

Unlocking Antimicrobial Peptides: In Silico Proteolysis and Artificial Intelligence-Driven Discovery from Cnidarian Omics

Ricardo Alexandre Barroso ^{1,2}, Guillermin Agüero-Chapin ^{1,2}, Rita Sousa ^{1,2}, Yovani Marrero-Ponce ^{3,4} and Agostinho Antunes ^{1,2,*}

Table S1: Overview of Cnidarian omics data used in this study

Table S2: BUSCO statistics for the 27 SRA-derived Transcriptomes

Table S3: Number of proteins in each database, provided in three stages: original dataset, after duplicate removal and following AMPir prediction filtering, with statistics included

Table S4: Statistics for each mining step (peptides, AMPs, and non-haemolytic and non-toxic AMPs)

Table S5 (Excel file): Characterization of the CnSA within the HSPN, detailing properties for each of the 3310 nodes, and the HC and HB values

Table S6 (Excel file): Overview of the 152 predicted ABPs obtained from omics data of Cnidaria.

Figure S1: Completeness scores of the 27 assembled transcriptomes obtained from SRA data (NCBI), assessed with BUSCO using the Metazoa dataset

Figure S2: Venn diagrams representing the number of predicted AMPs by AMPir, AMPlify, and Macrel from each protease–database combination

Figure S3: Venn diagrams representing the number of predicted non-haemolytic AMPs by HemoPi, Macrel, and MQSSM from each protease–database pair

Figure S4: Venn diagrams representing the number of predicted non-haemolytic and non-toxic AMPs by CAPTP, ToxinPred3, and ToxTeller from each protease–database pair

Figure S5: HSPN clustering of the 1017 AMPs obtained from the intersection of the Harmonic and Hub-Bridge datasets within the Cnidaria Singular AMPs (CnSA)

Figure S6: Strain-specific activity predictions for ABPs derived from the most representative CnSA datasets and Venn diagrams summarizing activity predictions for five bacterial strains—*Bacillus subtilis*, *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*

Figure S7: Distribution of the 152 predicted antimicrobial peptides (ABPs) across various categories

Figure S8: Antifungal and antiviral specific predictions for both intersection and union datasets in DBAASP

Table S1. Overview of Cnidarian omics data used in this study, including details on database sources, tissue types, and protein counts.

Species	Database	Subphylum	Group	UniProt/TSA/SRA Code	Date	Info	Protein Count
<i>Actinia tenebrosa</i>	1	Anthozoa	Hexacorallia, Actiniaria	UP000515163/GCF_009602425.1	06/11/2019	Non-specific / Whole Body	23842
<i>Nematostella vectensis</i>	1	Anthozoa	Hexacorallia, Actiniaria	UP000001593/GCA_000209225.1	22/08/2007	Non-specific / Whole Body	24435
<i>Desmophyllum pertusum</i>	1	Anthozoa	Hexacorallia, Scleractinia	UP001163046/GCA_029204205.1	17/04/2019	Non-specific / Whole Body	37379
<i>Pocillopora damicornis</i>	1	Anthozoa	Hexacorallia, Scleractinia	UP000275408/GCA_003704095.1	31/10/2018	Non-specific / Whole Body	25411
<i>Stylophora pistillata</i>	1	Anthozoa	Hexacorallia, Scleractinia	UP000225706/GCA_002571385.2	17/10/2017	Non-specific / Whole Body	24067
<i>Paramuricea clavata</i>	1	Anthozoa	Octocorallia, Malacalcyonacea	UP001152795/GCA_902702795.2	08/11/2022	Non-specific / Whole Body	69559
<i>Clytia hemisphaerica</i>	2	Medusozoa	Hydrozoa	UP000594262/GCA_902728285.1	04/02/2020	Non-specific / Whole Body	24557
<i>Hydra vulgaris</i>	2	Medusozoa	Hydrozoa	UP000694840/GCF_000004095.1	30/10/2009	Non-specific / Whole Body	21303
<i>Acanthogorgia aspera</i>	3	Anthozoa	Octocorallia, Malacalcyonacea	GEXC00000000.1	14/09/2016	Non-specific / Whole Body	81723
<i>Acropora cervicornis</i>	3	Anthozoa	Hexacorallia, Scleractinia	GJIG00000000.1	30/08/2022	Non-specific / Whole Body	1267328
<i>Acropora digitifera</i>	3	Anthozoa	Hexacorallia, Scleractinia	GIVI00000000.1	13/11/2020	Non-specific / Whole Body	748853
<i>Acropora gemmifera</i>	3	Anthozoa	Hexacorallia, Scleractinia	GFQE00000000.1	03/10/2017	Non-specific / Whole Body	268146
<i>Acropora hyacinthus</i>	3	Anthozoa	Hexacorallia, Scleractinia	GDIF00000000.1	29/07/2015	Non-specific / Whole Body	66872
<i>Acropora millepora</i>	3	Anthozoa	Hexacorallia, Scleractinia	GHGU00000000.1	02/04/2019	Non-specific / Whole Body	721367
<i>Acropora solitaryensis</i>	3	Anthozoa	Hexacorallia, Scleractinia	ICPH00000000.1	27/09/2019	Non-specific / Whole Body	121395
<i>Acropora tenuis</i>	3	Anthozoa	Hexacorallia, Scleractinia	IADL00000000.1	28/08/2018	Non-specific / Whole Body	365224
<i>Agaricia lamarcki</i>	3	Anthozoa	Hexacorallia, Scleractinia	GGLC00000000.3	28/08/2018	Non-specific / Whole Body	435762
<i>Alveopora japonica</i>	3	Anthozoa	Hexacorallia, Scleractinia	GGJR00000000.1	19/03/2018	Non-specific / Whole Body	961641
<i>Anthopleura elegantissima</i>	3	Anthozoa	Hexacorallia, Actiniaria	GBXJ00000000.1	20/11/2014	Non-specific / Whole Body	787875
<i>Astroides calycularis</i>	3	Anthozoa	Hexacorallia, Scleractinia	GIRZ00000000.1	23/07/2020	Non-specific / Whole Body	561718
<i>Briareum asbestinum</i>	3	Anthozoa	Octocorallia, Scleractyonacea	GHBD00000000.2	15/08/2019	Non-specific / Whole Body	613438
<i>Calyptrophora lyra</i>	3	Anthozoa	Octocorallia, Scleractyonacea	GJII00000000.1	17/03/2022	Non-specific / Whole Body	470134
<i>Chrysogorgia stellata</i>	3	Anthozoa	Octocorallia, Scleractyonacea	GJIJ00000000.1	17/03/2022	Non-specific / Whole Body	569607
<i>Clavularia sp.</i>	3	Anthozoa	Octocorallia, Malacalcyonacea	GHAW00000000.1	14/11/2018	Non-specific / Whole Body	614091
<i>Corynactis australis</i>	3	Anthozoa	Hexacorallia, Corallimorpharia	GELM00000000.1	01/11/2016	Non-specific / Whole Body	167069
<i>Ctenactis echinata</i>	3	Anthozoa	Hexacorallia, Scleractinia	GDZV00000000.1	10/11/2015	Non-specific / Whole Body	87019
<i>Cyphastrea serailia</i>	3	Anthozoa	Hexacorallia, Scleractinia	GETH00000000.1	16/06/2016	Non-specific / Whole Body	193524
<i>Edwardsiella carnea</i>	3	Anthozoa	Hexacorallia, Actiniaria	GGGD00000000.1	14/07/2018	Non-specific / Whole Body	475457
<i>Eleutherobia rubra</i>	3	Anthozoa	Octocorallia, Malacalcyonacea	GHFI00000000.1	12/02/2019	Non-specific / Whole Body	325194
<i>Favia/Dipsastraea lizadensis</i>	3	Anthozoa	Hexacorallia, Scleractinia	GDZU00000000.1	10/11/2015	Non-specific / Whole Body	110215
<i>Favites colemani</i>	3	Anthozoa	Hexacorallia, Scleractinia	GIVN00000000.1	13/11/2020	Non-specific / Whole Body	735877
<i>Fimbriaphyllia ancora</i>	3	Anthozoa	Hexacorallia, Scleractinia	ICQS00000000.1	16/09/2020	Non-specific / Whole Body	1082203
<i>Galaxea fascicularis</i>	3	Anthozoa	Hexacorallia, Scleractinia	GFAZ00000000.1	17/04/2017	Non-specific / Whole Body	1010746
<i>Heliopora coerulea</i>	3	Anthozoa	Octocorallia, Scleractyonacea	GFVH00000000.1	07/03/2018	Non-specific / Whole Body	299914

<i>Metridium senile</i>	3	Anthozoa	Hexacorallia, Actiniaria	GGGC00000000.1	27/02/2019	Non-specific / Whole Body	127132
<i>Montipora capitata</i>	3	Anthozoa	Hexacorallia, Scleractinia	GFRO00000000.1	17/07/2017	Non-specific / Whole Body	104929
<i>Montipora digitata</i>	3	Anthozoa	Hexacorallia, Scleractinia	GIVM00000000.1	13/11/2020	Non-specific / Whole Body	606544
<i>Palythoa caribaeorum</i>	3	Anthozoa	Hexacorallia, Zoantharia	GESO00000000.1	01/06/2016	Non-specific / Whole Body	459582
<i>Palythoa sp.</i>	3	Anthozoa	Hexacorallia, Zoantharia	GGUI00000000.1	09/10/2019	Non-specific / Whole Body	561450
<i>Pocillopora acuta</i>	3	Anthozoa	Hexacorallia, Scleractinia	GGRP00000000.1	10/07/2019	Non-specific / Whole Body	89222
<i>Pocillopora verrucosa</i>	3	Anthozoa	Hexacorallia, Scleractinia	GJZD00000000.1	10/01/2023	Non-specific / Whole Body	364512
<i>Polymyces wellsii</i>	3	Anthozoa	Hexacorallia, Scleractinia	GKDK00000000.1	10/07/2023	Non-specific / Whole Body	411660
<i>Porites lutea</i>	3	Anthozoa	Hexacorallia, Scleractinia	GGER00000000.1	19/01/2018	Non-specific / Whole Body	416366
<i>Protopalmytha variabilis</i>	3	Anthozoa	Hexacorallia, Zoantharia	GCVI00000000.1	11/04/2016	Non-specific / Whole Body	407283
<i>Renilla muelleri</i>	3	Anthozoa	Octocorallia, Scleractyonacea	GJZZ00000000.1	02/08/2022	Non-specific / Whole Body	119186
<i>Rhodactis indosinensis</i>	3	Anthozoa	Hexacorallia, Corallimorpharia	GEL00000000.1	01/11/2016	Non-specific / Whole Body	169678
<i>Rhodaniridogorgia sp.</i>	3	Anthozoa	Octocorallia, Scleractyonacea	GJIH00000000.1	17/03/2022	Non-specific / Whole Body	600342
<i>Ricordea yuma</i>	3	Anthozoa	Hexacorallia, Corallimorpharia	GELN00000000.1	01/11/2016	Non-specific / Whole Body	128218
<i>Scleronephthya gracillima</i>	3	Anthozoa	Octocorallia, Malacalcyonacea	GJHC00000000.1	28/07/2021	Non-specific / Whole Body	69225
<i>Scolanthus callimorphus</i>	3	Anthozoa	Hexacorallia, Actiniaria	GGGE00000000.1	27/02/2019	Non-specific / Whole Body	622171
<i>Seriatopora caliendrum</i>	3	Anthozoa	Hexacorallia, Scleractinia	GIVG00000000.1	13/11/2020	Non-specific / Whole Body	682598
<i>Stichodactyla helianthus</i>	3	Anthozoa	Hexacorallia, Actiniaria	GGNY00000000.1	22/05/2018	Non-specific / Whole Body	442313
<i>Xenia sp.</i>	3	Anthozoa	Octocorallia, Malacalcyonacea	GHBC00000000.1	28/01/2019	Non-specific / Whole Body	438621
<i>Zoanthus sp.</i>	3	Anthozoa	Hexacorallia, Zoantharia	GGTW00000000.1	07/09/2018	Non-specific / Whole Body	590885
<i>Alatina alata</i>	4	Medusozoa	Cubozoa	GEUJ00000000.1	05/07/2016	Non-specific / Whole Body	110225
<i>Calvadosia cruxmelitensis</i>	4	Medusozoa	Staurozoa	HAHC00000000.1	29/09/2017	Non-specific / Whole Body	286864
<i>Cassiopea andromeda</i>	4	Medusozoa	Scyphozoa	GJJJ00000000.1	03/10/2022	Non-specific / Whole Body	149941
<i>Chrysaora quinquecirrha</i>	4	Medusozoa	Scyphozoa	GILX00000000.1	29/04/2020	Non-specific / Whole Body	338074
<i>Copula sivickisi</i>	4	Medusozoa	Cubozoa	GHBG00000000.1	28/01/2019	Non-specific / Whole Body	417516
<i>Craterolophus convolvulus</i>	4	Medusozoa	Staurozoa	HAGZ00000000.1	29/09/2017	Non-specific / Whole Body	140294
<i>Dynamena pumila</i>	4	Medusozoa	Hydrozoa	GHMC00000000.1	29/05/2019	Non-specific / Whole Body	193769
<i>Haliclystus auricula</i>	4	Medusozoa	Staurozoa	HAHA00000000.1	29/09/2017	Non-specific / Whole Body	128662
<i>Haliclystus sanjuanensis</i>	4	Medusozoa	Staurozoa	HAHB00000000.1	29/09/2017	Non-specific / Whole Body	157217
<i>Hydra oligactis</i>	4	Medusozoa	Hydrozoa	GHUC00000000.1	08/09/2020	Non-specific / Whole Body	70952
<i>Hydractinia symbiolongicarpus</i>	4	Medusozoa	Hydrozoa	GJUZ00000000.1	17/03/2022	Non-specific / Whole Body	222167
<i>Lucernaria quadricornis</i>	4	Medusozoa	Staurozoa	HAHD00000000.1	29/09/2017	Non-specific / Whole Body	300790
<i>Millepora alcicornis</i>	4	Medusozoa	Hydrozoa	GKBI00000000.1	09/09/2022	Non-specific / Whole Body	138972
<i>Millepora complanata</i>	4	Medusozoa	Hydrozoa	GIXI00000000.1	30/08/2022	Non-specific / Whole Body	673826
<i>Millepora squarrosa</i>	4	Medusozoa	Hydrozoa	GFGU00000000.1	07/03/2018	Non-specific / Whole Body	797586
<i>Morbakka virulenta</i>	4	Medusozoa	Cubozoa	GHAF00000000.1	08/11/2018	Non-specific / Whole Body	197384
<i>Phialella quadrata</i>	4	Medusozoa	Hydrozoa	GKAK00000000.1	09/01/2023	Non-specific / Whole Body	64221

<i>Physalia physalis</i>	4	Medusozoa	Hydrozoa	GHBB00000000.1	28/01/2019	Non-specific / Whole Body	74421
<i>Podocoryna carnea</i>	4	Medusozoa	Hydrozoa	GJNL00000000.1	21/11/2022	Non-specific / Whole Body	811493
<i>Porpita porpita</i>	4	Medusozoa	Hydrozoa	GHBA00000000.1	28/01/2019	Non-specific / Whole Body	146355
<i>Rathkea octopunctata</i>	4	Medusozoa	Hydrozoa	GKAO00000000.1	09/01/2023	Non-specific / Whole Body	56744
<i>Rhopilema esculentum</i>	4	Medusozoa	Scyphozoa	GGZZ00000000.1	31/10/2019	Non-specific / Whole Body	185257
<i>Turritopsis sp.</i>	4	Medusozoa	Hydrozoa	IAAF00000000.1	17/08/2016	Non-specific / Whole Body	240888
<i>Velella velella</i>	4	Medusozoa	Hydrozoa	GHAZ00000000.1	28/01/2019	Non-specific / Whole Body	441513
<i>Anemonia viridis</i>	5	Anthozoa	Hexacorallia, Actiniaria	SRR1573633	11/09/2015	Tentacles	258361
<i>Catalaphyllia jardinei</i>	5	Anthozoa	Hexacorallia, Scleractinia	SRR18847906	21/04/2022	Tentacles	402397
<i>Ceriantheomorphe brasiliensis</i>	5	Anthozoa	Hexacorallia, Ceriantharia	SRR11802642	25/07/2020	Tentacles	214753
<i>Dofleinia cf. armata</i>	5	Anthozoa	Hexacorallia, Actiniaria	ERR6139597	10/11/2021	Tentacles	233127
<i>Entacmaea quadricolor</i>	5	Anthozoa	Hexacorallia, Actiniaria	GJFF00000000.1	19/04/2022	Tentacles	475573
<i>Exaiptasia diaphana</i>	5	Anthozoa	Hexacorallia, Actiniaria	SRR24433030	26/09/2023	Tentacles	1258938
<i>Goniopora lobata</i>	5	Anthozoa	Hexacorallia, Scleractinia	SRR18458550	24/03/2022	Tentacles	818587
<i>Goniopora norfolkensis</i>	5	Anthozoa	Hexacorallia, Scleractinia	SRR18458548	24/03/2022	Tentacles	943477
<i>Goniopora planulata</i>	5	Anthozoa	Hexacorallia, Scleractinia	SRR18458549	24/03/2022	Tentacles	931156
<i>Haliplanella luciae</i>	5	Anthozoa	Hexacorallia, Actiniaria	SRR7495168	13/11/2018	Tentacles	148450
<i>Heteractis aurora</i>	5	Anthozoa	Hexacorallia, Actiniaria	GJFG00000000.1	19/04/2022	Tentacles	458571
<i>Heteractis crispa</i>	5	Anthozoa	Hexacorallia, Actiniaria	GJFA00000000.1	18/04/2022	Tentacles	444825
<i>Heteractis magnifica</i>	5	Anthozoa	Hexacorallia, Actiniaria	GJFJ00000000.1	19/04/2022	Tentacles	488953
<i>Heterodactyla hemprichii</i>	5	Anthozoa	Hexacorallia, Actiniaria	SRR14115227	30/06/2021	Tentacles	503423
<i>Isarachmanthus nocturnus</i>	5	Anthozoa	Hexacorallia, Ceriantharia	SRR11802641	25/07/2020	Tentacles	199450
<i>Macrodictyla doreensis</i>	5	Anthozoa	Hexacorallia, Actiniaria	SRR14115222	30/06/2021	Tentacles	469869
<i>Oulactis sp.</i>	5	Anthozoa	Hexacorallia, Actiniaria	ERR2710216	03/03/2020	Tentacles	241076
<i>Pachycerianthus borealis</i>	5	Anthozoa	Hexacorallia, Ceriantharia	SRR11802643	25/07/2020	Tentacles	291992
<i>Pachycerianthus maua</i>	5	Anthozoa	Hexacorallia, Ceriantharia	SRR11802640	25/07/2020	Tentacles	273063
<i>Phyllo-discus semoni</i>	5	Anthozoa	Hexacorallia, Actiniaria	SRR14115230	30/06/2021	Tentacles	1255102
<i>Renilla koellikeri</i>	5	Anthozoa	Octocorallia, Scleractyonacea	SRR7495167	13/11/2018	Tentacles	307772
<i>Stichodactyla gigantea</i>	5	Anthozoa	Hexacorallia, Actiniaria	GJFK00000000.1	19/04/2022	Tentacles	583020
<i>Stichodactyla haddoni</i>	5	Anthozoa	Hexacorallia, Actiniaria	GJFH00000000.1	05/04/2022	Tentacles	663175
<i>Stichodactyla mertensii</i>	5	Anthozoa	Hexacorallia, Actiniaria	GJFE00000000.1	19/04/2022	Tentacles	427025
<i>Telmatactis stephensoni</i>	5	Anthozoa	Hexacorallia, Actiniaria	SRR14511809	14/11/2021	Tentacles	137386
<i>Aurelia aurita</i>	6	Medusozoa	Scyphozoa	SRR8089702	23/10/2018	Tentacles	205131
<i>Chironex fleckeri</i>	6	Medusozoa	Cubozoa	SRR1819888	09/08/2015	Tentacles	189330
<i>Chironex yamaguchii</i>	6	Medusozoa	Cubozoa	SRR8101946	01/01/2019	Tentacles	261771
<i>Chrysaora fuscescens</i>	6	Medusozoa	Cubozoa	SRR3180892	27/02/2016	Tentacles	118738
<i>Nemopilema nomurai</i>	6	Medusozoa	Scyphozoa	SRR7754710	20/09/2019	Tentacles	476585
<i>Sanderia malayensis</i>	6	Medusozoa	Scyphozoa	SRR6298216	07/12/2018	Tentacles	124838
<i>Tripedalia cystophora</i>	6	Medusozoa	Cubozoa	SRR8101526	01/01/2019	Tentacles	278958
<i>Cryptodendrum adhaesivum</i>	7	Anthozoa	Hexacorallia, Actiniaria	SRR14115232	30/06/2021	Nematosphere s	878716
<i>Heterodactyla hemprichii</i>	7	Anthozoa	Hexacorallia, Actiniaria	SRR14115226	30/06/2021	Nematosphere s	313286

Table S2. BUSCO statistics for the 27 SRA-derived transcriptomes. C – Complete; S – Singular Copies; D – Duplicated Copies; F – Fragment Copies; M – Missing Copies.

Database	Species	Code	C	S	D	F	M
5	<i>Anemonia viridis</i>	SRR1573633	91,30%	54,80%	36,50%	6,00%	2,70%
5	<i>Dofleinia cf. Armata</i>	ERR6139597	93,90%	59,40%	34,50%	3,90%	2,20%
5	<i>Exaiptasia diaphana</i>	SRR24433030	98,80%	17,90%	80,90%	0,90%	0,30%
5	<i>Haliplanella luciae</i>	SRR7495168	76,90%	49,60%	27,30%	13,40%	9,70%
5	<i>Heterodactyla hemprichii</i>	SRR14115227	87,50%	65,20%	22,30%	9,10%	3,40%
5	<i>Macroductyla doreensis</i>	SRR14115222	79,80%	53,00%	26,80%	11,80%	8,40%
5	<i>Oulactis sp.</i>	ERR2710216	94,30%	52,40%	41,90%	3,60%	2,10%
5	<i>Phyllodiscus semoni</i>	SRR14115230	98,20%	13,10%	85,10%	0,60%	1,20%
5	<i>Telmatactis stephensoni</i>	SRR14511809	55,50%	50,40%	5,10%	26,90%	17,60%
5	<i>Ceriantheomorpha brasiliensis</i>	SRR11802642	91,70%	66,00%	25,70%	4,80%	3,50%
5	<i>Isarachnanthus nocturnus</i>	SRR11802641	83,60%	55,70%	27,90%	9,40%	7,00%
5	<i>Pachycerianthus borealis</i>	SRR11802643	96,80%	50,90%	45,90%	1,30%	1,90%
5	<i>Pachycerianthus maua</i>	SRR11802640	82,40%	44,50%	37,90%	11,10%	6,50%
5	<i>Renilla koellikeri</i>	SRR7495167	96,30%	23,90%	72,40%	1,30%	2,40%
5	<i>Catalaphyllia jardinei</i>	SRR18847906	97,00%	46,10%	50,30%	1,80%	1,80%
5	<i>Goniopora lobata</i>	SRR18458550	96,60%	41,50%	55,50%	1,50%	1,50%
5	<i>Goniopora norfolkensis</i>	SRR18458548	96,60%	43,10%	53,60%	1,90%	1,50%
5	<i>Goniopora planulata</i>	SRR18458549	96,90%	44,70%	52,20%	1,90%	1,20%
6	<i>Chironex fleckeri</i>	SRR1819888	79,50%	68,30%	11,20%	10,50%	10,00%
6	<i>Chironex yamaguchii</i>	SRR8101946	89,40%	65,60%	23,80%	5,50%	5,10%
6	<i>Chrysaora fuscescens</i>	SRR3180892	58,30%	52,30%	6,00%	22,90%	18,80%
6	<i>Tripedalia cystophora</i>	SRR8101526	92,70%	71,80%	20,90%	4,00%	3,30%
6	<i>Aurelia aurita</i>	SRR8089702	73,80%	55,20%	18,60%	14,20%	12,00%
6	<i>Nemopilema nomurai</i>	SRR7754710	96,80%	26,60%	70,20%	1,30%	1,90%
6	<i>Sanderia malayensis</i>	SRR6298216	92,60%	72,20%	21,40%	3,20%	3,20%
7	<i>Cryptodendrum adhaesivum</i>	SRR14115232	90,10%	42,30%	47,80%	4,80%	5,10%
7	<i>Heterodactyla hemprichii</i>	SRR14115226	50,30%	43,50%	7,80%	31,00%	18,70%

Table S3. Number of proteins in each database, provided in three stages: original dataset, after duplicate removal (No_Dup) and following AMPir prediction filtering (No_Dup_AMPir). Statistics are included for each stage. No_Red_98% corresponds to protein counts after applying redundancy removal with CD-HIT at 0.98 sequence identity.

Database	Number of sequences	Sum of lengths	Minimum length	Average length	Maximum length
Original_1	204,693	87,052,731	8	425.3	22,539
Filtered_1_No_Dup	204,647	87,043,996	8	425.3	22,539
Filtered_1_No_Dup_AMPir	1473	147,689	45	100.3	627
Original_2	45,860	22,913,634	11	499.6	8,156
Filtered_2_No_Dup	45,859	22,913,498	11	499.7	8,156
Filtered_2_No_Dup_AMPir	191	23,419	55	112.6	360
Original_3	20,554,309	2,392,302,564	49	116.4	18,012
Filtered_3_No_Dup	15,491,809	1,917,236,786	49	123.8	18,012
Filtered_3_No_Dup_AMPir	1,045,220	70,691,396	49	67.6	606
Original_4	6,345,131	744,878,089	49	117.4	24,292
Filtered_4_No_Dup	5,018,612	617,800,624	49	123.1	24,292
Filtered_4_No_Dup_AMPir	326,045	21,851,044	49	67	607
Original_5	12,429,521	1,474,576,001	49	118.6	21,468
Filtered_5_No_Dup	8,338,288	1,079,249,660	49	129.4	21,468
Filtered_5_No_Dup_AMPir	579,611	38,945,446	49	67.2	609
Original_6	1,655,351	200,397,681	49	121.1	24,271
Filtered_6_No_Dup	1,132,735	149,146,332	49	131.7	24,271
Filtered_6_No_Dup_AMPir	96,921	6,412,716	49	66.2	586
Original_7	11,920,02	142,108,627	49	119.2	7,637
Filtered_7_No_Dup	925,924	117,808,010	49	127.2	7,637
Filtered_7_No_Dup_AMPir	53,236	3,632,718	49	68.2	305
NoAMPir_Composite_DB	31,157,874	3,991,198,906	8	128.1	24,292
NoAMPir_Composite_DB_No_Dup	30,626,016	3,942,328,738	8	128.7	24,292
NoAMPir_Composite_DB_No_Red_98%	25,264,871	3,009,412,726	11	119.1	24,292
AMPir_Composite_DB	2,102,697	141,704,428	45	67.4	627
AMPir_Composite_DB_No_Dup	2,072,171	139,648,678	45	67.4	627
AMPir_Composite_DB_No_Red_98%	1,939,076	130,783,896	45	67.4	627

Table S4. Statistics for each mining step (peptides, AMPs and non-haemolytic non-toxic AMPs). Min – Minimum. Av – Average. Max – Maximum. db – database; aspn - AspN, chym - Chymotrypsin, gluc - GluC, protk - Proteinase K and tryp – Trypsin.

Proteolysis Protocol	Peptides					AMPs					Non-Haemolytic and Non-Toxic AMPs				
	Number of sequences	Sum of lengths	Min. length	Av. length	Max. length	Number of sequences	Sum of lengths	Min. length	Av. length	Max. length	Number of sequences	Sum of lengths	Min. length	Av. length	Max. length
db1_aspn	4,376	81,416	11	18.6	40	214	4,188	11	19.6	38	9	184	12	20.4	33
db1_chym	2,696	40,761	11	15.1	40	35	639	11	18.3	36	3	64	16	21.3	31
db1_gluc	4,447	84,524	11	19	40	298	7,007	11	23.5	40	20	380	12	19	38
db1_protk	175	2,158	11	12.3	20	0	0	0	0	0	0	0	0	0	0
db1_tryp	4,318	78,289	11	18.1	40	49	857	11	17.5	34	4	54	12	13.5	17
db2_aspn	698	12,6	11	18.1	40	31	588	11	19	34	0	0	0	0	0
db2_chym	434	6,55	11	15.1	35	7	127	14	18.1	23	0	0	0	0	0
db2_gluc	744	14,292	11	19.2	40	47	1,115	11	23.7	38	1	18	18	18	18
db2_protk	25	318	11	12.7	22	1	14	14	14	14	0	0	0	0	0
db2_tryp	670	12,251	11	18.3	39	11	203	12	18.5	35	1	23	23	23	23
db3_aspn	1,820,209	33,705,927	11	18.5	40	134,007	2,591,695	11	19.3	40	5,055	102,719	11	20.3	40
db3_chym	818,622	11,676,951	11	14.3	40	14,534	247,057	11	17	40	946	16,993	11	18	40
db3_gluc	1,609,672	32,461,771	11	20.2	40	78,25	1,893,586	11	24.2	40	4,974	97,263	11	19.6	40
db3_protk	58,164	715,16	11	12.3	38	456	6,549	11	14.4	38	9	126	11	14	19
db3_tryp	1,849,069	33,133,654	11	17.9	40	38,029	651,794	11	17.1	40	4,023	65,665	11	16.3	40
db4_aspn	586,367	10,802,126	11	18.4	40	42,756	812,875	11	19	40	1,486	29,222	11	19.7	40
db4_chym	250,336	3,542,070	11	14.1	40	4,26	70,45	11	16.5	40	269	4,579	11	17	31
db4_gluc	519,668	10,417,513	11	20	40	24,7	585,475	11	23.7	40	1,586	29,954	11	18.9	40
db4_protk	17,08	208,892	11	12.2	37	111	1,539	11	13.9	22	3	46	12	15.3	20
db4_tryp	595,91	10,654,503	11	17.9	40	12,479	210,986	11	16.9	40	1,356	21,75	11	16	37
db5_aspn	996,612	18,503,524	11	18.6	40	72,434	1,394,948	11	19.3	40	2,633	53,244	11	20.2	40
db5_chym	436,127	6,201,383	11	14.2	40	3,929	66,532	11	16.9	40	234	4,082	11	17.4	38
db5_gluc	883,537	17,852,510	11	20.2	40	40,721	980,619	11	24.1	40	2,64	51,332	11	19.4	40
db5_protk	30,722	377,017	11	12.3	38	239	3,398	11	14.2	34	11	150	11	13.6	19
db5_tryp	1,014,421	18,203,744	11	17.9	40	20,694	352,647	11	17	40	2,134	34,286	11	16.1	40
db6_aspn	177,282	3,251,508	11	18.3	40	12,861	240,727	11	18.7	40	389	7,5	11	19.3	40
db6_chym	69,353	965,782	11	13.9	40	1,162	18,719	11	16.1	39	70	1,158	11	16.5	33
db6_gluc	153,954	3,102,010	11	20.1	40	6,591	153,405	11	23.3	40	416	7,489	11	18	38
db6_protk	4,704	56,995	11	12.1	23	21	297	11	14.1	23	0	0	0	0	0
db6_tryp	179,035	3,182,422	11	17.8	40	3,8	63,469	11	16.7	40	375	5,831	11	15.5	39
db7_aspn	97,943	1,804,776	11	18.4	40	6,914	135,987	11	19.7	40	299	6,292	11	21	39
db7_chym	49,536	717,004	11	14.5	40	937	16,47	11	17.6	38	54	1,064	11	19.7	35
db7_gluc	86,699	1,743,390	11	20.1	40	4,444	110,064	11	24.8	40	320	6,579	11	20.6	40
db7_protk	3,909	48,182	11	12.3	27	45	682	11	15.2	25	0	0	0	0	0
db7_tryp	100,524	1,809,687	11	18	40	2,029	35,385	11	17.4	40	208	3,577	11	17.2	34

BUSCO Assessment Results

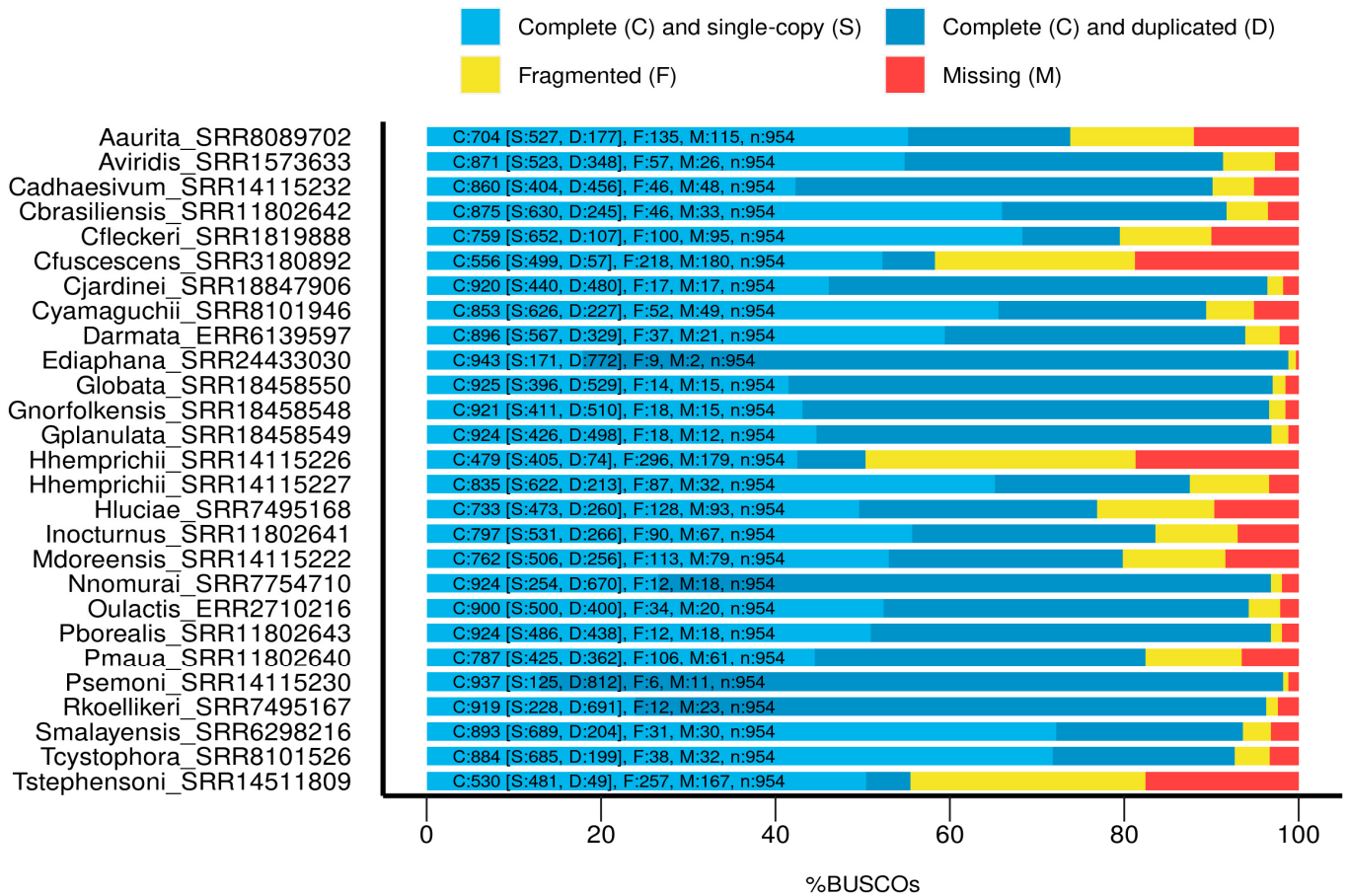


Figure S1. Completeness scores of the 27 assembled transcriptomes obtained from SRA data (NCBI), assessed with BUSCO using the Metazoa dataset. The chart shows the proportion of complete, fragmented, and missing gene copies for each transcriptome.



Figure S2. Venn Diagrams representing the number of predicted AMPs by AMPir (pink), AMPlify (blue) and Macrel (yellow) from each protease-database combination. db – database; aspn - AspN, chym - Chymotrypsin, gluc - GluC, protk - Proteinase K and tryp – Trypsin.



Figure S3. Venn Diagrams representing the number of predicted non-haemolytic AMPs by HemoPi (pink), MQSSM (blue) and Macrel (yellow) from each protease-database pair. db – database; aspN - AspN, chym - Chymotrypsin, gluc - GluC, protk - Proteinase K and trypt – Trypsin.



Figure S4. Venn Diagrams representing the number of predicted non-haemolytic and non-toxic AMPs by ToxinPred3 (pink), ToxTeller (blue) and CAPTP (yellow) from each protease-database pair. db – database; aspn - AspN, chym - Chymotrypsin, gluc - GluC, protk - Proteinase K and tryp – Trypsin.

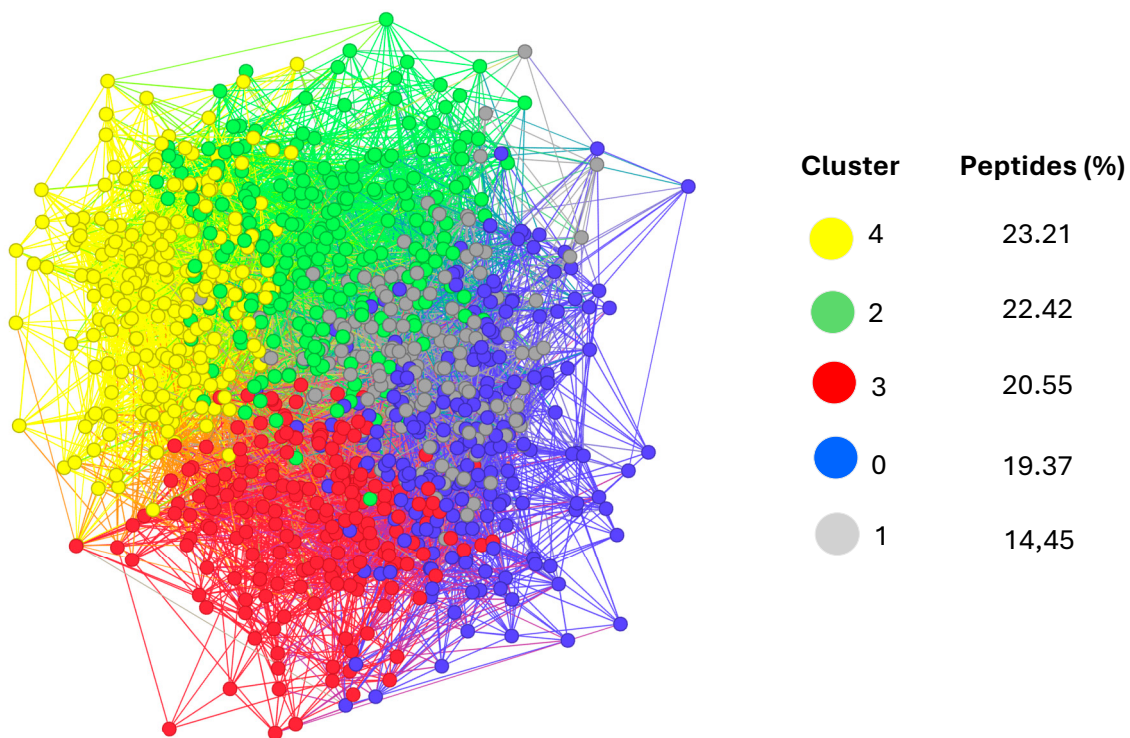
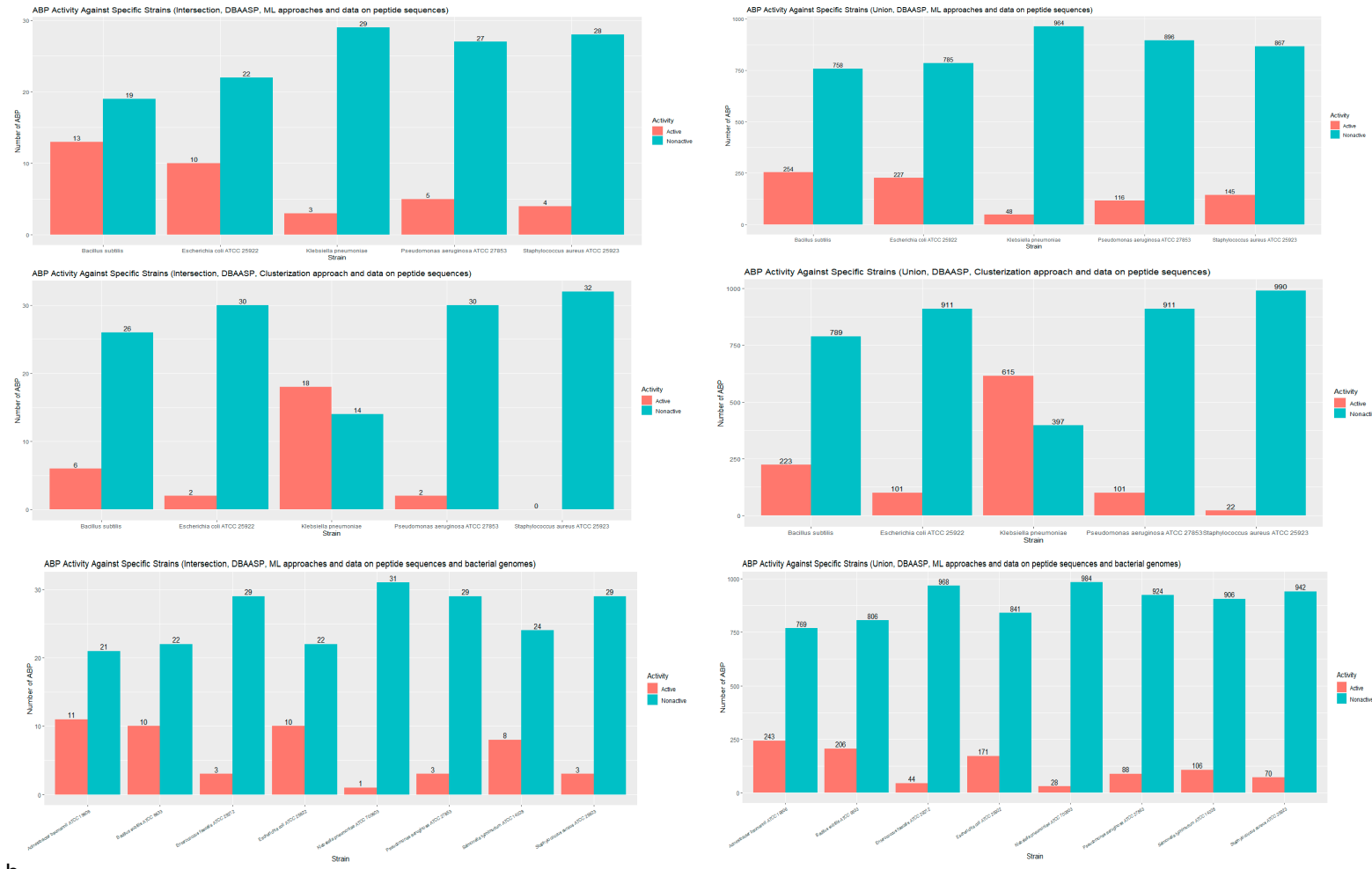


Figure S5. HSPN clustering of the 1,017 AMPs obtained from the intersection of the Harmonic and Hub-Bridge datasets within the Cnidaria Singular AMPs (CnSA). Distinct clusters are highlighted in different colours.

a



b



Figure S6. a) Strain-specific activity predictions for ABPs derived from the most representative CnSA datasets. Predictions were performed for both the union dataset (1,012 peptides) and the intersection dataset (32 peptides). Three models from DBAASP were used - Model I: ML predictions based on AMP sequence data; Model II: Cluster-based predictions using peptide sequence data; Model III: ML predictions integrating peptide and bacterial genome data. Results are represented in the boxplot, with red boxes indicating peptides predicted as active and blue boxes indicating non-active peptides for specific bacterial strains. **b)** Venn diagrams summarizing activity predictions for five bacterial strains - *Bacillus subtilis*, *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Staphylococcus aureus*. The diagrams illustrate the overlap of the strain-specific activity predictions from Figure 8a. These analyses were used to rank the potential activities of Cnidaria-derived ABPs against the selected strains.

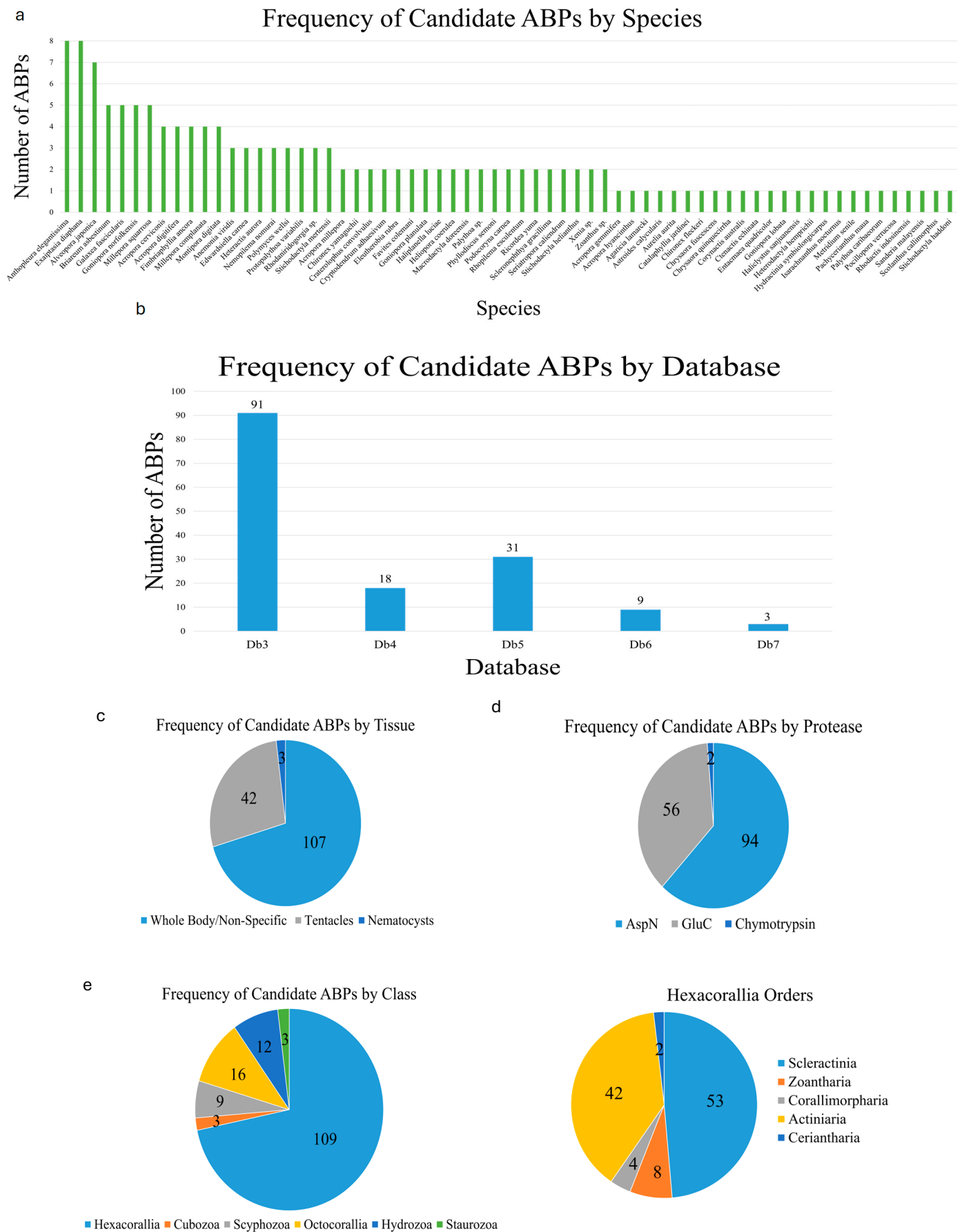
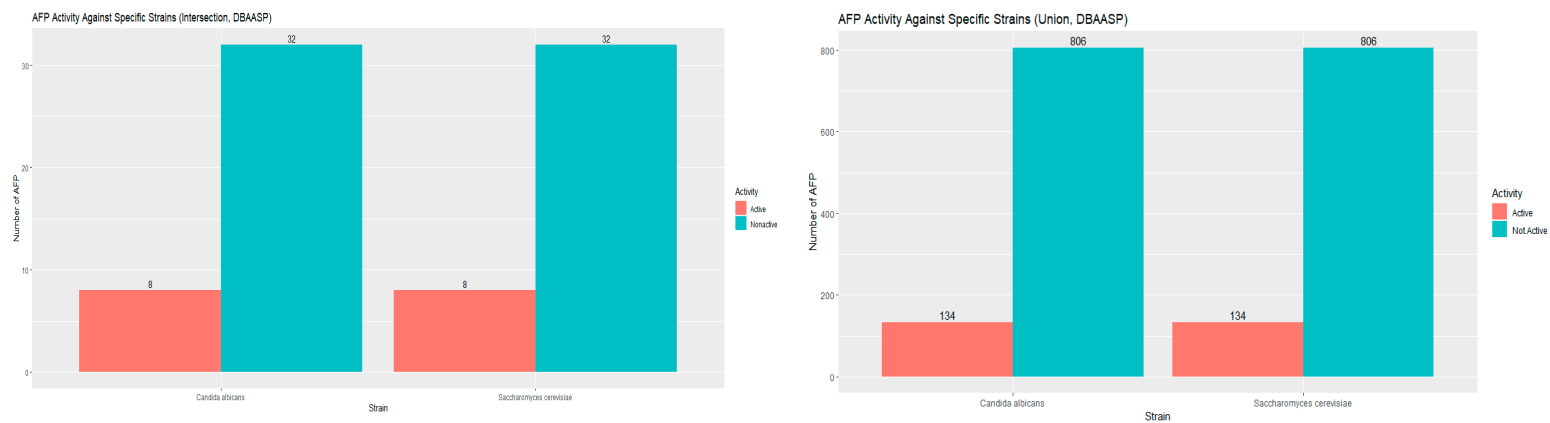


Figure S7. Distribution of the 152 predicted antimicrobial peptides (ABPs) across various categories: **a)** Cnidaria species; **b)** database origin; **c)** tissue type; **d)** protease associated with the in silico proteolysis; and **e)** Cnidarian classes, with a focus on Hexacorallia orders (Scleractinia, Zoantharia, Corallimorpharia, Actiniaria and Ceriantharia).

a



b



Figure S8. a) Antifungal specific predictions; and **b)** Antiviral specific predictions. Predictions were made for both intersection and union datasets in DBAASP.