pierre Pontarotti, D. Aurelle

## ► To cite this version:

Marine Pratlong, Corinne Rancurel, Pierre Pontarotti, D. Aurelle. Monophyly of Anthozoa (Cnidaria): why do nuclear and mitochondrial phylogenies disagree?. *Zoologica Scripta*, 2017, 46 (3), pp.363-371. 10.1111/zsc.12208 . hal-01474009

**HAL Id: hal-01474009**

**<https://hal.science/hal-01474009>**

Submitted on 26 Apr 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pratlong Marine

Adress: Station Marine d'Endoume  
Rue de la Batterie des Lions  
13007 Marseille  
FRANCE

Telephone number: 0033 4 91 04 16 31

email adress: [marine.pratlong@imbe.fr](mailto:marine.pratlong@imbe.fr)

## Monophyly of Anthozoa (Cnidaria): why do nuclear and mitochondrial phylogenies disagree?

PRATLONG Marine, RANCUREL Corinne, PONTAROTTI Pierre,  
AURELLE Didier

Analysis of phylogenomics studies of Cnidaria

Pratlong M., Rancurel C., *et al.*

Pratlong, M., Rancurel C., Pontarotti P., Aurelle D. (2016) Phylogenomics analysis of Cnidaria: discrepancy between mitochondrial and nuclear genes. *Zoologica Scripta*, 00, 000-000.

The phylum Cnidaria is usually divided into five classes: Anthozoa, Cubozoa, Hydrozoa, Scyphozoa and Staurozoa. The class Anthozoa is sub-divided into two subclasses: Hexacorallia and Octocorallia. Morphological and molecular studies based on nuclear rDNA and recent phylogenomic studies support the monophyly of Anthozoa. On the other hand, molecular studies based on mitochondrial markers, including two recent studies based on mitogenomic data, supported the paraphyly of Anthozoa, and positioned Octocorallia as sister group to Medusozoa (the monophyletic group of Cubozoa, Hydrozoa and Scyphozoa). On the basis of 51 nuclear orthologs from four hexacorallians, four octocorallians, two hydrozoans and one scyphozoan (with poriferans and *Homo sapiens* as outgroups), we built a multilocus alignment of 9 873 amino acids, which aimed at minimizing missing data and hidden paralogy, in order to understand the discrepancy between nuclear and mitochondrial phylogenies. Our phylogenetic analyses strongly supported the monophyly of Anthozoa. We compared the level of substitution saturation between our dataset, the datasets of two recent phylogenomic studies and one of a mitogenomic study. We found that mitochondrial DNA is more saturated than nuclear DNA at all the phylogenetic levels studied. Our results emphasize the need for a good evaluation of phylogenetic signal.

Per Pratlong, Aix Marseille Univ, Univ Avignon, CNRS, IRD, IMBE, Marseille, France; Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Équipe Evolution Biologique et Modélisation, Marseille, France. *E-mail* [marine.pratlong@imbe.fr](mailto:marine.pratlong@imbe.fr)

## Introduction

The phylum Cnidaria is usually divided into five classes: Anthozoa, Cubozoa, Hydrozoa, Scyphozoa and Staurozoa (Collins 2009; Daly *et al.* 2007). Historically, Octocorallia (soft corals, sea fans, sea pens and gorgonians) and Hexacorallia (hard corals or Scleractinia and sea anemones) were considered as sister groups, forming the clade of Anthozoa (Fig. 1. A; Berntson *et al.* 1999; Bridge *et al.* 1995; Cavalier-Smith *et al.* 1996; Collins 1998; Daly *et al.* 2007; Kim *et al.* 1999; Medina *et al.* 2001; Odorico & Miller 1997; Salvini-Plawen 1978; Won *et al.* 2001). The free-swimming medusa stage is absent in the Anthozoa clade, with no known exceptions. The life cycle of anthozoans is comprised of a sessile, reproductive, polyp stage with a larval stage. The remaining four classes belong to the clade of Medusozoa where the alternance between a sessile polyp stage and a free-swimming medusa stage is generally considered as an apomorphy lost several times (Collins *et al.* 2006; Daly *et al.* 2007). Recently, a phylogenomic study placed the clade of Myxozoa, a group of obligate endoparasites, as sister group to Medusozoa with good support (Chang *et al.* 2015). Understanding the phylogeny of Cnidaria is therefore of great interest to the study of the evolution of life cycles and life history traits in this group. The Anthozoa and Medusozoa clades are supported by nuclear phylogenies (Berntson *et al.* 1999; Cavalier-Smith *et al.* 1996; Chang *et al.* 2015; Collins 1998; Kim *et al.* 1999; Medina *et al.* 2001; Odorico & Miller 1997; Won *et al.* 2001; Zapata *et al.* 2015), by anatomical and life cycle features (Bridge *et al.* 1995; Collins *et al.* 2006; Daly *et al.* 2003, 2007; Marques & Collins 2004; Reft & Daly 2012; Salvini-Plawen 1978; Won *et al.* 2001). However, several studies based on mitochondrial genes have shown the paraphyly of Anthozoa, and the monophyly of a group constituted by Medusozoa and Octocorallia (Kayal *et al.* 2013; Kayal & Lavrov 2008; Osigus *et al.* 2013; Park *et al.* 2012; Shao *et al.* 2006; Fig. 1. B), without however being able to reject the hypothesis of anthozoan monophyly in Kayal *et al.* (2013). This clade (Octocorallia + Medusozoa), is not supported by morphological phylogenies (Won *et al.* 2001), and does not present evident synapomorphies. The results of Park *et al.* (2012) and Kayal *et al.* (2013) have been recently reanalyzed by Figueroa & Baco (2014). These authors observed that the level of saturation of

mitochondrial sequences was high, with a significant loss of information in the case of the deepest phylogenetic relationships. They suggested using nuclear gene phylogenies to resolve the high level phylogenetic relationships inside Cnidaria. Recently, phylogenomic data clarified the question of relationships among Hexacorallia, Octocorallia and Anthozoa by showing the monophyly of Anthozoa (Chang *et al.* 2015; Zapata *et al.* 2015). Nevertheless, we still lack studies dedicated to understanding the discrepancy between mitogenomics and phylogenomics studies. This discrepancy has been observed in other phylogenetic studies, including those of octocorals (Ament-Velázquez *et al.* 2016). It is well known today that the phylogeny of a specific gene or gene family does not necessary represent the phylogeny of the corresponding species because of incomplete lineage sorting (Hobolth *et al.* 2008), horizontal gene transfer (Gogarten *et al.* 2007), introgression (Rieseberg *et al.* 1996), hidden paralogy or lack of phylogenetic information (because of long branch attraction or substitution saturation, for example; Figueroa & Baco 2014) (Brocchieri 2001; Galtier & Daubin 2008). Today, Next Generation Sequencing methods allow to fill the gap between mitochondrial and nuclear gene sampling in cnidarian class-level phylogeny. It is now pertinent to use phylogenomics in order to better understand the origins of different phylogenetic signals between mitochondrial and nuclear markers (Fig. 1). We used a method comparable to the one of Zapata *et al.* (2015) to build a new phylogeny of the Cnidaria phylum with 51 nuclear genes for 11 cnidarian species in which we attempted to maximize our control on hidden paralogy. First, we confirmed the congruence of our phylogeny using a concatenated gene alignment with previous phylogenomics results. In a second step, we used the individual alignments to check if the alternative scenario (i.e. anthozoan paraphyly) was well supported at the individual gene level. Finally, we tried to understand the differences between nuclear and mitochondrial phylogenies by analyzing our dataset and the datasets of Kayal *et al.* (2013), Chang *et al.* (2015) and Zapata *et al.* (2015) in light of substitution saturation which is a mechanism known to lead to incongruent phylogenies.

## Material and Methods

### Markers selection

Our strategy aimed at identifying a set of markers with the lowest levels of missing data and sequence saturation, as well as minimizing putative paralogs. Therefore we used stringent criteria in the choice of the markers. We used OrthoMCL (Li *et al.* 2003) to search for potential ortholog reciprocal relationship between the available ESTs banks or transcriptomes of four hexacorallians (Cnidaria, Anthozoa): *Acropora palmata* (order Scleractinia), *Pocillopora damicornis* (order Scleractinia), *Anemonia viridis* (order Actiniaria) and *Nematostella vectensis* (order Actiniaria); four octocorallians (Cnidaria, Anthozoa): *Corallium rubrum* (order Alcyonacea), *Eunicella verrucosa* (order Alcyonacea), *Leptogorgia sarmentosa* (order Alcyonacea), *Gorgonia ventalina* (order Alcyonacea); two hydroidolins (Cnidaria, Hydrozoa): *Clytia hemisphaerica* (order Leptothecata) and *Hydra vulgaris* (order Anthoathecata); one Discomedusa (Cnidaria, Scyphozoa): *Aurelia aurita* (order Semaestomeae). We downloaded directly amino acid sequences from Ensembl and Compagen (Cunningham *et al.* 2015). Concerning all remaining species, we downloaded nucleic acid sequences and found the longest open reading frame using the getorf EMBOSS tool (Rice *et al.* 2000). The characteristics and references of each dataset are indicated in Table 1. First, we manually chose orthologous groups containing at least one sequence of each species. Then, we aligned all the sequences for each orthologous group using MUSCLE with default parameters (Edgar 2004) and we built individual maximum likelihood trees using FASTTREE (Price *et al.* 2009). Concerning inparalogs (sequences of the same species forming a monophyletic group), we kept the sequence with the shortest branch length in order to select genes with the lowest evolution rate (and therefore minimize substitution saturation). In order to avoid the problem of unrecognized paralogy, we used PhyloTreePruner (Kocot *et al.* 2013) that kept only orthologous groups forming a monophyletic group containing one sequence per species. At this step, we lacked an outgroup, which we did not include in the orthoMCL analysis for reasons of memory limitation. We performed a blastp (Altschul *et al.* 1990) search of the sequences of each orthologous groups against the peptides banks of the sponges *Amphimedon queenslandica* (Porifera, Demospongiae, Heteroscleromorpha, order Haplosclerida),

*Ephydatia muelleri* (Porifera, Demospongiae, Heteroscleromorpha, order Spongillida), *Sycon ciliatum* (Porifera, Calcarea, Calcaronea, order Leucosolemida), *Oscarella carmela* (Porifera, Homoscleromorpha, order Homosclerophorida), and of *Homo sapiens* (Chordata, Mammalia, order Primates). These outgroups were chosen because they represent two main Metazoan lineages: sponges and bilaterians. Inside bilaterians we chose *H. sapiens*, for which a wealth of genetic data is available, in order to maximize the probability of identifying genes orthologous to those analyzed in Cnidarians. We used an e.value of  $10^{-20}$  and kept all the hits. Finally, we performed the same procedure as before for the control of paralogy.

### *Phylogenetic analyses*

For each individual alignment, we selected conserved regions using Gblock (Castresana 2000) with the option allowing gaps for all positions. Next, we concatenated all individual alignments and obtained a final alignment of 9 873 positions (with 8.9 % of missing data; Appendix 1). ProtTest (Abascal *et al.* 2005) was used to choose the model for protein evolution that best fitted our dataset. We performed Maximum-likelihood phylogenetic analyzes of our final alignment using RaxML (Stamatakis 2014) with 1 000 fast bootstraps iterations and a PROTGAMMAILG model according to the ProtTest output. A Bayesian phylogenetic analysis was performed for the final concatenated alignment and for each individual alignment / gene using MrBayes (Ronquist *et al.* 2012) (number of generations: 200 000; sample frequencies: 100; burnin=25 %; number of chains: 4; model of protein evolution: GAMMA-I-LG).

Using the resulting tree files of the individual alignments analyzes, we performed a Bayesian concordance analysis with Bucky (Larget *et al.* 2010) which estimates a concordance tree from clades supported by the largest proportions of genes. We then applied three commonly used tests to test phylogenetic hypotheses for output trees from RaxML: the Approximately Unbiased (AU) test (Shimodaira 2002), the Hasegawa (KH) test (Kishino & Hasegawa 1989), and the Shimodaira-Hasegawa (SH) test (Shimodaira & Hasegawa 1999). We specifically tested the following hypotheses: 1: monophyly of Anthozoa, 2: monophyly of Octocorallia and Medusozoa, 3: monophyly of

Hexacorallia and Medusozoa. All these tests were performed with the software CONSEL (Shimodaira & Hasegawa 2001).

#### *Substitution saturation analysis*

Because we aimed to compare the levels of substitution saturation between datasets from different studies with different number of species, we first analyzed the effect of the number of species on two commonly used analyses of substitution saturation (Appendix 2). We showed that the statistical test of Steel *et al.* (1993) implemented in DAMBE (Xia & Xie 2001) was sensible to the number of species in the dataset and thus was not appropriate in the case of our comparative analysis. This was not the case for the slope of the regression of the plot of phylogenetic distances against observed distances. Therefore, we chose to keep only this last analysis of substitution saturation for the following analysis.

We drew the plot of phylogenetic distances against observed distances for the individual alignments. Phylogenetic and observed distances were calculated with PATRISTIC (Fourment & Gibbs 2006) using the RaxML trees, and PAUP software (Swofford 2004) respectively. For each plot, we calculated the linear regression using R (R Development Core Team 2008). Additionally, we performed this analysis for the mitochondrial dataset of Kayal *et al.* (2013). For this second dataset, we built individual ML trees using RaxML and the PROTCATGTR model used in the original paper. In order to compare the nuclear and mitochondrial concatenated alignments, we performed a hierarchical substitution saturation analysis in a similar way than Figueroa and Baco (2014) using our dataset and the datasets of Kayal *et al.* (2013), Zapata *et al.* (2015) and Chang *et al.* (2015) (characteristics of these datasets are detailed in Table 2). We plotted phylogenetic distances against observed distances for the concatenated alignments of these four studies using i) Cnidaria + Porifera datasets, ii) Cnidaria datasets, iii) Anthozoa datasets, iv) Hexacorallia datasets, v) Octocorallia datasets, vi) Medusozoa datasets. In each case, we calculated the linear regression using R (R Development Core Team 2008). Because divergent clades could mislead our comparison if they are not present in all considered studies, we removed Myxozoa from the dataset of Chang *et al.*



(2015) and Cerianths from the dataset of Kayal *et al.* (2013) for this analysis. We used here original trees and alignments provided by the authors.

## Results and discussion

### *Nuclear phylogeny of Cnidaria*

We found 335 orthologous groups between the eleven Cnidaria species included in our orthoMCL analysis. From these 335 groups, 123 were kept after the PhyloTreePruner filtration, which reduced the inaccurate orthology assignment bias. Finally, 51 orthologous groups were retained after the addition of the outgroup sequences.

The Maximum Likelihood (ML) and Bayesian trees based on the concatenation of the sequences of these 51 genes were identical with the exception of the relationships between poriferans (Fig. 2, Appendix 3). In the two cases, we found the monophyly of Cnidaria, Octocorallia, Hexacorallia and Medusozoa which were supported by most of the previously published phylogenies (Berntson *et al.* 1999; Chang *et al.* 2015; Chen *et al.* 1995; Collins 2009; Daly *et al.* 2007; France *et al.* 1996; Kayal *et al.* 2013; Kayal & Lavrov 2008; Park *et al.* 2012; Ryan *et al.* 2013; Song & Won 1997; Won *et al.* 2001; Zapata *et al.* 2015).

Regarding the relationships between the main cnidarian groups, the ML and Bayesian trees gave high support to the monophyly of Anthozoa (Fig. 2, Appendix 3; ML bootstrap support of 99 and Bayesian posterior probability of 1), which confirmed the results obtained in other studies based on nuclear markers (Cavalier-Smith *et al.* 1996; France *et al.* 1996; Odorico & Miller 1997; Zapata *et al.* 2015). The two alternative hypotheses (i.e. grouping Octocorallia – Medusozoa and grouping Medusozoa – Hexacorallia) were strongly rejected by all AU, KH and SH tests (Table 3). However, our study did not take into account the possible paraphyly of Hexacorallia due to the uncertain position of Ceriantharia, which is placed either as sister taxon to Hexacorallia, to Octocorallia or to Anthozoa (Stampar *et al.* 2014; Zapata *et al.* 2015). Furthermore, these taxonomic results should be taken cautiously due to the limited taxa sampling and lack of any basal taxa at the level of the phylum, classes and subclasses. More than

strictly resolving cnidarian phylogeny, we highlighted here the congruence of our phylogeny in order to explore the differences between individual gene based phylogenies as well as differences between nuclear and mitochondrial phylogenies.

The topology of the Bucky's primary concordance tree (using individual – gene alignment – Bayesian phylogenies) was identical to the one resolved with the ML and Bayesian phylogenies based on concatenated sequences (Fig. 2. A). The monophyly of Anthozoa was supported here by 32 % [24 ; 41] of individual trees while the two alternative hypotheses had lower support (9 % [2 ; 14] for the monophyly of (Octocorallia – Medusozoa) and 6 % [2 ; 10] for the monophyly of (Hexacorallia – Medusozoa)) (Fig. 2. B-C). Our results therefore confirmed other phylogenies obtained using nuclear markers and underlined the incongruence between mitochondrial and nuclear phylogenies. Moreover, this result raised limitations concerning the use of a limited number of genes in phylogenetic analyses. In our case, although the main signal was congruent with the accepted cnidarian phylogeny, 15 % of our genes showed a different signal.

Concerning nuclear markers, all concatenated and individual plots of patristic distance versus observed distance demonstrated substitution saturation signal with different strengths (Fig. 3): for small distances, the phylogenetic distances followed the observed distances but for larger distances, phylogenetic distances increased faster than observed distances. As a complementary test, we also built a new phylogeny using a concatenated alignment of the 20 % less saturated genes (i.e. those with the highest slopes). This phylogeny again supported the monophyly of Anthozoa with a bootstrap value of 84. Therefore our results based on nuclear genes do not seem to be strongly biased by saturation.

#### *Possible origin of the discrepancy between nuclear and mitochondrial phylogenies*

Substitution saturation could be one of the mechanisms explaining the discrepancy between our results and previous results based on mitochondrial data (Figuerola & Baco 2014). To test this hypothesis, we compared the slope of the linear regression of the phylogenetic vs observed distances plots performed on hierarchical datasets (Fig. 4,

Appendix 4). A low slope is the signal of saturated sequences. Our hierarchical substitution saturation analysis showed that in all studies, the slope of the linear regression was low in the dataset Cnidaria + Porifera, increased in Cnidaria, Anthozoa, Hexacorallia, and Octocorallia where it reached its maximum. In Medusozoa, the slope was weaker than in Octocorallia and Hexacorallia. This result is congruent with divergent times proposed for Cnidarian clades: around 600 Ma for Medusozoa, 300 Ma for Octocorallia and Hexacorallia (Park *et al.* 2012). The difference in the saturation levels observed between Hexacorallia and Octocorallia, despite similar divergence times, could be due to a slower evolution rate in Octocorallia, for nuclear and mitochondrial markers. In all datasets, the slope was much higher in our study and in the study of Zapata *et al.* (2015) than in the mitochondrial study of Kayal *et al.* (2013). This confirms that the mitochondrial sequences are globally more saturated than nuclear sequences in Cnidaria, with an increase of the level of saturation with taxonomic level. The slope of the regression in our study was always higher than the one in the study of Zapata *et al.* (2015). This may be the result of our conservative control of paralogy resulting in the selection of slowly evolving genes. The difference between the nuclear and mitochondrial slope was very small in Octocorallia. This clade seems to be the only one where mitochondrial sequences could be used with the same effectiveness as nuclear sequences. In the case of the study of Chang *et al.* (2015), the slope of the regression was also higher than for the mitochondrial study, but with much less difference than the two other nuclear studies. This could be the consequence of the presence in the concatenated alignment of Chang *et al.* (2015) of 27 % of ribosomal and mitochondrial genes. As ribosomal genes have been reported to be more saturated than other nuclear genes and may contain a different phylogenetic signal (Nosenko *et al.* 2013), Chang *et al.* (2015) state that their results was not influenced by these genes, but they still could induce a decrease of the saturation level of their dataset.

Finally, we compared the slope of the linear regression of the phylogenetic vs observed distances plots between the individual genes, in our study and in the study of Kayal *et al.* (2013), in order to check if some of the mitochondrial genes are less saturated than others. The 13 mitochondrial genes used in Kayal *et al.* (2013) showed signal of substitution saturation, because of their relatively low slopes. The slopes of the linear

regression of nuclear genes were significantly higher than those of mitochondrial genes ( $p = 2.4e-08$ ; Fig. 3). This confirms that mitochondrial DNA is more saturated than nuclear DNA at the gene level in Cnidaria. The level of saturation of all mitochondrial genes is similar, with less variation than in nuclear genes. The gene COB (cytochrome b) was the least saturated with a slope of 0.366.

When substitution saturation is high, similarity between sequences does not necessarily reflect phylogenetic relationships (Steel *et al.* 1993; Xia *et al.* 2003; Xia & Lemey 2009). The results observed here suggest that mitochondrial sequences may not be appropriate to resolve deep cnidarian phylogeny, as suggested by Figueroa & Baco (2014), Osigus *et al.* (2013) and Van Iten *et al.* (2014). The paraphyly of Anthozoa, observed in several studies based on mitochondrial markers (Kayal *et al.* 2013; Kayal & Lavrov 2008; Osigus *et al.* 2013; Park *et al.* 2012; Shao *et al.* 2006), could indeed be the result of substitution saturation. Here, on the basis on an extended gene dataset, we confirm the analysis of Figueroa and Baco (2014), namely that nuclear markers are less affected by substitution saturation and more appropriate to resolve deep phylogeny of Cnidaria. Mitochondrial DNA is usually considered as poorly informative for anthozoan phylogeny at the species and genus level because of its slow evolution rate (Chen *et al.* 2009; France & Hoover 2002; McFadden *et al.* 2011; Shearer *et al.* 2002). Stampar *et al.* (2014) argued that the most parsimonious scenario for this slow mtDNA evolution in Anthozoa is that the rate of evolution was high in the common ancestor of Cnidaria and Bilateria, and decreased in anthozoans. Because we highlighted here that mitochondrial DNA seems to be much more saturated than nuclear DNA in Cnidaria, including Anthozoa, it seems indeed that the slow mitochondrial evolution rate appeared in the branch leading to Anthozoa.

## Conclusion

The validation of this Cnidaria tree of life has major impact for the study of morphological and life history traits among Cnidaria. We argue here that nuclear markers should be favored for resolving cnidarian class-level phylogeny. More generally, our results confirm the caution concerning the use of a reduced number of

markers for phylogenetic analyzes. Indeed, even within nuclear markers, different phylogenies were inferred depending on the gene. The high diversity of phylogenetic histories and levels of informations between genes therefore requires the use of enough markers and adequate statistical tests before validating a phylogenetic result.

## Appendices

The following supplementary material is available:

Appendix 1. Amino acid alignment used in this study.

Appendix 2. Analysis of the sensitivity to the number of species of two markers of substitution saturation.

Appendix 3. Phylogenetic relationships among cnidarians based on A. Maximum-likelihood analysis; B. Bayesian analysis. Bootstrap supports in A were all 100 % apart from the two indicated values. In B, all posterior probabilities were equal to 1.

Appendix 4. Saturation plots of the hierarchical substitution saturation analysis.

## Acknowledgements

This work is a contribution to the Labex OT-Med (n° ANR-11-LABX-0061) funded by the French Government “Investissements d’Avenir” program of the French National Research Agency (ANR) through the A\*MIDEX project (n° ANR-11-IDEX-0001-02). This project has been funded by the ADACNI program of the French National Research Agency (ANR) (project n°ANR-12-ADAP-0016; <http://adacni.imbe.fr>). We thank ECCOREV Research Federation (FR 3098) for the financial support of this study. The authors thank Xegen Company ([www.xegen.fr](http://www.xegen.fr)) for their advice. We thank Anne Haguenaier for the development of *Corallium rubrum* transcriptome, Etienne Danchin for his collaboration for the OrthoMCL analysis and Abigail E. Cahill for her contribution to the correction of the manuscript.

## References

- Abascal, F., Zardoya, R., & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, 21(9), 2104–2105.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.
- Ament-Velásquez, S. L., Breedy, O., Cortés, J., Guzman, H. M., Wörheide, G., & Vargas, S. (2016). Homoplasious colony morphology and mito-nuclear phylogenetic discordance among Eastern Pacific octocorals. *Molecular Phylogenetics and Evolution*, 98, 373–381.
- Berntson, E. A., France, S. C., & Mullineaux, L. S. (1999). Phylogenetic relationships within the class Anthozoa (phylum Cnidaria) based on nuclear 18S rDNA sequences. *Molecular Phylogenetics and Evolution*, 13(2), 417–433.
- Bridge, D., Cunningham, C. W., DeSalle, R., & Buss, L. W. (1995). Class-level relationships in the phylum Cnidaria: molecular and morphological evidence. *Molecular Biology and Evolution*, 12(4), 679–689.
- Brocchieri, L. (2001). Phylogenetic inferences from molecular sequences: review and critique. *Theoretical Population Biology*, 59(1), 27–40.
- Burge, C. A., Mouchka, M. E., Harvell, C. D., & Roberts, S. (2013). Immune response of the Caribbean sea fan, *Gorgonia ventalina*, exposed to an *Aplanochytrium* parasite as revealed by transcriptome sequencing. *Frontiers in Physiology*, 4.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4), 540–552.
- Cavalier-Smith, T., Allsopp, M. T. E. P., Chao, E. E., Boury-Esnault, N., & Vacelet, J. (1996). Sponge phylogeny, animal monophyly, and the origin of the nervous system: 18S rRNA evidence. *Canadian Journal of Zoology*, 74(11), 2031–2045.

- Chang, E. S., Neuhof, M., Rubinstein, N. D., Diamant, A., Philippe, H., Huchon, D., & Cartwright, P. (2015). Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. *Proceedings of the National Academy of Sciences*, 112(48), 14912–14917.
- Chen, C. A., Odorico, D. M., Tenlohuis, M., Veron, J. E. N., & Miller, D. J. (1995). Systematic relationships within the Anthozoa (Cnidaria: Anthozoa) using the 5'-end of the 28S rDNA. *Molecular Phylogenetics and Evolution*, 4(2), 175–183.
- Chen, I.-P., Tang, C.-Y., Chiou, C.-Y., Hsu, J.-H., Wei, N. V., Wallace, C. C., Muir, P., et al. (2009). Comparative analyses of coding and noncoding DNA regions indicate that *Acropora* (Anthozoa: Scleractina) possesses a similar evolutionary tempo of nuclear vs. mitochondrial genomes as in plants. *Marine Biotechnology*, 11(1), 141–152.
- Collins, A. G. (1998). Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence. *Proceedings of the National Academy of Sciences*, 95(26), 15458–15463.
- Collins, A. G. (2009). Recent insights into cnidarian phylogeny. *Smithsonian Contributions to Marine Sciences*, 38, 139–149.
- Collins, A. G., Schuchert, P., Marques, A. C., Jankowski, T., Medina, M., & Schierwater, B. (2006). Medusozoan phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. *Systematic Biology*, 55(1), 97–115.
- Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., et al. (2015). Ensembl 2015. *Nucleic Acids Research*, 43(D1), D662–D669.
- Daly, M., Brugler, M. R., Cartwright, P., Collins, A. G., Dawson, M. N., Fautin, D. G., France, S. C., et al. (2007). The phylum Cnidaria: a review of phylogenetic

- patterns and diversity 300 years after Linnaeus. *Zootaxa*, 127–182.
- Daly, M., Fautin, D. G., & Cappola, V. A. (2003). Systematics of the Hexacorallia (Cnidaria: Anthozoa). *Zoological Journal of the Linnean Society*, 139(3), 419–437.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797.
- Figueroa, D. F., & Baco, A. R. (2014). Octocoral mitochondrial genomes provide insights into the phylogenetic history of gene order rearrangements, order reversals, and cnidarian phylogenetics. *Genome Biology and Evolution*, 7(1), 391–409.
- Fourment, M., & Gibbs, M. J. (2006). PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evolutionary Biology*, 6(1), 1.
- France, S. C., & Hoover, L. L. (2002). DNA sequences of the mitochondrial COI gene have low levels of divergence among deep-sea octocorals (Cnidaria: Anthozoa). *Hydrobiologia*, 471(1–3), 149–155.
- France, S. C., Rosel, P. E., Agenbrood, J. E., Mullineaux, L. S., & Kocher, T. D. (1996). DNA sequence variation of mitochondrial large-subunit rRNA provides support for a two-subclass organization of the Anthozoa (Cnidaria). *Molecular Marine Biology and Biotechnology*, 5(1), 15–28.
- Galtier, N., & Daubin, V. (2008). Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1512), 4023–4029.
- Gogarten, J. P., Doolittle, W. F., & Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution*, 19(12), 2226–2238.
- Hobolth, A., Christensen, O. F., Mailund, T., & Schierup, M. H. (2007). Genomic



- relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet*, 3(2), e7.
- Kayal, E., & Lavrov, D. V. (2008). The mitochondrial genome of *Hydra oligactis* (Cnidaria, Hydrozoa) sheds new light on animal mtDNA evolution and cnidarian phylogeny. *Gene*, 410(1), 177–186.
- Kayal, E., Roure, B., Philippe, H., Collins, A. G., & Lavrov, D. V. (2013). Cnidarian phylogenetic relationships as revealed by mitogenomics. *BMC Evolutionary Biology*, 13(1), 5.
- Kim, J., Kim, W., & Cunningham, C. W. (1999). A new perspective on lower metazoan relationships from 18S rDNA sequences. *Molecular Biology and Evolution*, 16(3), 423–427.
- Kishino, H., & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution*, 29(2), 170–179.
- Kocot, K. M., Citarella, M. R., Moroz, L. L., & Halanych, K. M. (2013). PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evolutionary Bioinformatics Online*, 9, 429–435.
- Larget, B. R., Kotha, S. K., Dewey, C. N., & Ané, C. (2010). BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics*, 26(22), 2910–2911.
- Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13(9), 2178–2189.
- Marques, A. C., & Collins, A. G. (2004). Cladistic analysis of Medusozoa and cnidarian evolution. *Invertebrate Biology*, 123(1), 23–42.
- McFadden, C. S., Benayahu, Y., Pante, E., Thoma, J. N., Nevarez, P. A., & France, S. C.

- (2011). Limitations of mitochondrial gene barcoding in Octocorallia. *Molecular Ecology Resources*, 11(1), 19–31.
- Medina, M., Collins, A. G., Silberman, J. D., & Sogin, M. L. (2001). Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proceedings of the National Academy of Sciences*, 98(17), 9707–9712.
- Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., Maldonado, M., et al. (2013). Deep metazoan phylogeny: when different genes tell different stories. *Molecular Phylogenetics and Evolution*, 67(1), 223–233.
- Odorico, D. M., & Miller, D. J. (1997). Internal and external relationships of the Cnidaria: implications of primary and predicted secondary structure of the 5'-end of the 23S-like rDNA. *Proceedings of the Royal Society B: Biological Sciences*, 264(1378), 77–82.
- Osigus, H.-J., Eitel, M., Bernt, M., Donath, A., & Schierwater, B. (2013). Mitogenomics at the base of Metazoa. *Molecular Phylogenetics and Evolution, Mitogenomics and Metazoan Evolution*, 69(2), 339–351.
- Park, E., Hwang, D.-S., Lee, J.-S., Song, J.-I., Seo, T.-K., & Won, Y.-J. (2012). Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Molecular Phylogenetics and Evolution*, 62(1), 329–345.
- Pratlong, M., Haguénauer, A., Chabrol, O., Klopp, C., Pontarotti, P., & Aurelle, D. (2015). The red coral (*Corallium rubrum*) transcriptome: a new resource for population genetics and local adaptation studies. *Molecular Ecology Resources*, 15(5), 1205–1215.
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and*

- Evolution*, 26(7), 1641–1650.
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Reft, A. J., & Daly, M. (2012). Morphology, distribution, and evolution of apical structure of nematocysts in hexacorallia. *Journal of Morphology*, 273(2), 121–136.
- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(6), 276–277.
- Rieseberg, L. H., Whitton, J., & Linder, C. R. (1996). Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica*, 45(3), 243-262.
- Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., et al. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526), 261–263.
- Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D. L., Darling, A., Höhna, S., Larget, B., et al. (2012). MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542.
- Ryan, J. F., Pang, K., Schnitzler, C. E., Nguyen, A.-D., Moreland, R. T., Simmons, D. K., Koch, B. J., et al. (2013). The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science*, 342(6164).
- Salvini-Plawen, L. V. (1978). On the origin and evolution of the lower Metazoa. *Journal of Zoological Systematics and Evolutionary Research*, 16(1), 40–87.
- Shao, Z., Graf, S., Chaga, O. Y., & Lavrov, D. V. (2006). Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): a linear DNA molecule

- encoding a putative DNA-dependent DNA polymerase. *Gene*, 381, 92–101.
- Shearer, T. L., van Oppen, M. J. H., Romano, S. L., & Wörheide, G. (2002). Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology*, 11(12), 2475–2487.
- Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, 51(3), 492–508.
- Shimodaira, H., & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, 16, 1114–1116.
- Shimodaira, H., & Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, 17(12), 1246–1247.
- Song, J., & Won, J. H. (1997). Systematic relationship of the anthozoan orders based on the partial nuclear 18S rDNA sequences. *Korean Journal of Biological Sciences*, 1(1), 43–52.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
- Stampar, S. N., Maronna, M. M., Kitahara, M. V., Reimer, J. D., & Morandini, A. C. (2014). Fast-evolving mitochondrial DNA in Ceriantharia: a reflection of Hexacorallia paraphyly? *PLoS ONE*, 9(1).
- Steel, M. A., Lockhart, P. J., & Penny, P. (1993). Confidence in evolutionary trees from biological data. *Nature*, 364, 440–442.
- Swofford, D. L. (2004). *Paup 4.0 for Macintosh: phylogenetic analysis using parsimony (software and user's Book for Macintosh)*. Sinauer Associates, Incorporated.
- Van Iten, H., Marques, A. C., Leme, J. de M., Pacheco, M. L. A. F., & Simões, M. G. (2014). Origin and early diversification of the phylum Cnidaria Verrill: major

- developments in the analysis of the taxon's Proterozoic–Cambrian history. *Palaeontology*, 57(4), 677–690.
- Vidal-Dupiol, J., Zoccola, D., Tambutté, E., Grunau, C., Cosseau, C., Smith, K. M., Freitag, M., et al. (2013). Genes related to ion-transport and energy production are upregulated in response to CO<sub>2</sub>-driven pH decrease in corals: new insights from transcriptome analysis. *PLoS ONE*, 8(3), e58652.
- Won, J., Rho, B., & Song, J. (2001). A phylogenetic study of the Anthozoa (phylum Cnidaria) based on morphological and molecular characters. *Coral Reefs*, 20(1), 39–50.
- Xia, X., & Lemey, P. (2009). Assessing substitution saturation with DAMBE. *The Phylogenetic Handbook* (Second edition.). Cambridge University Press.
- Xia, X., & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity*, 92(4), 371–373.
- Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, 26(1), 1–7.
- Zapata, F., Goetz, F. E., Smith, S. A., Howison, M., Siebert, S., Church, S. H., Sanders, S. M., et al. (2015). Phylogenomic analyses support traditional relationships within Cnidaria. *PLoS ONE*, 10(10), e0139068.

### Tables and Figure Legends

Fig. 1. Hypotheses of the cnidarian phylogeny tested in this study. A. The monophyly of Anthozoa. B. The paraphyly of Anthozoa.

Fig. 2. A. Bayesian primary concordance tree resulting from the analysis of MrBayes results (bayesian analysis) using the program Bucky (Larget *et al.*, 2010). Numbers on branches represent the average concordance factor. The 95 % confidence intervals are included. B-C. Results of the average concordance factors for the two alternative trees for the nodes of interest.

Fig. 3. Boxplot of the slopes' distributions of the saturation plots for this dataset and the dataset of Kayal *et al.* (2013).

Fig. 4. Slopes of the linear regression of the saturation plots from the hierarchical substitution saturation analysis.

Table 1. Characteristics and references of each dataset used in the phylogeny.

Table 2. Details of the characteristics of the datasets of this study and of those of Kayal *et al.* (2013), Zapata *et al.* (2015) and Chang *et al.* (2015).

Table 3. Results of the tests of alternative topologies for the ML analyses (only p values are indicated). AU : Approximately Unbiased test; KH : Kishino-Hasegawa test; SH : Shimodaira-Hasegawa test. The topologies tested were the monophyly of Anthozoa (1), the monophyly of [Octocorallia + Medusozoa] (2), the monophyly of [Hexacorallia + Medusozoa] (3).

Tables and Figures

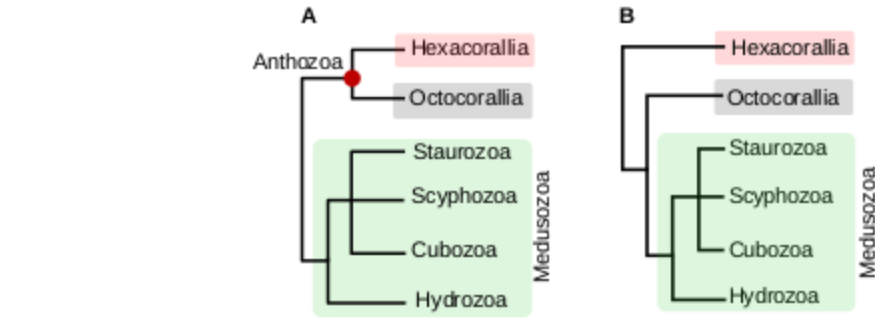
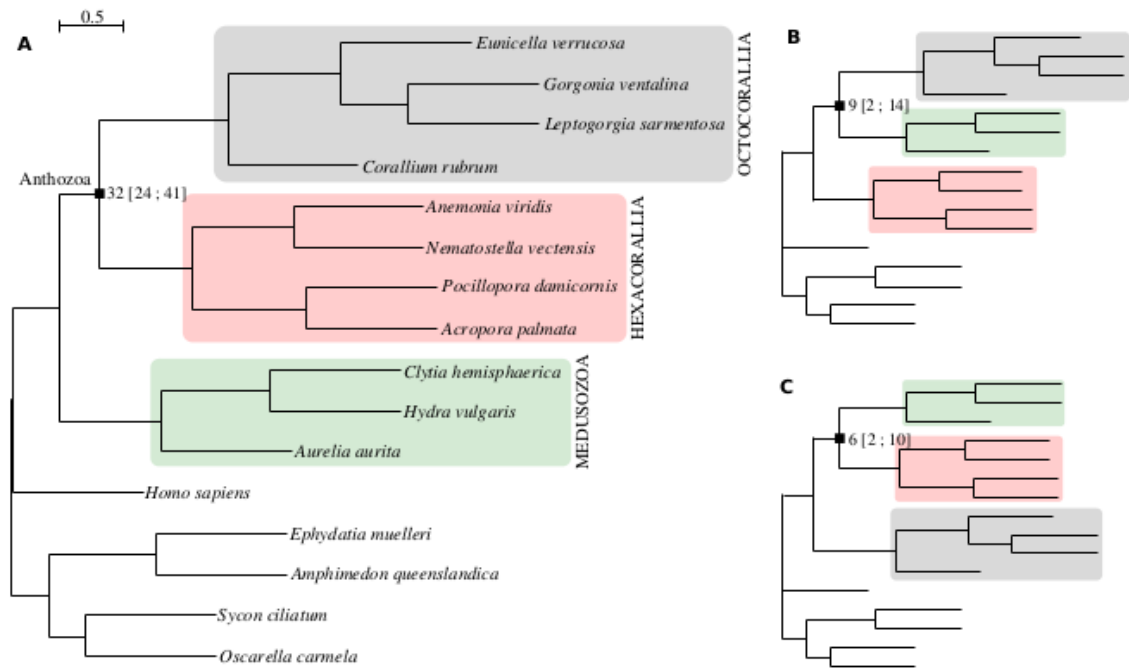


Figure 1.

Figure 2.



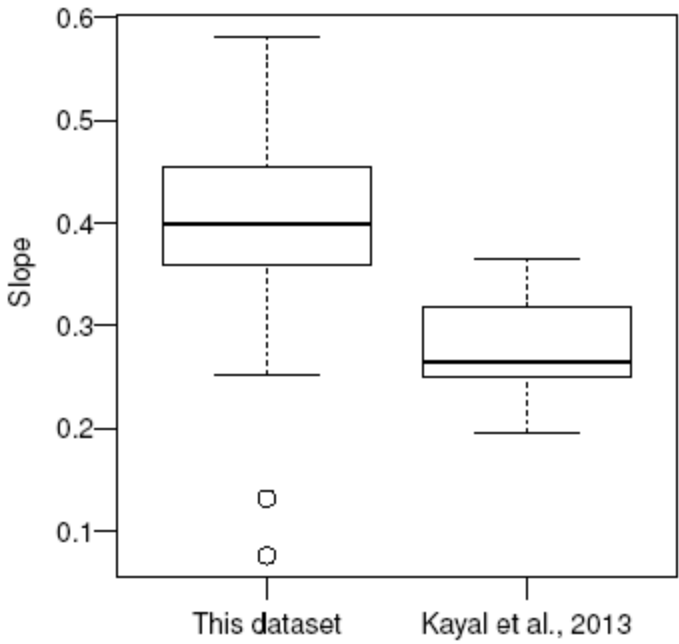


Figure 3.



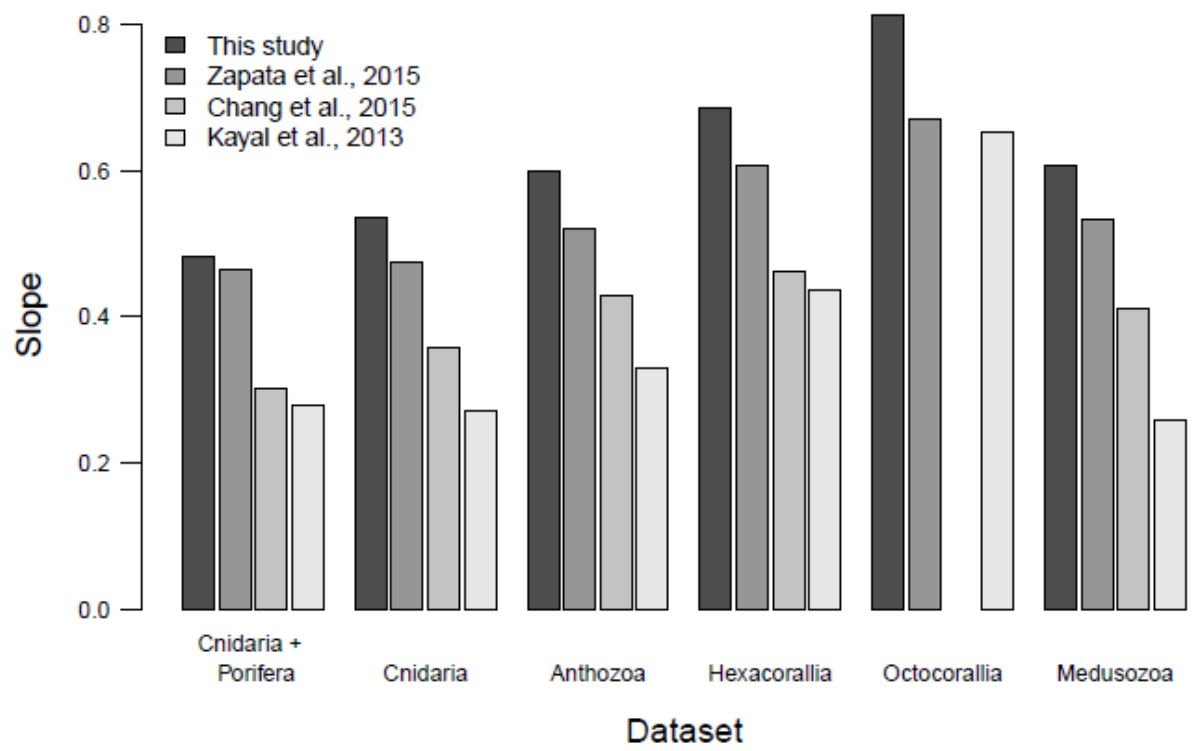


Figure 4.

Table 1.

Classification				Species	Source	Number of ESTs
Phylum	Class	Sub Class	Order			
Cnidaria	Anthozoa	Hexacorallia	Scleractinia	<i>Acropora palmata</i>	NCBI	43 150
				<i>Pocillopora damicornis</i>	Vidal-Dupiol <i>et al.</i> 2013	72 890
			Actiniaria	<i>Anemonia viridis</i>	NCBI – EST Bank	39 939
				<i>Nematostella vectensis</i>	NCBI – EST Bank	163 314
		Octocorallia	Alcyonacea	<i>Eunicella verrucosa</i>	Romiguier <i>et al.</i> 2014	15 461
				<i>Gorgonia ventalina</i>	Burge <i>et al.</i> 2013	90 230
				<i>Leptogorgia sarmentosa</i>	Romiguier <i>et al.</i> , 2014	13 299
				<i>Corallium rubrum</i>	Pratlong <i>et al.</i> 2015	48 074
	Hydrozoa	Hydroidolina	Anthoathecata	<i>Hydra vulgaris</i>	NCBI – EST Bank	184 731
			Leptothecata	<i>Clytia hemisphaerica</i>	NCBI – EST Bank	85 991

	Scyphozoa	Discomedusae	Semaeostomeae	<i>Aurelia aurita</i>	<a href="http://www.compagen.org">www.compagen.org</a>	62 632
Chordata	Mamalia		Primates	<i>Homo sapiens</i>	Ensembl	99 436
Porifera	Heteroscleromorpha	Spongillida	Spongillida	<i>Ephidatia muelleri</i>	<a href="http://www.compagen.org">www.compagen.org</a>	29 154
			Haplosclerida	<i>Amphimedon queenslandica</i>	NCBI	29 883
	Calcarea	Calcaronea	Leucosolenida	<i>Sycon ciliatum</i>	<a href="http://www.compagen.org">www.compagen.org</a>	50 731
	Homoscleromorpha		Homosclerophorida	<i>Oscarella carmela</i>	<a href="http://www.compagen.org">www.compagen.org</a>	29 220

Table 2.

Characteristics				This study	Zapata <i>et al.</i> 2015	Chang <i>et al.</i> 2015	Kayal <i>et al.</i> 2013
Missing data on complete dataset (%)				8.9	43	12	5
Number of amino acids				9 873	365 159	51 940	3 485
Number of sequences	Porifera			4	1	13	21
	Cnidaria	Anthozoa	Hexacorallia	4	9	10	30
			Octocorallia	4	4	2	19
		Meduzosoa		3	17	10	29

Table 3.

Topology	AU	KH	WKH	SH	WSH
(Med,(Hex,Oct))	1	1	1	1	1
(Hex,(Med,Oct))	$8 \times 10^{-12}$	0	0	0	0
(Oct,(Med,Hex))	$7 \times 10^{-7}$	0	0	0	0