






Review

# Nonribosomal Peptide Synthetases in Animals

Wouter Suring<sup>1,2</sup> , Dylan Hoogduin<sup>1</sup>, Giang Le Ngoc<sup>1,3</sup> , Abraham Brouwer<sup>4</sup> , Nico M. van Straalen<sup>1,\*</sup>  and Dick Roelofs<sup>1,5</sup> 

- <sup>1</sup> A-LIFE Ecology and Evolution, Faculty of Science, Vrije Universiteit Amsterdam, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands
- <sup>2</sup> Department of Academy Technology & Innovation, NHL Stenden University of Applied Sciences, Rengerslaan 8-10, 8917 DD Leeuwarden, The Netherlands
- <sup>3</sup> Biomedical Primate Research Centre, Lange Kleiweg 161, 2282 GJ Rijswijk, The Netherlands
- <sup>4</sup> BioDetection Systems, Science Park 406, 1098 XH Amsterdam, The Netherlands
- <sup>5</sup> Keygene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands
- \* Correspondence: n.m.van.straalen@vu.nl

**Abstract:** Nonribosomal peptide synthetases (NRPSs) are a class of cytosolic enzymes that synthesize a range of bio-active secondary metabolites including antibiotics and siderophores. They are widespread among both prokaryotes and eukaryotes but are considered rare among animals. Recently, several novel NRPS genes have been described in nematodes, schistosomes, and arthropods, which led us to investigate how prevalent NRPS genes are in the animal kingdom. We screened 1059 sequenced animal genomes and showed that NRPSs were present in 7 out of the 19 phyla analyzed. A phylogenetic analysis showed that the identified NRPSs form clades distinct from other adenylate-forming enzymes that contain similar domains such as fatty acid synthases. NRPSs show a remarkably scattered distribution over the animal kingdom. They are especially abundant in rotifers and nematodes. In rotifers, we found a large variety of domain architectures and predicted substrates. In the nematode *Plectus sambesii*, we identified the beta-lactam biosynthesis genes L-δ-(α-amino adipoyl)-L-cysteinyl-D-valine synthetase, isopenicillin N synthase, and deacetoxy-cephalosporin C synthase that catalyze the formation of beta-lactam antibiotics in fungi and bacteria. These genes are also present in several species of Collembola, but not in other hexapods analyzed so far. In conclusion, our survey showed that NRPS genes are more abundant and widespread in animals than previously known.

**Keywords:** NRPS; nonribosomal; peptide; synthetase; animals; Metazoa; distribution; evolution; phylogeny; beta-lactam



**Citation:** Suring, W.; Hoogduin, D.; Le Ngoc, G.; Brouwer, A.; van Straalen, N.M.; Roelofs, D. Nonribosomal Peptide Synthetases in Animals. *Genes* **2023**, *14*, 1741. <https://doi.org/10.3390/genes14091741>

Academic Editor: Alexander S. Graphodatsky

Received: 23 July 2023  
Revised: 25 August 2023  
Accepted: 28 August 2023  
Published: 30 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Nonribosomal peptide synthetases (NRPSs) synthesize a range of peptide products with a wide spectrum of biological functions including antibiotic and siderophore activities. They are used in industrial biotechnology to produce various pharmaceuticals such as cytostatics and immunosuppressants. NRPSs are widespread among both prokaryotes and eukaryotes, but they have only been identified in a limited number of animal species [1–4]. Recently, Shou et al. (2016) showed that the nematode *Caenorhabditis elegans* uses two NRPSs to produce the compound nemamide that promotes larval survival [2]. Homologs of these NRPSs exist in over twenty nematode species [2,5]. Furthermore, in schistosomes (Trematoda, Platyhelminthes), an NRPS is involved in the production of a male factor promoting female development [4]. In springtails (Collembola, Arthropoda), we identified the NRPS L-δ-(α-amino adipoyl)-L-cysteinyl-D-valine synthetase that is known to catalyze the first step in beta-lactam antibiotic biosynthesis [3]. It is part of a beta-lactam biosynthesis gene cluster and is suggested to be present in over 60% of springtail families, especially in soil-living species [6]. These findings primed us to investigate how widespread NRPSs are in the animal kingdom as a whole.

NRPSs can synthesize a large range of compounds due to their modular architecture. Each module adds one amino acid to the peptide that is produced in the cytosol (nonribosomally). The enzymatic center typically consists of three domains. The adenylation domain (A domain) activates the aminoacyl substrate before transferring it to a pantetheine cofactor that is covalently attached to the peptide carrier protein domain (PCP domain) of the NRPS [7]. Next, the condensation domain (C domain) catalyzes the formation of a peptide bond between the amino acid substrates of adjacent modules. NRPS enzymes can contain many modules, resulting in a great variety of possible peptide products [8]. In addition, modules can incorporate nonproteinogenic amino acids (e.g., in the case of penicillin biosynthesis). Furthermore, NRPSs can contain additional domains that oxidize, circularize, or otherwise modify the peptide product and can even form hybrid enzymes with polyketide synthases.

NRPSs require activation through the attachment of phosphopantetheine groups to their peptidyl carrier domains by phosphopantetheine transferases (PPTases). Microorganisms utilize multiple PPTases for different groups of enzymes [9,10]. PPTases are also present in animals because they are required for the activation of acyl carrier proteins involved in fatty acid metabolism. In the springtail *Folsomia candida*, in which an NRPS was previously identified, only a single PPTase was found [11]. This PPTase was coregulated with the NRPS during stress, suggesting that a native arthropod enzyme was recruited to activate the NRPS obtained by horizontal gene transfer.

Other types of NRPSs are also known to exist in animals. These are the single-module Ebony [12] (in *Drosophila melanogaster* and other insects) and acyl CoA-synthetase family member 4 (ACSF4-U26) (e.g., in mammals) [13]. However, these enzymes use other domains, not the condensation domains found in the NRPSs investigated here [14]. The Ebony protein is involved in pigmentation and other processes [15]. ACSF4-U26 is a beta-alanine-activating enzyme that may be involved in post-translational modification [13].

While the diversity of bacterial and fungal NRPSs is staggering and receives a lot of attention, due to their great biotechnological potential, a comparative study of NRPSs in the animal kingdom has not yet been conducted. In this study, we analyzed 1059 sequenced animal genomes for NRPSs to map the distribution and abundance of metazoan NRPSs and provide insight into their evolutionary history. We identified candidate NRPS genes in 86 species from 7 different phyla. NRPSs appeared to be especially abundant in rotifers, where we found both NRPS-PKS and NRPS genes with many different predicted substrates. Finally, we describe a second potential case of beta-lactam biosynthesis capacity in an animal genome, the nematode *Plectus sambei*.

## 2. Materials and Methods

### 2.1. NRPS Identification

We analyzed 1059 genomes from 19 animal phyla (Supplementary Table S1) for putative NRPS genes using antiSMASH software version 4.1.0 [16]. Both the bacterial and fungal versions of antiSMASH were used for the analysis. For each scaffold or contig containing a putative NRPS, we performed a check for contamination, since genome assemblies sometimes contain sequences from microorganisms. We extracted the scaffold sequences from the genomes and ran a BLASTN against the NCBI nonredundant database. The idea is that the putative NRPS may resemble microbial sequences but that (parts of) the surrounding area in the genome should be similar to metazoan sequences.

We manually analyzed the results and considered a scaffold as an animal sequence when at least one hit with an e-value  $< 1 \times 10^{-10}$  against another animal was found. When no such hit was found, but instead a hit against a microorganism with an e-value  $< 1 \times 10^{-10}$  was returned, we considered the scaffold potentially contaminated. When no hits below the threshold against any organism were found, we manually analyzed the scaffold by BLASTing predicted protein sequences from the scaffold against the NCBI nonredundant database. In some cases, the nature of the scaffold (microbial or metazoan) could not be

reliably determined (usually when the scaffold was very short). These cases were noted as inconclusive and were not used for further analysis.

Since antiSMASH did not identify the NRPSs previously reported in Platyhelminthes [4] (probably due to their specific domain architecture) we used the *Schistosoma mansoni* NRPS (XP\_018648700) to analyze other animal phyla for these types of NRPSs. We conducted a TBLASTN against the nr/nt database of NCBI and included each hit with an e-value  $<1 \times 10^{-10}$  in our analysis. To visualize the distribution of the analyzed genomes and identified putative NRPSs, we extracted a phylogenetic tree of the animal phyla from the open tree of life version 13.4 [17] using the rotl package version 3.0.12 [18] in R 4.1.0 (R Core Team, 2021).

## 2.2. Phylogenetic Analyses

Our phylogenetic analysis of NRPSs in animals was wholly based on the adenylation domains; however, these domains are also found in other enzymes present in animals (e.g., in acyl-coenzyme A synthetases). To make sure that the identified putative NRPSs in animals were not from this group, we included other metazoan adenylate-forming enzymes in our analysis. The amino acid sequences of the adenylation domains identified by antiSMASH were extracted. A set of metazoan adenylate-forming enzymes (Supplementary Table S2) was extracted from UniProt [19] to be used as the outgroup. InterProScan [20] was used to predict the adenylation domains of the outgroup proteins as well as the proteins identified using the *S. mansoni* NRPS (Sm-NRPS), and these domain sequences were extracted.

Only the extracted adenylation domains  $\geq 300$  amino acids were used in the phylogenetic analysis. The adenylation domains of all animals except rotifers (which were used in a separate analysis) were aligned using Clustal Omega version 1.2.4 [21] using default settings. Columns with  $<70\%$  coverage were trimmed using trimAl version 1.4.rev15 [22]. The resulting alignment was manually inspected and sequences with gaps over a large stretch of the alignment and sequences that contained Xs were removed before realigning and trimming. A phylogenetic tree was inferred by maximum likelihood using IQTREE version 2.2.3 [23] using the Q.pfam + F + R7 model with 1000 bootstrap replicates.

An ACVS-based phylogenetic tree was constructed using IQ-TREE version 2.0.6 [23]. The JTT + I + G4 model was used with 1000 bootstrap replicates. The IPNS-based phylogenetic tree was constructed using the LG + G4 model with 1000 bootstrap replicates. The cefEF-based phylogenetic was constructed using the WAG + I + G4 model with 1000 bootstrap replicates.

A phylogenetic analysis of rotifer sequences was performed using IQTREE version 2.0.6 [23] using the LG + F + R5 model with 1000 bootstrap replicates. The phylogenetic trees were visualized using iTOL version 6 [24].

## 2.3. NRPSs in Rotifers

Because so many putative NRPSs were identified in rotifers, we did a separate analysis for this group. The adenylation domains were extracted from these sequences and aligned to the adenylation domains of a selection of other rotifer adenylate-forming enzymes (Supplementary Table S2). These domains were aligned using Clustal Omega version 1.2.4 [21] using default settings, and the alignment was manually inspected. Using trimAl version 1.4.rev15 [22], columns with  $<70\%$  coverage were trimmed. The substrate predictions (based on profile hidden Markov models) for the adenylation domains of putative rotifer NRPSs were extracted from the antiSMASH output.

## 2.4. Beta-Lactam Biosynthesis Genes in *Plectus Sambesii*

AntiSMASH identified a putative NRPS gene in the nematode *Plectus sambesii* that resembled L- $\delta$ -( $\alpha$ -aminoadipoyl)-L-cysteinyl-D-valine synthetase (ACVS) (approximately 50% identity on the amino acid level). ACVS catalyzes the first step in beta-lactam biosynthesis in bacteria and fungi and potentially also in springtails. Because this would represent

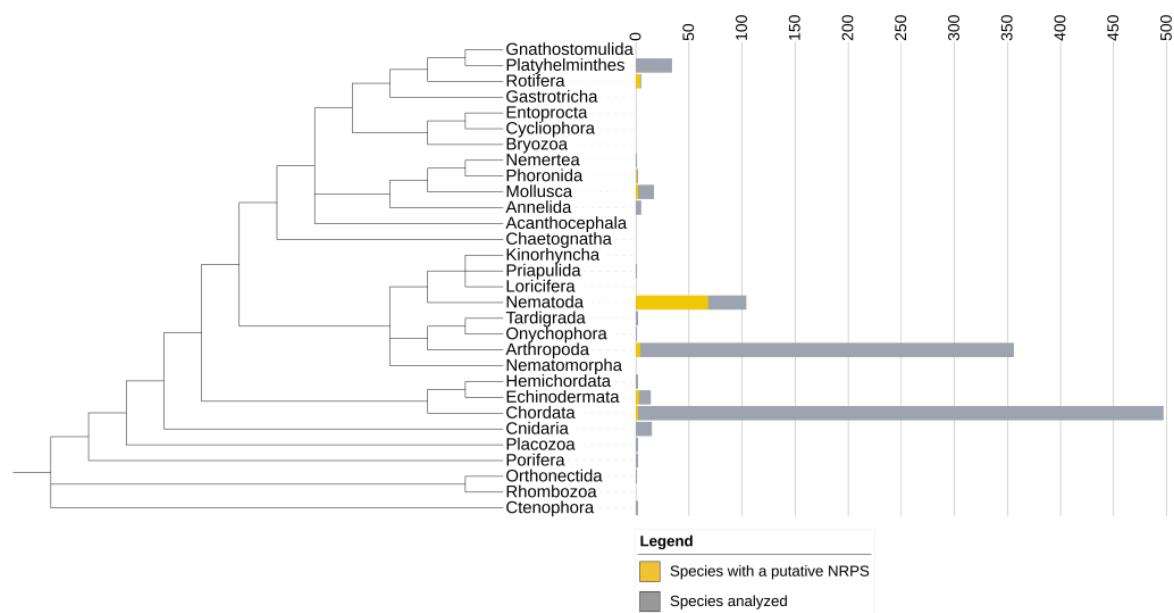
another remarkable case (in addition to Collembola) of beta-lactam biosynthesis genes in animals, we did a separate analysis for these genes in nematodes. We identified other beta-lactam biosynthesis genes in *P. sambesii* using TBLASTN with a selection of beta-lactam biosynthesis genes from microorganisms (Supplementary Table S3). Genome regions with a hit below the threshold of  $1 \times 10^{-10}$  were translated to proteins and used in a BLASTP against the NCBI nonredundant database. When they had a lower e-value against a non-beta-lactam protein, they were excluded from further analysis.

Phylogenetic trees of ACVS, IPNS and cefEF proteins from bacteria, fungi, spring-tails, and *P. sambesii* were inferred to see whether the nematode proteins form monophyletic clades with other animals. The proteins were aligned using Clustal Omega version 1.2.4 [21] with default settings, and the resulting alignments were manually adjusted (Supplementary File S1).

### 3. Results

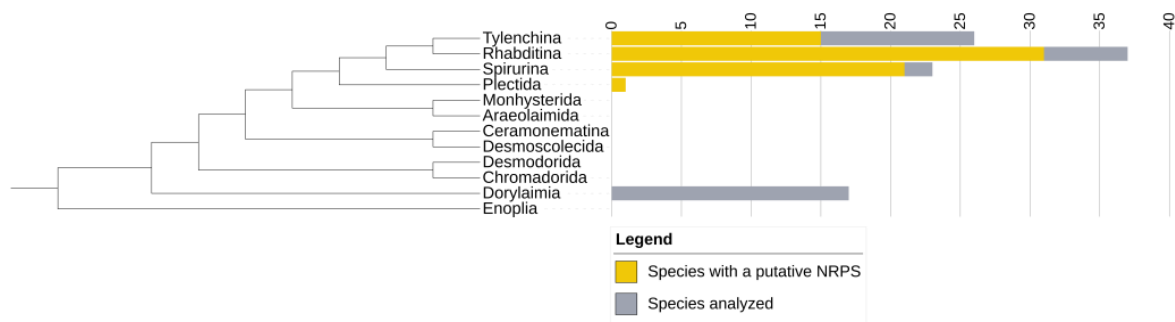
#### 3.1. The Distribution of NRPS Genes in Animals

Genomic data were available for 19 animal phyla, with most genomes representing Chordata, Nematoda, and Arthropoda. We identified NRPS clusters in over 8% of the 1059 animal genomes analyzed (Figure 1a, Supplementary Table S4). For each NRPS cluster identified by antiSMASH, we conducted a contamination check to analyze whether the putative NRPSs were truly part of the animal genome by evaluating whether the surrounding genes had best BLAST hits to other animals. AntiSMASH identified 772 putative NRPS clusters, of which 199 passed the contamination check and are likely true animal sequences (Supplementary Table S4). The occurrence of NRPSs showed a scattered distribution over metazoan phyla: Annelida, Arthropoda, Chordata, Cnidaria, Echinodermata, Hemichordata, Mollusca, Nematoda, Nemertea, Phoronida, Platyhelminthes, Priapulida, and Rotifera. Within the phylum Nematoda, genomes were available from Rhabditina, Tylenchina, Spirurina, Plectida, and Dorylaimia. NRPS and NRPS–polyketide synthases (NRPS–PKS) were identified in many species, but not in the early-diverging Dorylaimia (Figure 1b).



(a)

Figure 1. Cont.



(b)

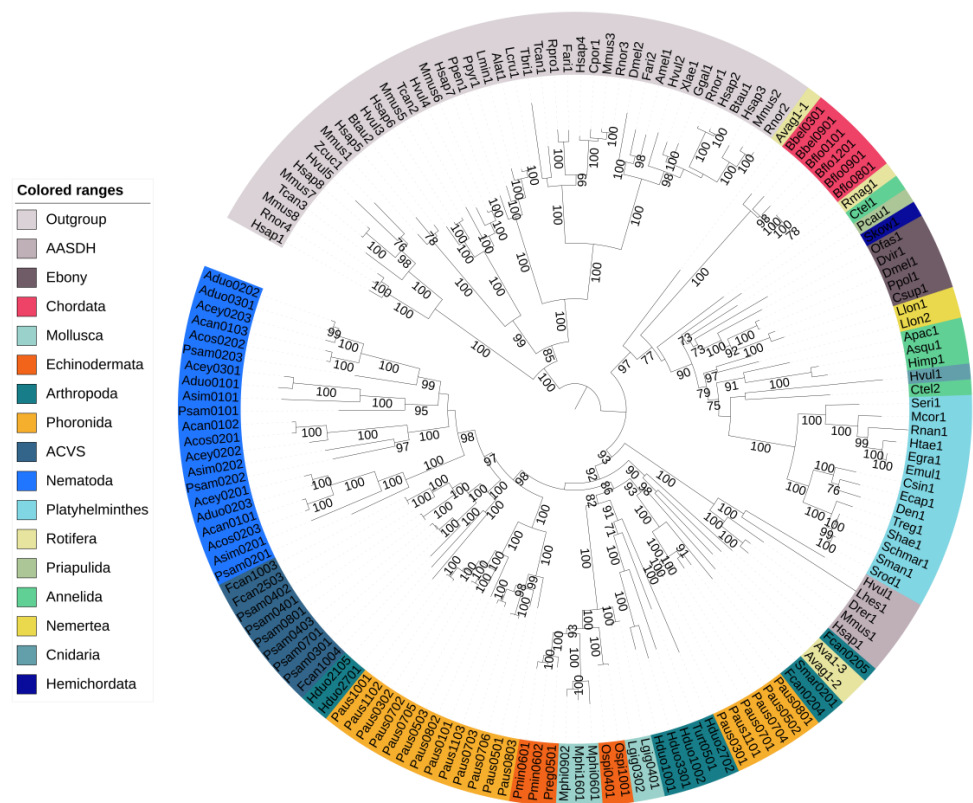
**Figure 1.** Distribution of identified putative nonribosomal peptide synthetases (NRPSs) across the animal kingdom. The sums of the yellow and grey bars represent the number of genomes analyzed. The yellow bars represent the number of genomes that contained at least one putative NRPS identified by antiSMASH. (a) NRPSs were scattered over the animal phyla; they were present in both early diverging phyla and phyla that evolved later. Most sequenced genomes were from Arthropoda, Chordata, and Nematoda. Putative NRPSs were identified in Nematoda, Rotifera, Arthropoda, Chordata, Phoronida, Mollusca and Echinodermata. NRPSs not recognized by antiSMASH (e.g. in Platyhelminthes) are not included here (b) Species from five nematode clades were analyzed in more detail. NRPSs were identified in Rhabditina, Tylenchina, Spirurina, and Plectida. No putative NRPSs were identified in Dorylaimia. The nematode phylogeny was based on [25].

Figure 2 shows the phylogenetic relationships among all animal NRPS sequences. NRPSs from Platyhelminthes clustered with Ebony as expected. The A domains from Annelida, Priapulida, Hemichordata, Nemertea and Cnidaria identified using Sm-NRPS were also part of this clade (Figure 2 and Table S5). In this clade, we also found *Branchiostoma belcheri* and *Branchiostoma floridae* (Chordata). In contrast to Ebony proteins, the *Branchiostoma* proteins did appear to contain regular condensation domains. A rotifer NRPS that showed similarity to Sm-NRPS contained one adenylation domain that clustered with Ebony and two adenylation domains that clustered with NRPSs identified by antiSMASH.

The NRPSs in Nematoda formed a monophyletic clade except for some *Plectus sambesii* adenylation domains. This is discussed below under beta-lactam biosynthesis genes in *P. sambesii*. A tree with all nematode adenylation domains is available in Supplementary Figure S1. In the springtail *F. candida* (Arthropoda), we identified a second NRPS, in addition to the two copies of ACVS reported previously, on scaffold LNIX01000001.1. Furthermore, in another springtail, *Holacanthella duospinosa*, we also found an NRPS that was different from ACVS.

The NRPSs identified in Echinodermata (*Ophiothrix spiculata*, *Patiria miniate*, *Patiriella regularis*) and Mollusca (*Lottia gigantea*, *Modiolus philippinarum*) formed a monophyletic clade, even though the phyla themselves were not related at all (echinoderms are deuterostomes, molluscs are protostomes) (Figure 1a). It should be noted that all NRPSs in molluscs and echinoderms are located on short scaffolds, which makes it more difficult to achieve absolute certainty regarding the absence of contamination. Lastly, a clade of arthropod NRPSs consisting of *Tetranychus urticae* (Acari) and *Holacanthella duospinosa* (Collembola) was present in the tree.





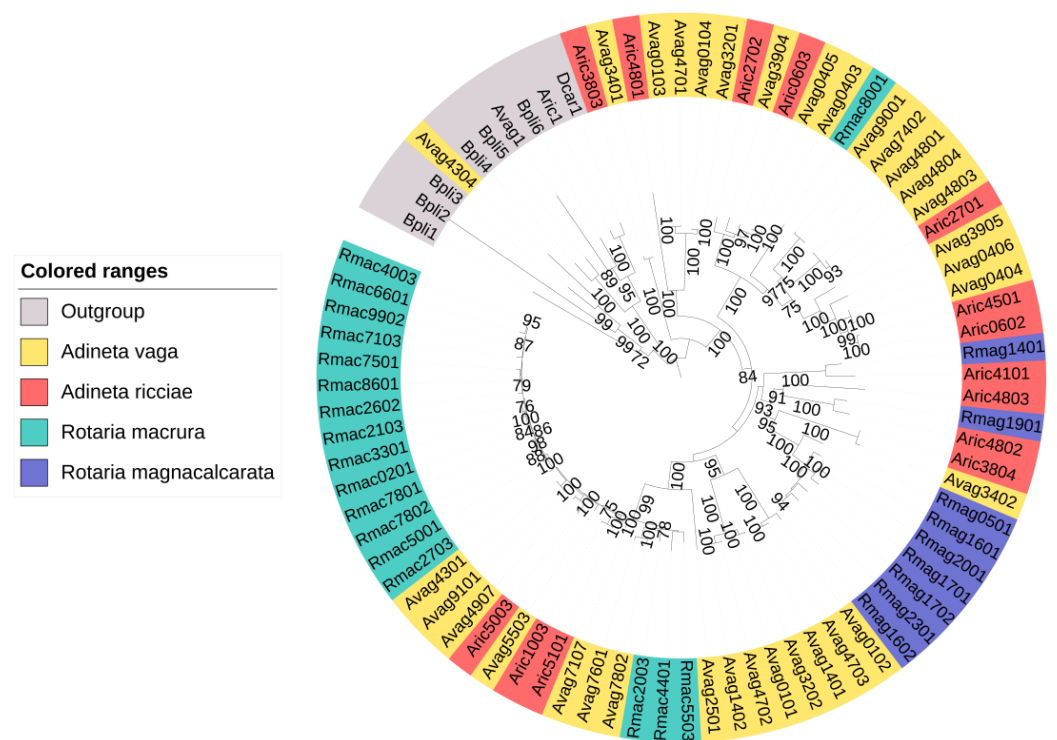
**Figure 2.** Phylogenetic tree of putative nonribosomal peptide synthetases and other adenylate-forming enzymes in animals. The outgroup proteins formed a monophyletic clade. Ebony clustered with NRPSs from Platyhelminthes and other proteins identified based on their similarity to Sm-NRPS. The chordata are also part of this clade. AASDH clustered with putative NRPSs identified by antiSMASH. NRPSs identified in Mollusca and Echinodermata were found in a single clade in the tree. The nematode NRPSs formed a monophyletic clade except for several *Plectus sambesii* adenylation domains, which clustered in the ACVS clade. Bootstrap support  $\geq 70\%$  is shown.

3.2. Putative NRPSs Are Abundant in Rotifers

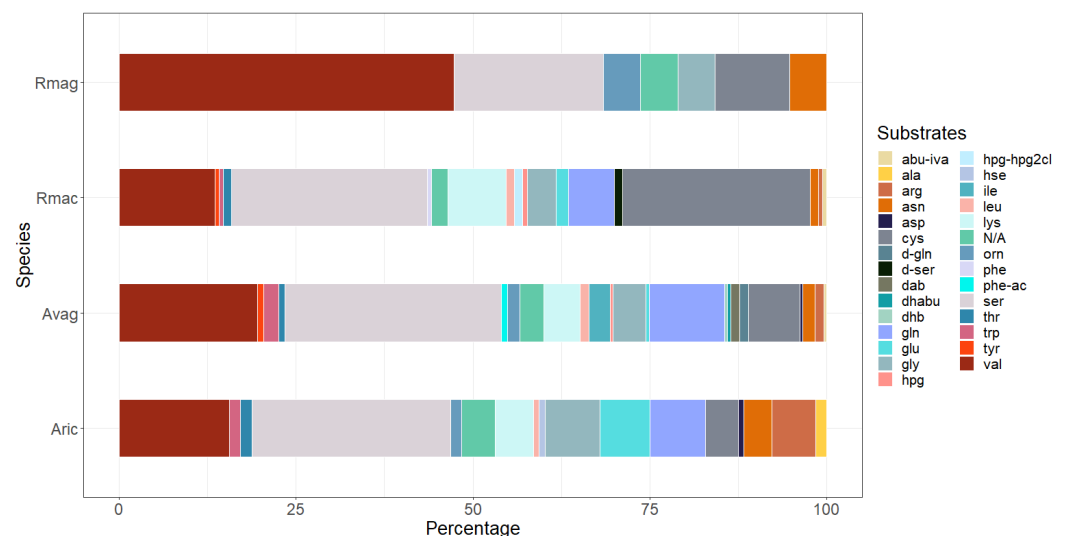
In rotifers, we identified up to 79 NRPS clusters per species (Table 1). One of these clustered with non-NRPS adenylating enzymes in the phylogenetic analysis (Figure 3). The putative NRPSs of the four rotifer species did not cluster in separate groups, but instead formed several clades together. We found both pure NRPSs and NRPS-PKS hybrids in rotifers. Moreover, large differences were observed between species in both the number of NRPS clusters and their predicted substrates (Table 1 and Figure 4). No NRPSs were identified in *Brachionus calyciflorus*, a rotifer that is classified outside the subclass Bdelloidea, to which the other four rotifers belong.

**Table 1.** Nonribosomal peptide synthetases (NRPSs) and hybrid nonribosomal peptide synthetase-polyketide synthases (NRPS-PKSs) in rotifers. The bdelloid rotifers from the genera *Adineta* and *Rotaria* contained many NRPSs, while they appeared to be completely absent in the genome of the non-bdelloid *Brachionus calyciflorus*.

Species	NRPS	NRPS-PKS
<i>Adineta ricciae</i>	46	2
<i>Adineta vaga</i>	69	5
<i>Rotaria macrura</i>	79	0
<i>Rotaria magnacalcarata</i>	12	1
<i>Brachionus calyciflorus</i>	0	0



**Figure 3.** Phylogeny of the identified putative rotifer NRPSs and other rotifer adenylate-forming enzymes (outgroup). One of the putative NRPSs clustered with the outgroup proteins. The others formed separate clades. Bootstrap support  $\geq 70\%$  is shown.



**Figure 4.** Substrate predictions for putative nonribosomal peptide synthetases (NRPSs) in rotifers. Rotifer NRPSs are predicted to use a wide variety of amino acid substrates including nonproteinogenic amino acids. Abbreviations: Rmag: *Rotaria magnacalcarata*; Rmac: *Rotaria macrura*; Avag: *Adineta vaga*; Aric: *Adineta ricciae*.

It should be noted that some rotifer NRPS clusters were rather short and might have in fact together formed a single NRPS. This would reduce the total number of NRPSs reported in Table 1. The smaller NRPS clusters are often located on small scaffolds that do not contain other genes. However, for *Adineta vaga*, a new chromosome-level assembly has been published [26], which contains most of the NRPS clusters that were identified here in the more fragmentary assemblies (Supplementary Table S4).

### 3.3. Beta-Lactam Biosynthesis Genes in *P. sambesii*

Several of the adenylation domains of the nematode *P. sambesii* clustered with spring-tail L- $\delta$ -( $\alpha$ -aminoadipoyl)-L-cysteinyl-D-valine synthetase (ACVS) (Figure 2). ACVS catalyzes the first step of beta-lactam antibiotic biosynthesis in microorganisms. Hence, we analyzed whether the genome of *P. sambesii* also contained other enzymes from the beta-lactam biosynthesis pathways. In addition to ACVS, we identified an isopenicillin N synthase (IPNS), which catalyzes the second step of the pathway that forms the beta-lactam ring and a deacetoxycephalosporin C synthase (*cefEF*, DAOC synthase).

The *P. sambesii* IPNS and deacetoxycephalosporin C synthase are found on the same scaffold (GenBank ID NIUQ01002120.1). The ACVS gene was scattered over multiple short scaffolds, and it could not be determined whether the assembly contained these scaffolds due to contamination (Supplementary Table S4). However, phylogenetic analyses were used to investigate whether the *P. sambesii* ACVS, IPNS, and DAOC synthase proteins would cluster within a clade of microorganisms or within the animal clade. All three proteins formed monophyletic clades with the previously identified ACVS, IPNS, and DAOC synthase proteins of springtails, indicating that this is an animal (nematode) ACVS (Figure 5).

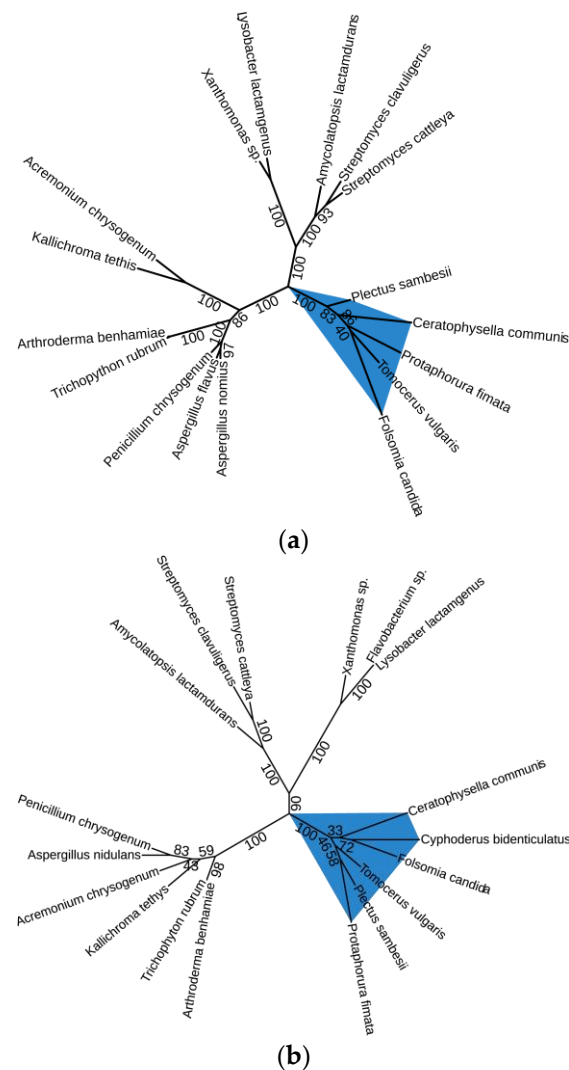
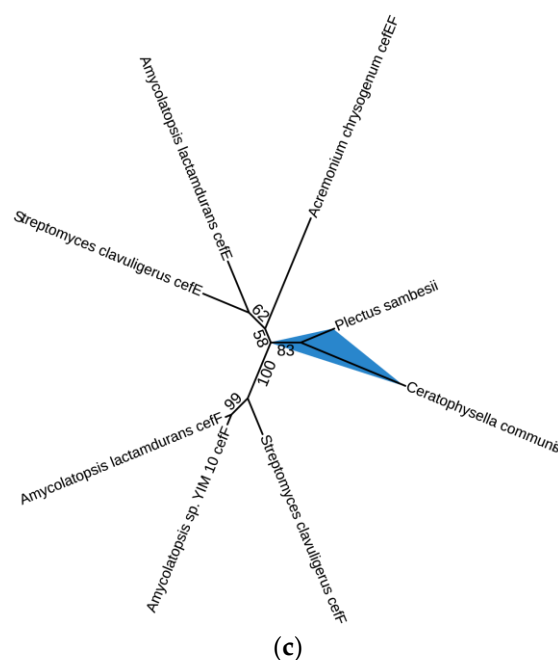


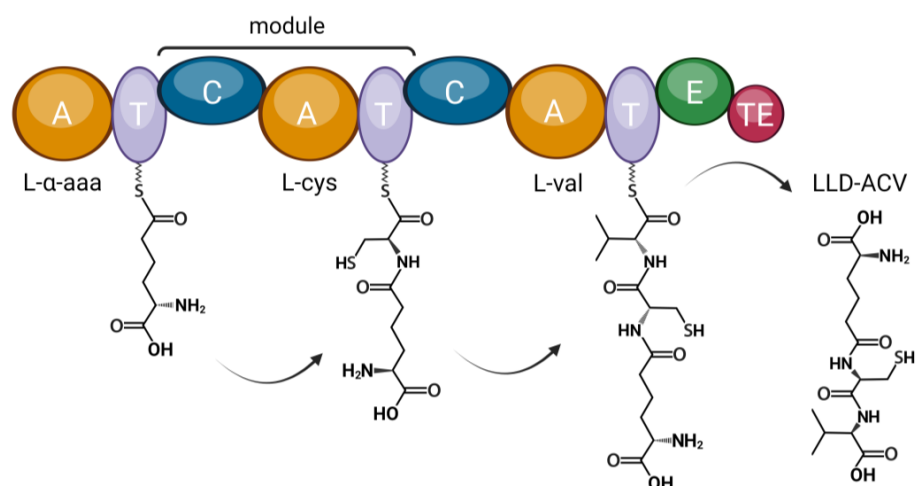
Figure 5. Cont.





**Figure 5.** Phylogenetic analyses show that *Plectus sambesii* beta-lactam biosynthesis proteins form monophyletic groups with other animals (springtails). (a) L- $\delta$ -( $\alpha$ -aminoadipoyl)-L-cysteinyl-D-valine synthetase (ACVS), (b) isopenicillin N synthase (IPNS), and (c) deacetoxycephalosporin C synthase (cefEF).

*P. sambesii* ACVS shares approximately 50% amino acid identity with *F. candida* ACVS, while *P. sambesii* IPNS has an identity percentage of approximately 65% to *F. candida* IPNS. The predicted substrates for the three adenylation domains of the *P. sambesii* protein are aminoadipic acid, cysteine, and valine, which are the expected substrates for all known ACVS proteins (Figure 6). *P. sambesii* was the only plectid nematode that was analyzed in this study.



**Figure 6.** The domain architecture of ACVS. The structure and substrate predictions of a *Plectus sambesii* putative NRPS resemble that of bacterial, fungal and springtail ACVS. The predicted substrates based on the adenylation domains are: aminoadipic acid (aad), cysteine (cys), and valine (val). The final product is a tripeptide (LLD-ACV). The amino acid substrates are activated by the adenylation (A) domains and attached to phosphopantetheine groups that are bound to the thiolation (T) domains. Amide bonds between neighbouring substrates are formed by the condensation (C) domains. The epimerization (E) domain epimerizes L-valine while the thioesterase (TE) domain catalyzes the release of the peptide. Created with [BioRender.com](https://www.biorender.com).

#### 4. Discussion

This study is the first kingdom-wide analysis of NRPSs in animals. We found that NRPSs are widespread in animals and in some cases very abundant. Over 8% of species and 30% of animal phyla that we analyzed contained NRPSs. For some phyla, we analyzed only a small number of genomes. Yet, in some of these phyla we still found putative NRPSs, indicating they may be abundant in these clades. Their presence appeared scattered over the animal tree as we found them in Chordata, Mollusca, Echinodermata, Phoronida, Arthropoda, Nematoda, and Rotifera.

The function and evolutionary history of NRPSs in animals is still not clear. Previously, it has been suggested that some of these genes may be the result of horizontal gene transfer events from bacteria to animals [1,6]. An analysis of an NRPS-PKS of *C. elegans* showed that based on its polyketide synthase domains (AT, KR and KS), it was most likely derived from an animal fatty acid synthase [5]. However, a phylogenetic analysis of its TE domain showed that it formed a group with bacterial NRPS-PKS TE domains [2].

If NRPSs were acquired by animals through horizontal gene transfer, then we might expect to find them in only few lineages, or at least restricted to certain clades, unless the transfer event happened deep in the Metazoa. What we observe is a scattered distribution over many animal phyla. This suggests that, if NRPSs were acquired through horizontal gene transfer, then this must have happened early in the evolution of the clade, maybe even at its origin. The scattered distribution would then be the result of gene loss.

Why rotifers, nematodes, and springtails are among the animal clades with the highest prevalence of NRPSs remains unclear. A property common to all three animal groups is the ability of some species to undergo anhydrobiosis, a phenomenon best investigated in tardigrades. The extreme dehydration during anhydrobiosis is known to be accompanied by DNA degradation [27–32]. Reassembly of the genome during rehydration could predispose anhydrobiotic animals to HGT. However, in the two species of tardigrade, we found no NRPS genes. Several NRPSs were found in phoronids (horseshoe worms, a small phylum of lophophorate marine invertebrates related to bryozoans), but what aspect of their ecology would make them prone to recruit NRPSs is another riddle.

NRPSs are particularly abundant in rotifers. Both their domain architecture and substrate specificity are highly variable. Rotifers exert the highest rate of horizontal gene transfer found to date in animals (approximately 7.5%) [30], and many horizontally transferred genes have been acquired recently [33]. Interestingly, we found no NRPSs in the only non-bdelloid species in our analysis. Hence, if NRPSs in rotifers have been acquired through horizontal gene transfer, then they may have been (1) transferred after the divergence of bdelloid rotifers, (2) transferred earlier in rotifer evolution but subsequently lost in some species, or (3) only one or a few NRPSs are ancestral, and they were lost in some species but expanded in bdelloids. As more non-bdelloid genomes are sequenced, it will be interesting to see if they contain NRPS genes.

The identified metazoan NRPSs did not cluster with adenylate-forming enzymes in the outgroup except for Ebony and AASDH. In fungi, something similar was observed as amino adipate reductases grouped with NRPSs, suggesting they are more similar to NRPSs than to other adenylate-forming enzymes [34]. We found that *Branchiostoma* putative NRPSs grouped with Ebony, indicating that these proteins may have evolved from a common ancestral gene. However, in the identified *Branchiostoma* sequences, we found condensation domains that were not present in Ebony.

Shou et al. (2016) found that nematodes utilize a multimodule NRPS and a multimodule NRPS-PKS to produce nemamide, a compound that promotes larval survival [2]. Homologs of these genes are present in more than twenty species of nematodes. In addition to the species previously identified to contain an NRPS, we identified 38 other species with NRPS genes. In most nematodes, we found an NRPS and an NRPS-PKS hybrid. Nematode NRPSs form a monophyletic group, suggesting a possible single origin and subsequent expansion into two types: a hybrid and a pure NRPS. Due to a lack of genomes from several clades of nematodes in our dataset, it was not possible to completely assess

which nematode clades contained NRPSs. However, these genes appeared to be absent from Dorylaimia, suggesting that the NRPSs that produce nemamide could be exclusive to nematode clades that diverged more recently.

Beta-lactam antibiotics are produced by a subset of fungal and bacterial clades. The most famous beta-lactam is penicillin, which is produced by the fungus *Penicillium chrysogenum*. Recently, we identified the first beta-lactam biosynthesis pathway in an animal [3]. The springtail *F. candida* contains a unique combination of four beta-lactam biosynthesis genes, while other springtails contain different sets of beta-lactam genes [6]. In our previous analysis, we did not find beta-lactam genes in non-springtail animals, including its closest relatives Diplura and Protura. Here, we showed that beta-lactam biosynthesis genes are also present in the plectid nematode *P. sambesii*. These genes appeared to be limited to a small clade of nematodes since we did not find them in Rhabditina or Dorylaimia. Since only a single plectid genome was available, it is currently unclear how prevalent they are. It is interesting to observe that the combination of beta-lactam genes in *P. sambesii* was identical to the combination we found in several springtail species [6]. It was distinct from the beta-lactam gene clusters of fungi and bacteria. The first two genes that catalyze the formation of the core penem structure appeared to be included. This core structure involves a heterobicyclic ring, which is central to the antimicrobial activity of the beta-lactam compound [35].

It has been suggested that the beta-lactam biosynthesis pathway of bacteria may have been horizontally transferred from *Streptomyces* to fungi [36]. Subsequently, fungi may have recruited native enzymes to produce new beta-lactam compounds. For example, in the case of penicillin production—which is restricted to fungi—the acyltransferase that catalyzes the final step in the biosynthesis of penicillin appears to be derived from a native fungal gene [37]. Hence, it is possible that springtails and nematodes have also added native enzymes to their beta-lactam biosynthesis pathways.

Recent reports have shown that the diversity of secondary metabolites in animals may be much higher than thought earlier [38]. While previously only a limited number of NRPS genes were identified in animals, here we demonstrated that they are present in over 30% of the animal phyla that were analyzed. NRPSs are especially abundant in rotifers, where we found up to 79 clusters per species. Metazoan NRPSs showed a variety of domain architectures including NRPS-PKS hybrids. The data provided may open up a largely unexplored field of research: the functional significance of NRPS genes in animals. Such NRPSs could potentially be used for designing biotechnological products such as novel beta-lactam compounds.

## 5. Conclusions

Many more animal genomes than previously assumed contain nonribosomal peptide synthetase genes. The capacity of animals to synthesize secondary compounds playing a role in signaling, antibiotic synthesis, and other functions might be much greater than thought before. This could also lead to discovery of new NRPSs with biotechnological applications. They are scattered across the metazoan tree and some clades contain many more NRPSs than others. These genes are often assumed to have entered animal genomes by horizontal gene transfer from bacteria, but since they are not restricted to specific lineages, it seems unlikely that all of them represent HGT events, although some do for sure. In addition, it is unclear which ecological traits predispose an animal to recruit these genes. A more detailed analysis, concentrated on specific lineages, could reveal the underlying mechanisms.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14091741/s1>.

**Author Contributions:** Conceptualization, W.S.; methodology, W.S.; software, W.S., D.H. and G.L.N.; formal analysis, W.S. and D.H.; investigation, W.S. and D.H.; writing—original draft preparation, W.S. and N.M.v.S.; writing—review and editing, N.M.v.S., A.B., D.R. and W.S.; visualization, W.S.; supervision, D.R. and N.M.v.S.; project administration, D.R.; funding acquisition, D.R., A.B. and N.M.v.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a grant from the Dutch Biotechnology Based Ecologically Balanced Sustainable Industrial Consortium (BE-Basic Foundation, <http://www.be-basic.org> (accessed on 27 August 2023) (grant number F07.003.05).

**Data Availability Statement:** The analyzed genomes are publicly available through <https://www.ncbi.nlm.nih.gov/> (accessed 2021–2023). Other datasets are available as Supplementary Files.

**Acknowledgments:** We would like to thank Friso Heeg and Ruben Doorn for their help with data analysis. We are grateful to SURF for using the Lisa cluster.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Wang, H.; Fewer, D.P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9259–9264. [CrossRef] [PubMed]
- Shou, Q.; Feng, L.; Long, Y.; Han, J.; Nunnery, J.K.; Powell, D.H.; A Butcher, R. A hybrid polyketide-nonribosomal peptide in nematodes that promotes larval survival. *Nat. Chem. Biol.* **2016**, *12*, 770–772. [CrossRef] [PubMed]
- Roelofs, D.; Timmermans, M.J.T.N.; Hensbergen, P.; Van Leeuwen, H.; Koopman, J.; Faddeeva, A.; Suring, W.; de Boer, T.E.; Mariën, J.; Boer, R.; et al. A functional isopenicillin N synthase in an animal genome. *Mol. Biol. Evol.* **2013**, *30*, 541–548. [CrossRef]
- Chen, R.; Wang, J.; Gradinaru, I.; Vu, H.S.; Geboers, S.; Naidoo, J.; Ready, J.M.; Williams, N.S.; DeBerardinis, R.J.; Ross, E.M.; et al. A male-derived nonribosomal peptide pheromone controls female schistosome development. *Cell* **2022**, *185*, 1506–1520. [CrossRef] [PubMed]
- O'Brien, R.V.; Davis, R.W.; Khosla, C.; Hillenmeyer, M.E. Computational identification and analysis of orphan assembly-line polyketide synthases. *J. Antibiot.* **2014**, *67*, 89–97. [CrossRef] [PubMed]
- Suring, W.; Meusemann, K.; Blanke, A.; Mariën, J.; Schol, T.; Agamennone, V.; Faddeeva-Vakhrusheva, A.; Berg, M.P.; Brouwer, A.; van Straalen, N.M.; et al. Evolutionary ecology of beta-lactam gene clusters in animals. *Mol. Ecol.* **2017**, *26*, 3217–3229. [CrossRef] [PubMed]
- Miller, B.R.; Gulick, A.M. Structural biology of nonribosomal peptide synthetases. *Methods Mol. Biol.* **2016**, *1401*, 3–29. [PubMed]
- Flissi, A.; Ricart, E.; Campart, C.; Chevalier, M.; Dufresne, Y.; Michalik, J.; Jacques, P.; Flahaut, C.; Lisacek, F.; Leclère, V.; et al. Norine: Update of the nonribosomal peptide resource. *Nucleic Acids Res.* **2020**, *48*, D465–D469. [PubMed]
- Lambalot, R.H.; Gehring, A.M.; Flugel, R.S.; Zuber, P.; LaCelle, M.; Marahiel, M.A.; Reid, R.; Khosla, C.; Walsh, C.T. A new enzyme superfamily—The phosphopantetheinyl transferases. *Chem. Biol.* **1996**, *3*, 923–936. [CrossRef]
- Mootz, H.D.; Finking, R.; Marahiel, M.A. 4'-Phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. *J. Biol. Chem.* **2001**, *276*, 37289–37298. [CrossRef]
- Suring, W.; Mariën, J.; Broekman, R.; Van Straalen, N.M.; Roelofs, D. Biochemical pathways supporting beta-lactam biosynthesis in the springtail *Folsomia candida*. *Biol. Open* **2016**, *5*, 1784–1789. [CrossRef]
- Richardt, A.; Kemme, T.; Wagner, S.; Schwarzer, D.; Marahiel, M.A.; Hovemann, B.T. Ebony, a novel nonribosomal peptide synthetase for  $\beta$ -alanine conjugation with biogenic amines in *Drosophila*. *J. Biol. Chem.* **2003**, *278*, 41160–41166. [CrossRef] [PubMed]
- Drozak, J.; Veiga-da-Cunha, M.; Kadziolka, B.; Van Schaftingen, E. Vertebrate Acyl CoA synthetase family member 4 (ACSF4-U26) is a  $\beta$ -alanine-activating enzyme homologous to bacterial non-ribosomal peptide synthetase. *FEBS J.* **2014**, *281*, 1585–1597. [CrossRef]
- Izoré, T.; Tailhades, J.; Hansen, M.H.; Kaczmarek, J.A.; Jackson, C.J.; Cryle, M.J. *Drosophila melanogaster* nonribosomal peptide synthetase Ebony encodes an atypical condensation domain. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 2913–2918. [CrossRef]
- Wittkopp, P.J.; True, J.R.; Carroll, S.B. Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* **2002**, *129*, 1849–1858. [CrossRef]
- Blin, K.; Wolf, T.; Chevrete, M.G.; Lu, X.; Schwalen, C.J.; Kautsar, S.A.; Suarez Duran, H.G.; De Los Santos, E.L.; Kim, H.U.; Nave, M.; et al. AntiSMASH 4.0—Improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017**, *45*, W36–W41. [CrossRef] [PubMed]
- Hinchliff, C.E.; Smith, S.A.; Allman, J.F.; Burleigh, J.G.; Chaudhary, R.; Coghill, L.M.; Crandall, K.A.; Deng, J.; Drew, B.T.; Gazis, R.; et al. Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12764–12769. [CrossRef] [PubMed]

18. Michonneau, F.; Brown, J.W.; Winter, D.J. rotl: An R package to interact with the Open Tree of Life data. *Methods Ecol. Evol.* **2016**, *7*, 1476–1481. [\[CrossRef\]](#)
19. The UniProt Consortium. UniProt: The Universal Protein knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–D531. [\[CrossRef\]](#)
20. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [\[CrossRef\]](#)
25. Ahmed, M.; Roberts, N.G.; Adediran, F.; Smythe, A.B.; Kocot, K.M.; Holovachov, O. Phylogenomic Analysis of the Phylum Nematoda: Conflicts and Congruences with Morphology, 18S rRNA, and Mitogenomes. *Front. Ecol. Evol.* **2022**, *9*, 769565. [\[CrossRef\]](#)
26. Simion, P.; Narayan, J.; Houtain, A.; Derzelle, A.; Baudry, L.; Nicolas, E.; Arora, R.; Cariou, M.; Cruaud, C.; Gaudray, F.R.; et al. Chromosome-level genome assembly reveals homologous chromosomes and recombination in asexual rotifer *Adineta vaga*. *Sci. Adv.* **2021**, *7*, eabg4216. [\[CrossRef\]](#)
27. Rebecchi, L.; Cesari, M.; Altiero, T.; Frigieri, A.; Guidetti, R. Survival and DNA degradation in anhydrobiotic tardigrades. *J. Exp. Biol.* **2009**, *212*, 4033–4039. [\[CrossRef\]](#)
28. Barra, J.A.; Poinot-Balaguer, N. Modifications ultrastructurales accompagnant l'anhydrobiose chez un Collembole: *Folsomides variabilis*. *Rev. Ecol. Biol. Sol* **1977**, *14*, 189–197.
29. Gladyshev, E.; Meselson, M. Extreme resistance of bdelloid rotifers to ionizing radiation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5139–5144. [\[CrossRef\]](#)
30. Crisp, A.; Boschetti, C.; Perry, M.; Tunnacliffe, A.; Micklem, G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **2015**, *16*, 50. [\[CrossRef\]](#)
31. Gladyshev, E.A.; Arkhipova, I.R. Genome structure of bdelloid rotifers: Shaped by asexuality or desiccation? *J. Hered.* **2010**, *101* (Suppl. S1), S85–S93. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Hespeels, B.; Knapen, M.; Hanot-Mambres, D.; Heuskin, A.; Pineux, F.; Lucas, S.; Koszul, R.; van Doninck, K. Gateway to genetic exchange? DNA double-strand breaks in the bdelloid rotifer *Adineta vaga* submitted to desiccation. *J. Evol. Biol.* **2014**, *27*, 1334–1345. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Eyres, I.; Boschetti, C.; Crisp, A.; Smith, T.P.; Fontaneto, D.; Tunnacliffe, A.; Barraclough, T.G. Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species from desiccating habitats. *BMC Biol.* **2015**, *13*, 90. [\[CrossRef\]](#)
34. Bushley, K.E.; Turgeon, B.G. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol. Biol.* **2010**, *10*, 26. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Lang, M.; Schneider, P.; Tosch, W.; Scartazzini, R.; Zak, O. The penems, a new class of  $\beta$ -lactam antibiotics. 7. Synthesis and antimicrobial activity of 2-heterocyclylmercaptoalkyl derivatives. *J. Antibiot.* **1986**, *39*, 525–534. [\[CrossRef\]](#)
36. Gutiérrez, S.; Fierro, F.; Casqueiro, J.; Martín, J.F. Gene organization and plasticity of the  $\beta$ -lactam genes in different filamentous fungi. *Antonie Van Leeuwenhoek* **1999**, *75*, 81–94. [\[CrossRef\]](#)
37. García-Estrada, C.; Vaca, I.; Ullán, R.V.; Van Den Berg, M.A.; Bovenberg RAI Martín, J.F. Molecular characterization of a fungal gene paralogue of the penicillin penDE gene of *Penicillium chrysogenum*. *BMC Microbiol.* **2009**, *9*, 104. [\[CrossRef\]](#)
38. Torres, J.P.; Schmidt, E.W. The biosynthetic diversity of the animal world. *J. Biol. Chem.* **2019**, *294*, 17684–17692. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.