

## **Title**

Transcriptome sequencing of three *Pseudo-nitzschia* species reveals comparable gene sets and the presence of Nitric Oxide Synthase genes in diatoms

## **Author names and affiliations**

Valeria Di Dato, Francesco Musacchia, Giuseppe Petrosino, Shrikant Patil, Marina Montresor, Remo Sanges, Maria Immacolata Ferrante\*.

## **Affiliation and Address**

Stazione Zoologica Anton Dohrn, Villa Comunale 1, 80121, Naples, Italy

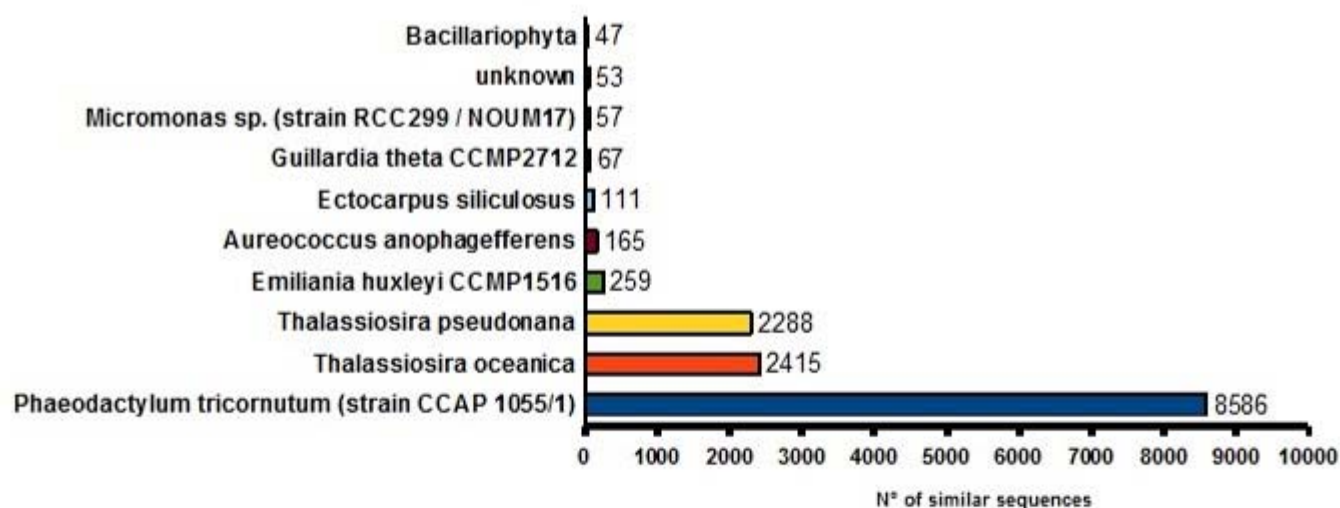
## **Supplementary Material**

### **This PDF file contains**

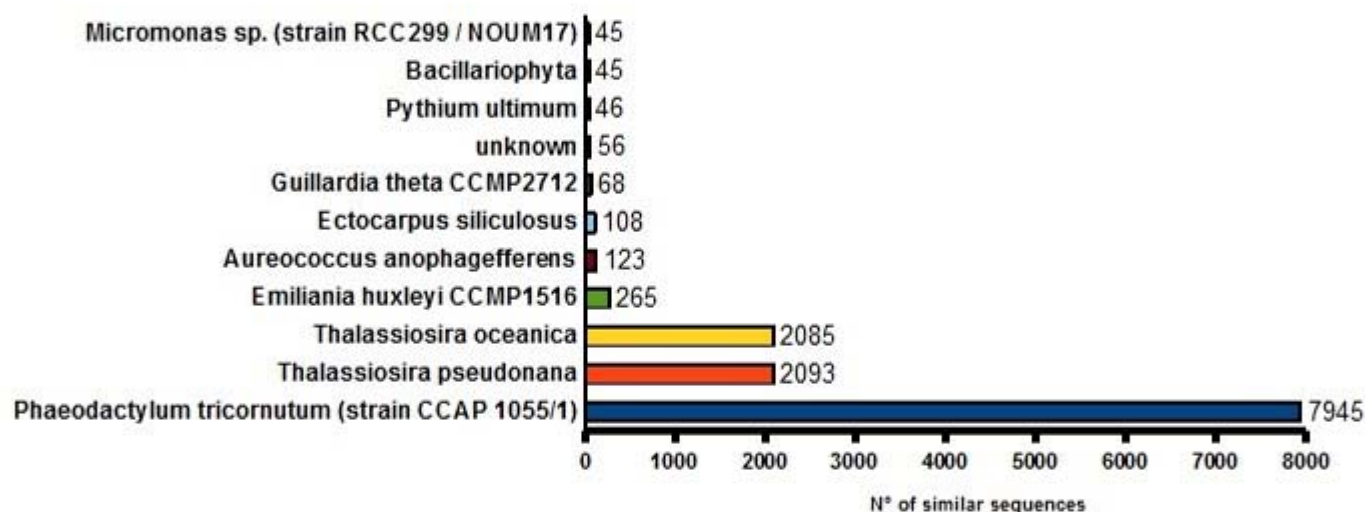
Supplementary Figure S1  
Supplementary Figure S2  
Supplementary Figure S3  
Supplementary Figure S4  
Supplementary Table S1  
Supplementary Table S8  
Supplementary Table S9  
Supplementary Note

## Supplementary Figure S1

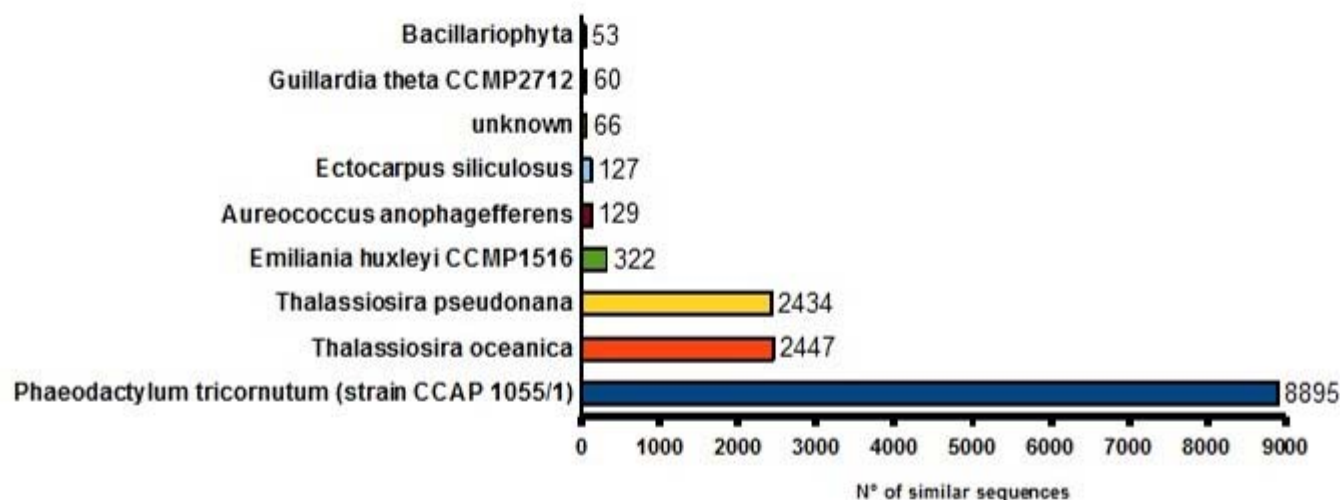
***Pseudo-nitzschia arenysensis* closer organisms abundance**



***Pseudo-nitzschia delicatissima* closer organisms abundance**



***Pseudo-nitzschia multistriata* closer organisms abundance**

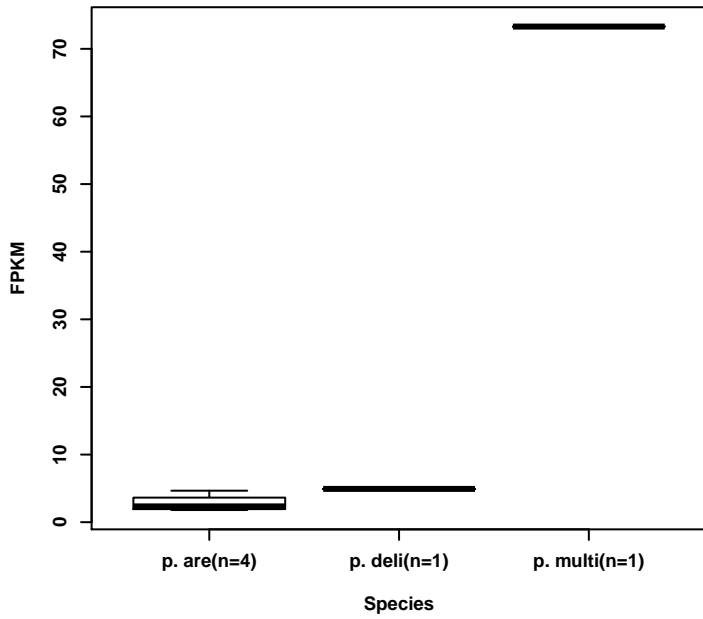
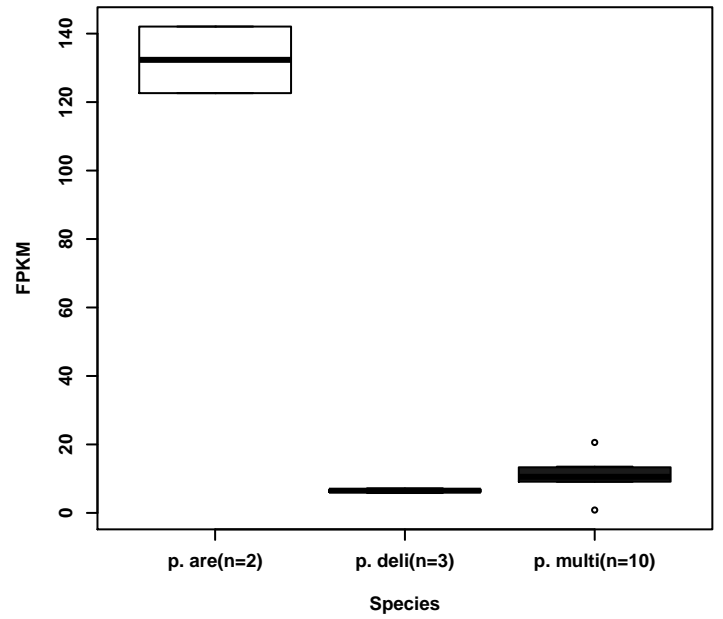
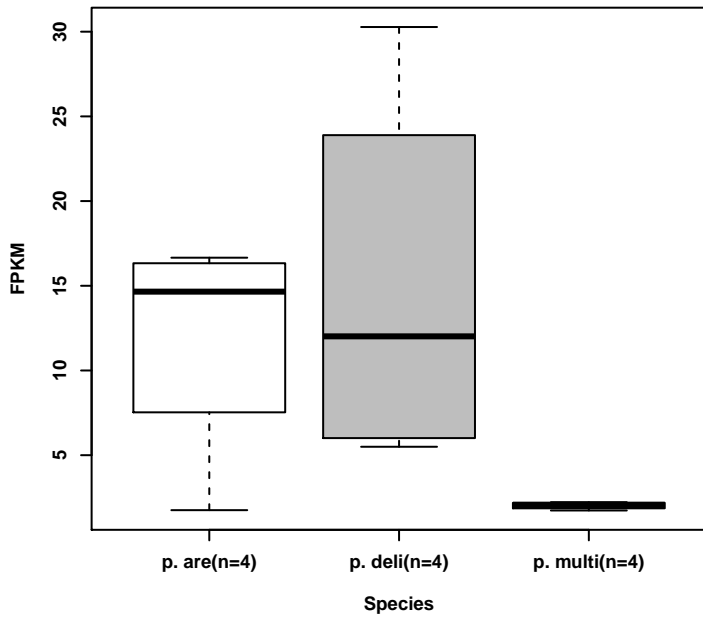
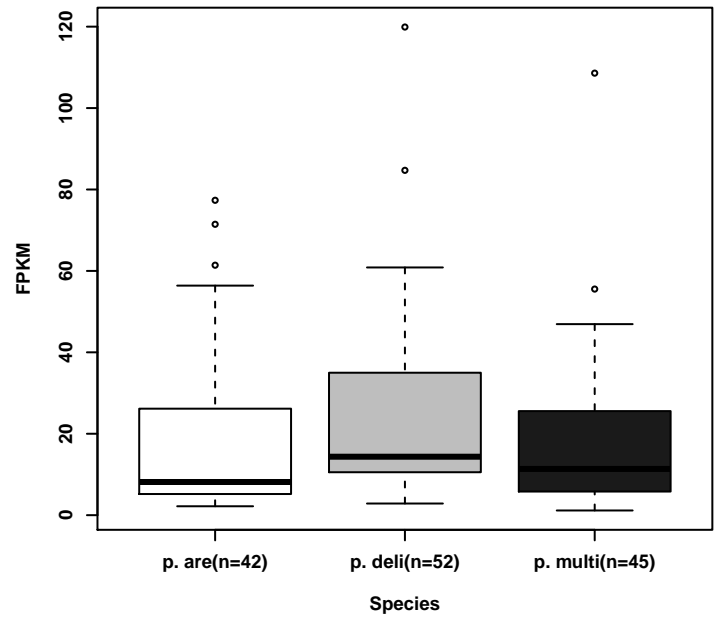
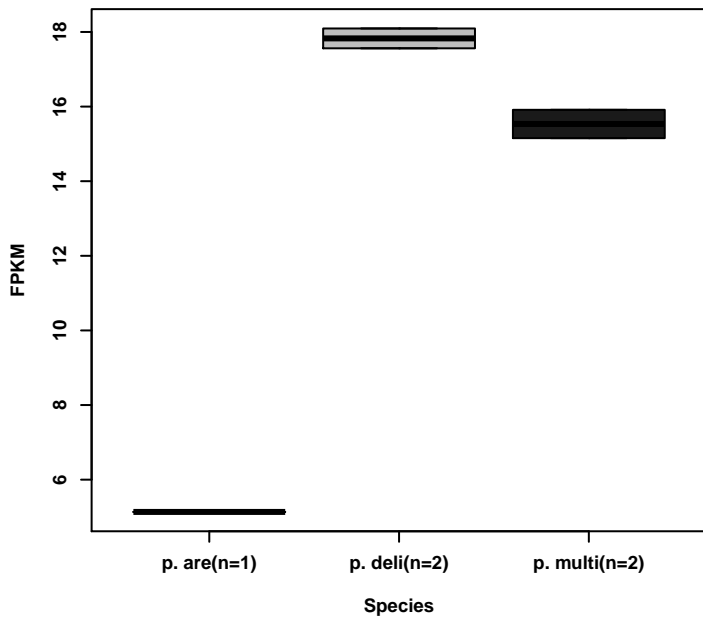
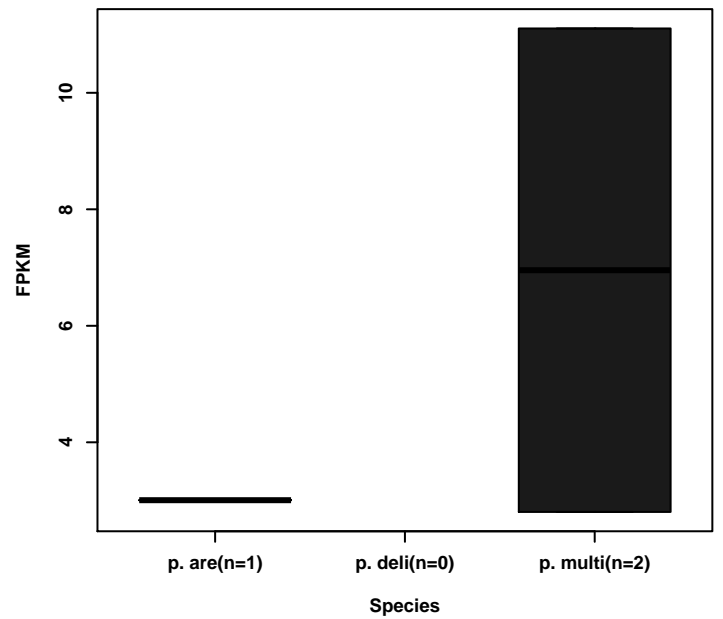


**Supplementary Figure 1. *Pseudo-nitzschia* closer organisms abundance.**

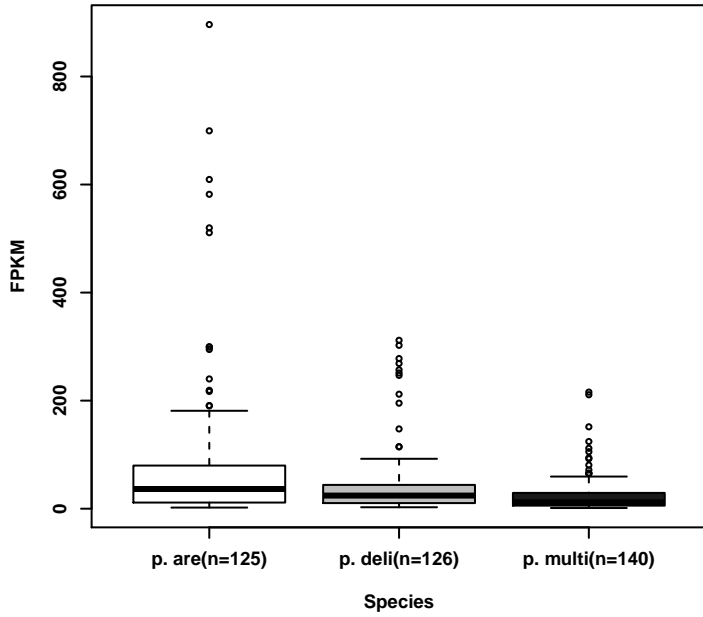
The graphs display the top ten organisms scoring the BlastP hits for each *Pseudo-nitzschia* proteome. Numbers on the side of each species-associated bar indicate the number of scoring sequences.

**Supplementary Figure S2. Pathways expression based on the first level of annotation.**

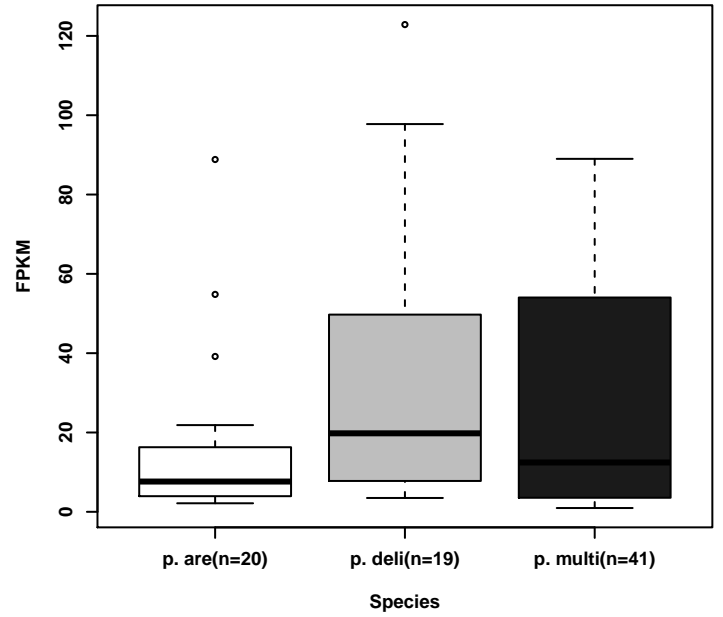
Each graph shows the expression of the general pathways for each species. The levels of expression are represented with box plots, each box represents the distribution of the expression values of the sequences associated to the relative pathway for each species. The black bar in each rectangle represents the median values of each sequence population. Points outside the main rectangle represent the value out-layers. Numbers in parenthesis indicate the number of sequences associated to the relative pathway for each species.

**Alkaloid biosynthesis****Alkaloid degradation****Alkene biosynthesis****Amine and polyamine biosynthesis****Amine and polyamine degradation****Amine and polyamine metabolism**

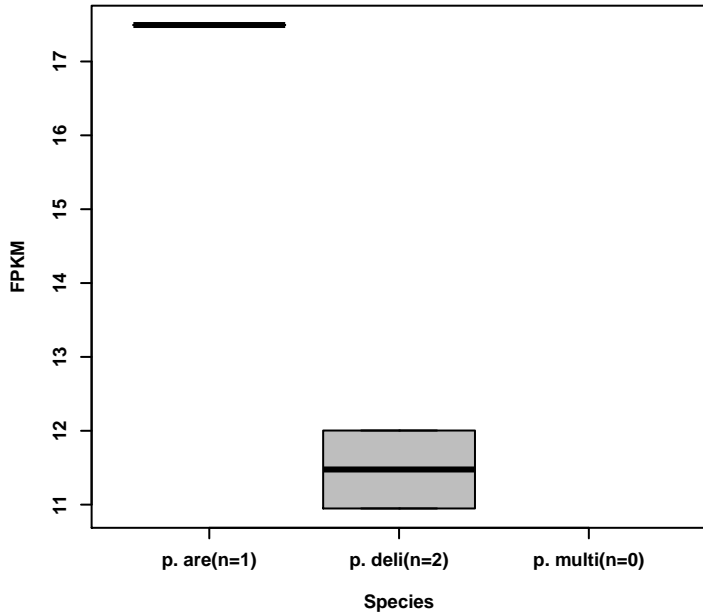
**Amino-acid biosynthesis**



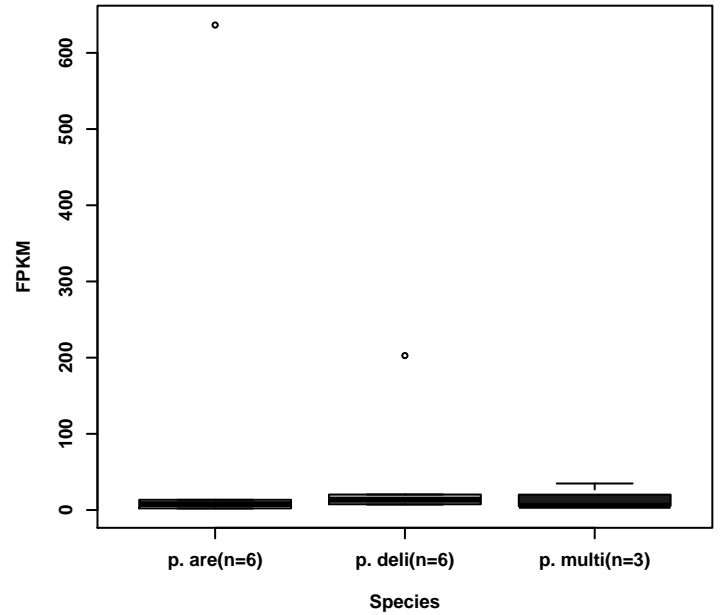
**Amino-acid degradation**



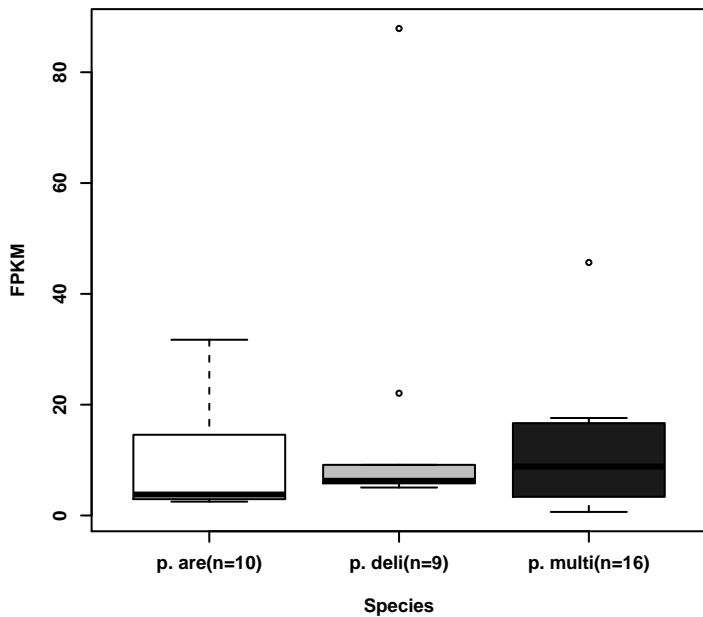
**Amino-acid metabolism**



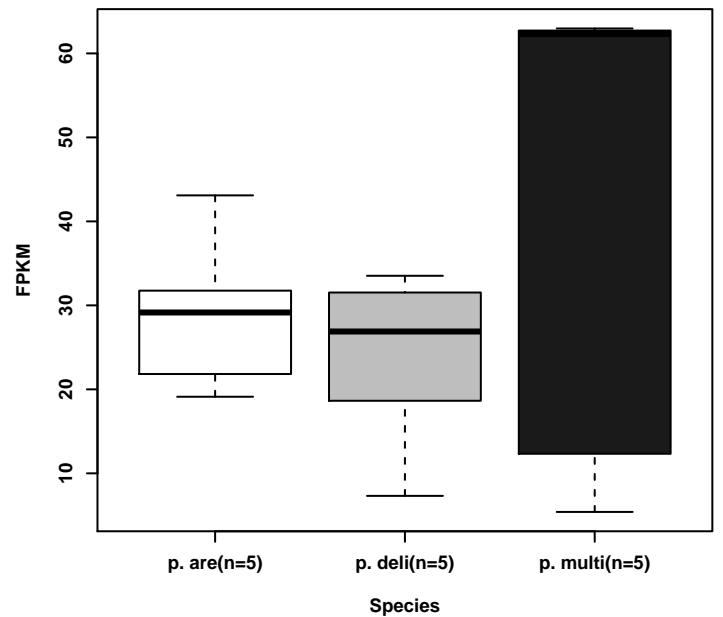
**Aminoacyl-tRNA biosynthesis**



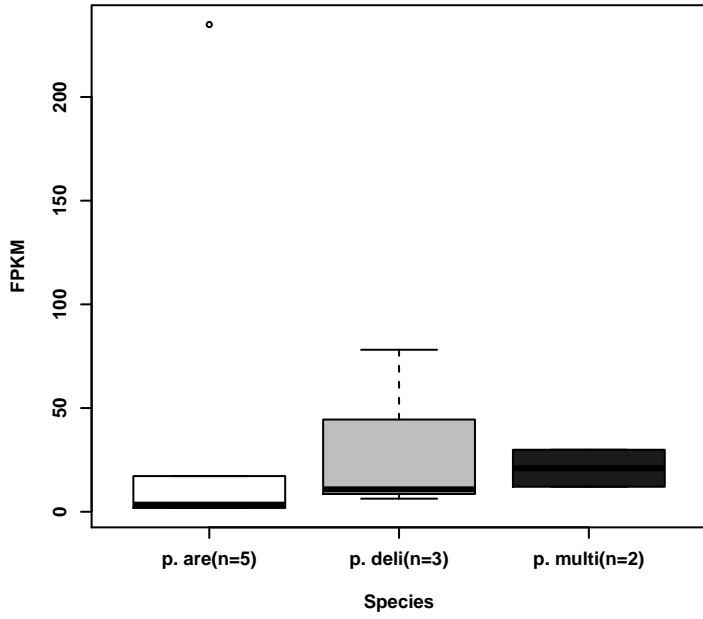
**Antibiotic biosynthesis**



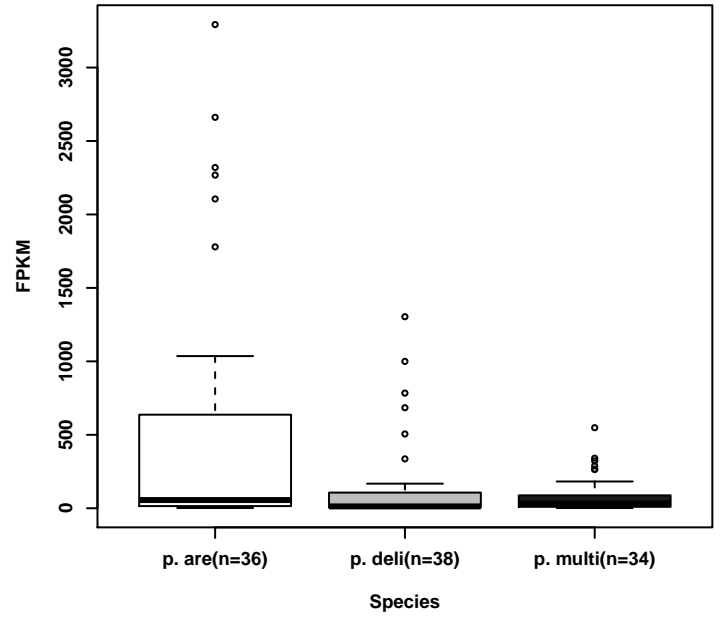
**Aromatic compound metabolism**



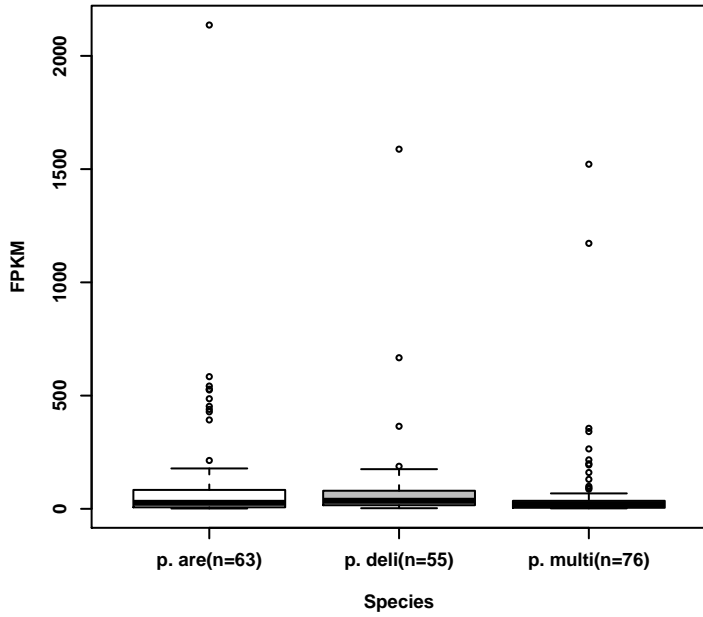
**Carbohydrate acid metabolism**



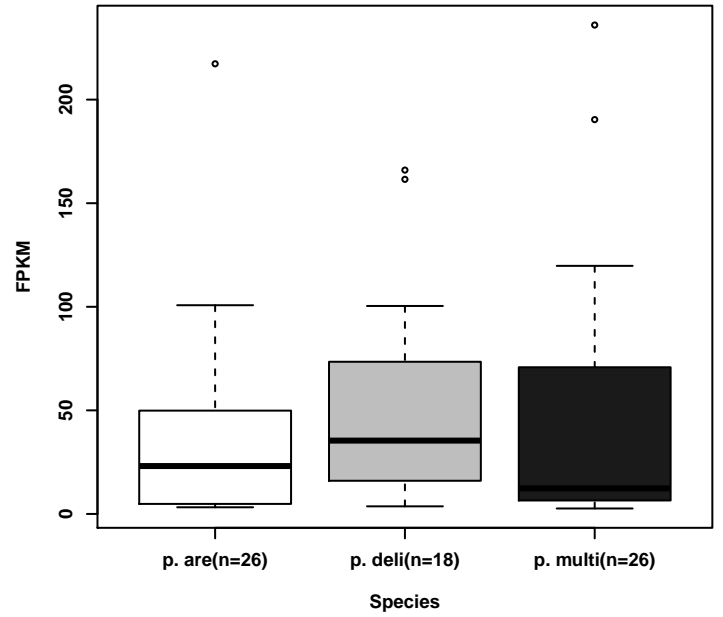
**Carbohydrate biosynthesis**



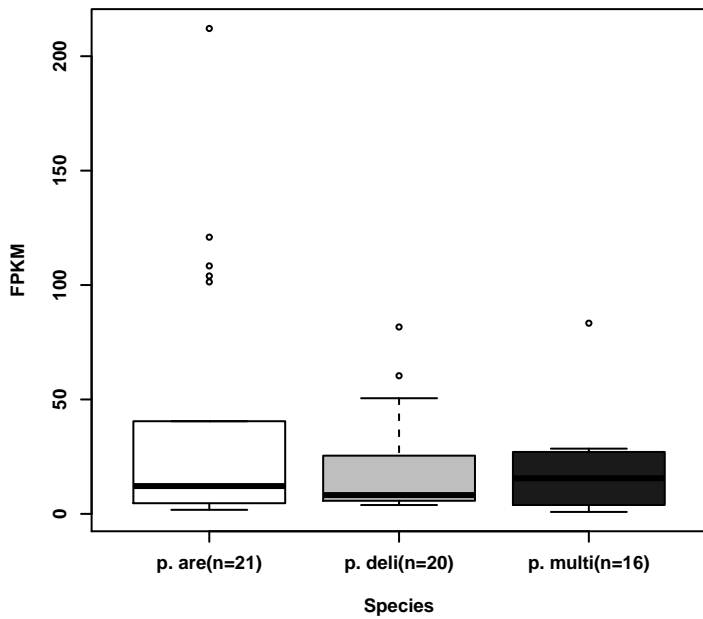
**Carbohydrate degradation**



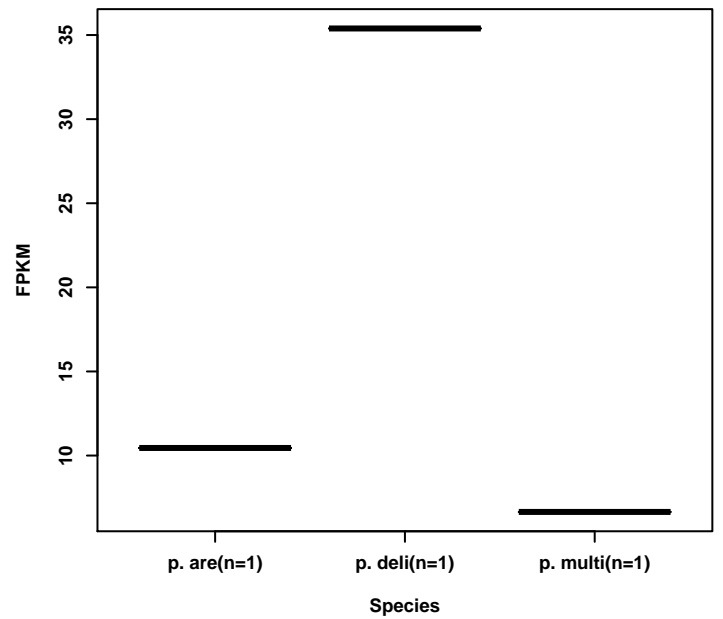
**Carbohydrate metabolism**

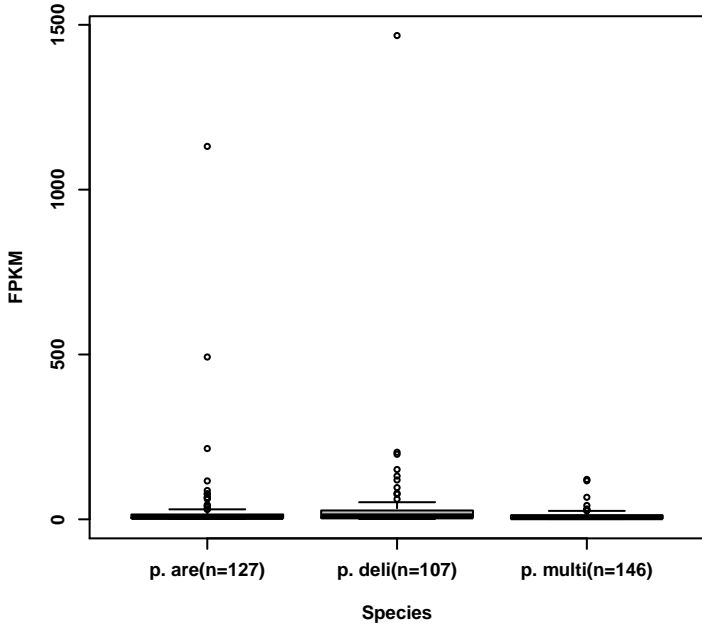
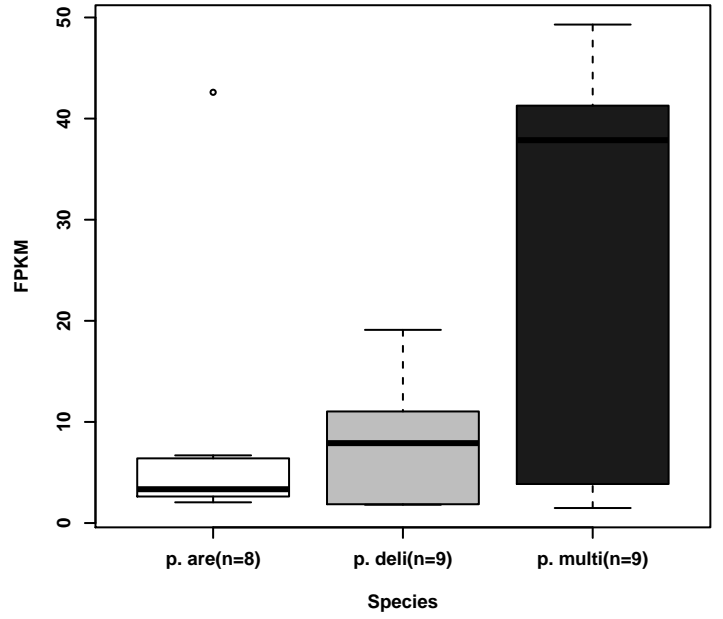
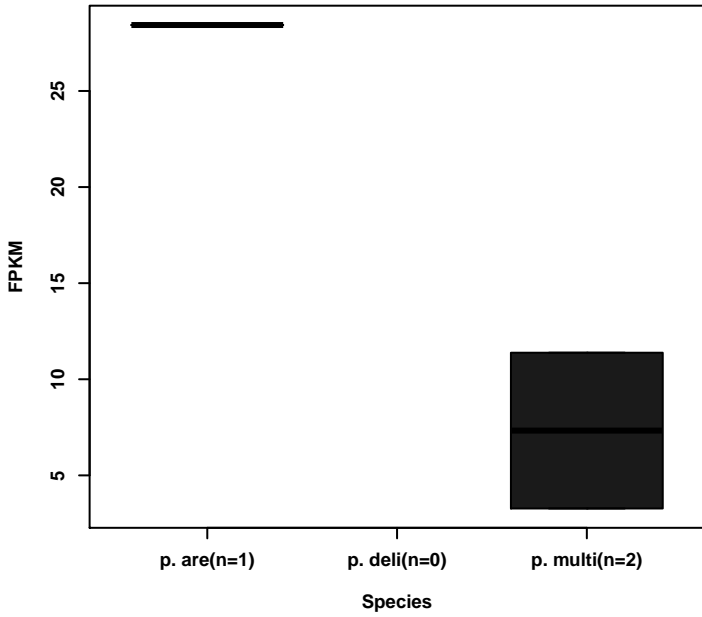
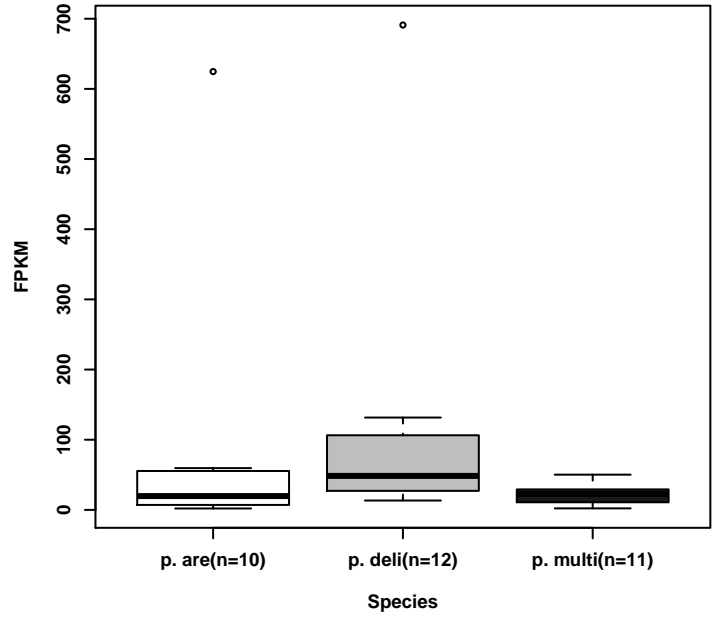
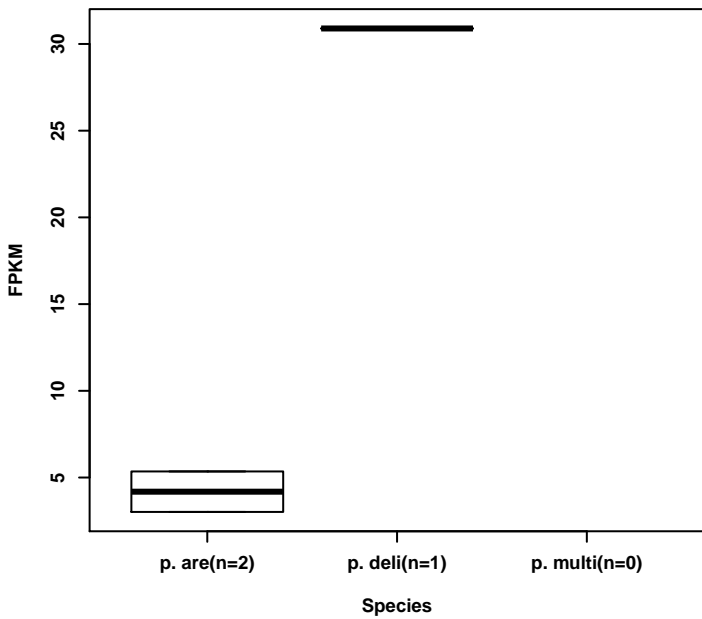
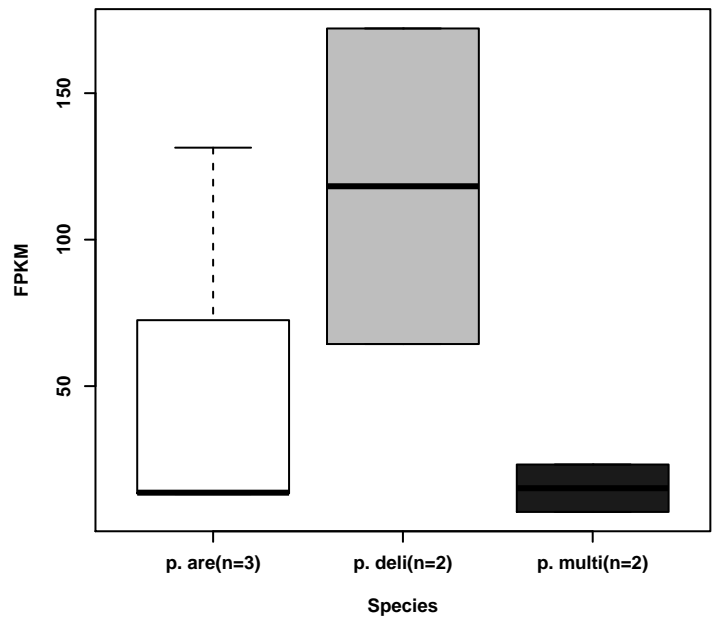


**Carotenoid biosynthesis**



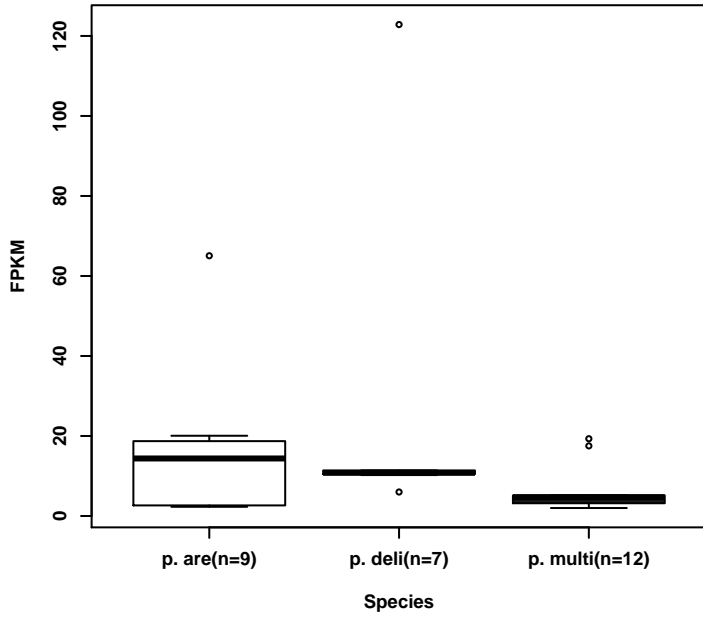
**Cell wall biogenesis**



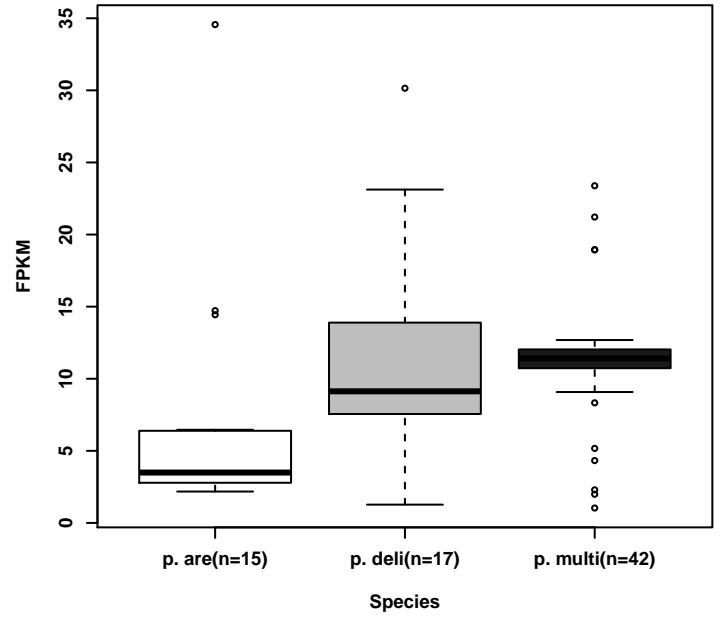
**Cofactor biosynthesis****Cofactor degradation****Cofactor metabolism****Energy metabolism****Fermentation****Flavonoid metabolism**



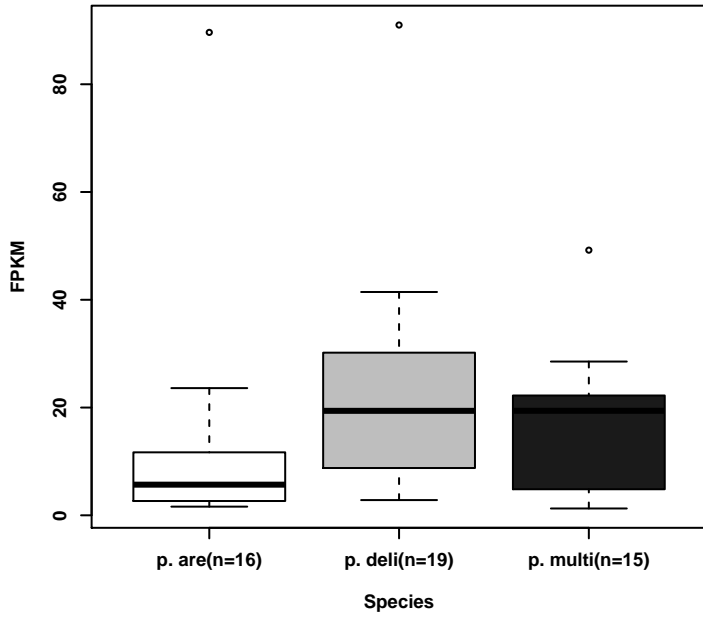
**Genetic information processing**



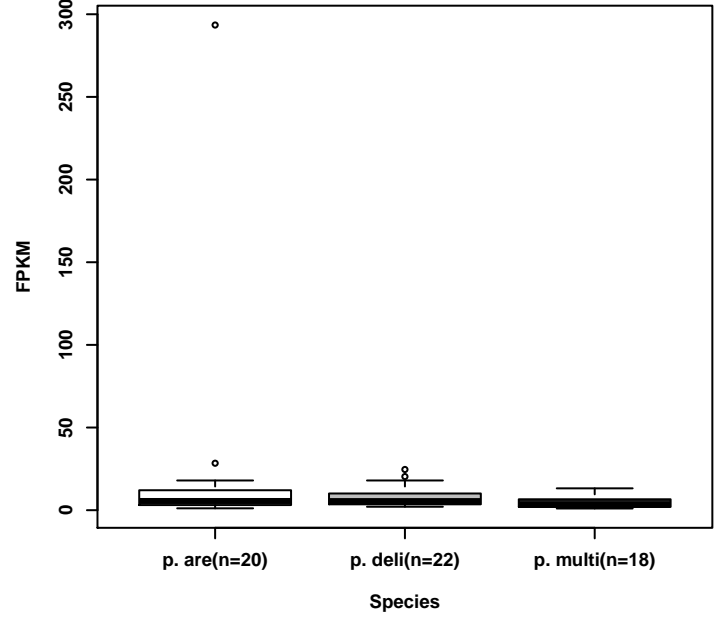
**Glycan metabolism**



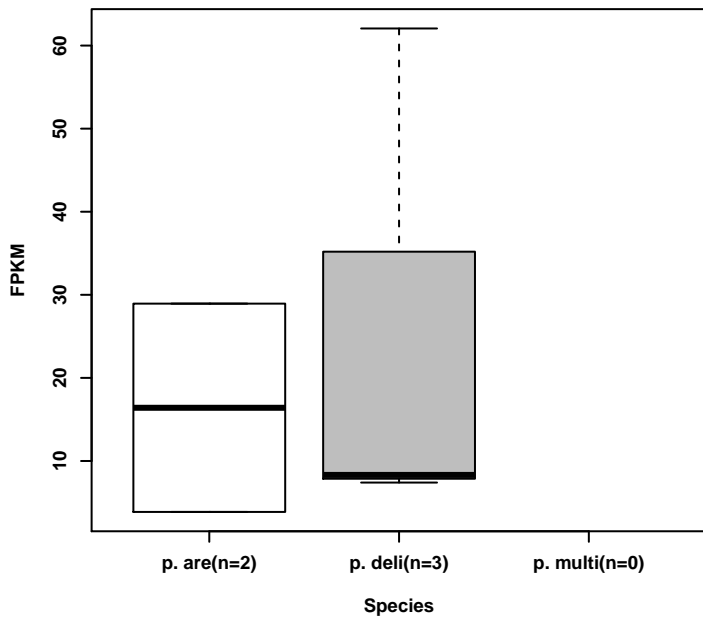
**Glycerolipid metabolism**



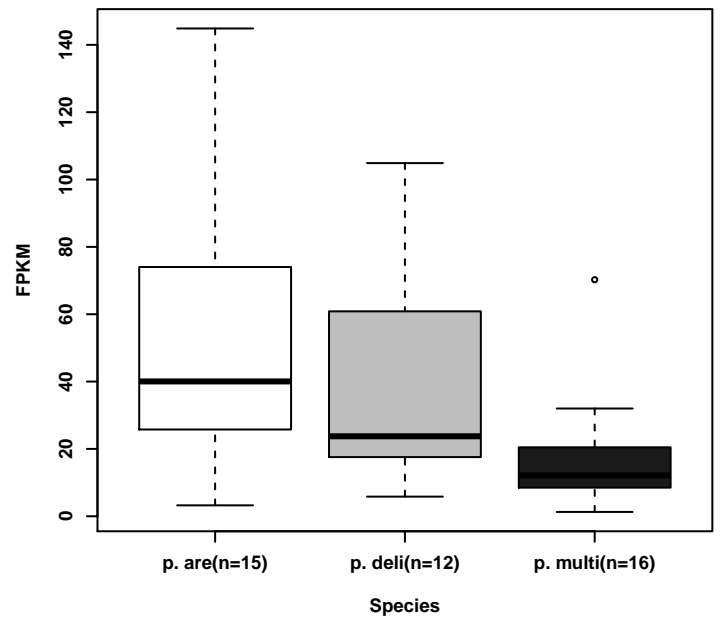
**Glycolipid biosynthesis**



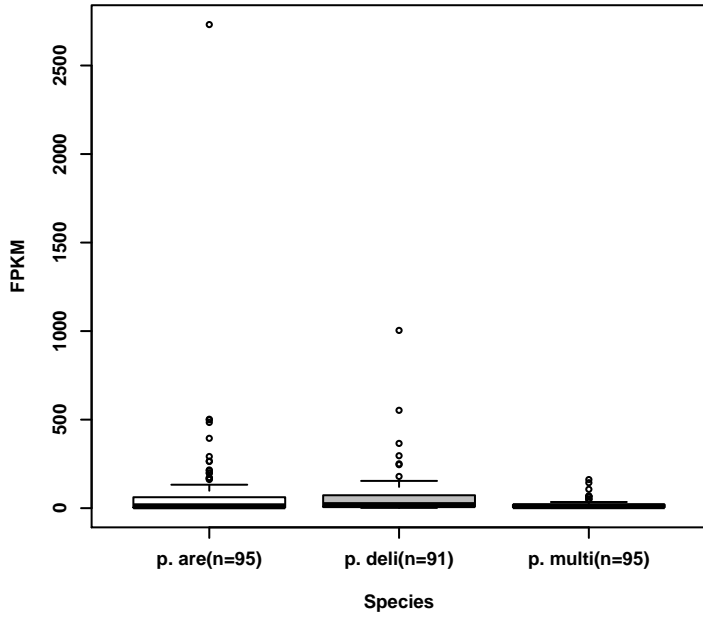
**Hormone biosynthesis**



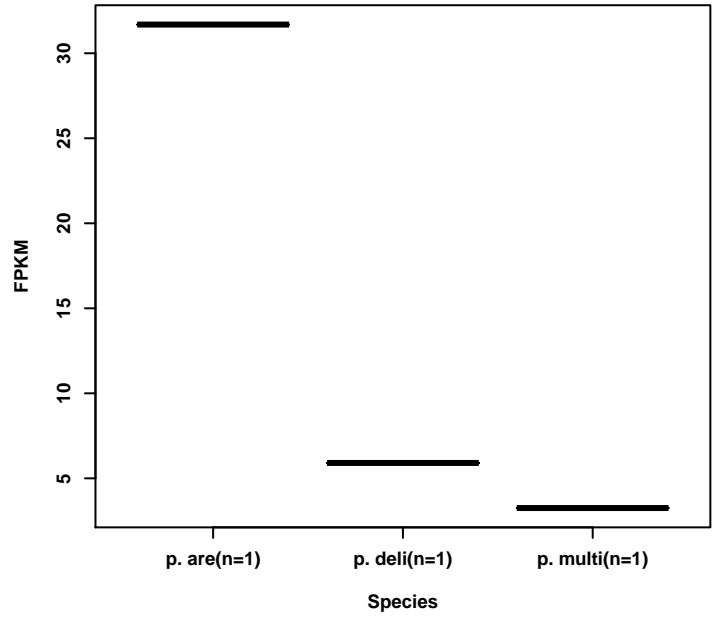
**Isoprenoid biosynthesis**



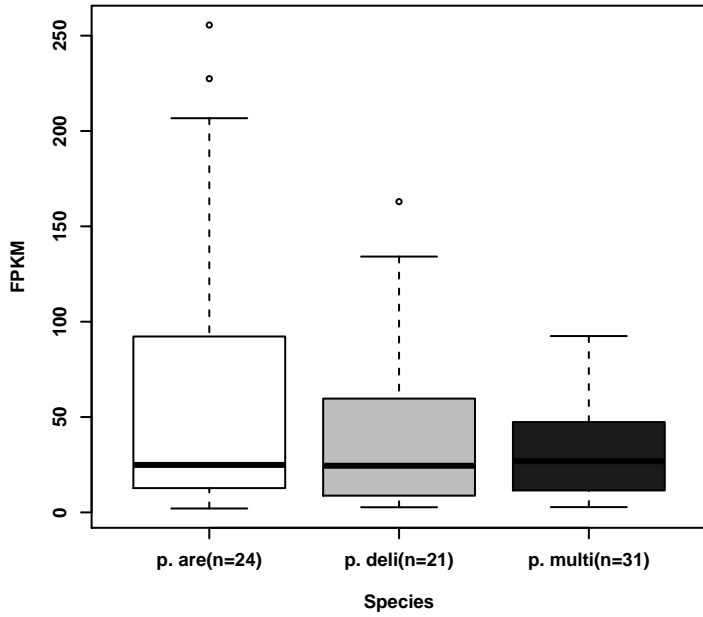
**Lipid metabolism**



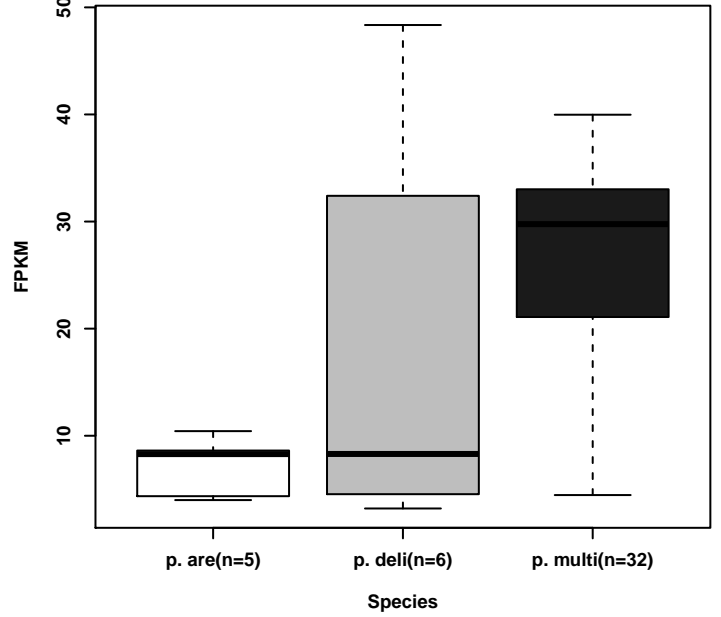
**Membrane lipid metabolism**



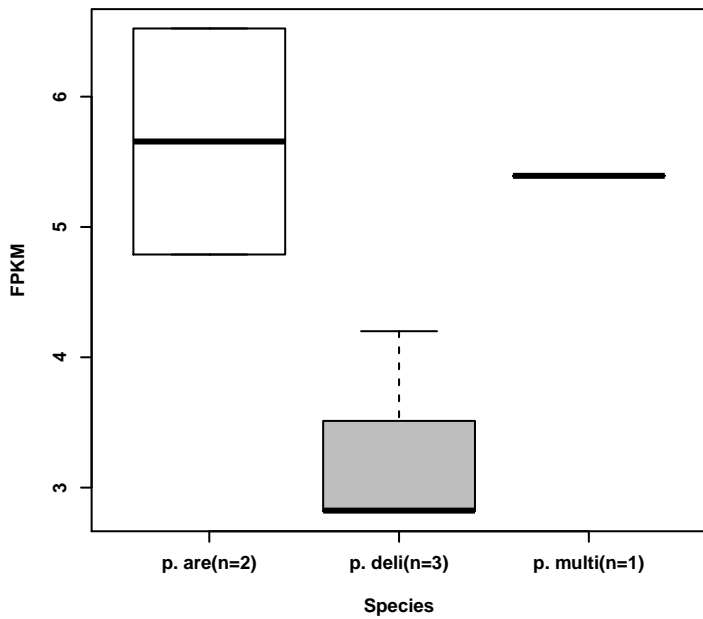
**Metabolic intermediate biosynthesis**



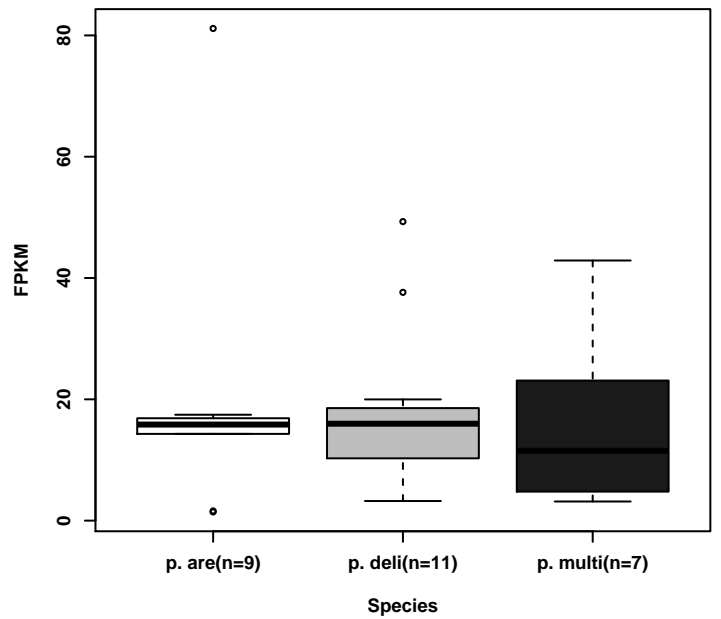
**Metabolic intermediate metabolism**

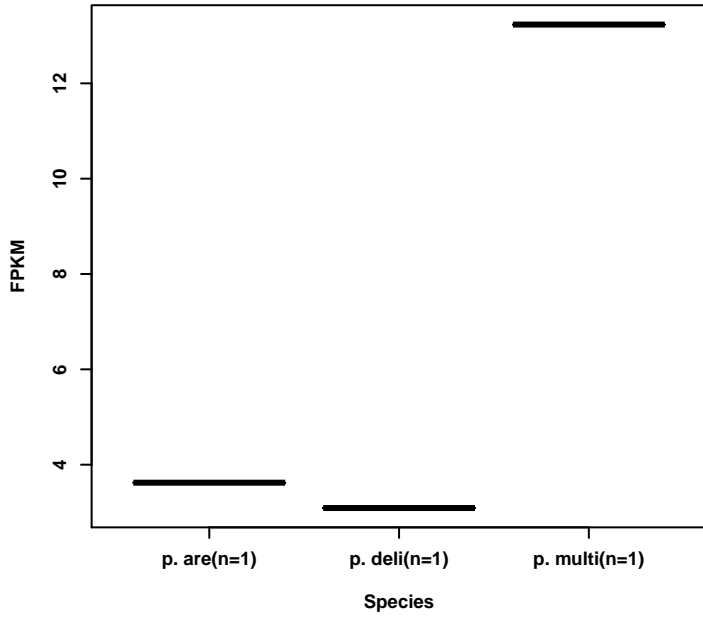
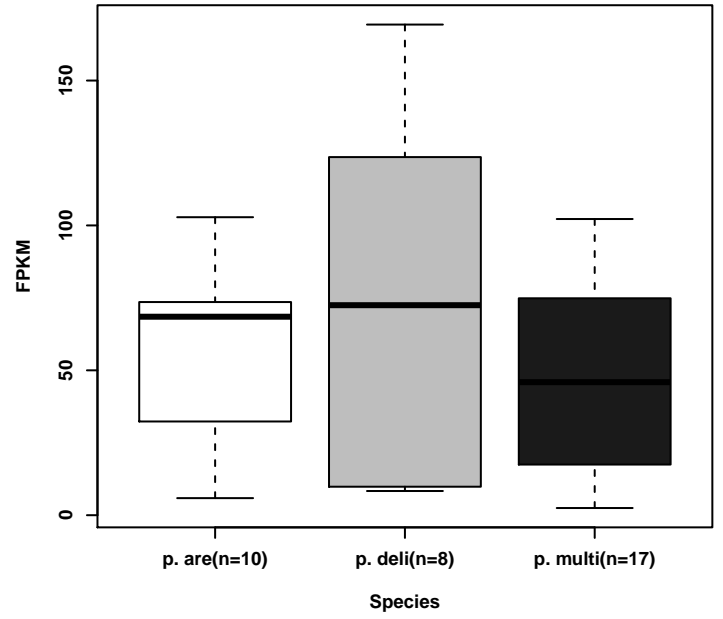
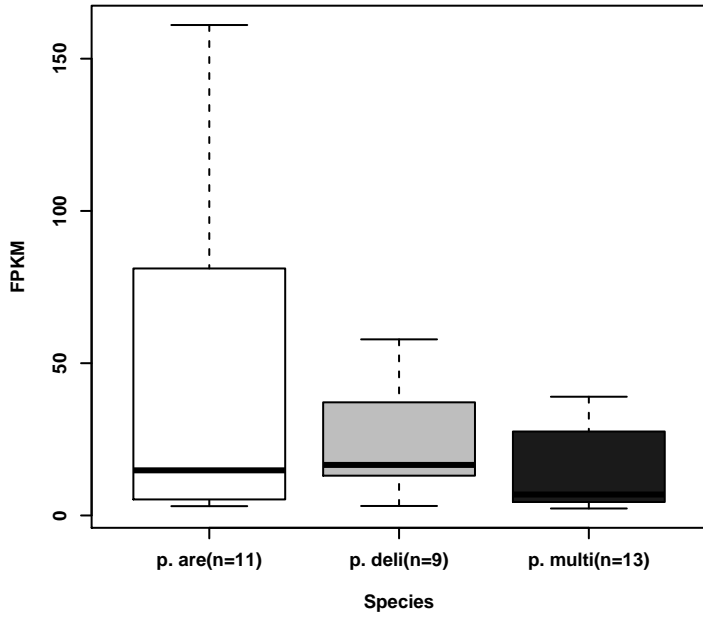
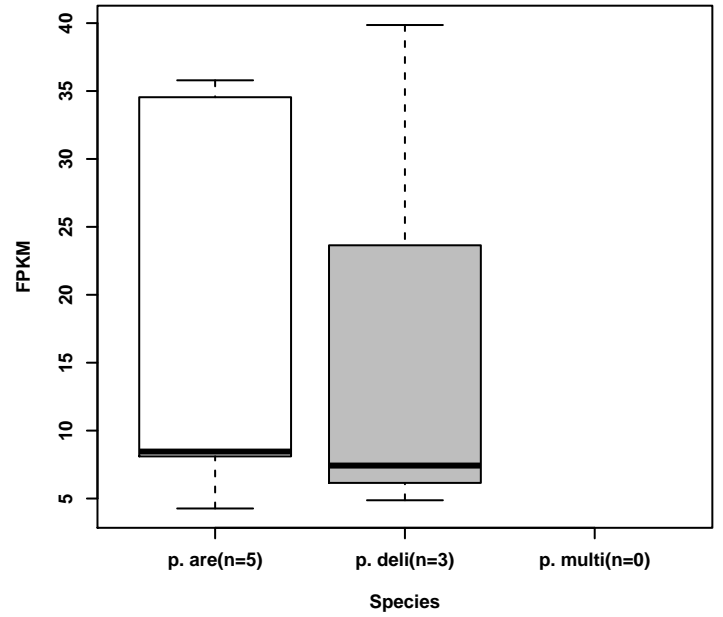
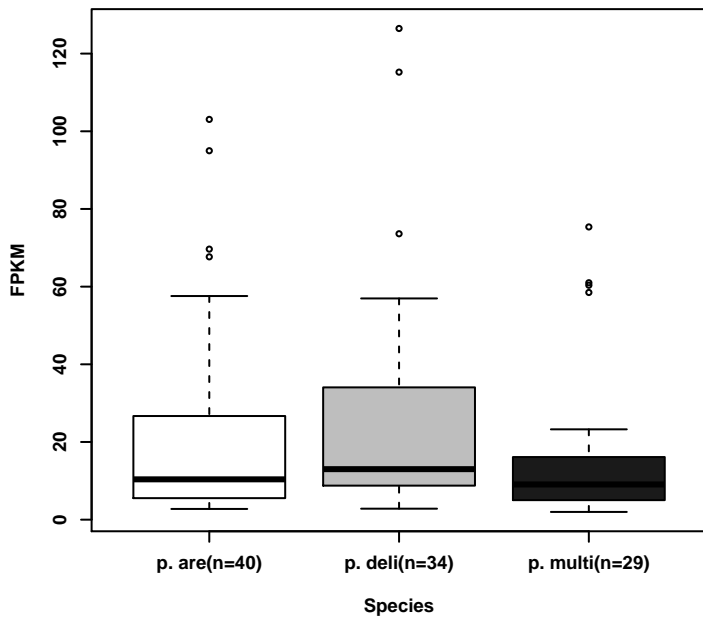
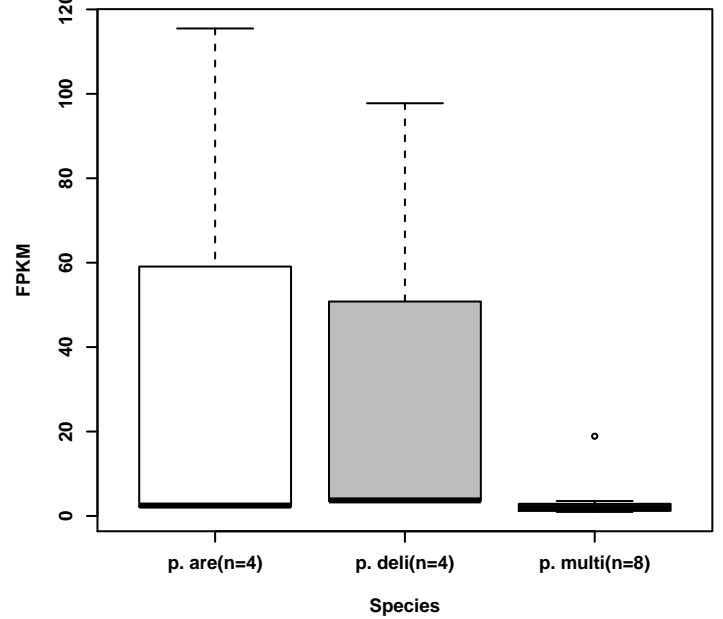


**Mycotoxin biosynthesis**

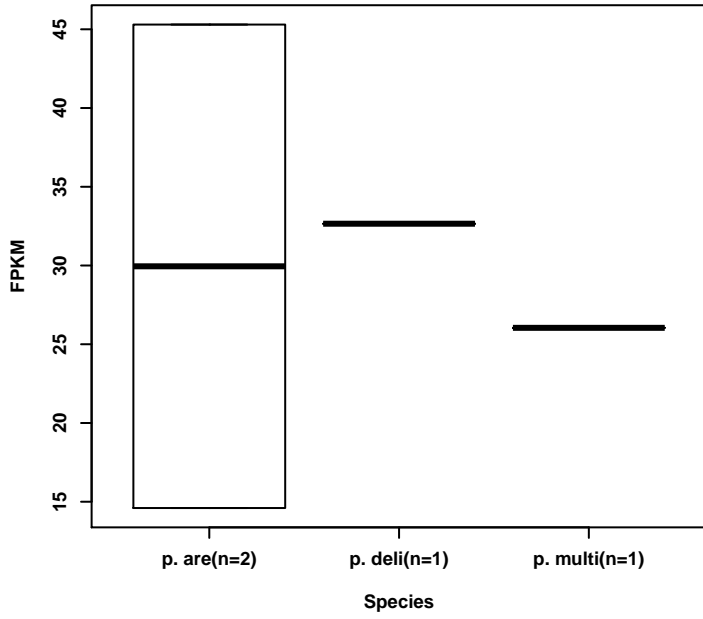


**Nitrogen metabolism**

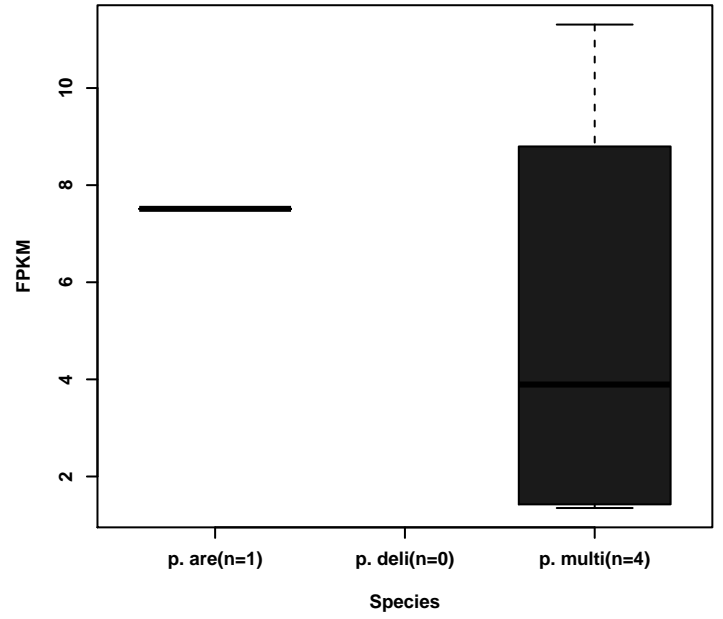


**Nucleoside biosynthesis****Nucleotide-sugar biosynthesis****One-carbon metabolism****Organic acid metabolism****Phospholipid metabolism****Photosynthesis**

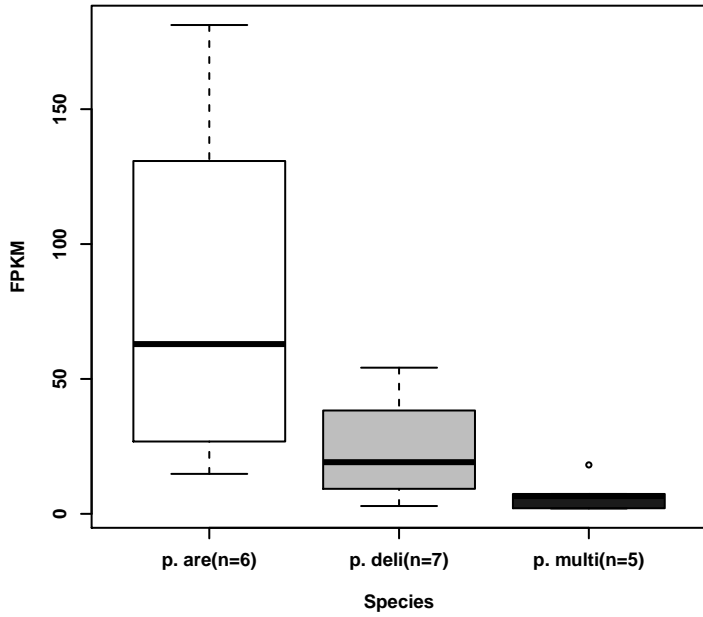
Phytoalexin biosynthesis



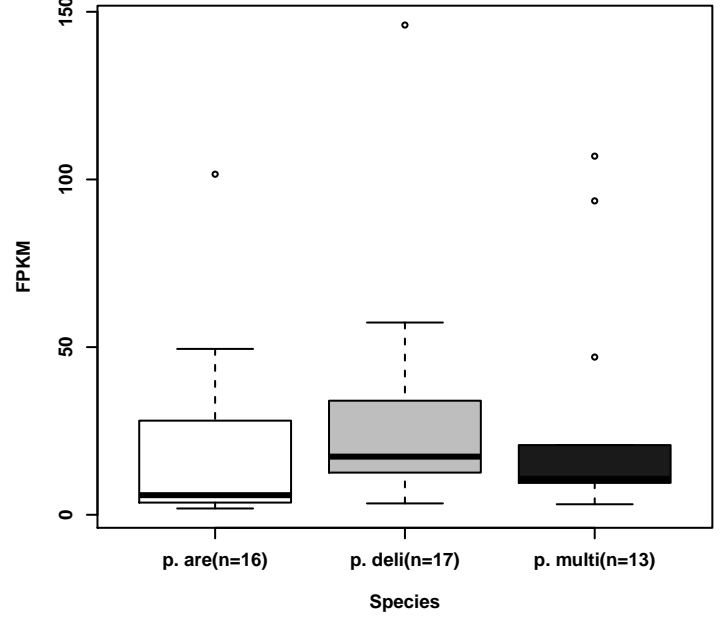
Pigment biosynthesis



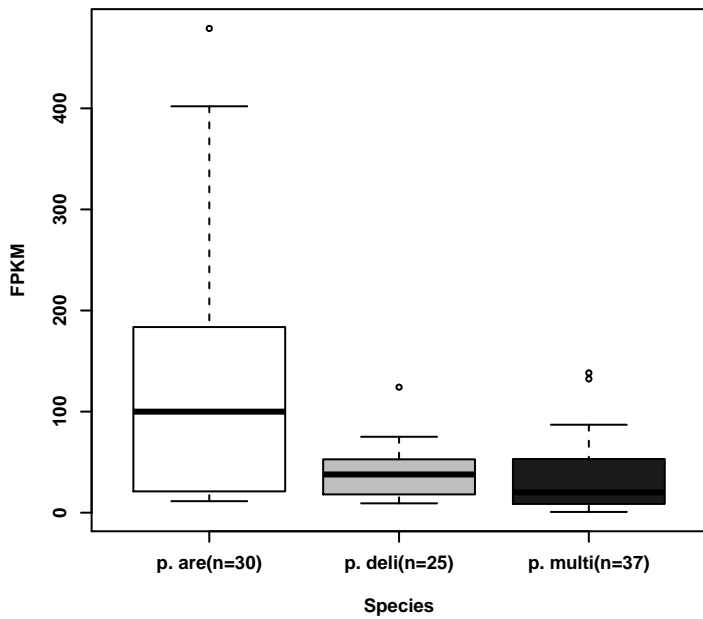
Plant hormone biosynthesis



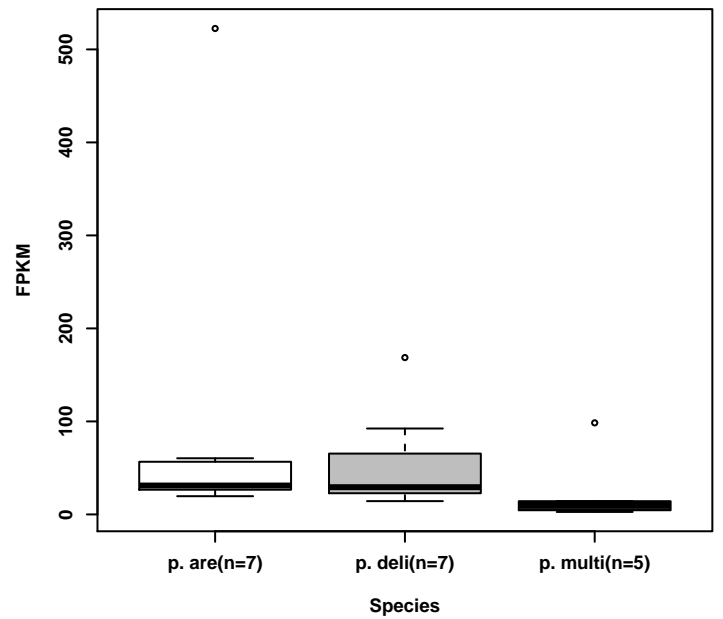
Polyol metabolism

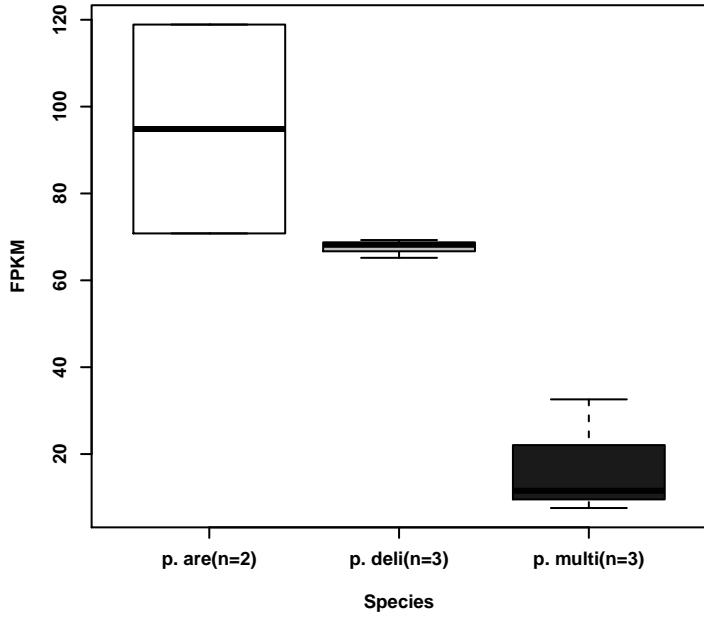
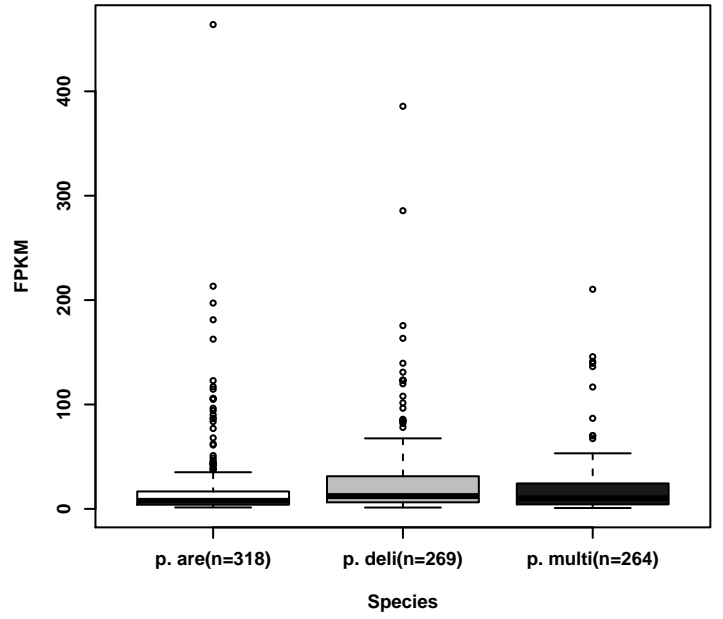
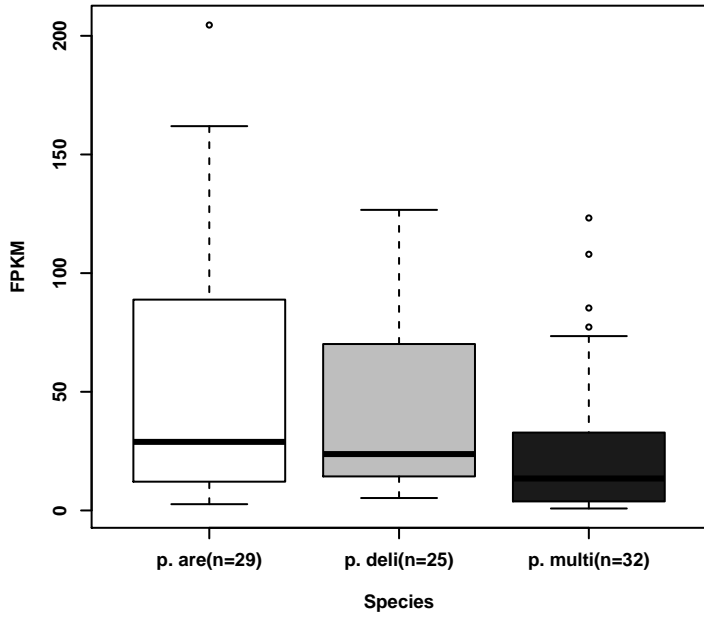
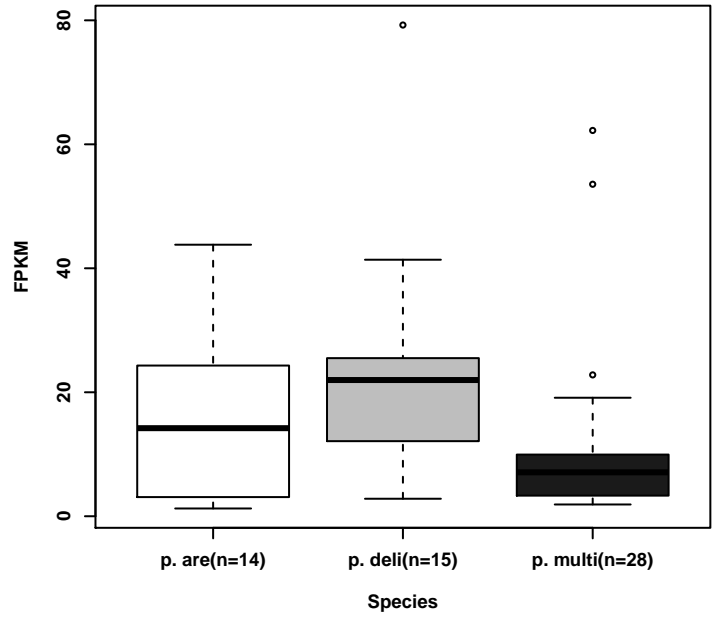
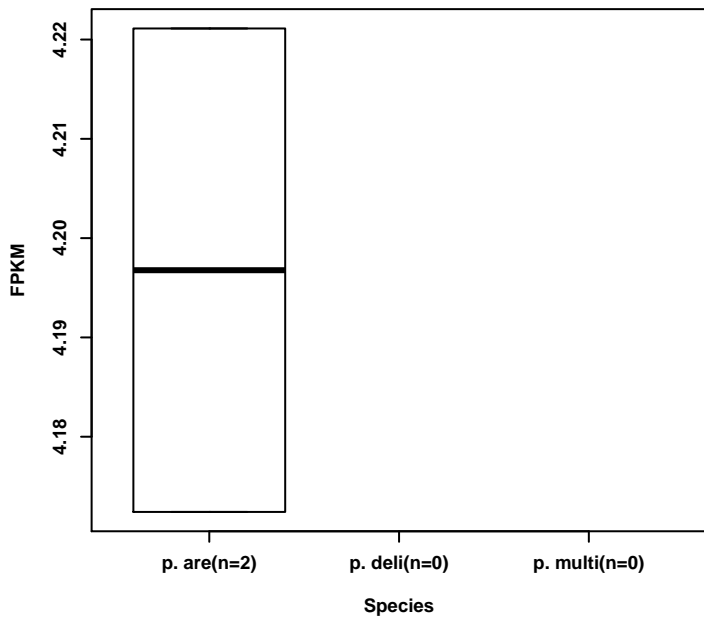
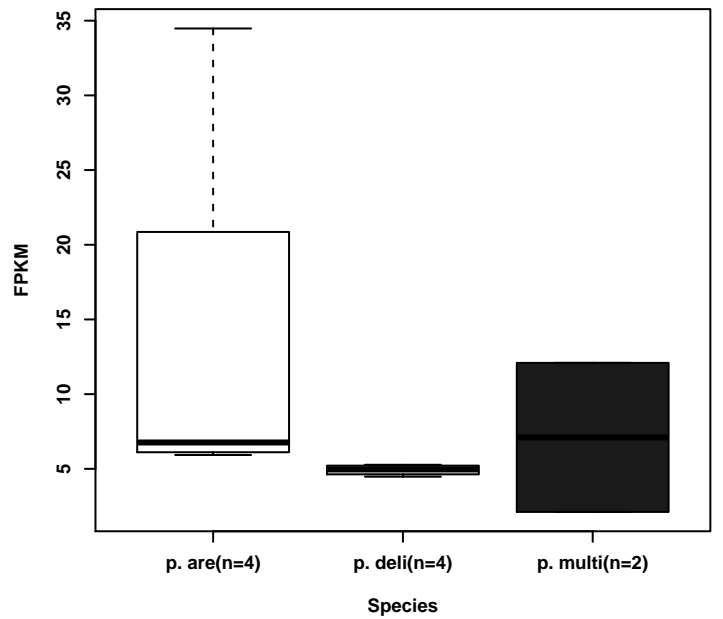


Porphyrin-containing compound metabolism

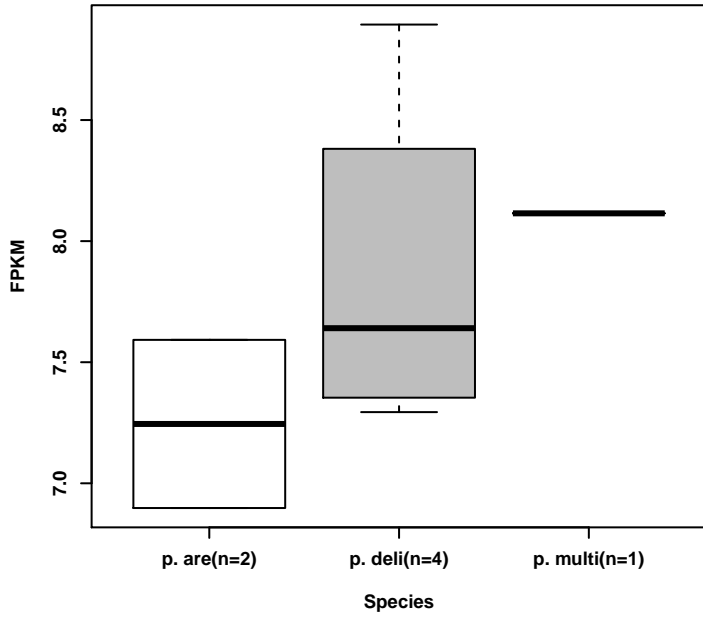


Protein biosynthesis

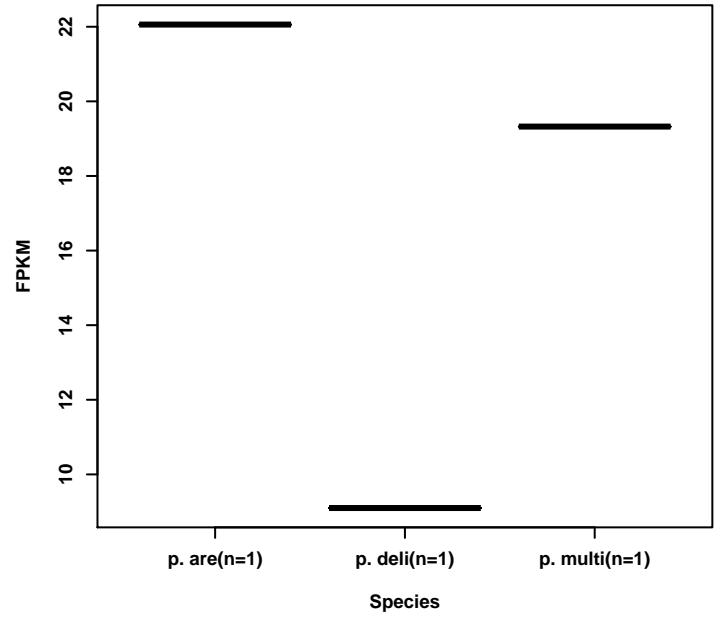


**Protein degradation****Protein modification****Purine metabolism****Pyrimidine metabolism****Secondary metabolite biosynthesis****Secondary metabolite metabolism**

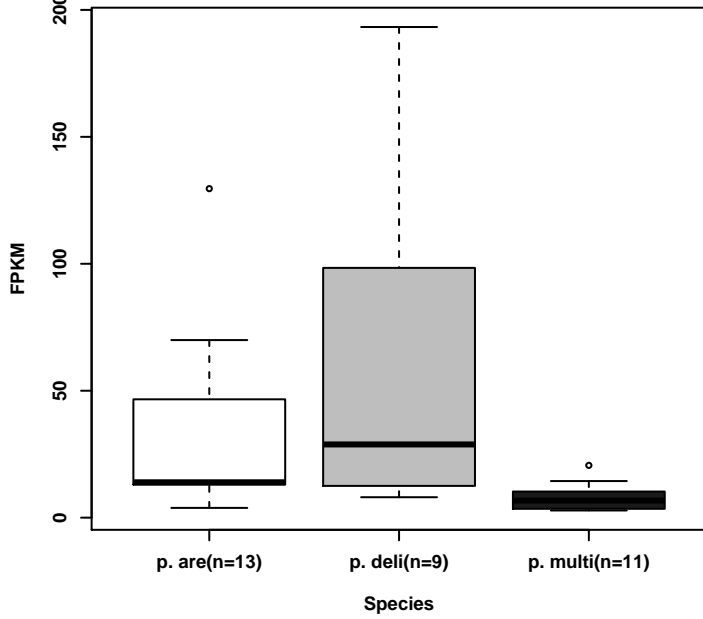
Signal transduction



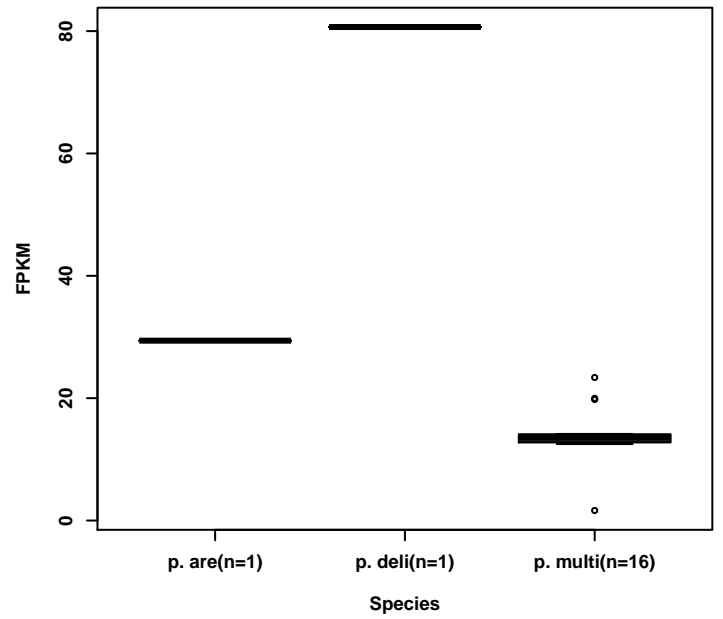
Spore coat biogenesis



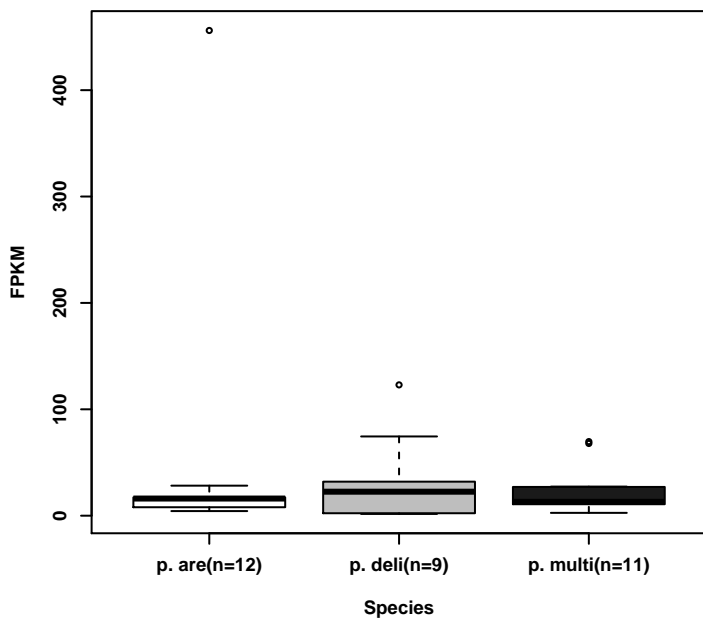
Steroid biosynthesis



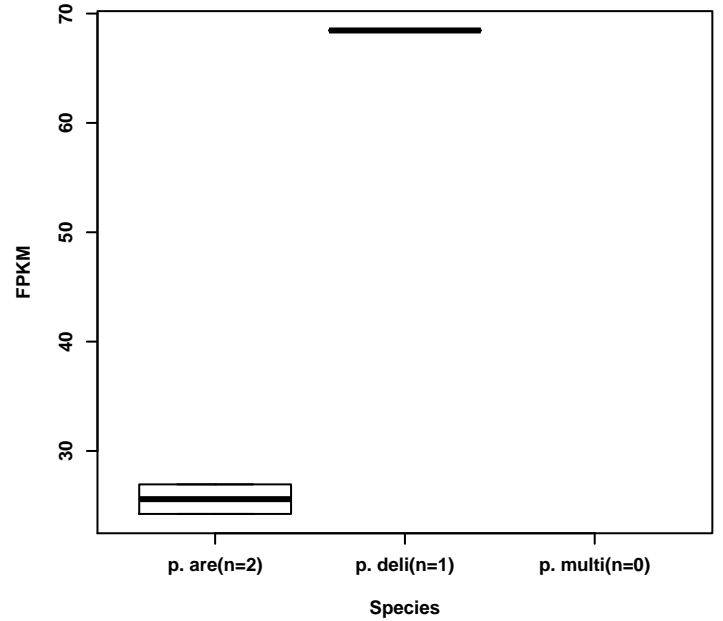
Steroid metabolism



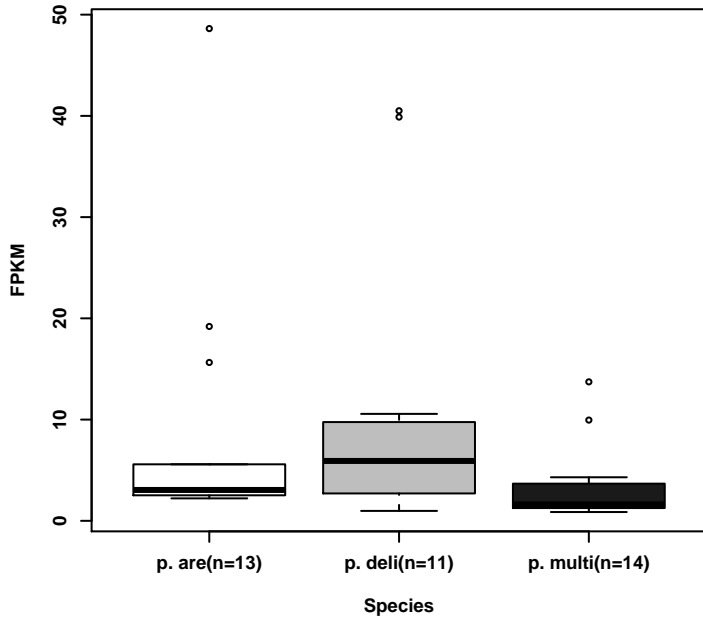
Sulfur metabolism



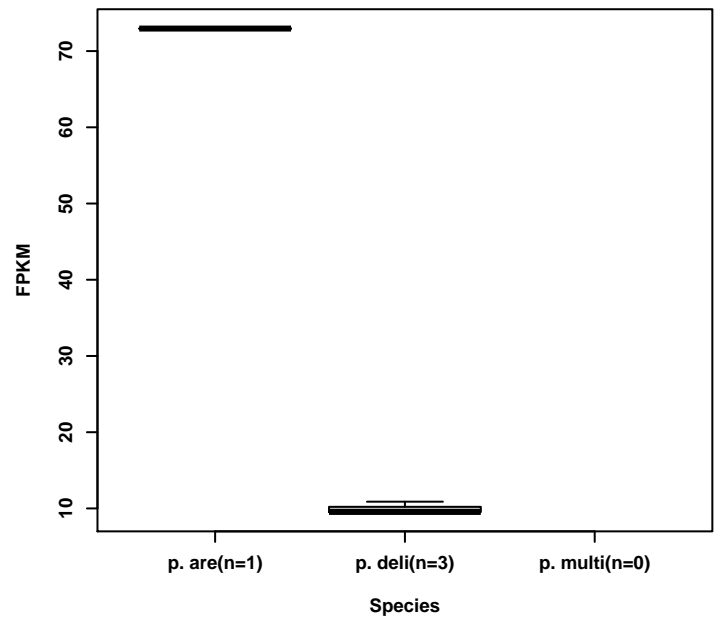
Terpene metabolism



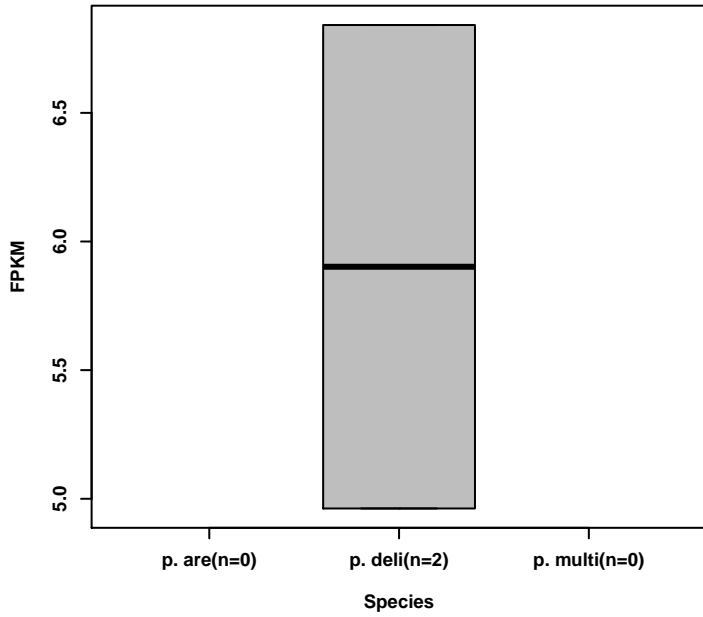
**tRNA modification**



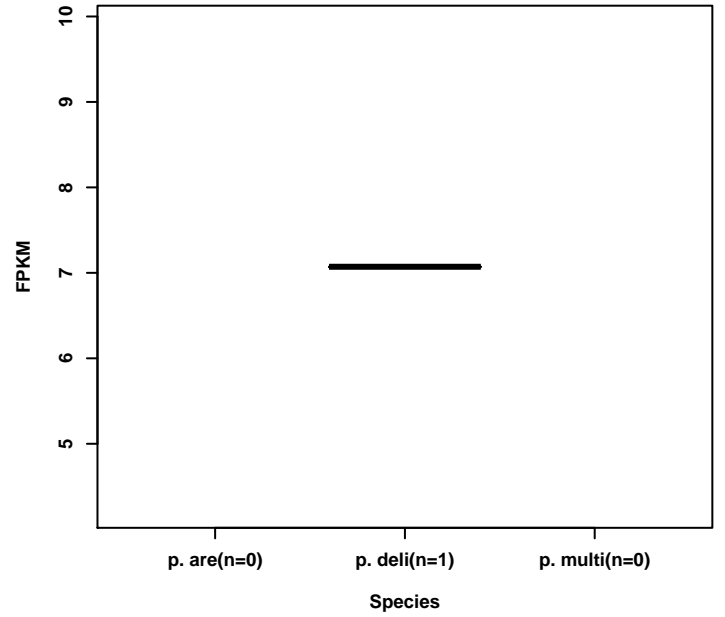
**Xenobiotic degradation**



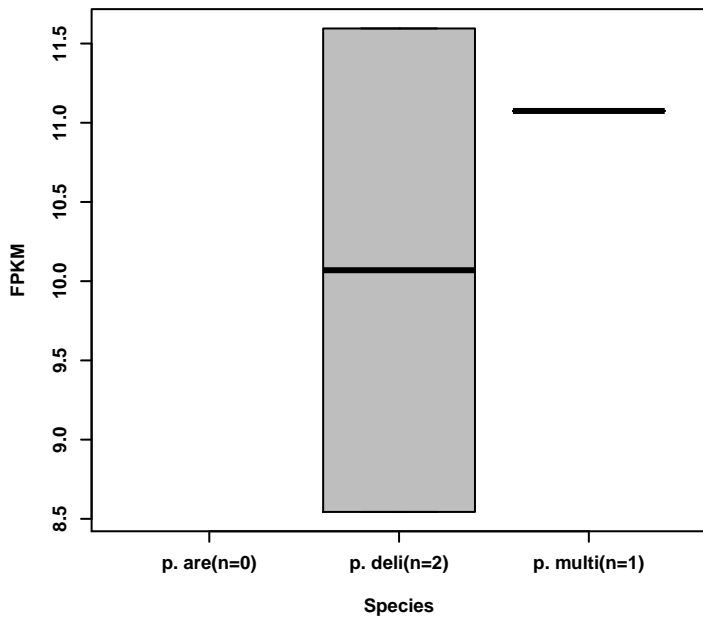
**Capsule biogenesis**



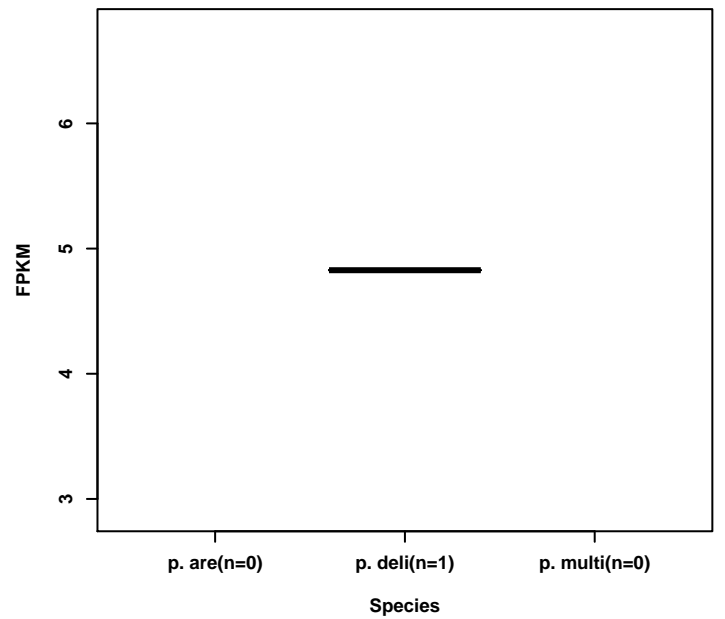
**Glycan degradation**



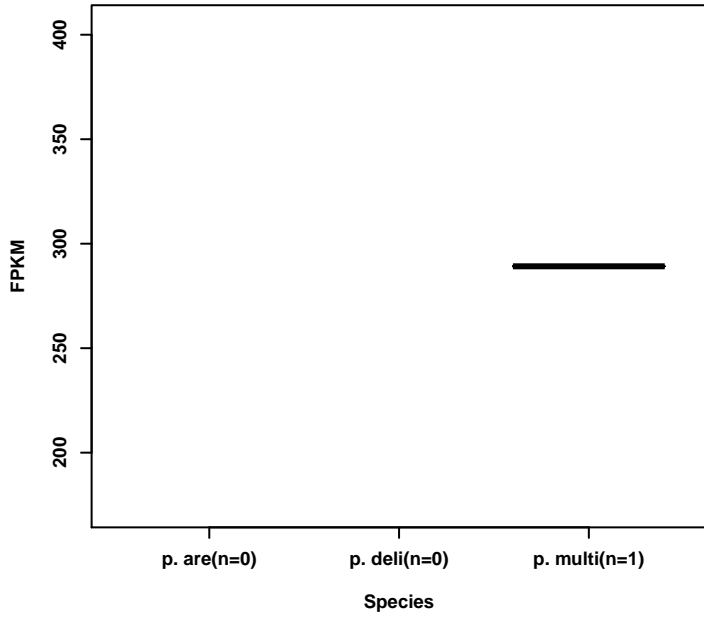
**Organosulfur biosynthesis**



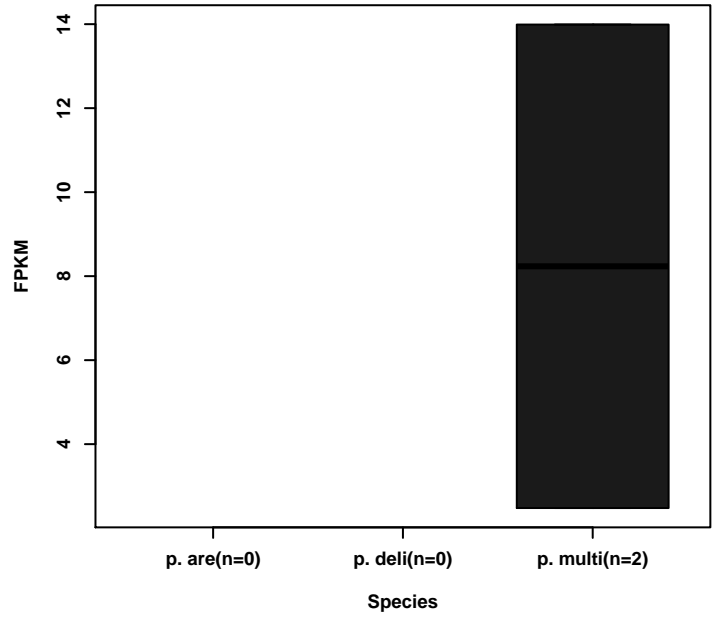
**Siderophore biosynthesis**



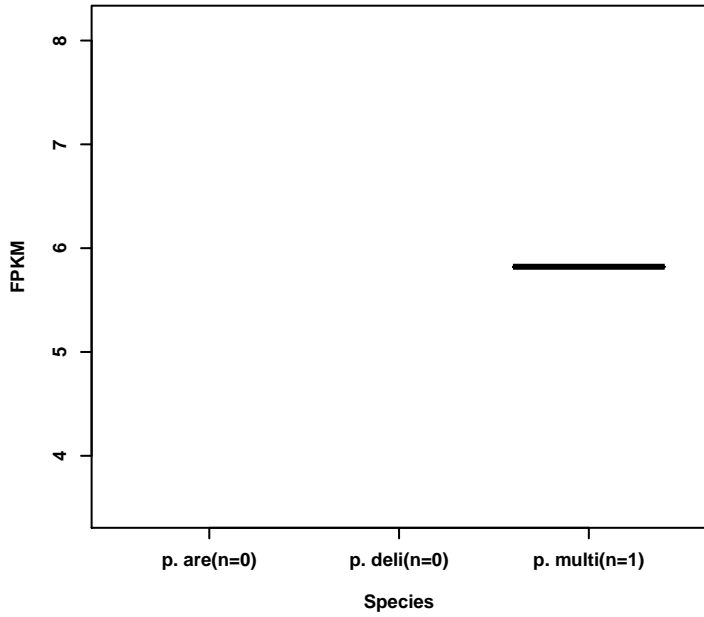
Alcohol metabolism



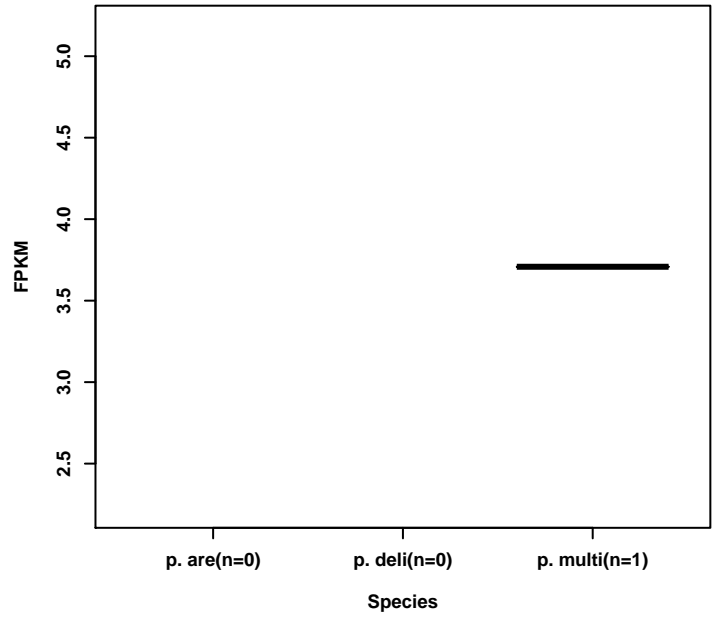
Bacterial outer membrane biogenesis



Hydrocarbon metabolism



Plant hormone metabolism

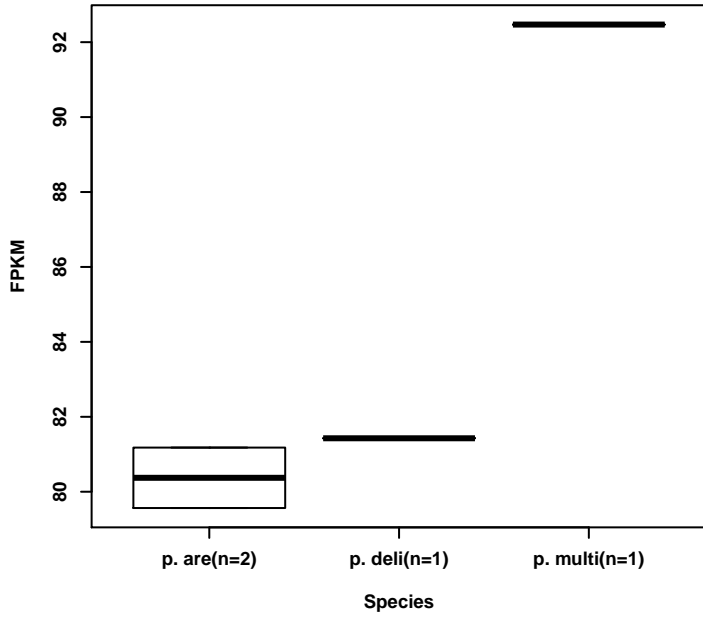




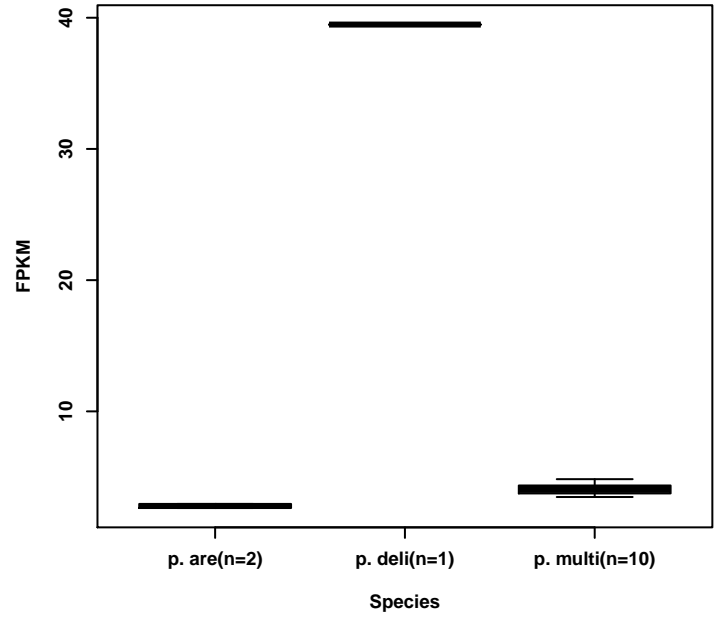
**Supplementary Figure S3. Pathways expression based on the second level of annotation.**

Each graph shows the expression of one of the specific pathways included in the more general pathways, showed in Figure S2, for each species. See Figure S2 legend for details.

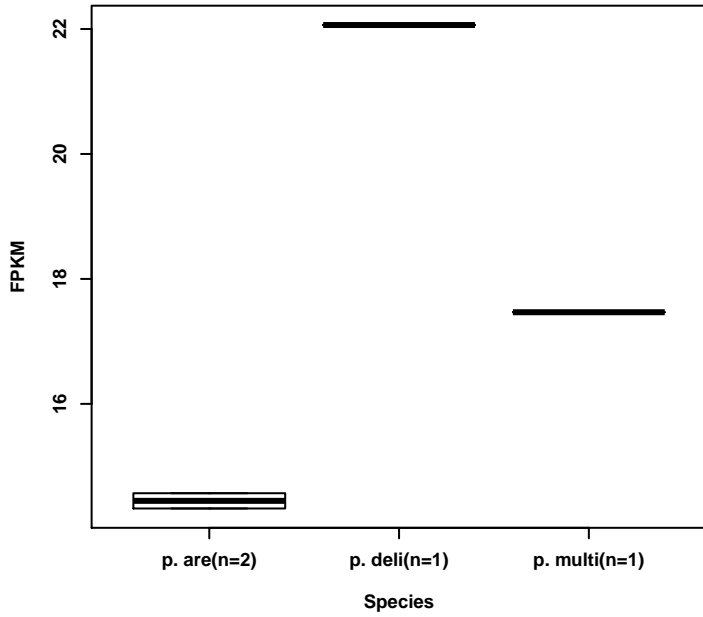
1-deoxy-D-xylulose 5-phosphate biosynthesis



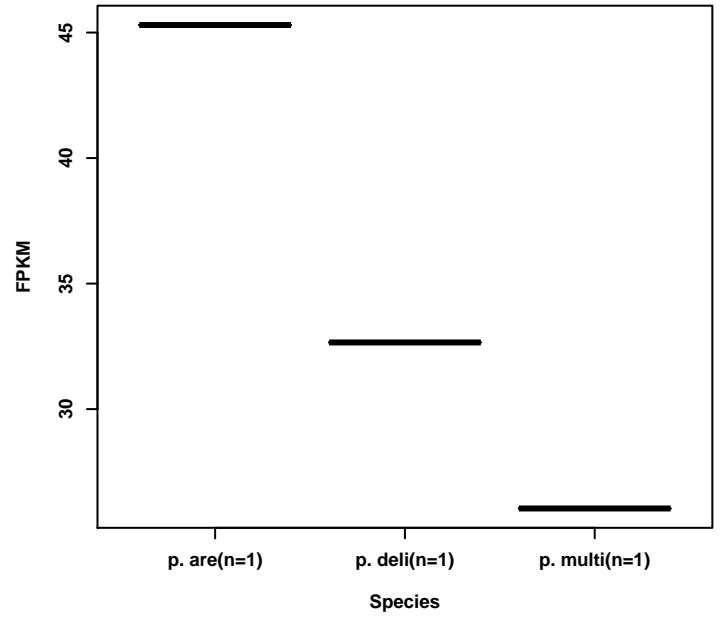
2-deoxy-D-ribose 1-phosphate degradation



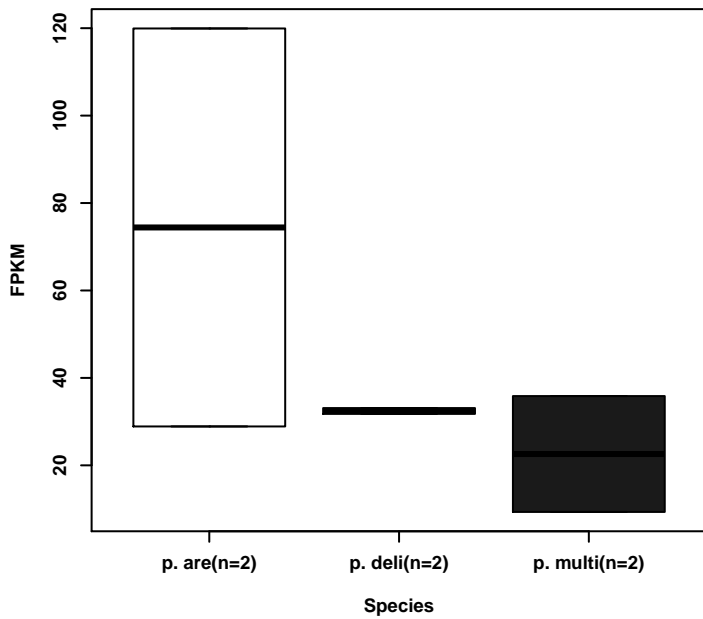
2-deoxystreptamine biosynthesis



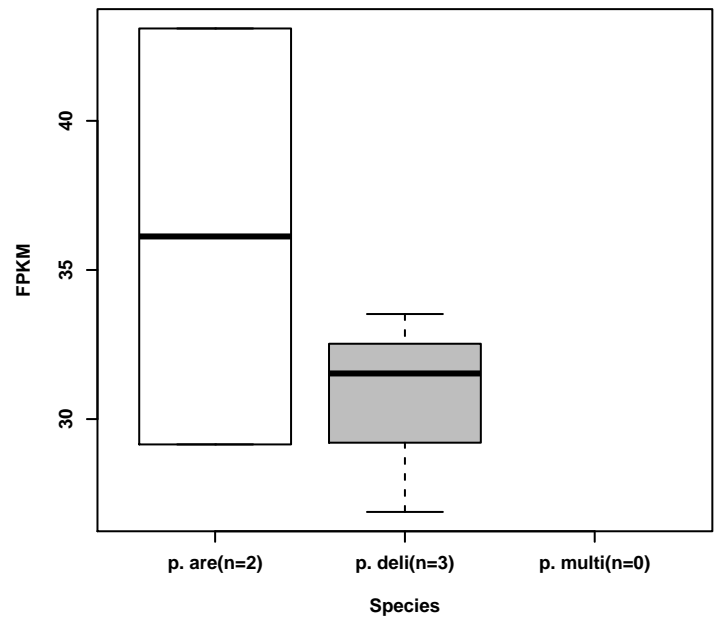
3,4',5-trihydroxystilbene biosynthesis



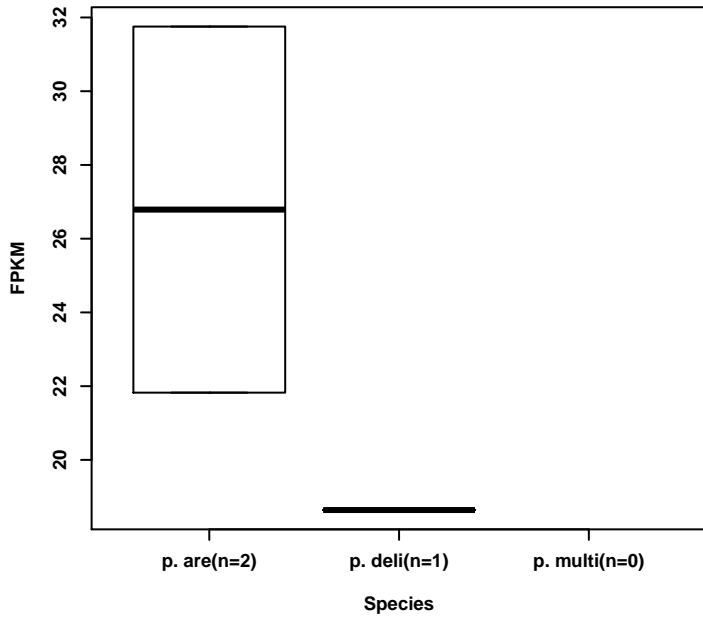
3',5'-cyclic AMP degradation



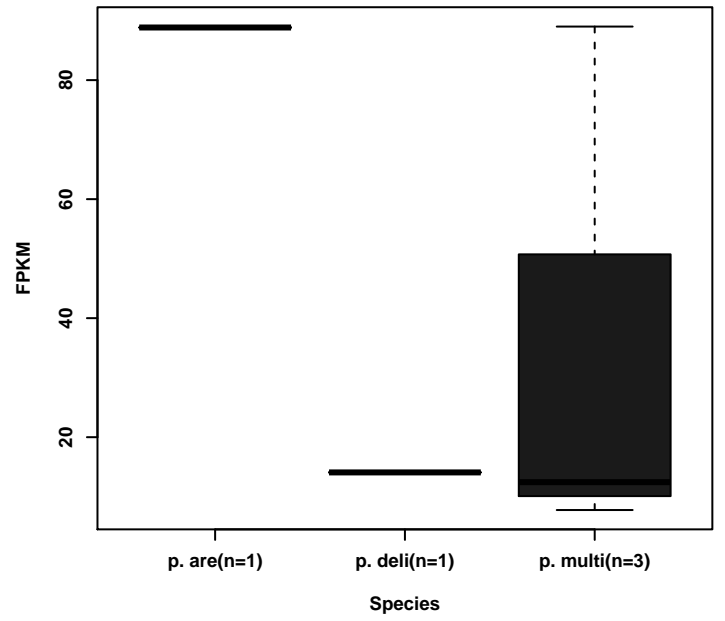
3-chlorocatechol degradation



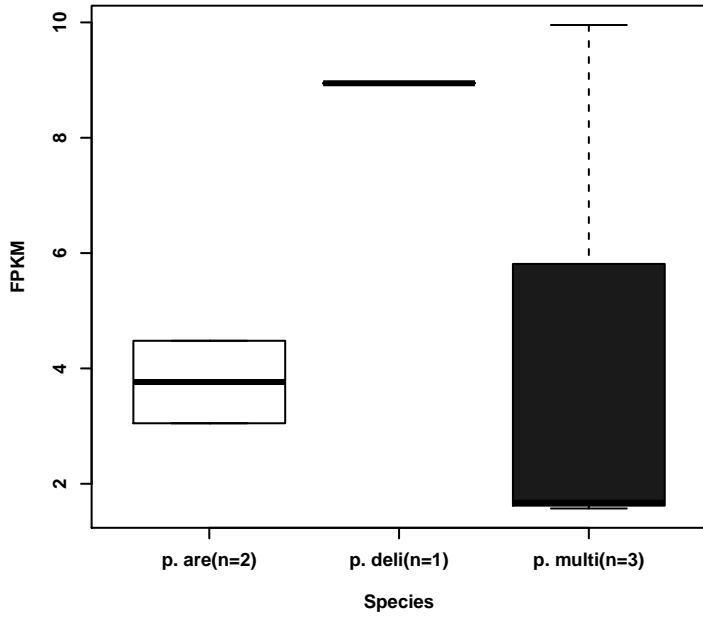
3-phenylpropanoate degradation



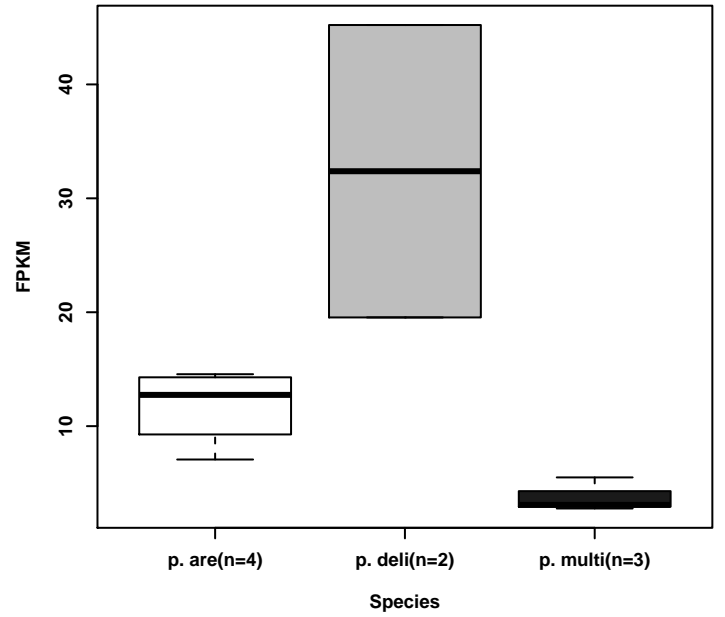
4-aminobutanoate degradation



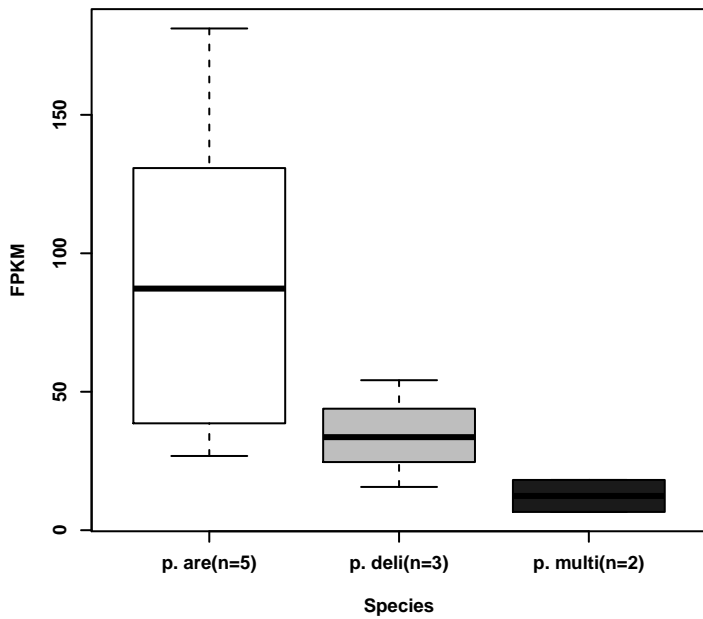
5-methoxycarbonylmethyl-2-thiouridine-tRNA biosynthesis



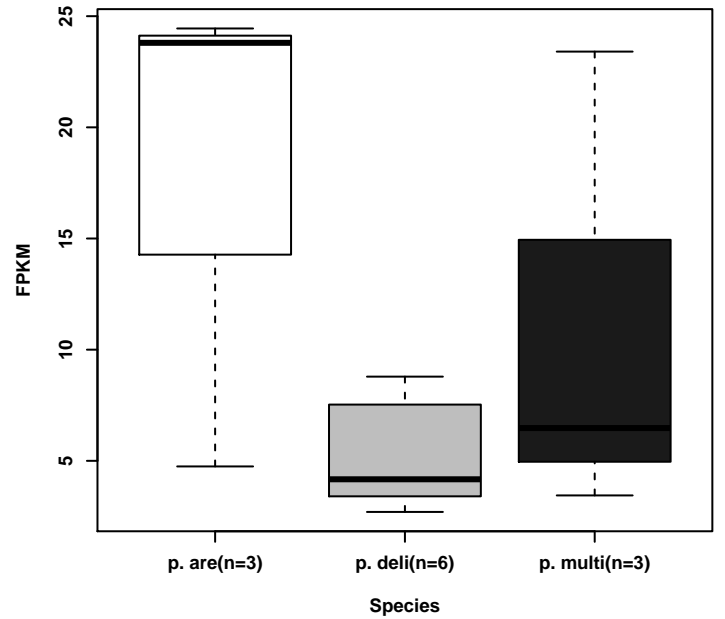
5-phospho-alpha-D-ribose 1-diphosphate biosynthesis



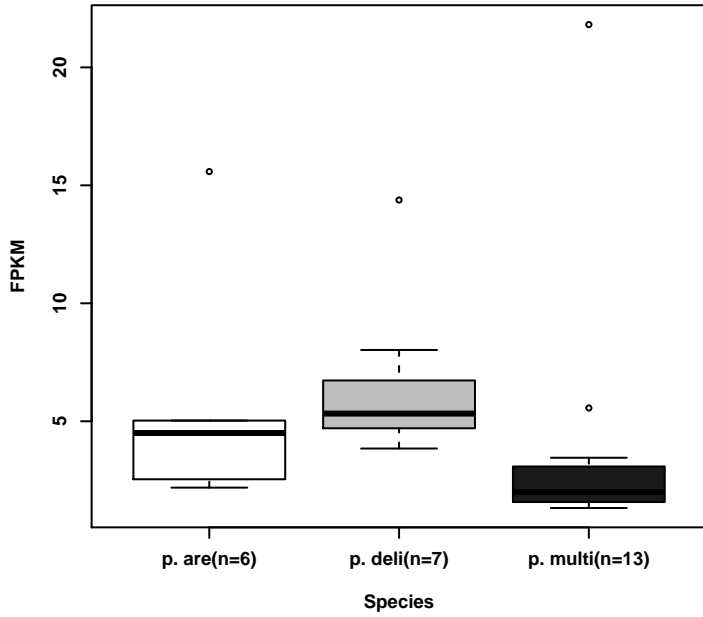
abscisate biosynthesis



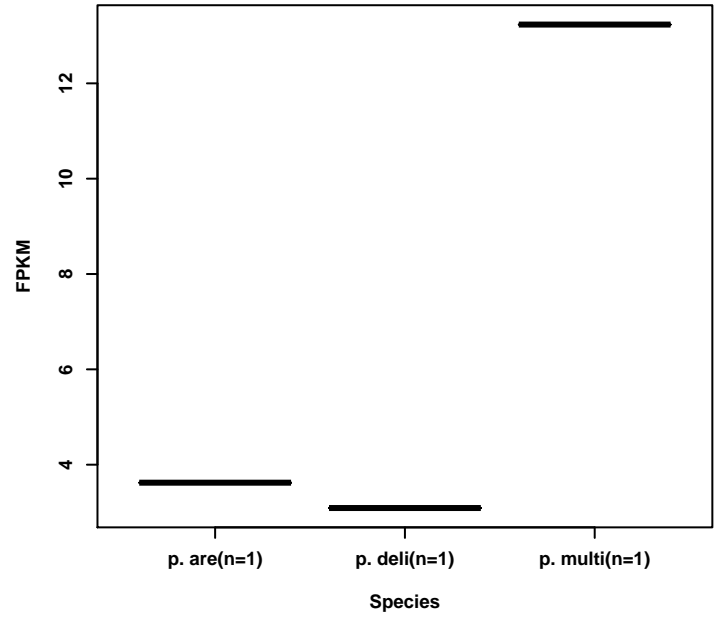
acetyl-CoA biosynthesis



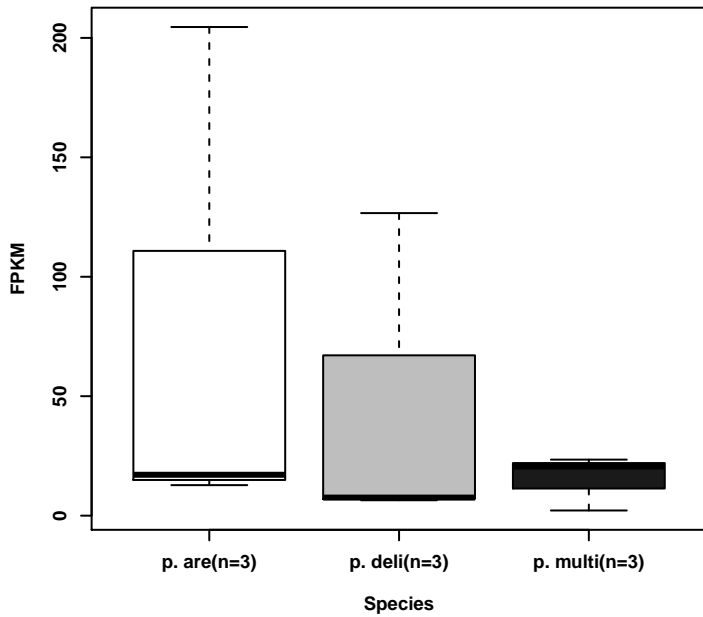
adenosylcobalamin biosynthesis



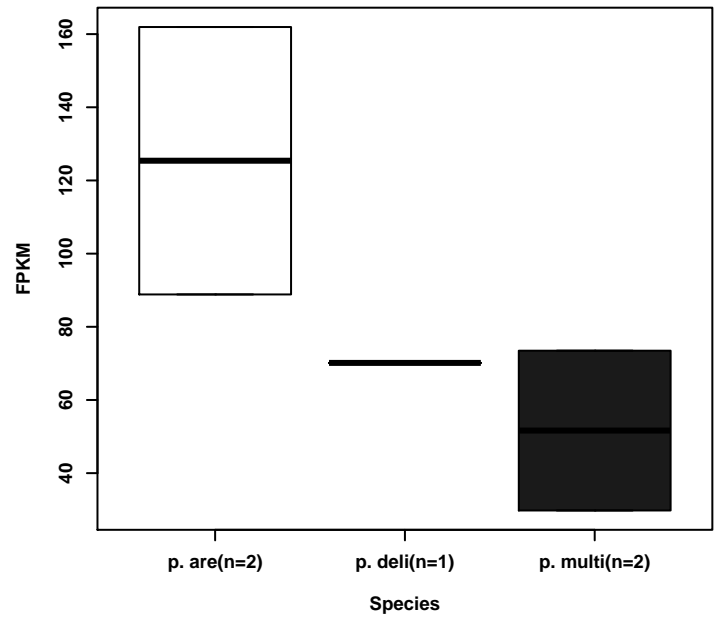
alpha-ribazole biosynthesis



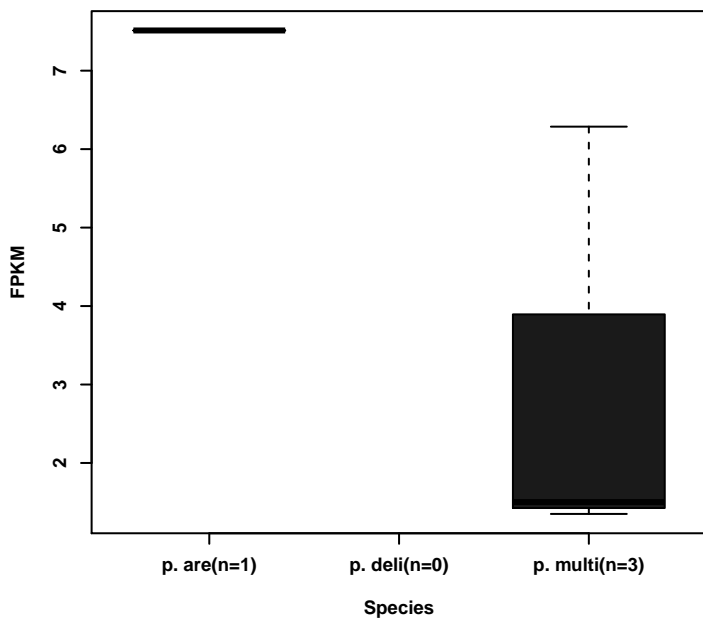
AMP biosynthesis via de novo pathway



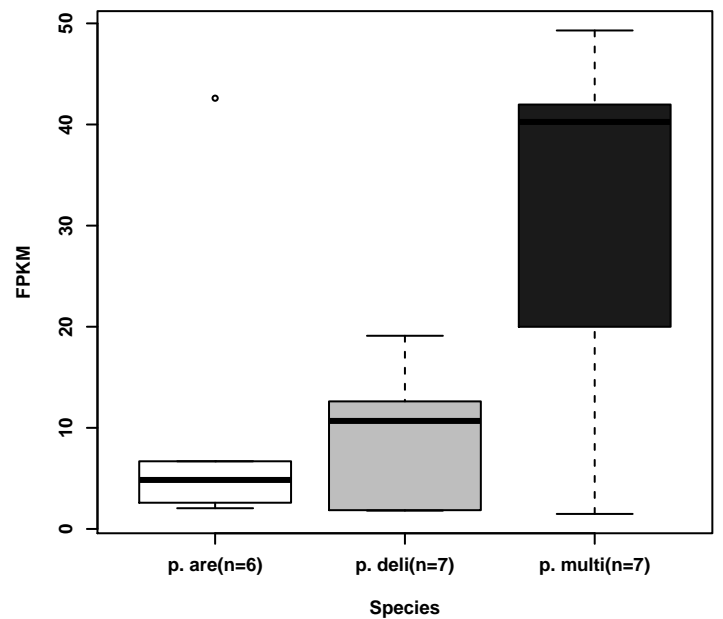
AMP biosynthesis via salvage pathway



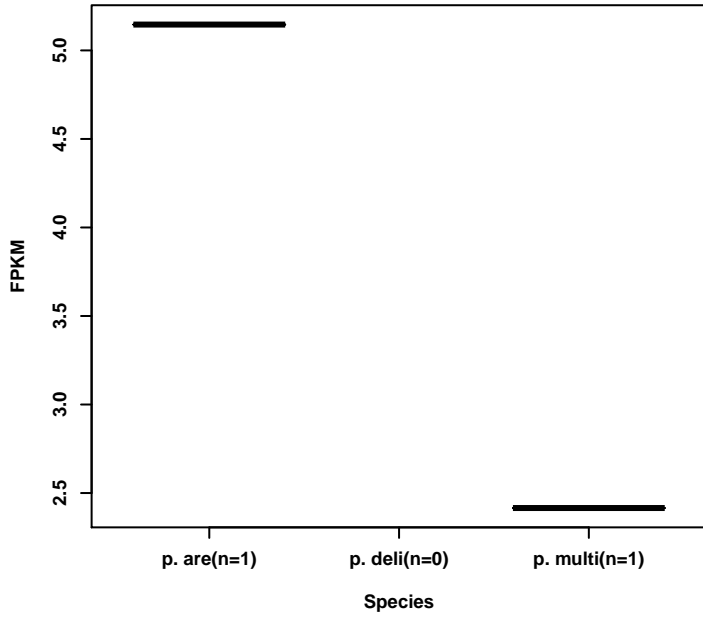
anthocyanin biosynthesis



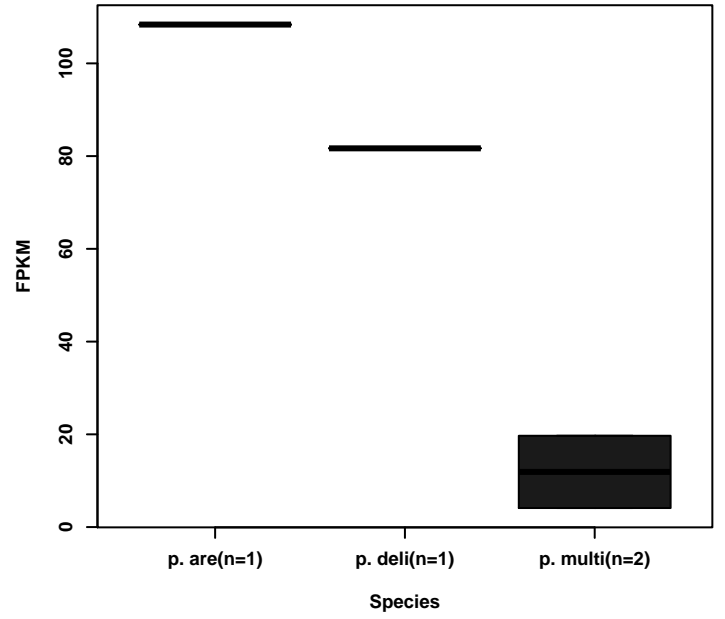
B6 vitamer degradation



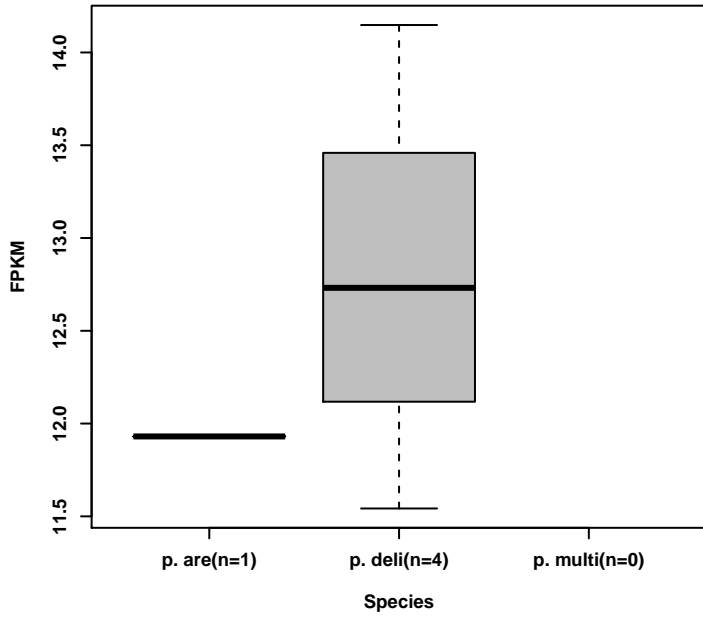
**B6 vitamer interconversion**



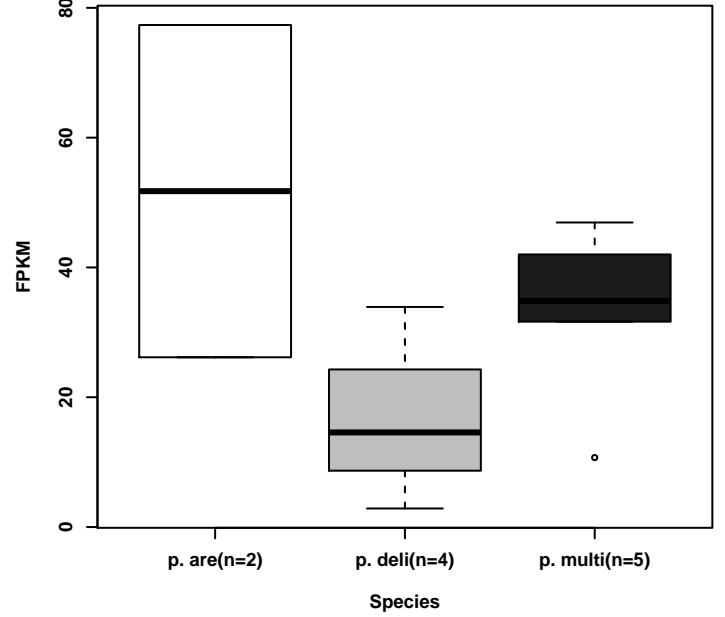
**beta-carotene biosynthesis**



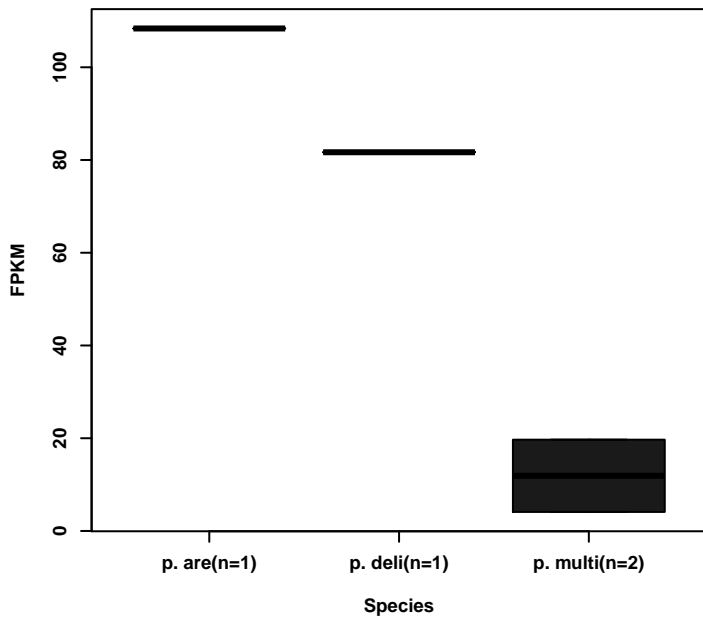
**betaine biosynthesis via choline pathway**



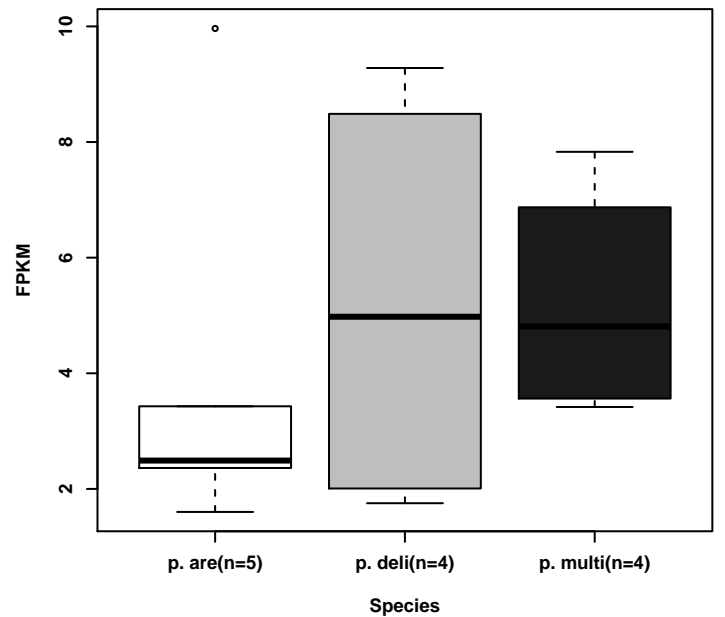
**betaine biosynthesis via glycine pathway**



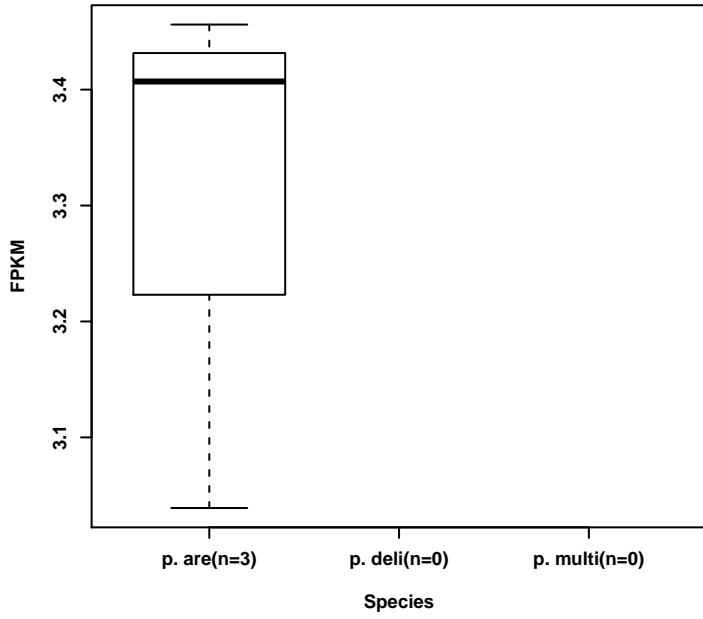
**beta-zeacarotene biosynthesis**



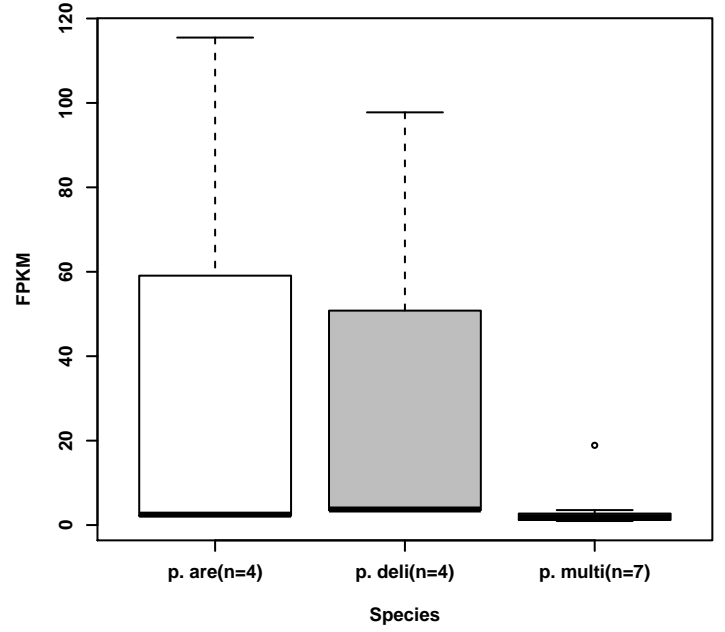
**biotin biosynthesis**



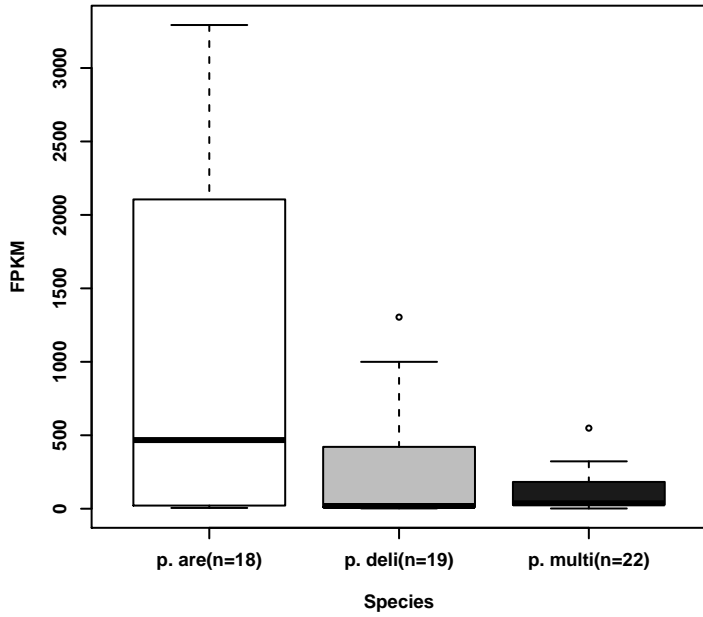
**C21-steroid hormone metabolism**



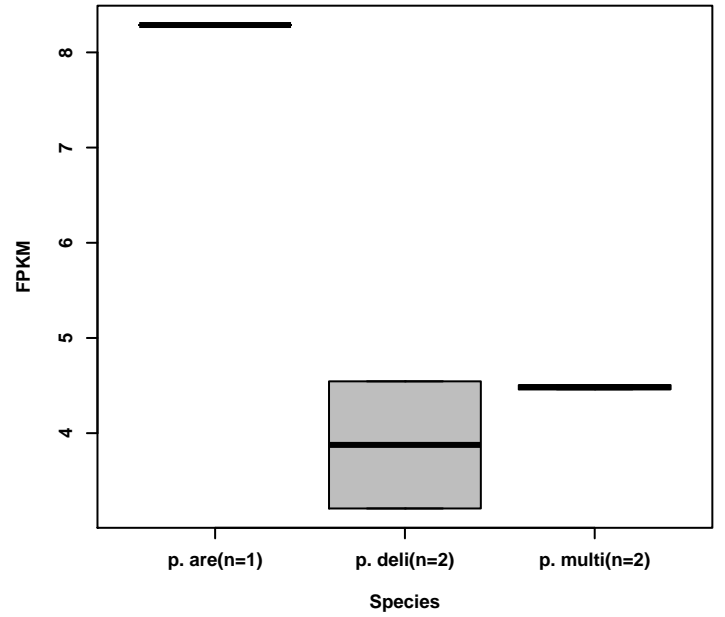
**C4 acid pathway**



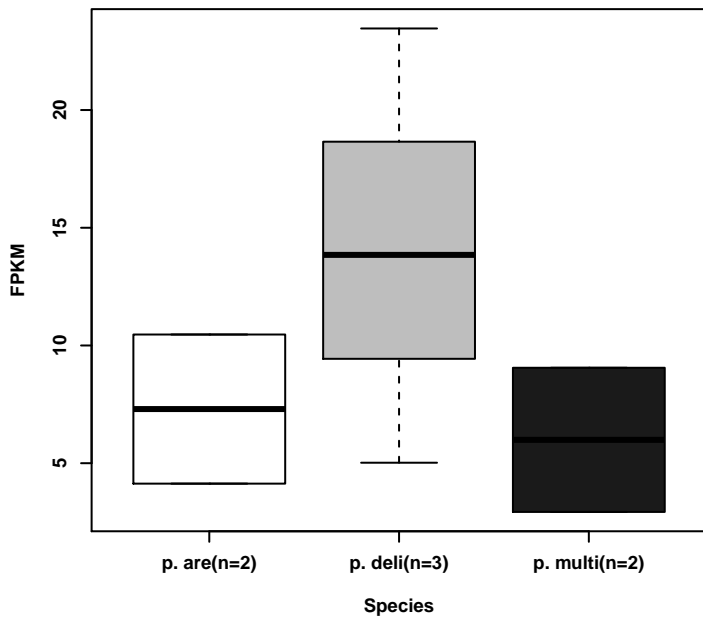
**Calvin cycle**



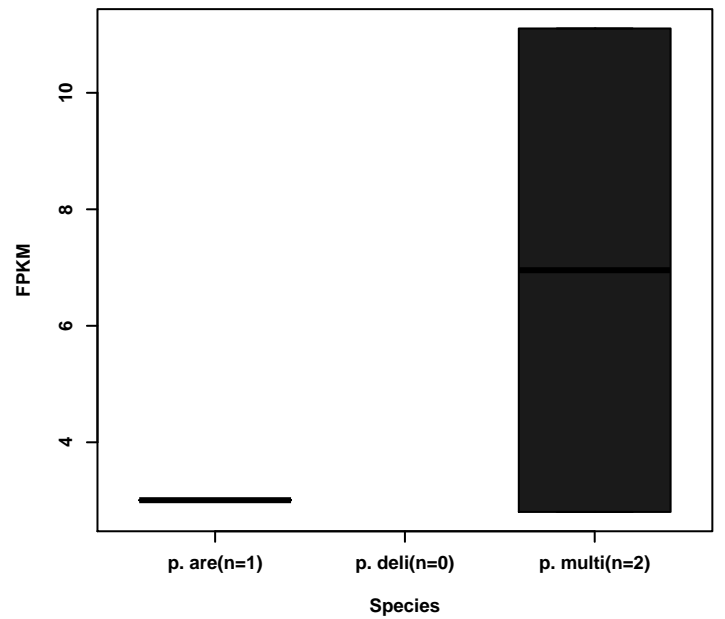
**carbamoyl phosphate degradation**



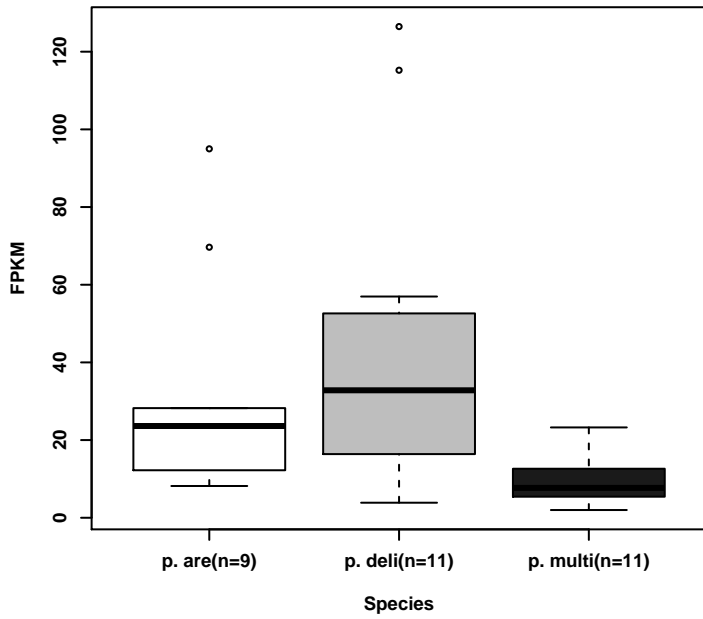
**carnitine biosynthesis**



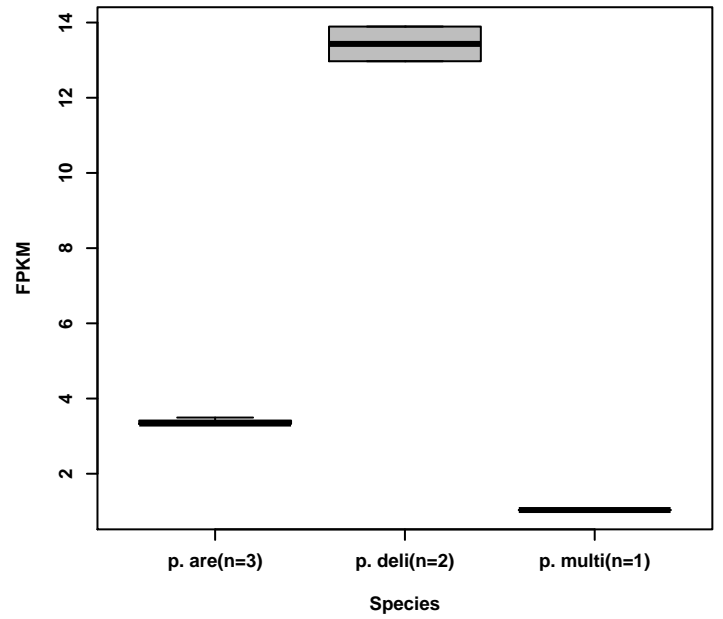
**carnitine metabolism**



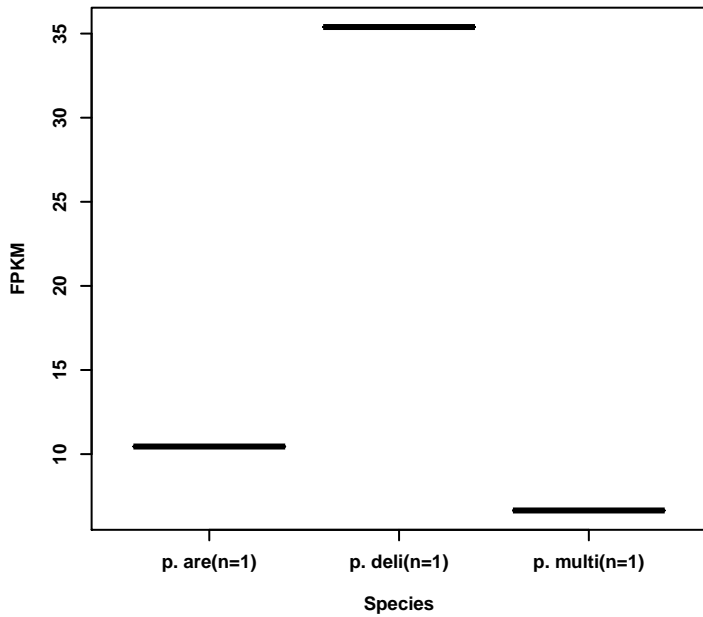
**CDP–diacylglycerol biosynthesis**



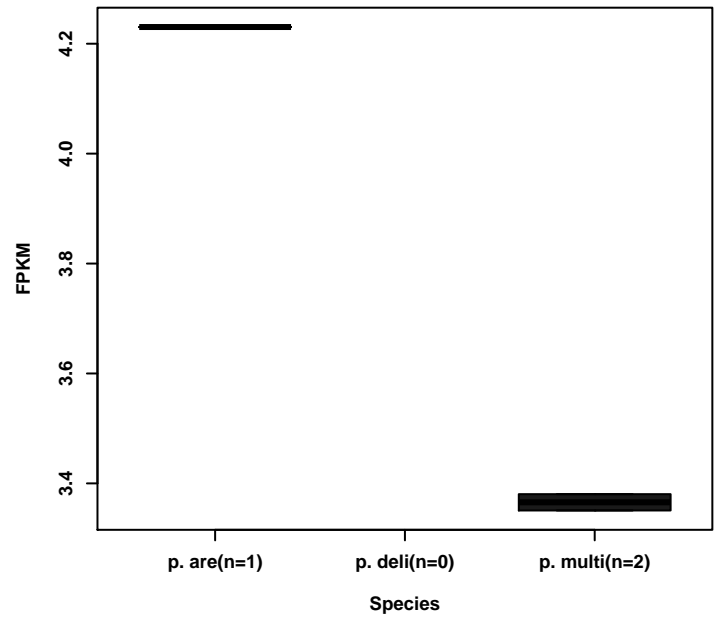
**cellulose degradation**



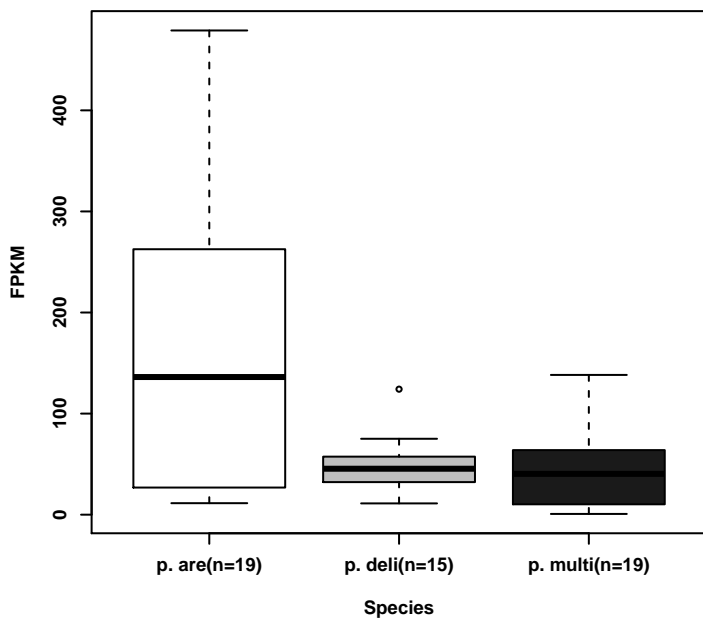
**cell wall polysaccharide biosynthesis**



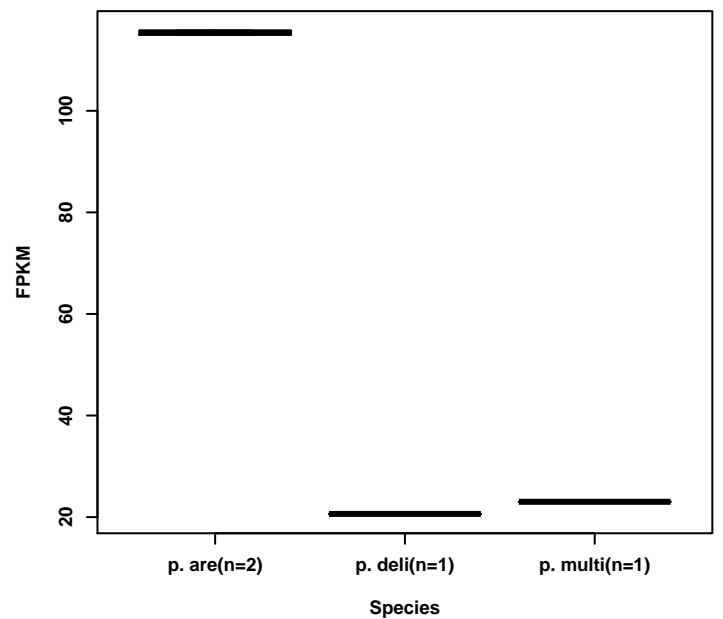
**cephalosporin C biosynthesis**



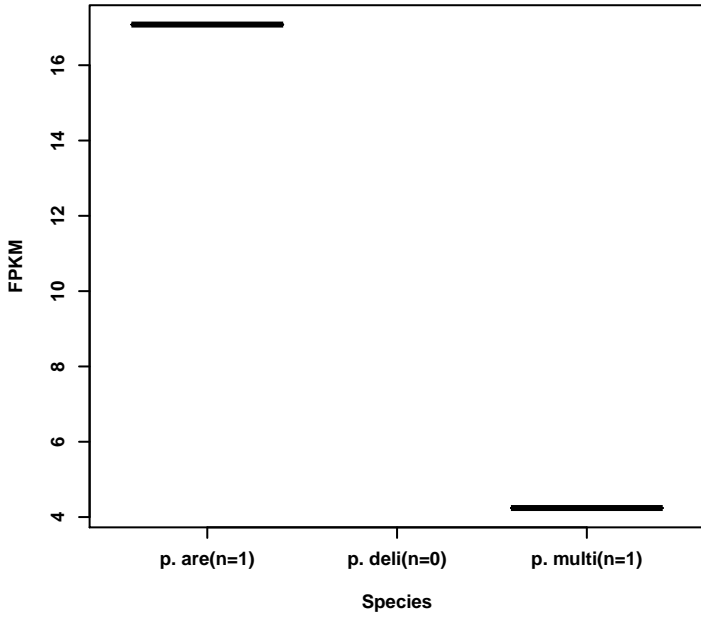
**chlorophyll biosynthesis**



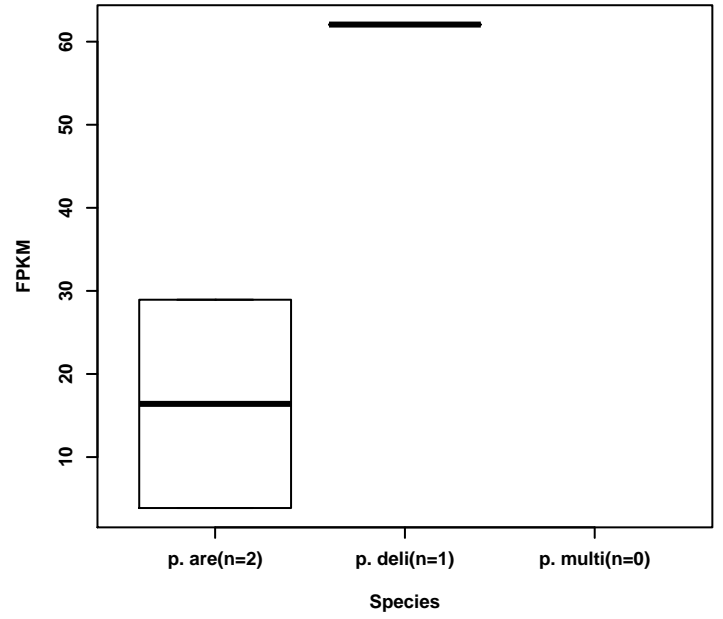
**chlorophyll biosynthesis (light-independent)**



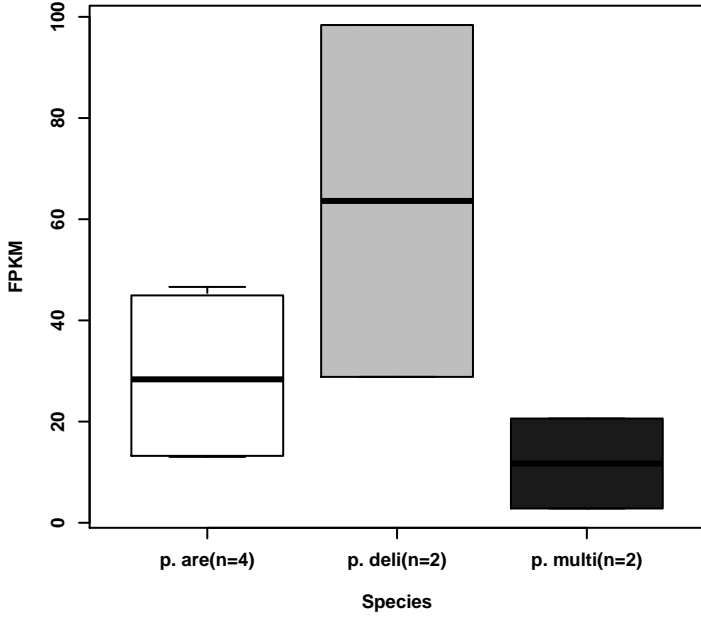
chlorophyll degradation



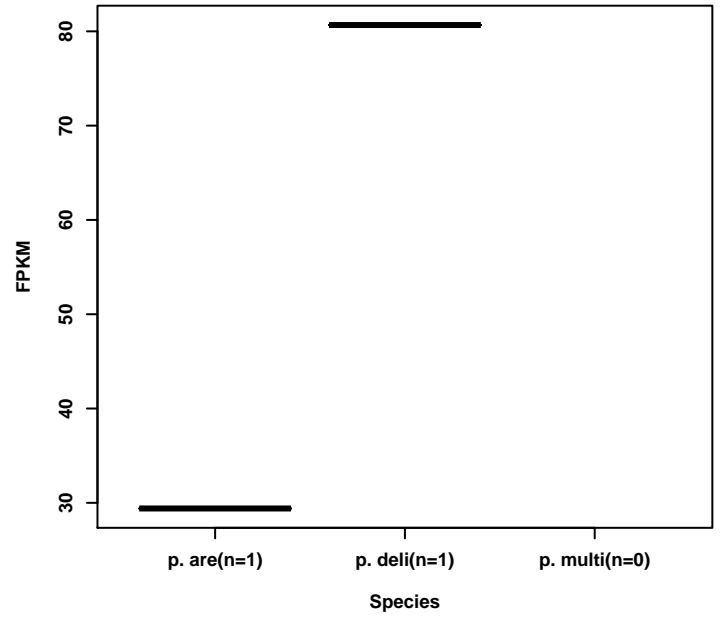
cholecalciferol biosynthesis



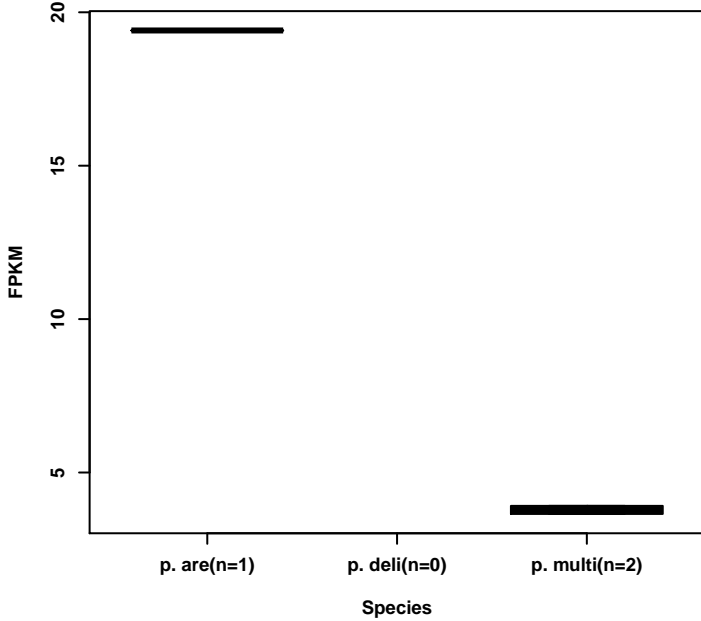
cholesterol biosynthesis



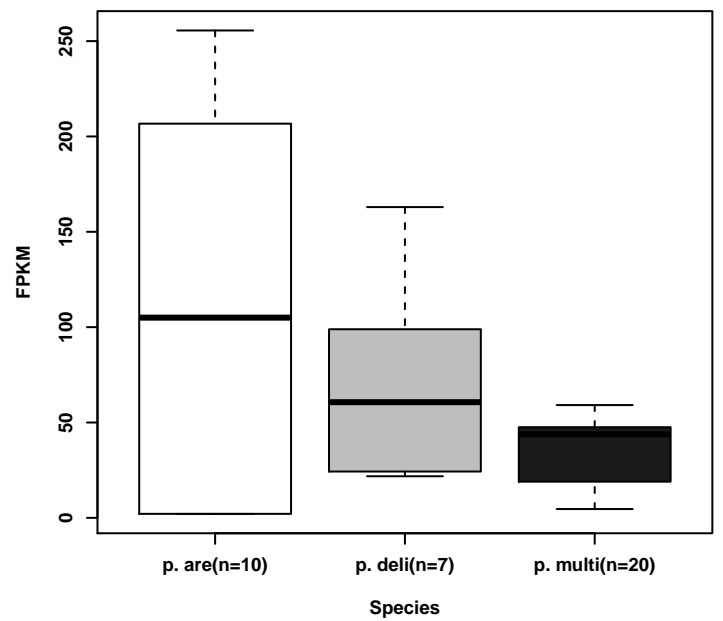
cholesterol metabolism



choline biosynthesis

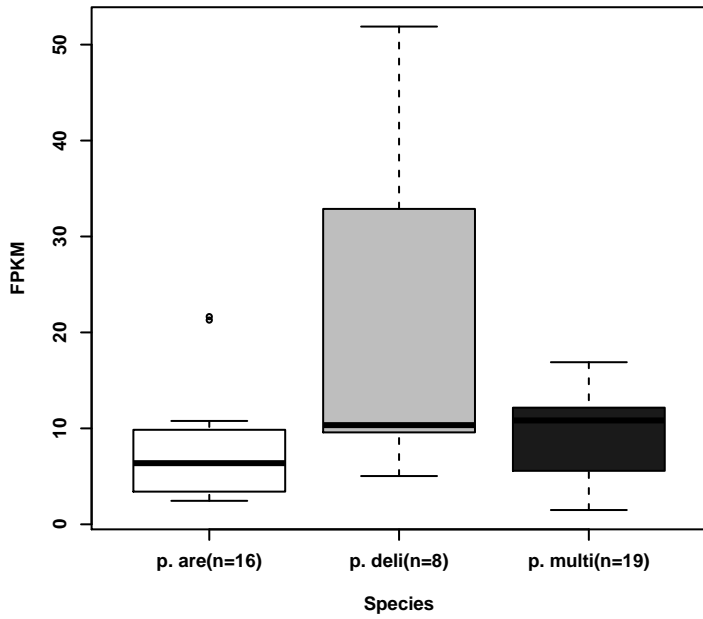


chorismate biosynthesis

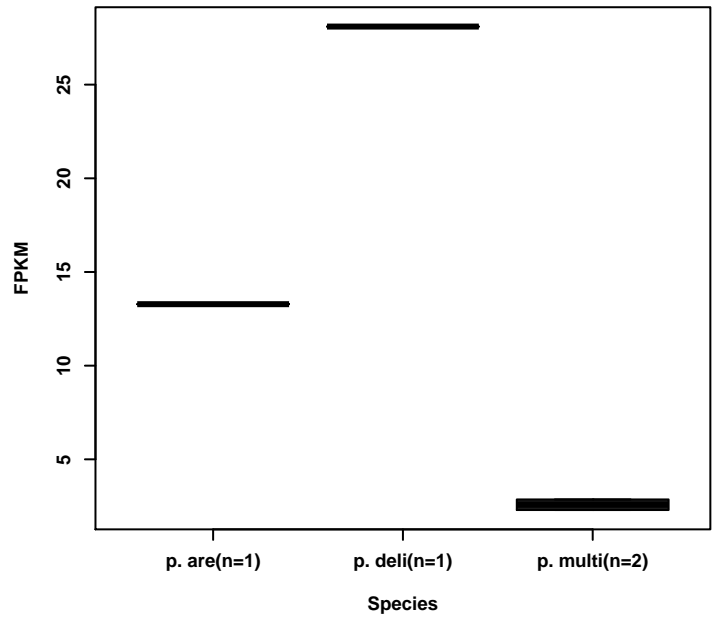




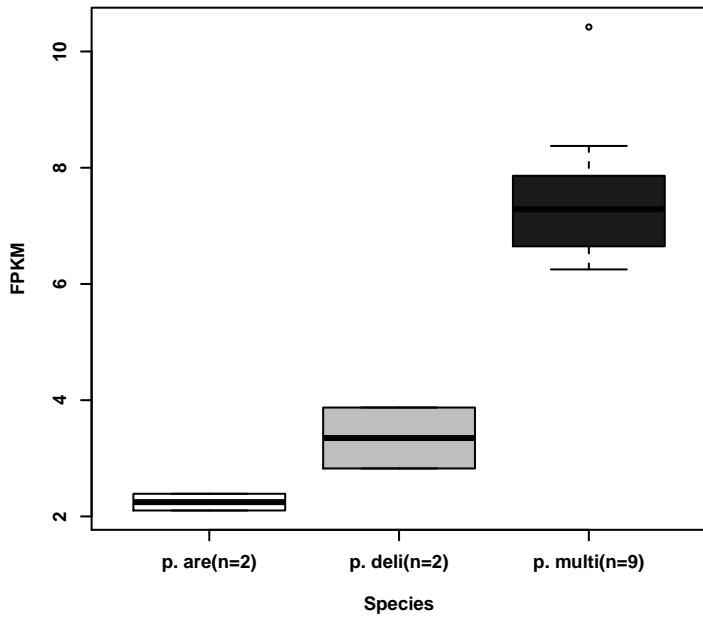
coenzyme A biosynthesis



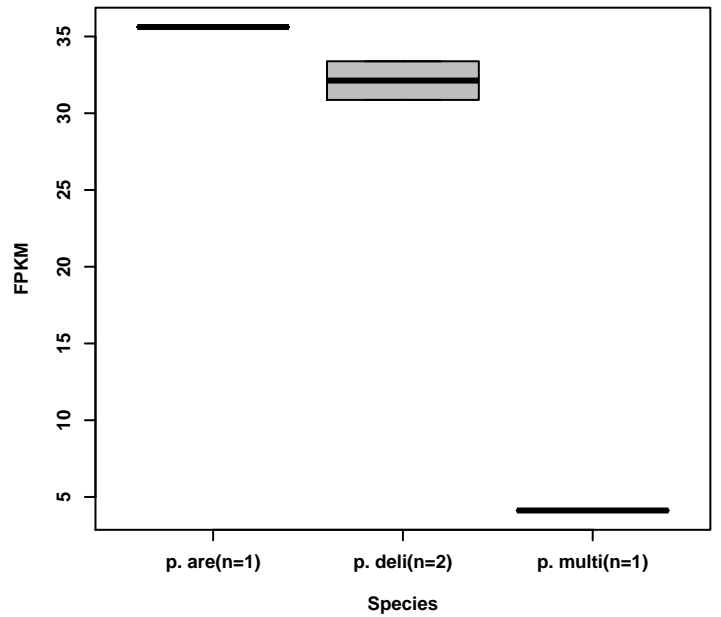
CTP biosynthesis via de novo pathway



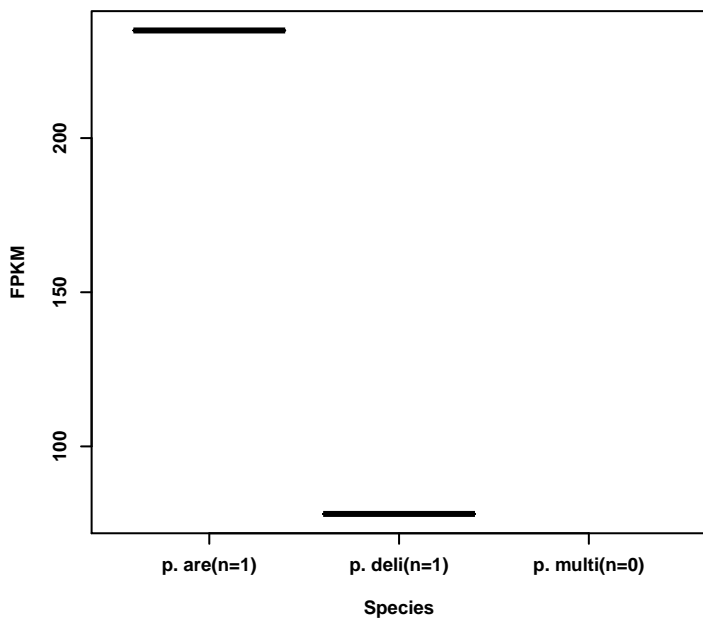
CTP biosynthesis via salvage pathway



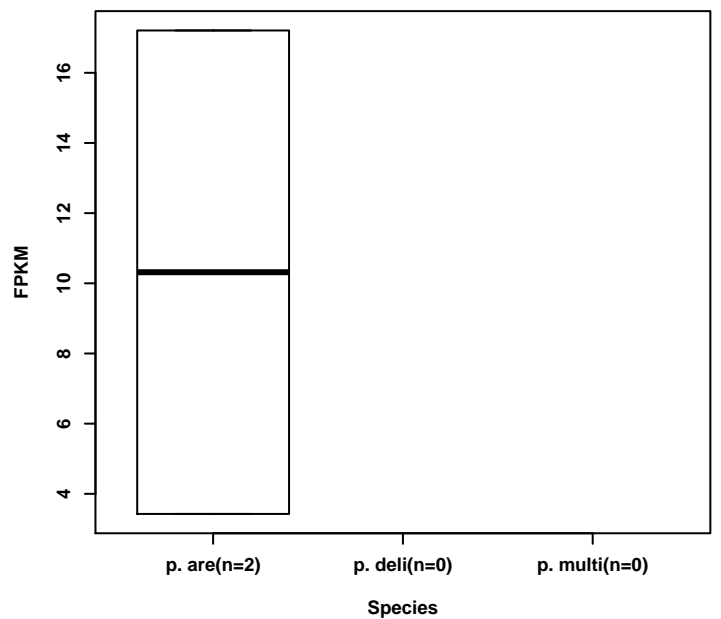
D-alanine biosynthesis



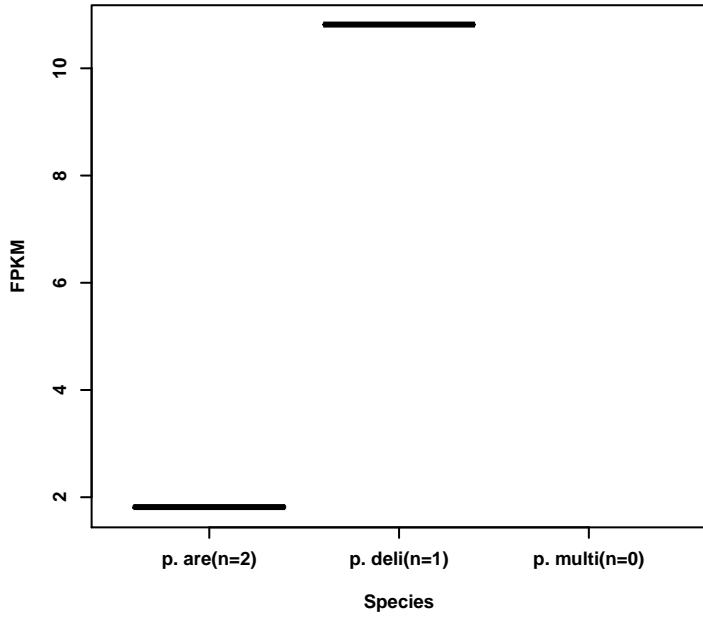
D-galactonate degradation



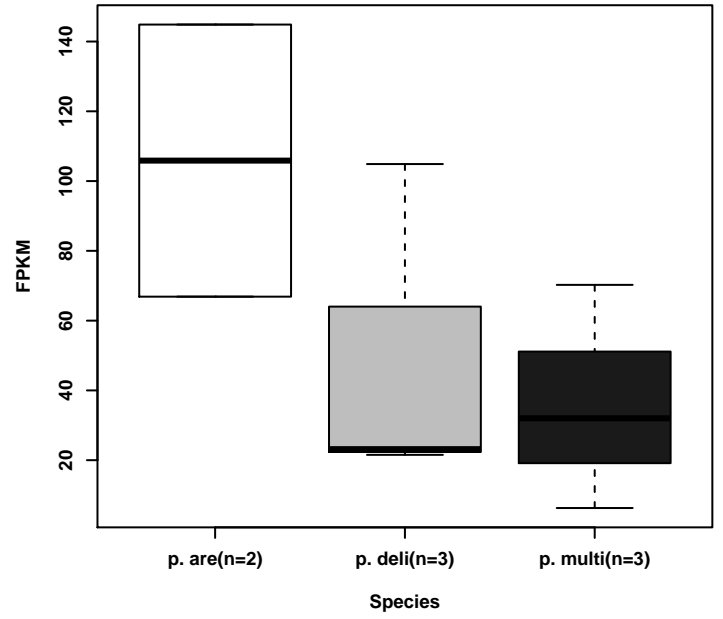
D-gluconate biosynthesis



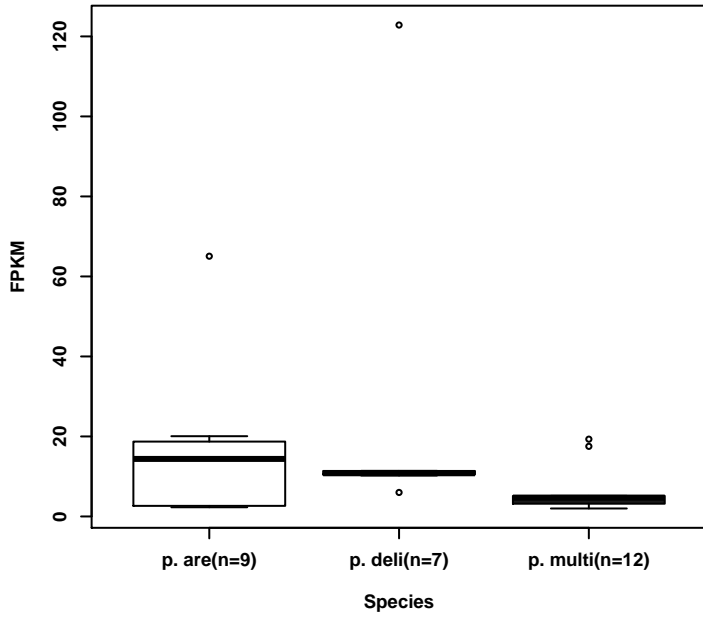
D-gluconate degradation



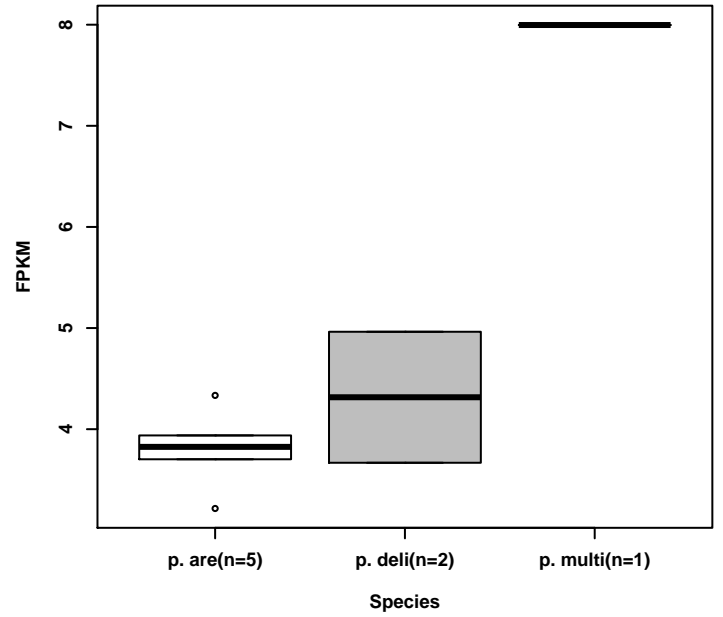
dimethylallyl diphosphate biosynthesis



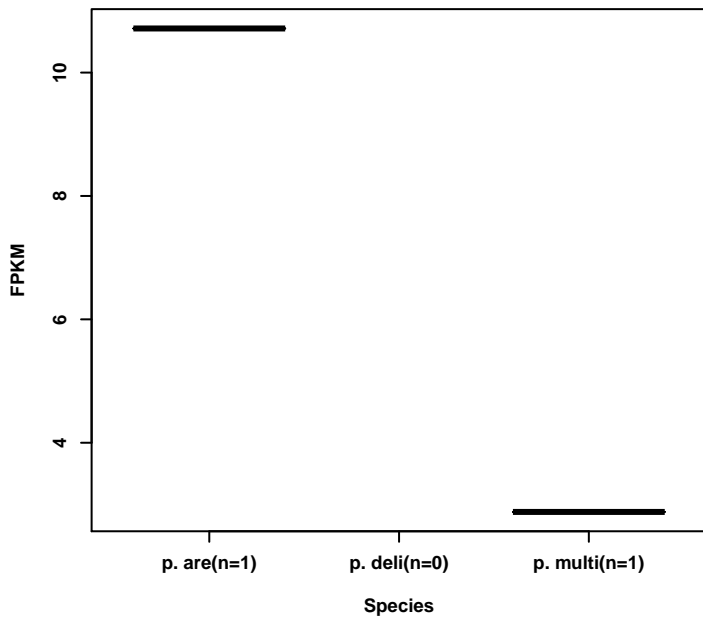
DNA replication



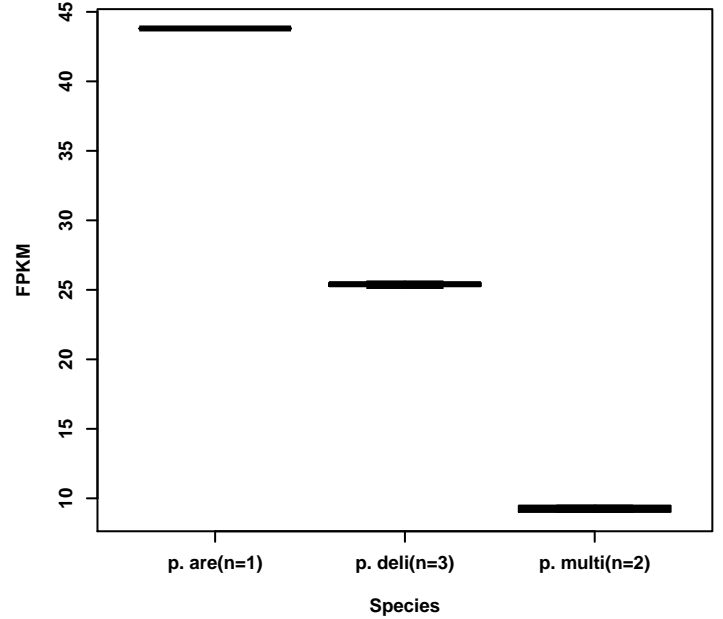
D-ribose degradation



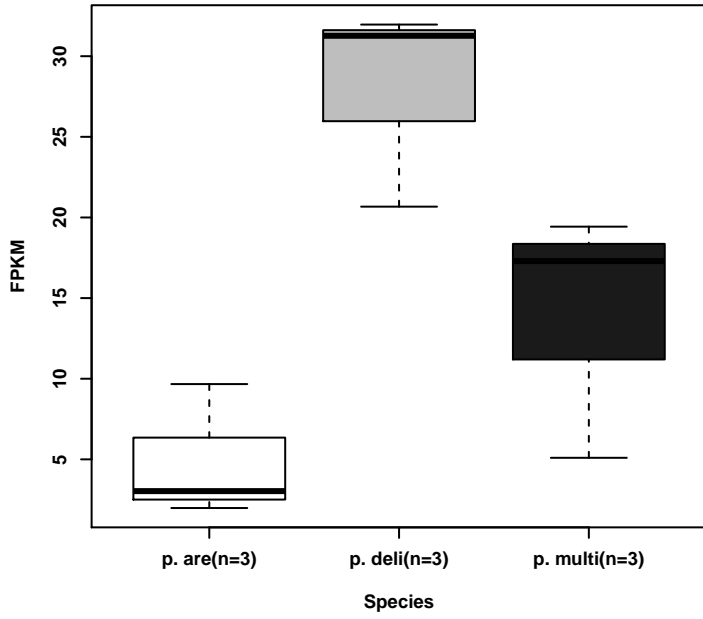
dTTP biosynthesis



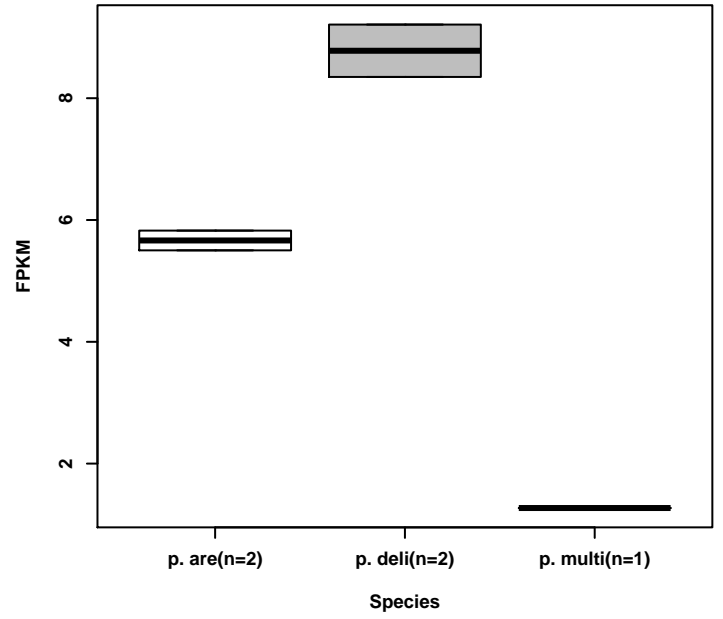
dUMP biosynthesis



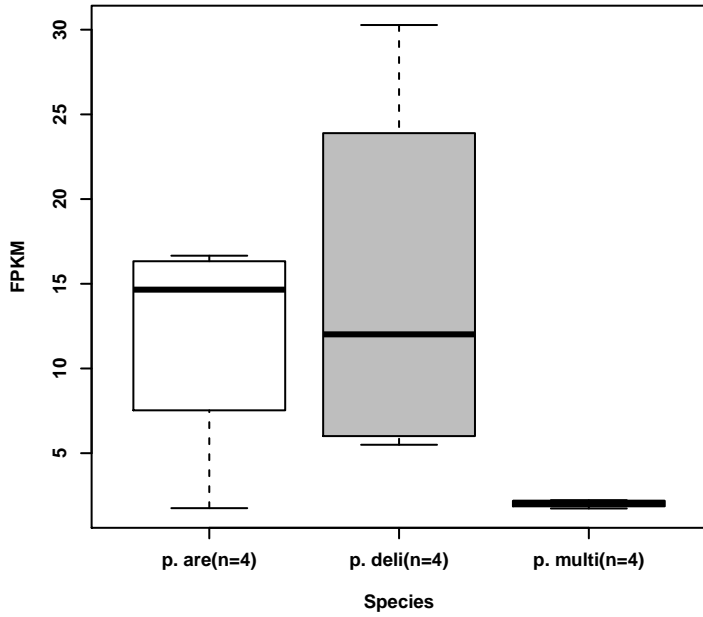
**eIF5A hypusination**



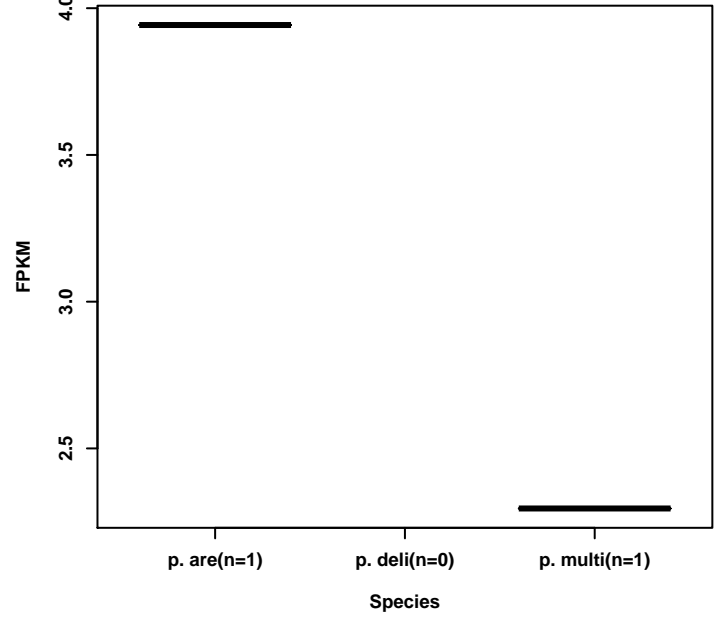
**ether lipid biosynthesis**



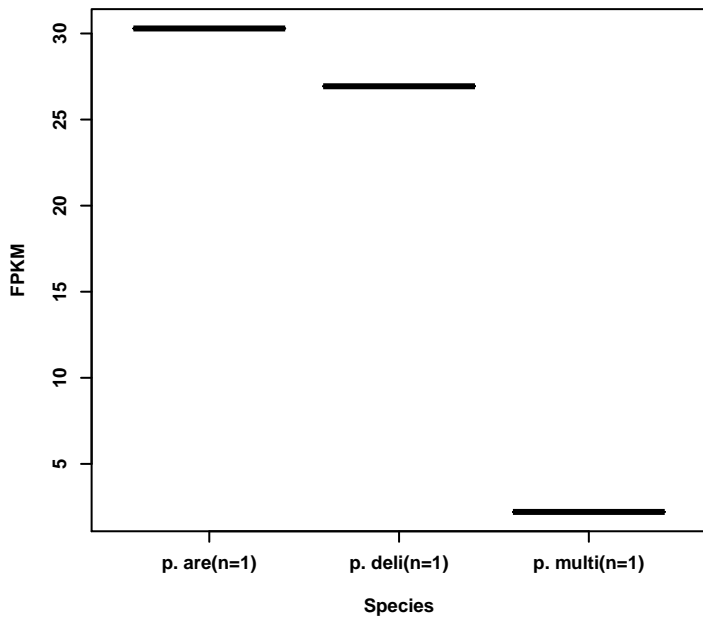
**ethylene biosynthesis via S-adenosyl-L-methionine**



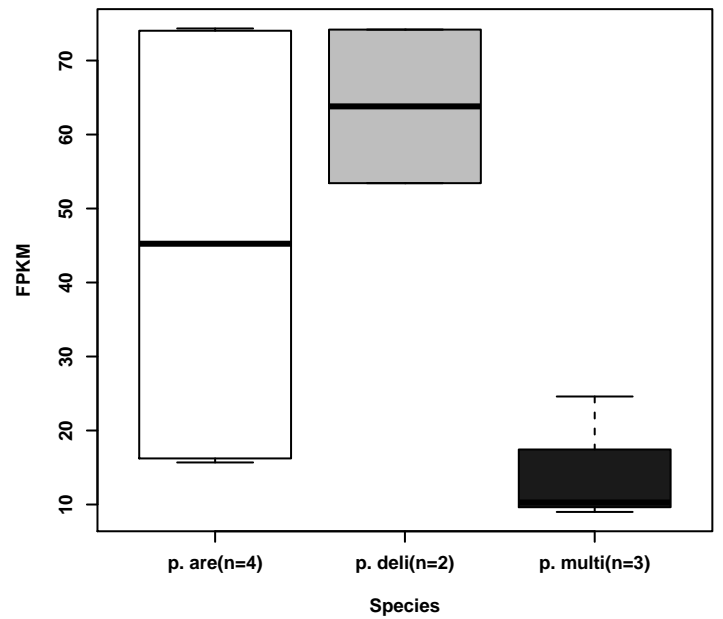
**exopolysaccharide biosynthesis**



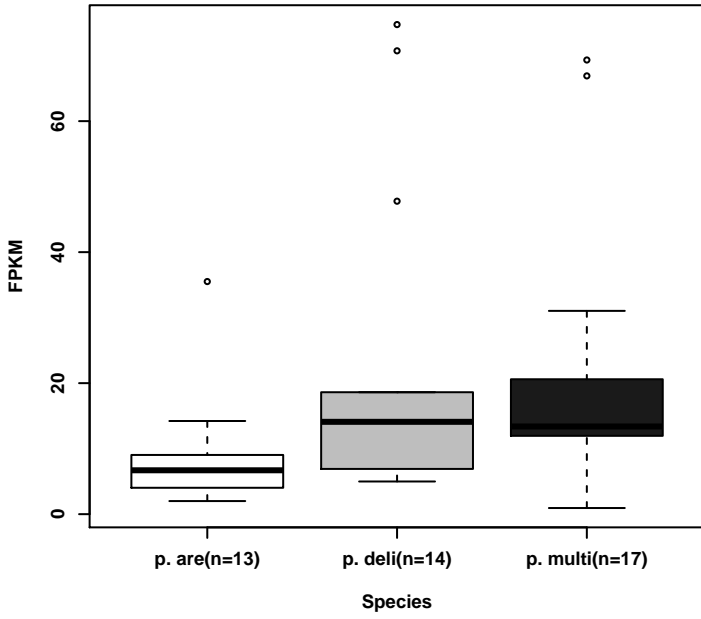
**FAD biosynthesis**



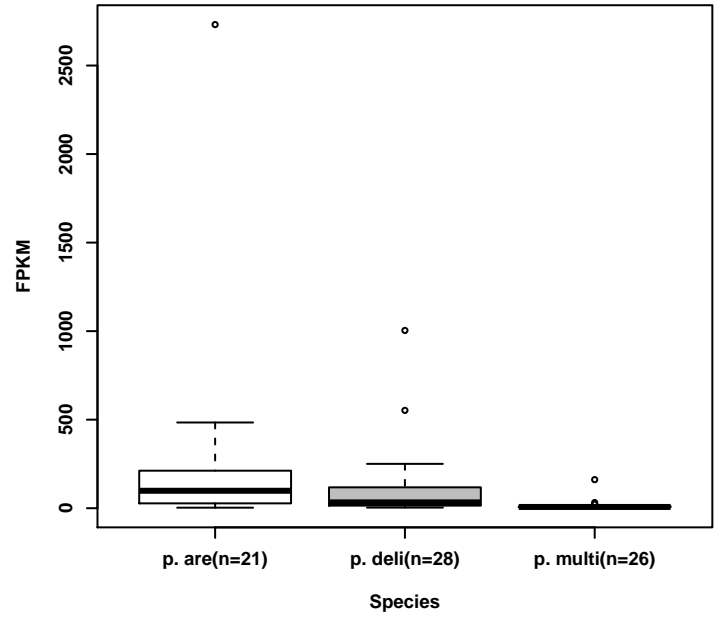
**farnesyl diphosphate biosynthesis**



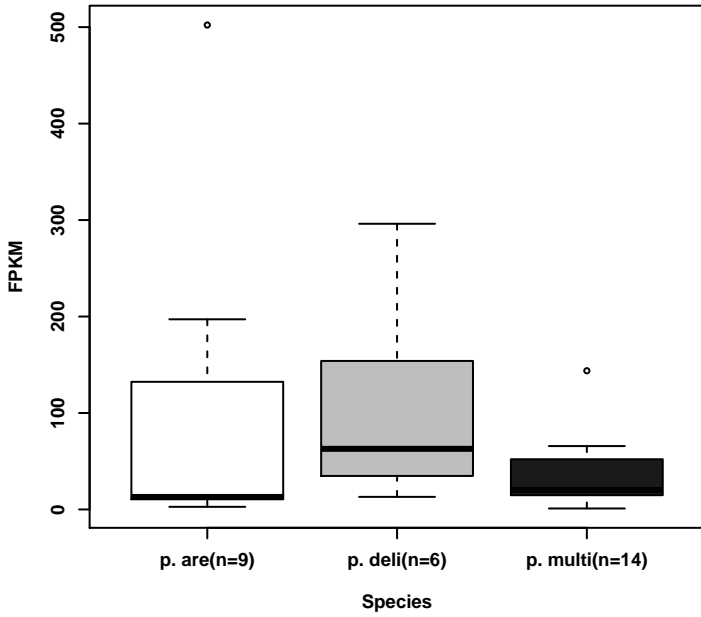
**fatty acid beta-oxidation**



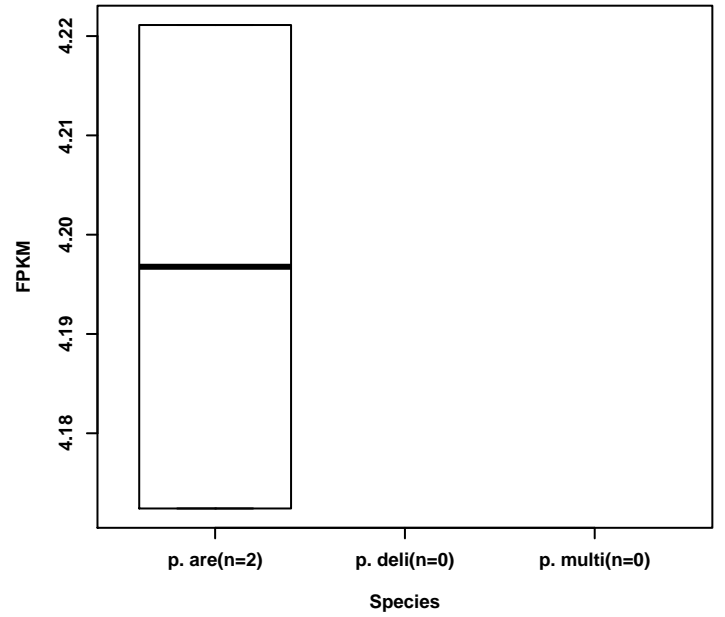
**fatty acid biosynthesis**



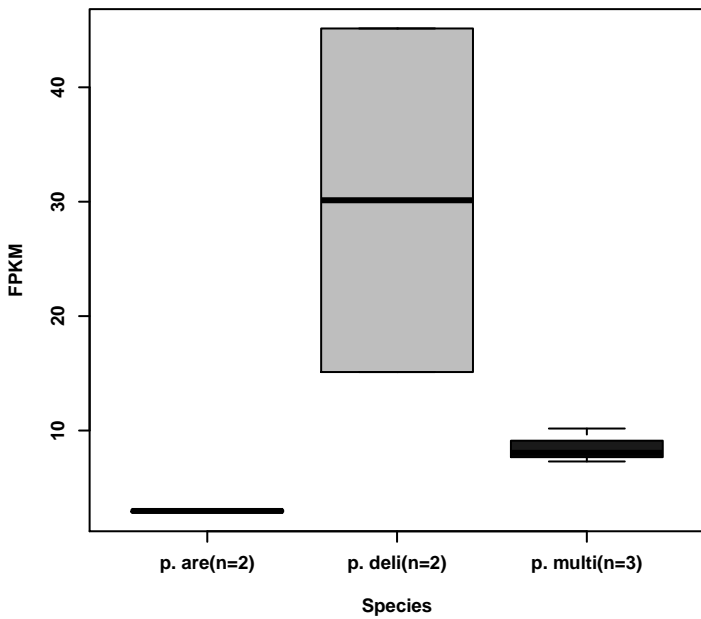
**fatty acid metabolism**



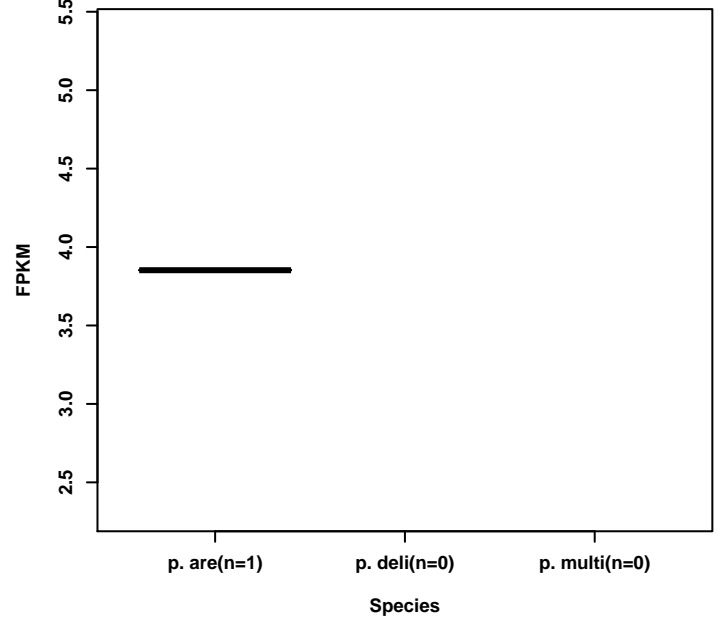
**flavonoid biosynthesis**



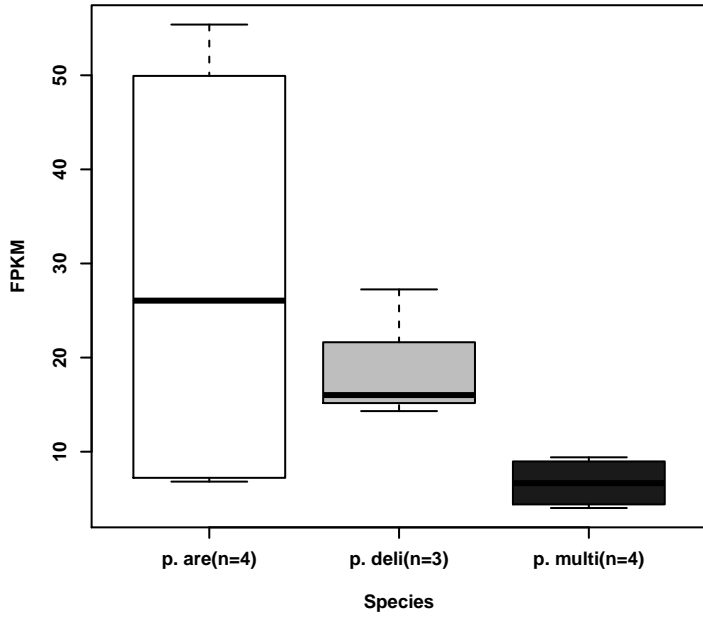
**FMN biosynthesis**



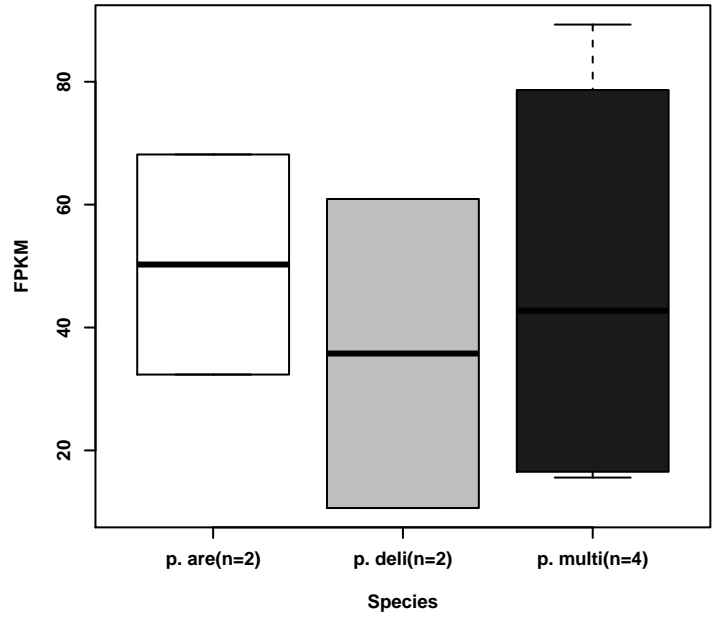
**formaldehyde assimilation via serine pathway**



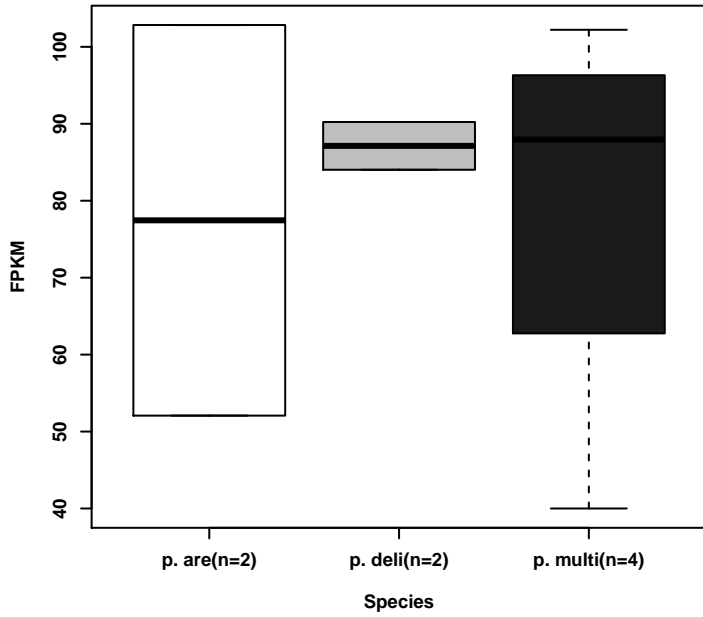
galactose metabolism



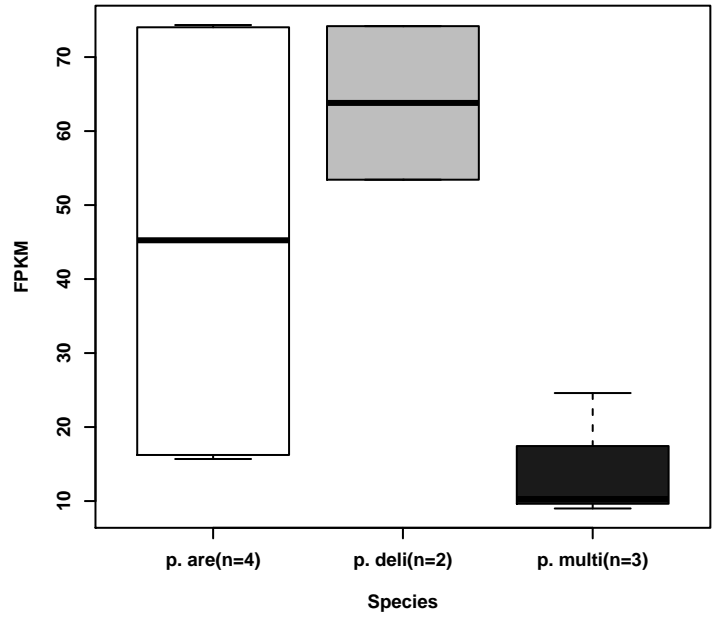
GDP- $\alpha$ -D-mannose biosynthesis



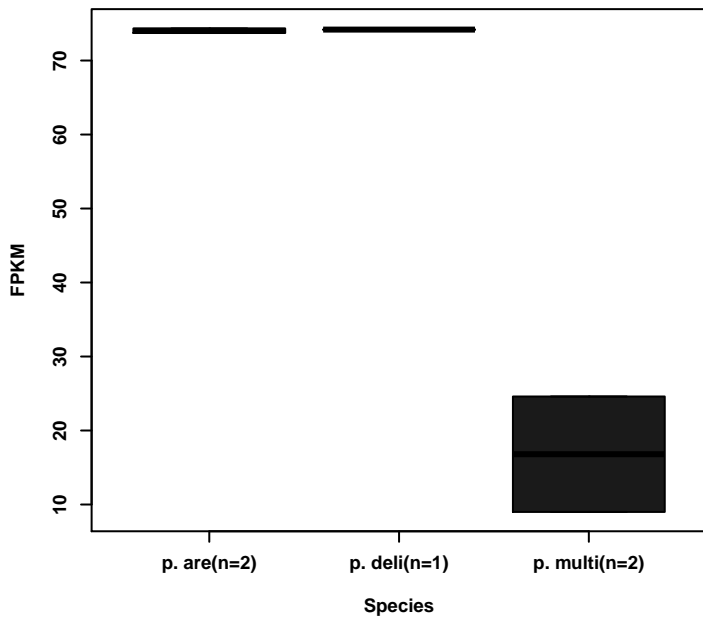
GDP-L-fucose biosynthesis via de novo pathway



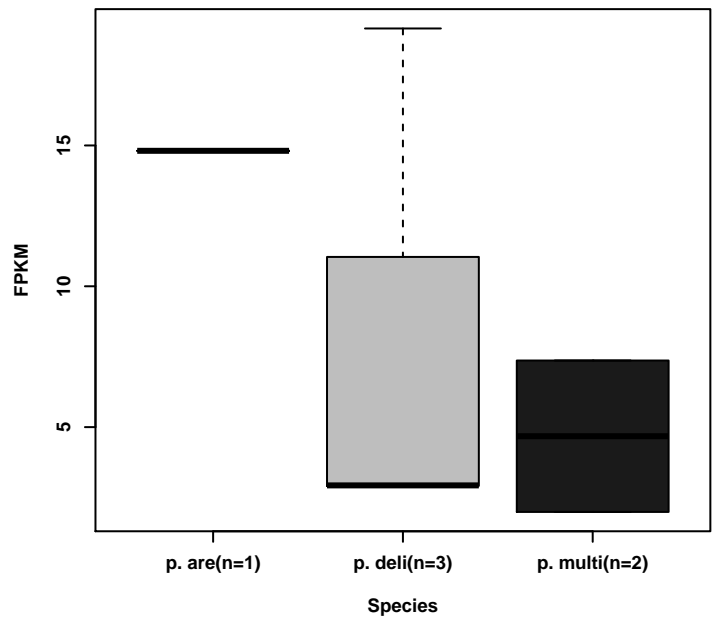
geranyl diphosphate biosynthesis



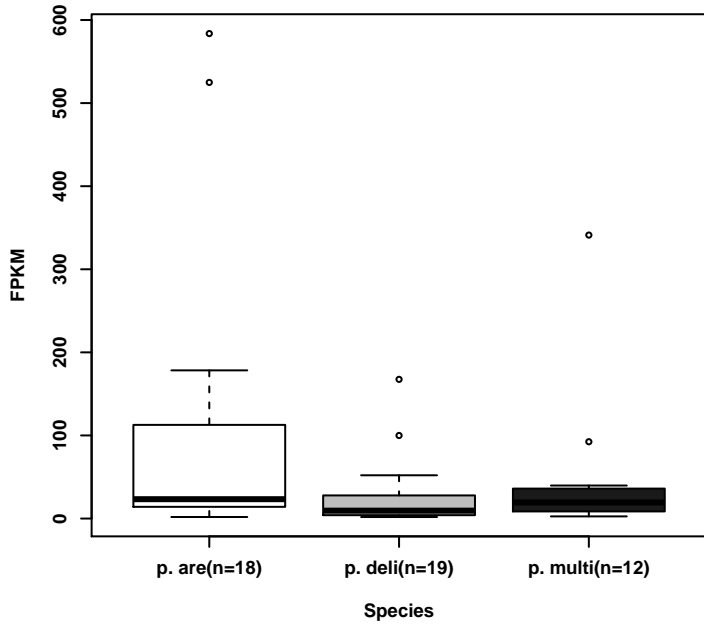
geranylgeranyl diphosphate biosynthesis



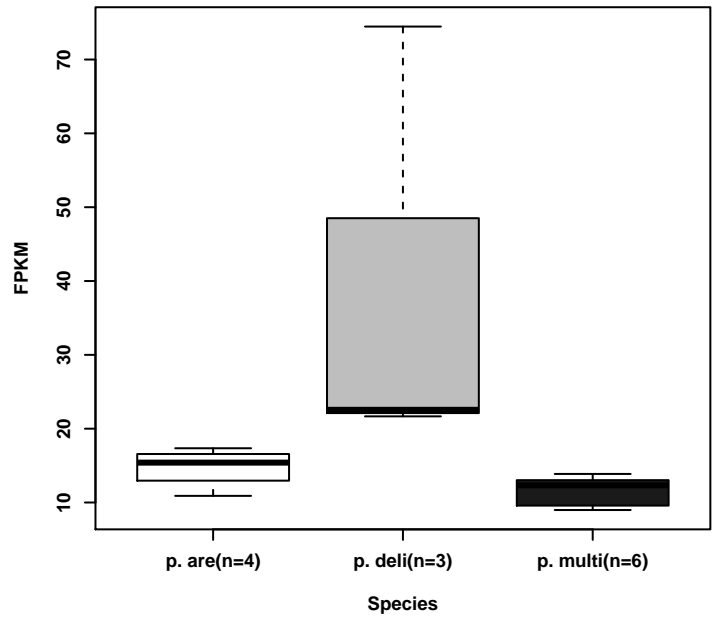
gibberellin biosynthesis



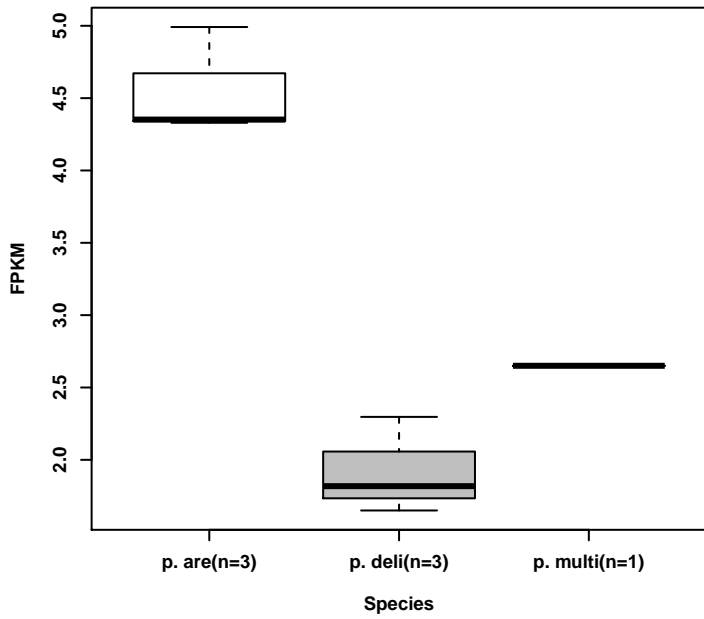
**gluconeogenesis**



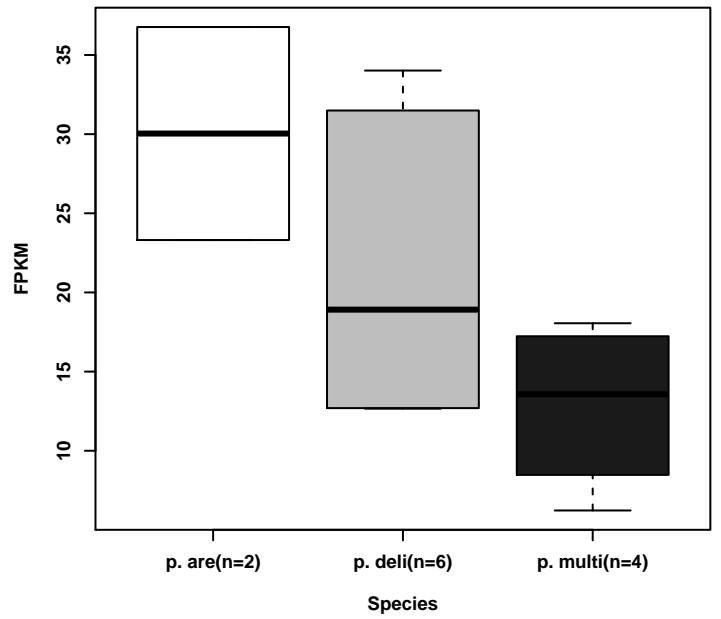
**glutathione biosynthesis**



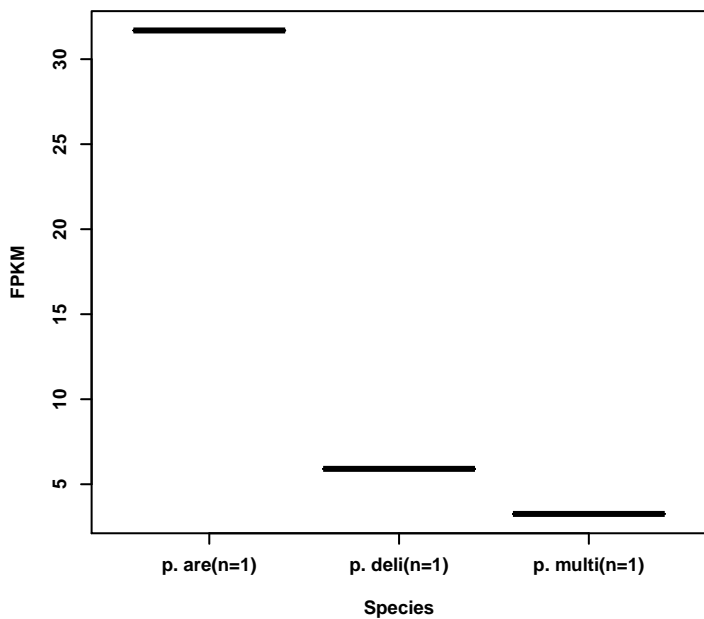
**glutathione metabolism**



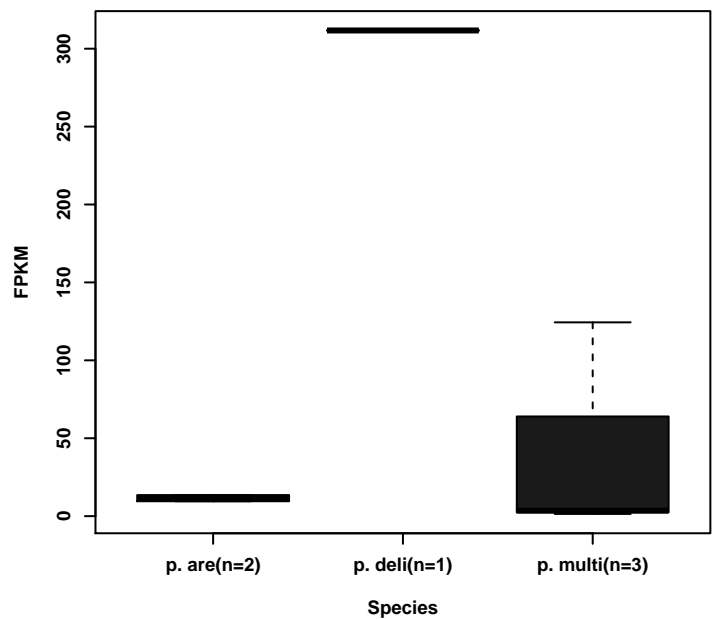
**glycerol degradation via glycerol kinase pathway**



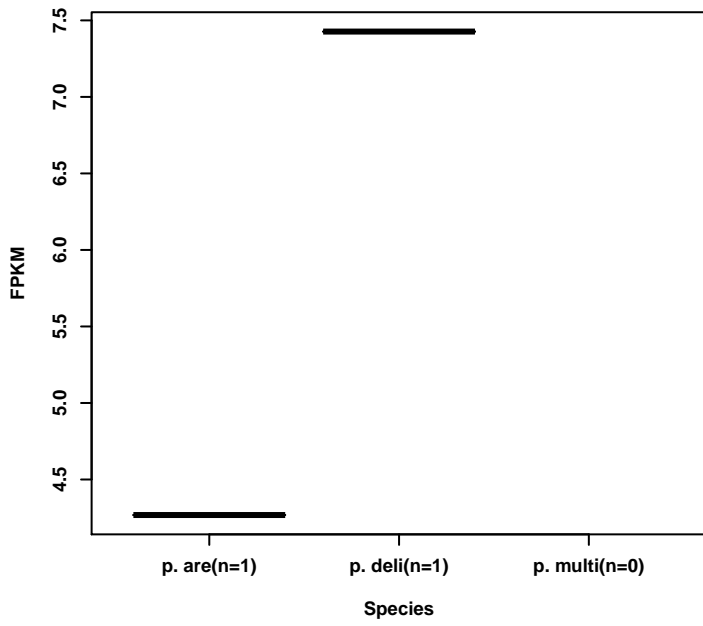
**glycerophospholipid metabolism**



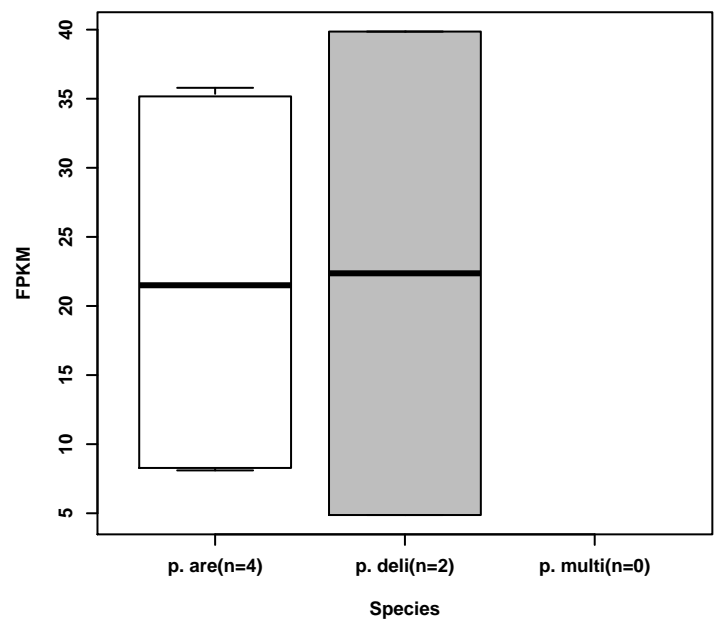
**glycine biosynthesis**



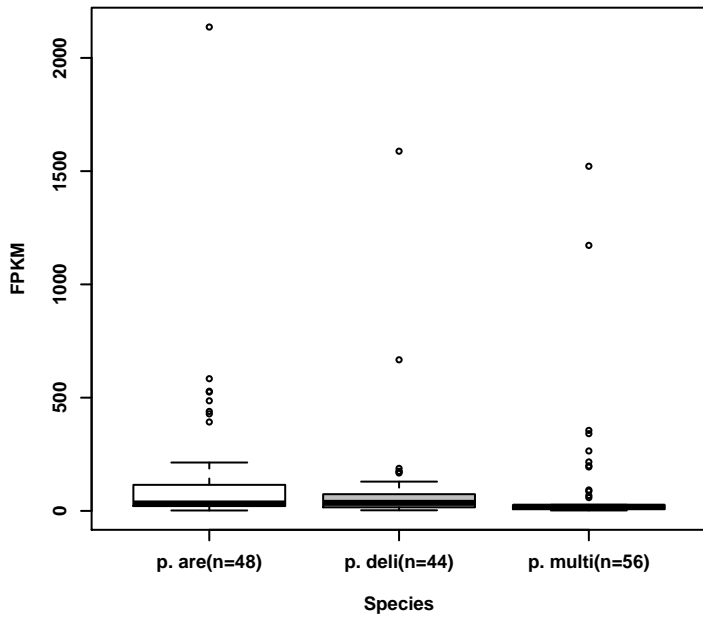
**glycolate biosynthesis**



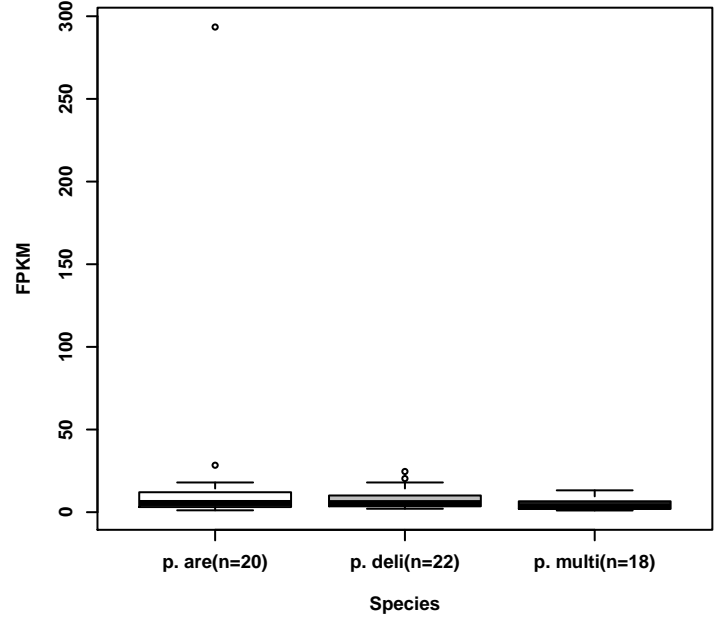
**glycolate degradation**



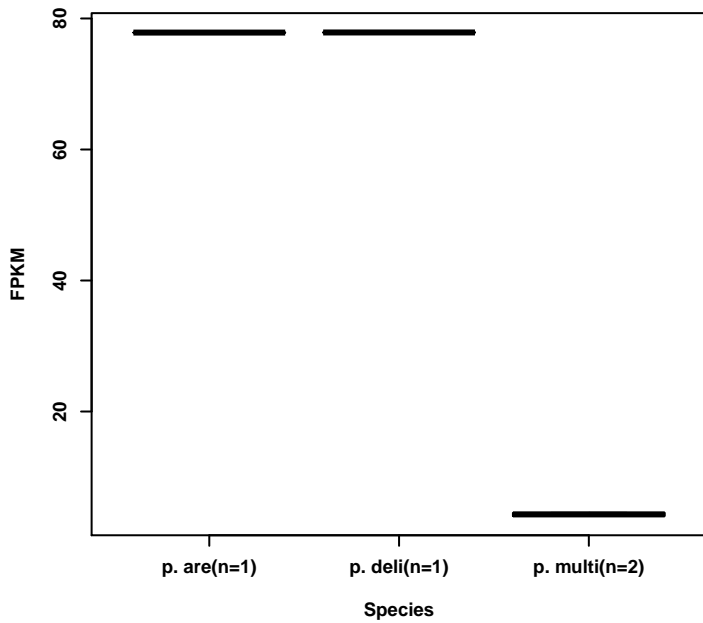
**glycolysis**



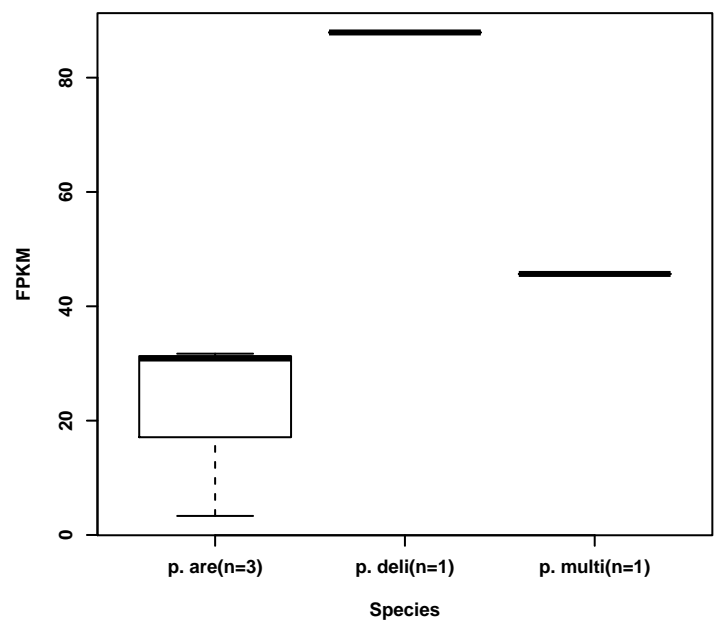
**glycosylphosphatidylinositol-anchor biosynthesis**



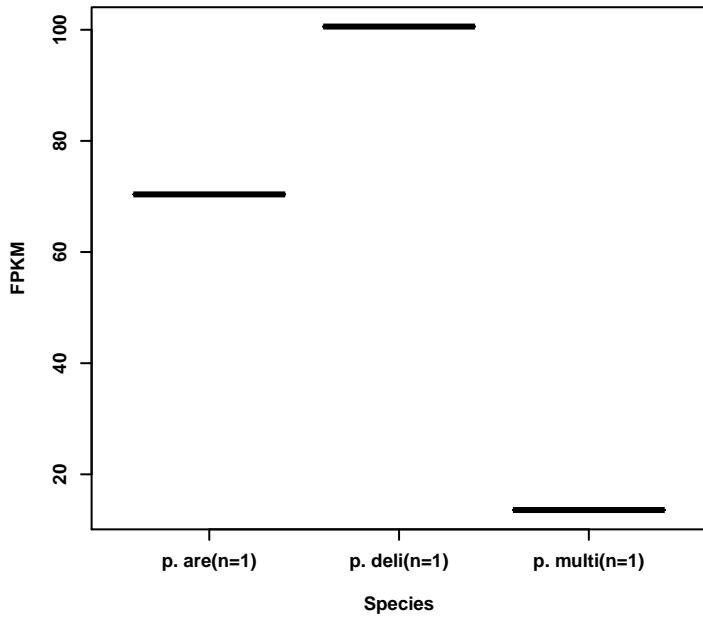
**GMP biosynthesis**



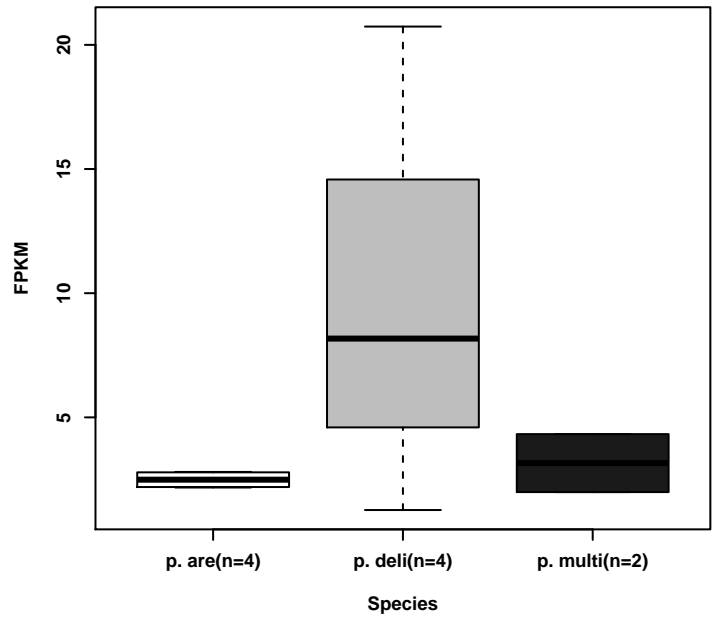
**gramicidin S biosynthesis**



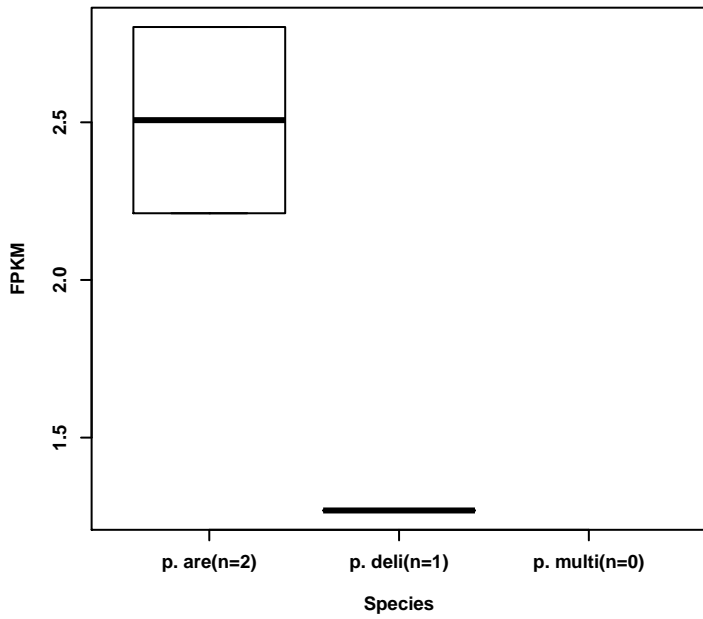
guanine degradation



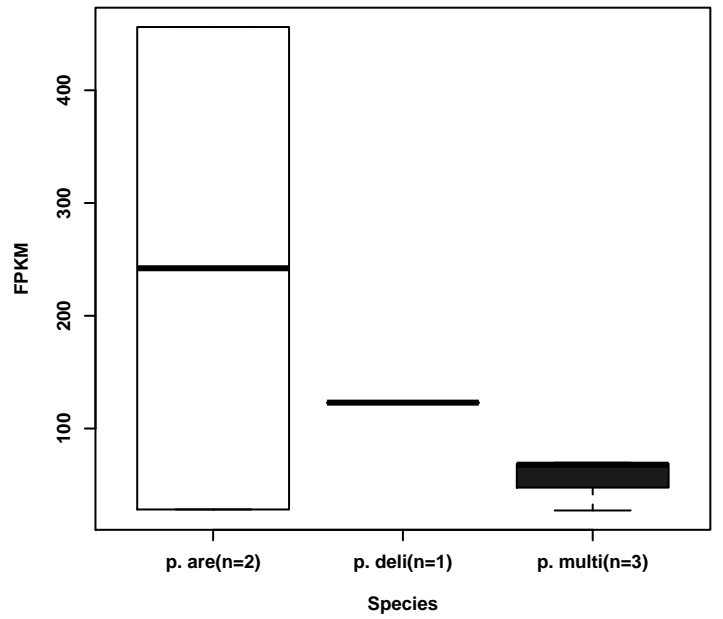
heparan sulfate biosynthesis



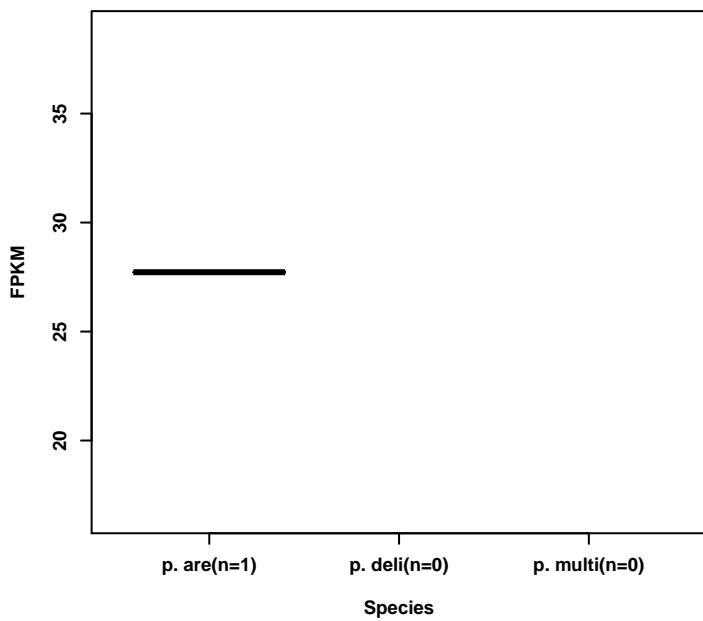
heparin biosynthesis



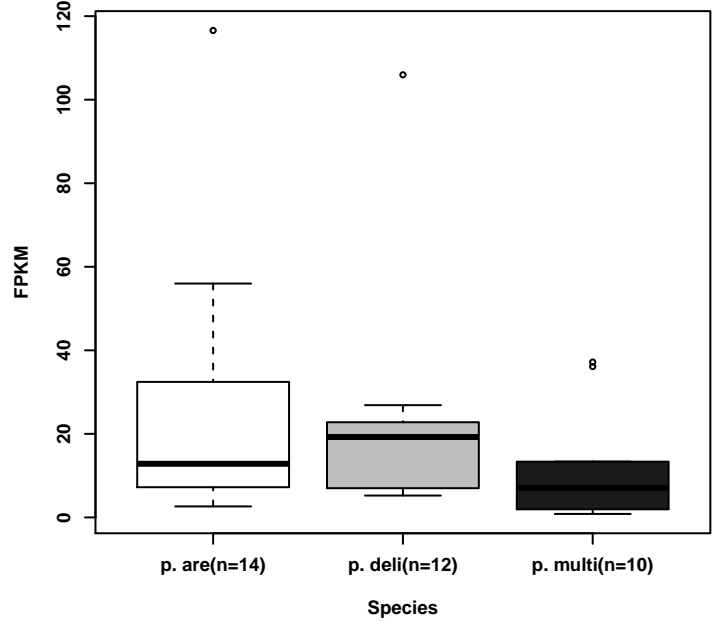
hydrogen sulfide biosynthesis



hydroperoxy eicosatetraenoic acid biosynthesis

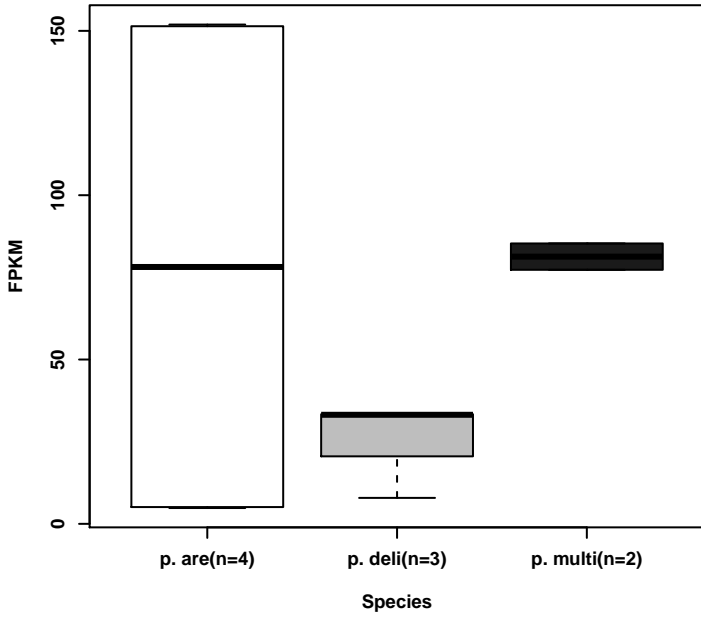


IMP biosynthesis via de novo pathway

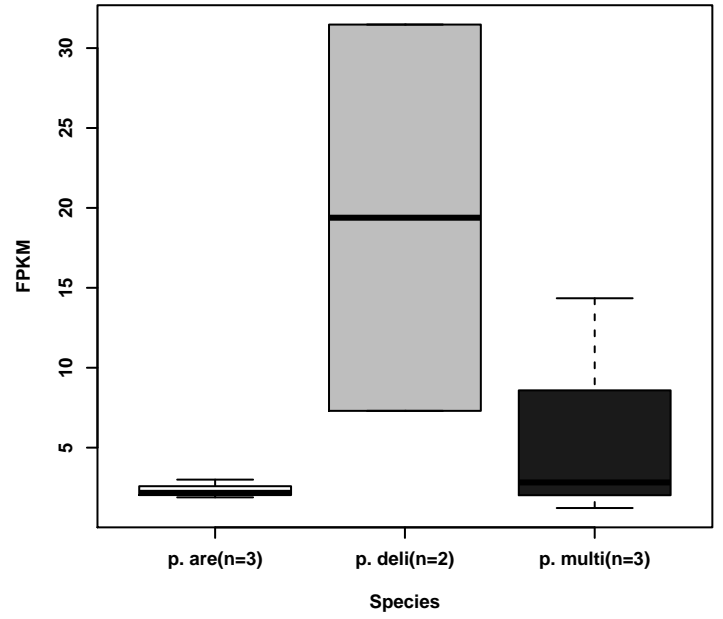




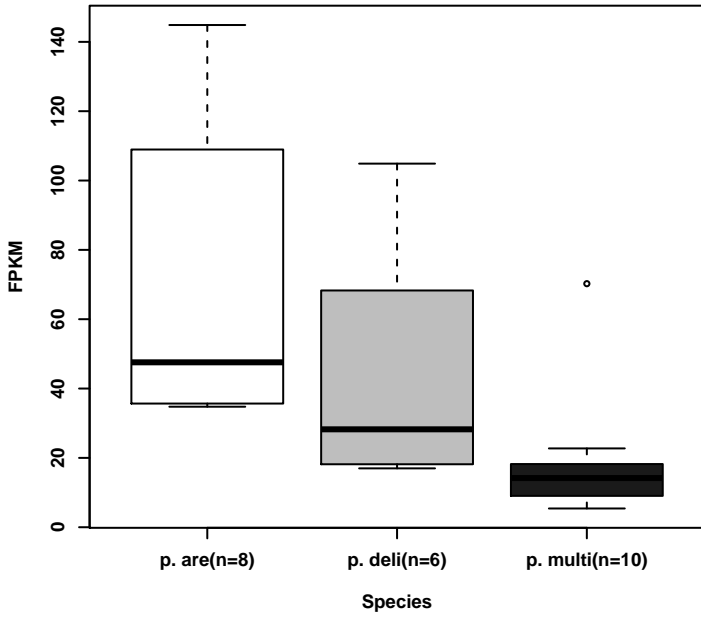
**IMP biosynthesis via salvage pathway**



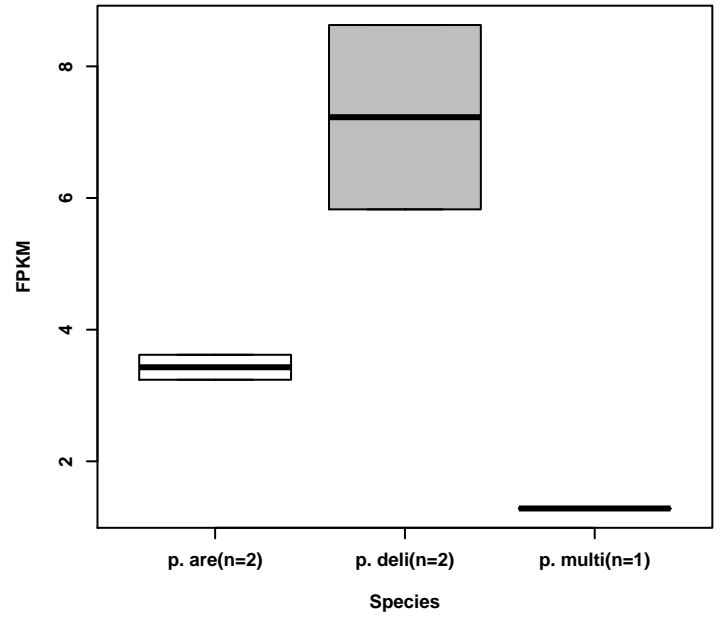
**iron-sulfur cluster biosynthesis**



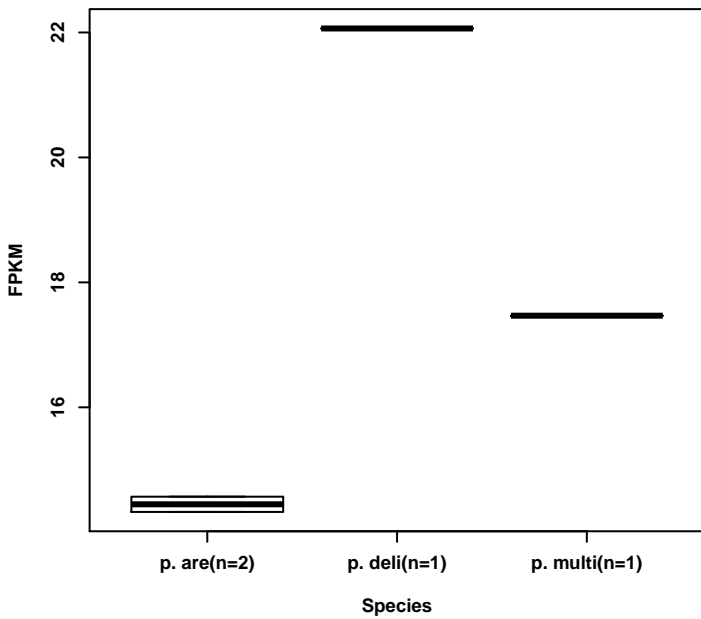
**isopentenyl diphosphate biosynthesis via DXP pathway**



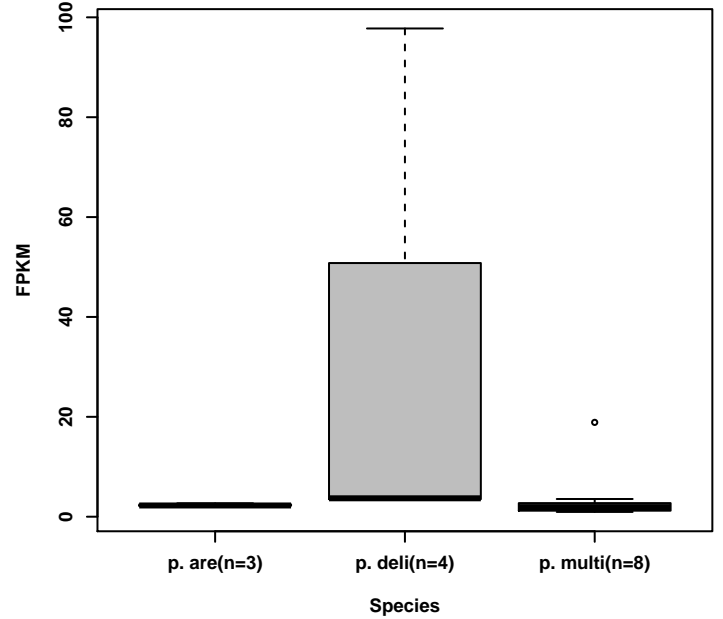
**isopentenyl diphosphate biosynthesis via mevalonate pathway**



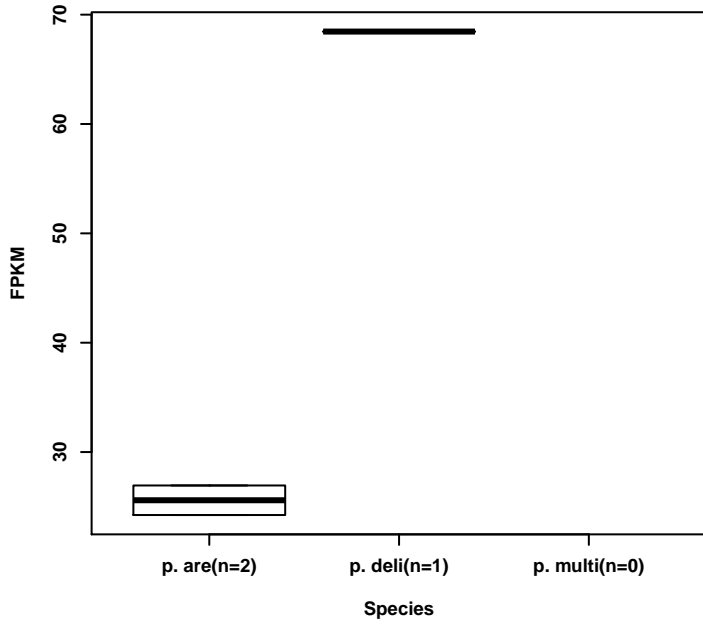
**kanamycin biosynthesis**



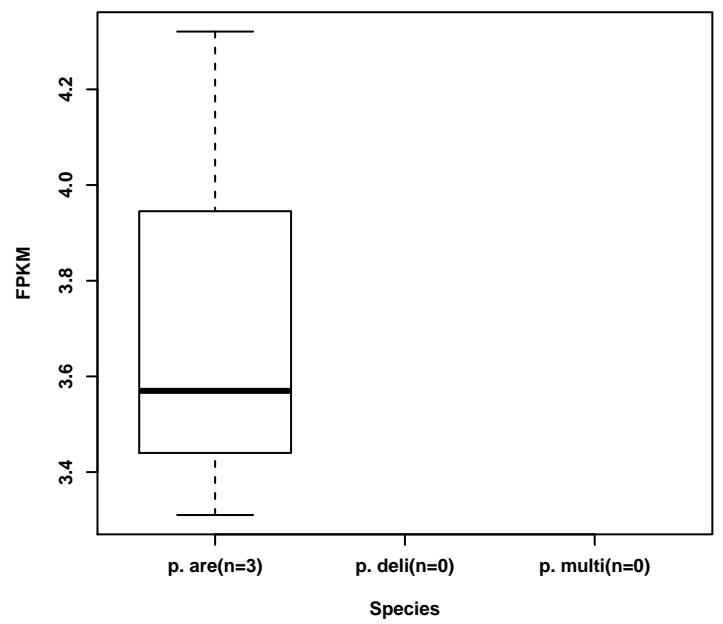
**L-alanine degradation via transaminase pathway**



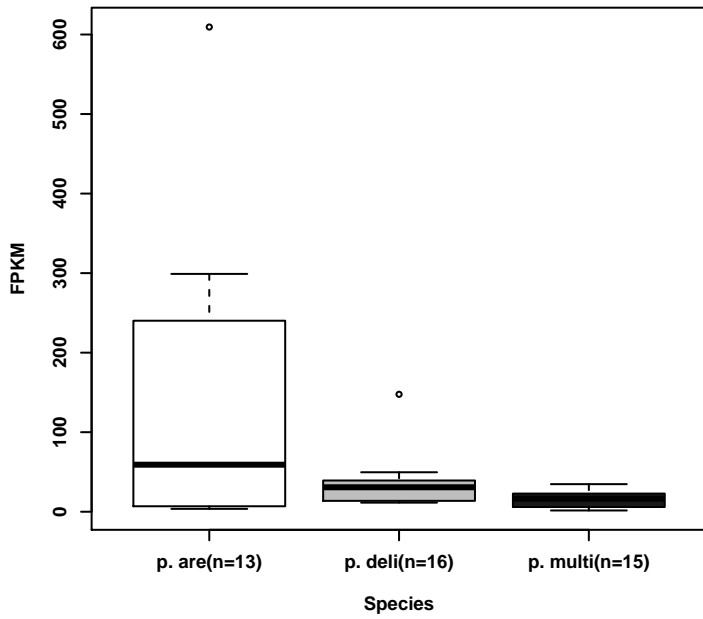
**Ianosterol biosynthesis**



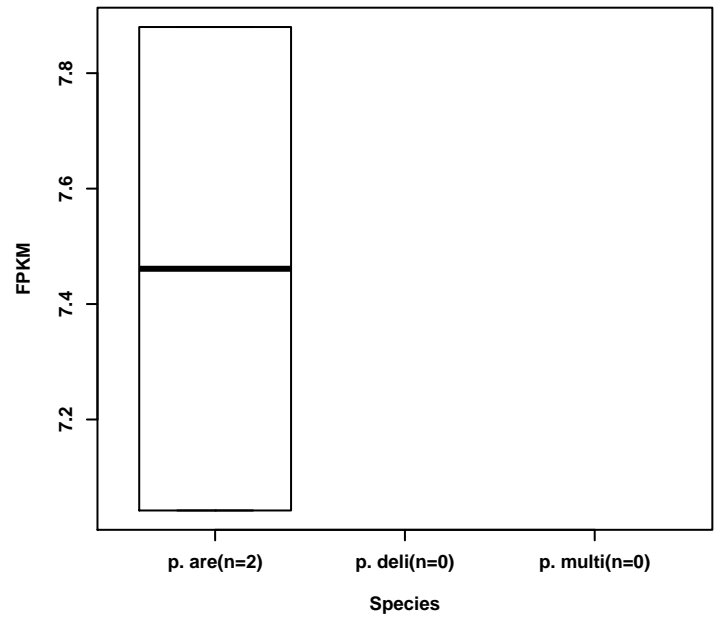
**L-arabinose degradation via L-ribulose**



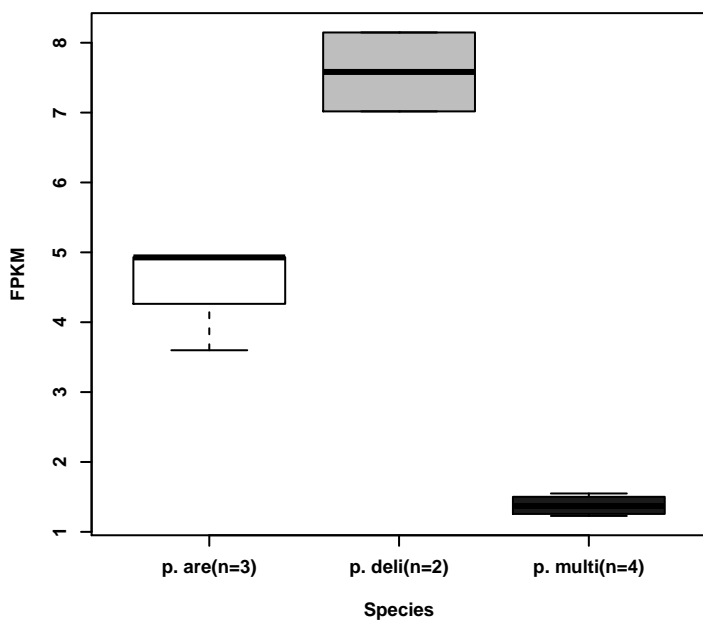
**L-arginine biosynthesis**



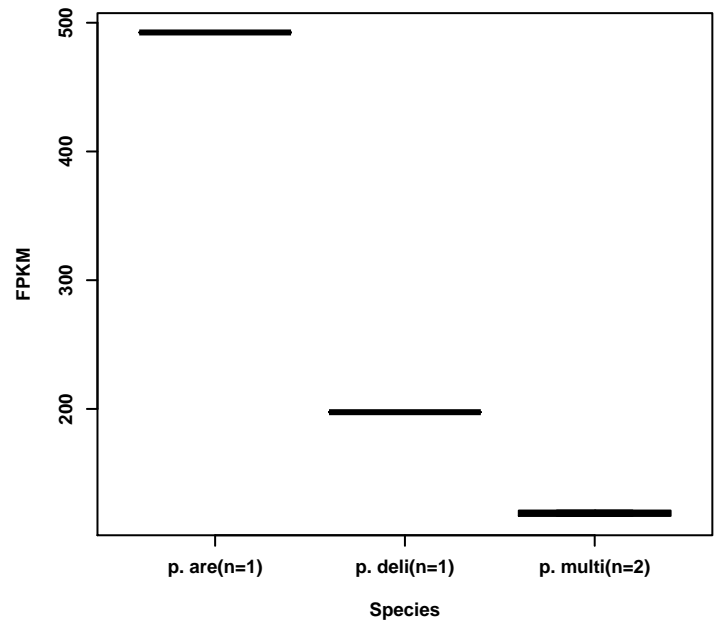
**L-arginine degradation [regulation]**



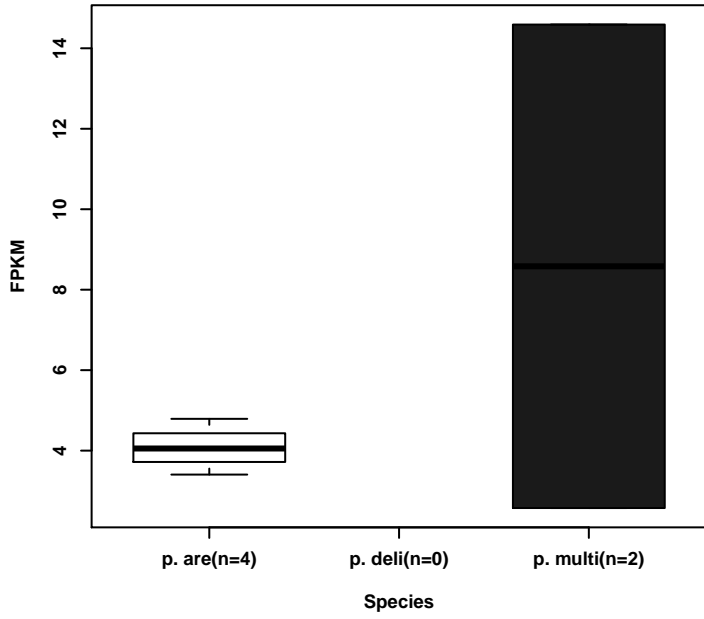
**L-ascorbate biosynthesis**



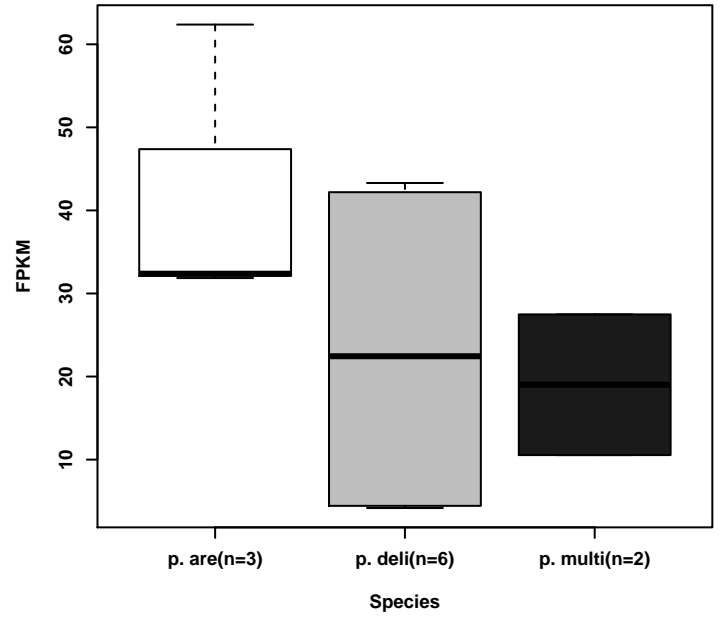
**L-ascorbate biosynthesis via GDP-alpha-D-mannose pathwa**



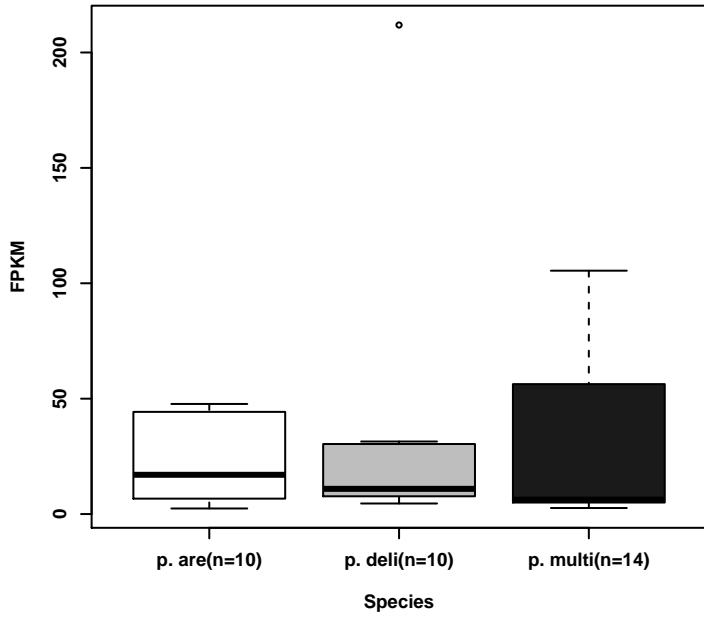
L-ascorbate biosynthesis via UDP- $\alpha$ -D-glucuronate pathw



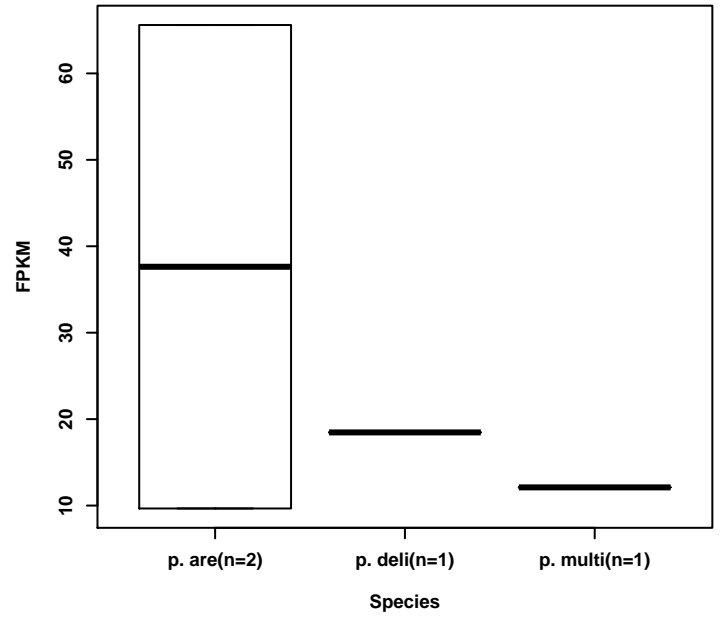
L-asparagine biosynthesis



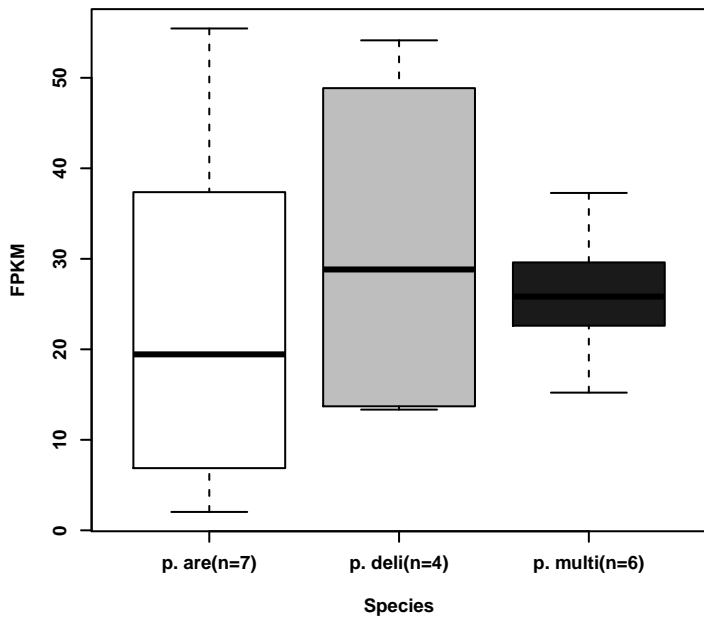
L-cysteine biosynthesis



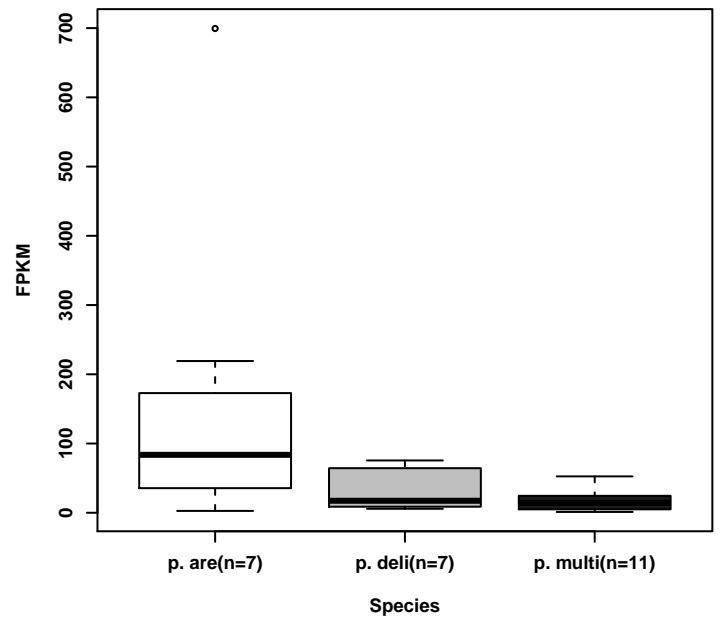
leukotriene B4 biosynthesis



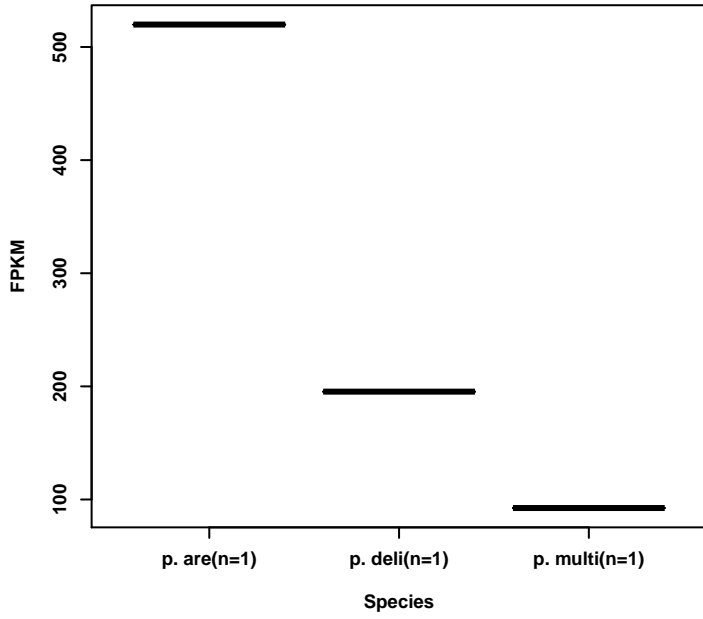
L-glutamate biosynthesis via GLT pathway



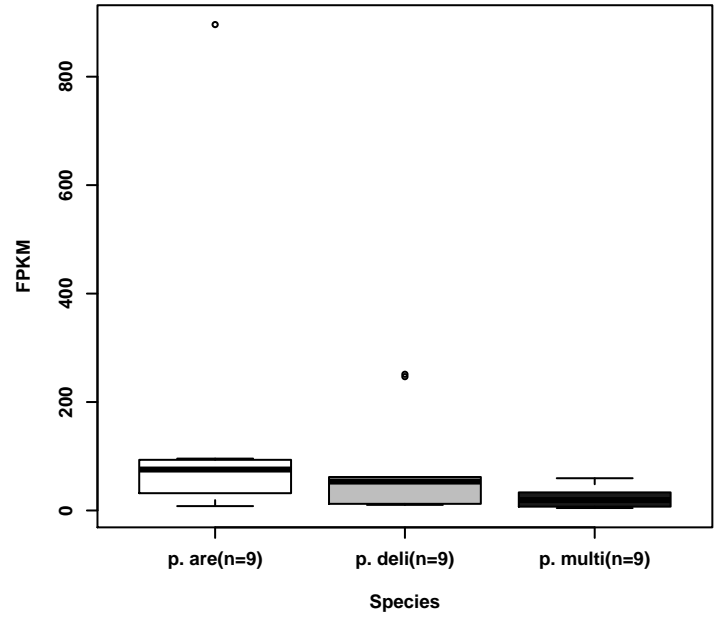
L-histidine biosynthesis



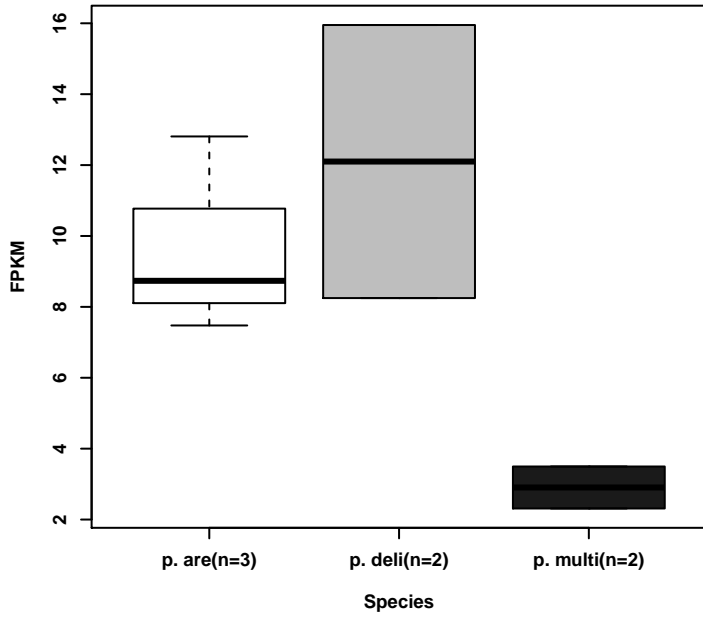
L-homocysteine biosynthesis



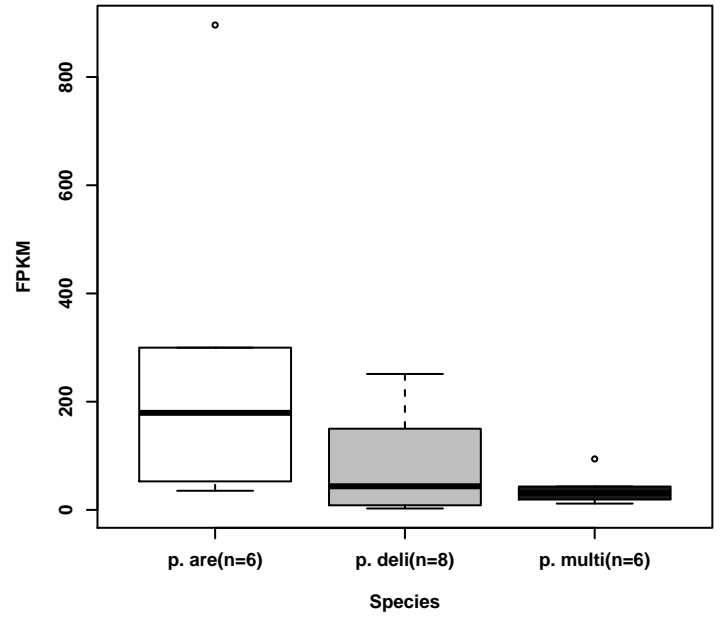
L-isoleucine biosynthesis



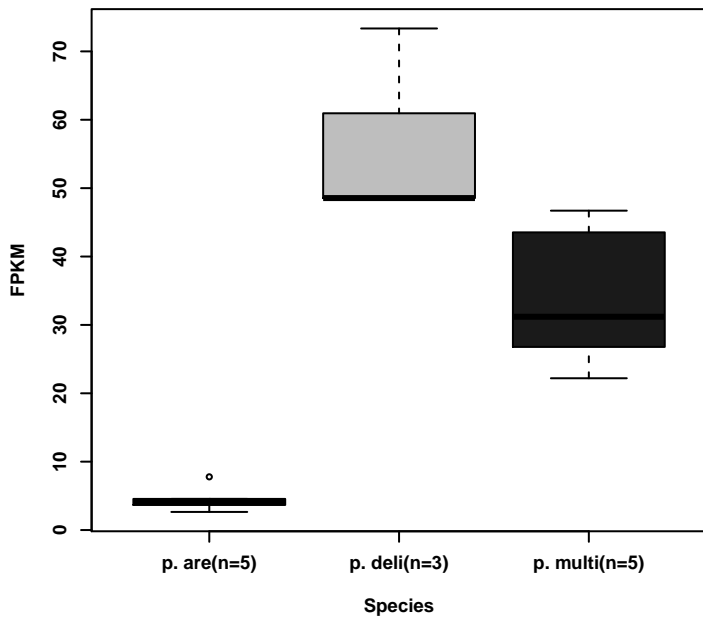
L-kynurenine degradation



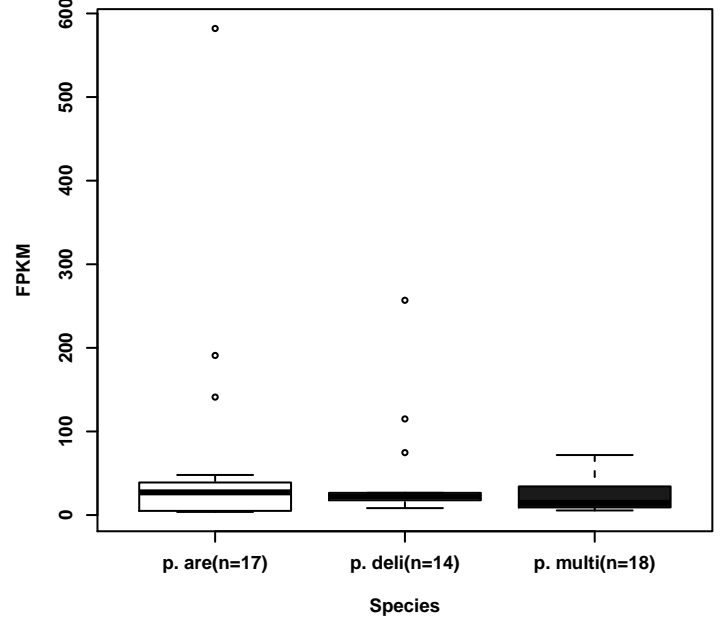
L-leucine biosynthesis



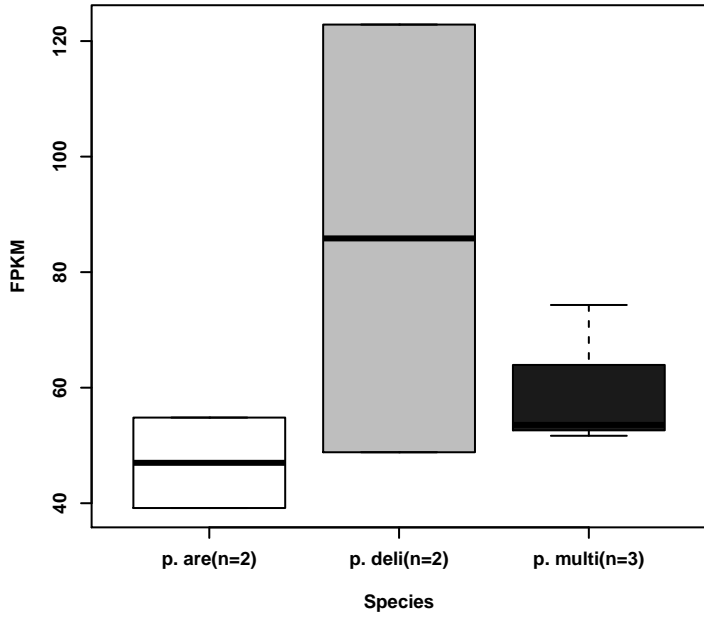
L-leucine degradation



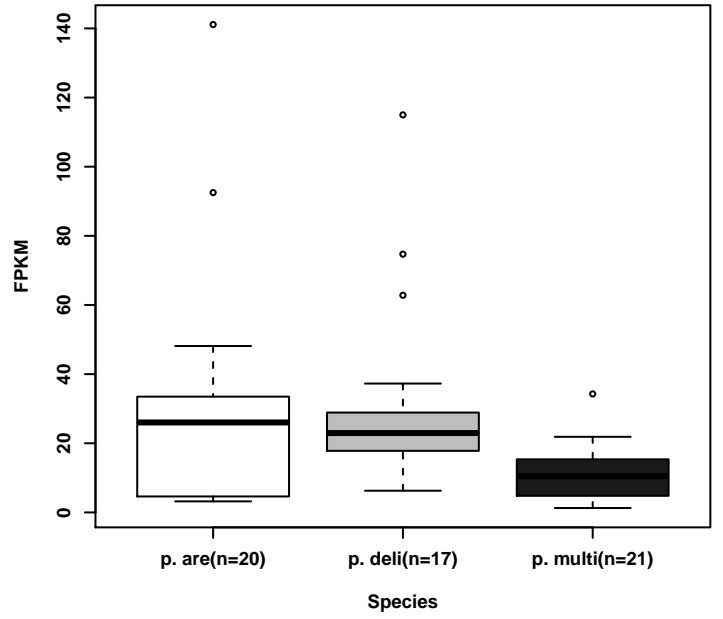
L-lysine biosynthesis via DAP pathway



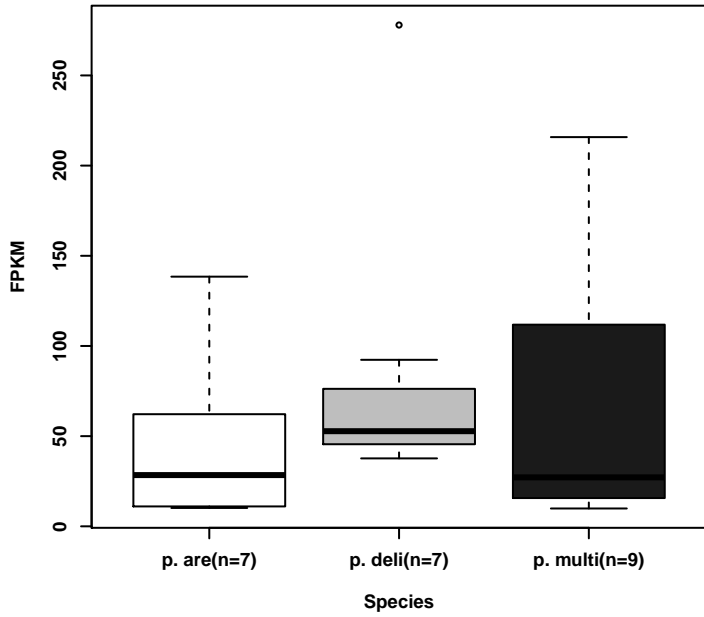
L-lysine degradation via saccharopine pathway



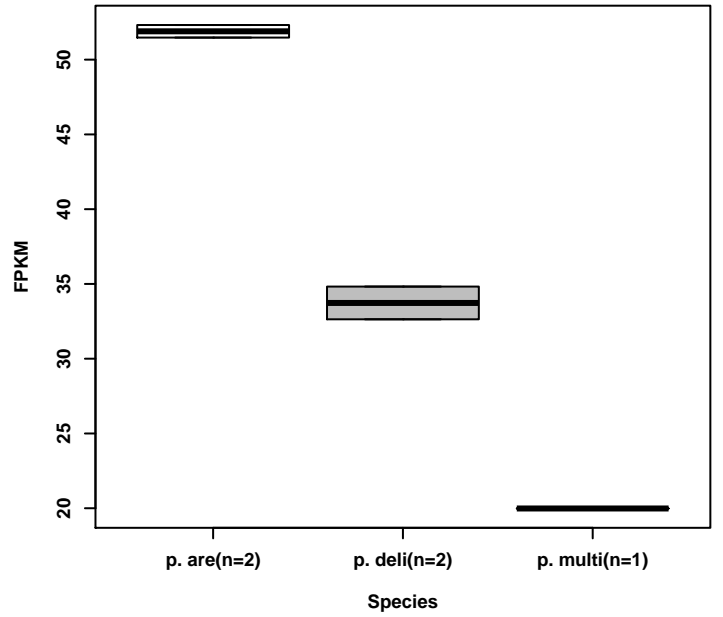
L-methionine biosynthesis via de novo pathway



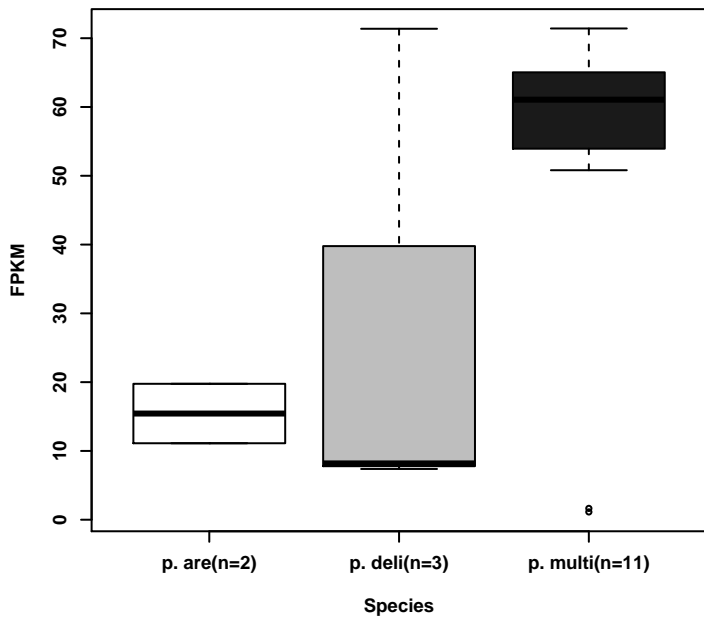
L-methionine biosynthesis via salvage pathway



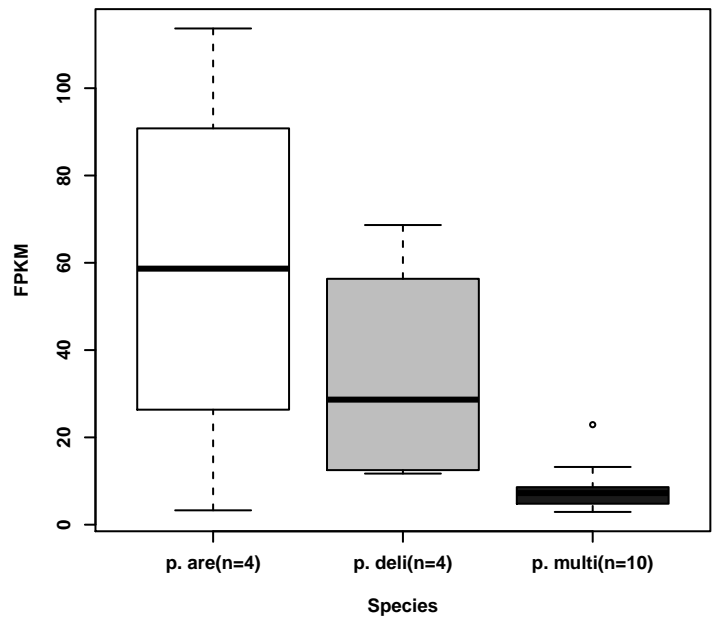
L-phenylalanine biosynthesis



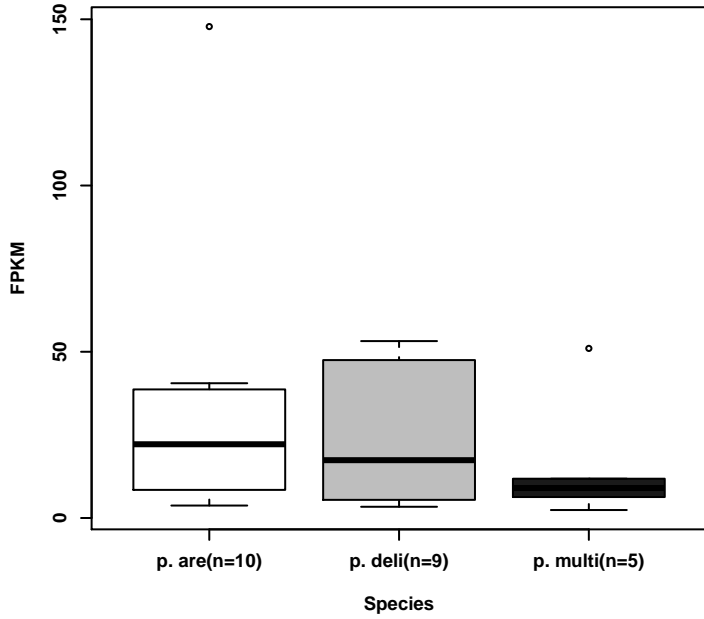
L-phenylalanine degradation



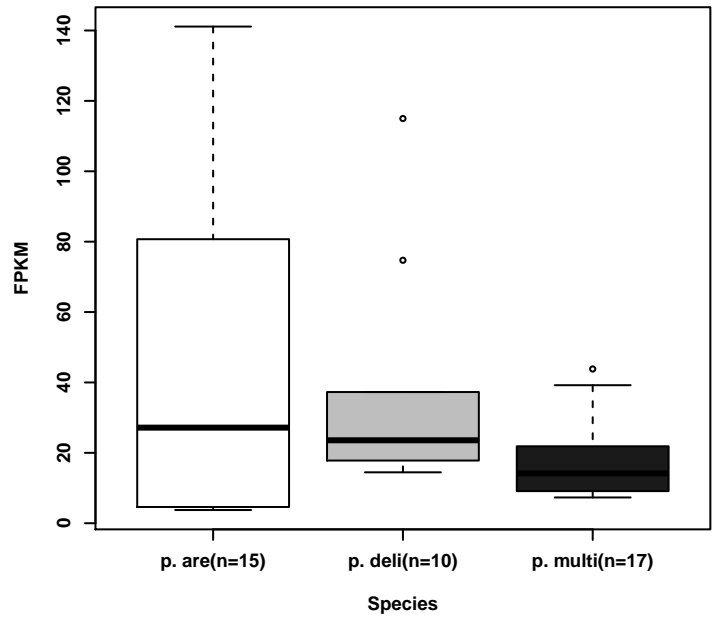
L-proline biosynthesis



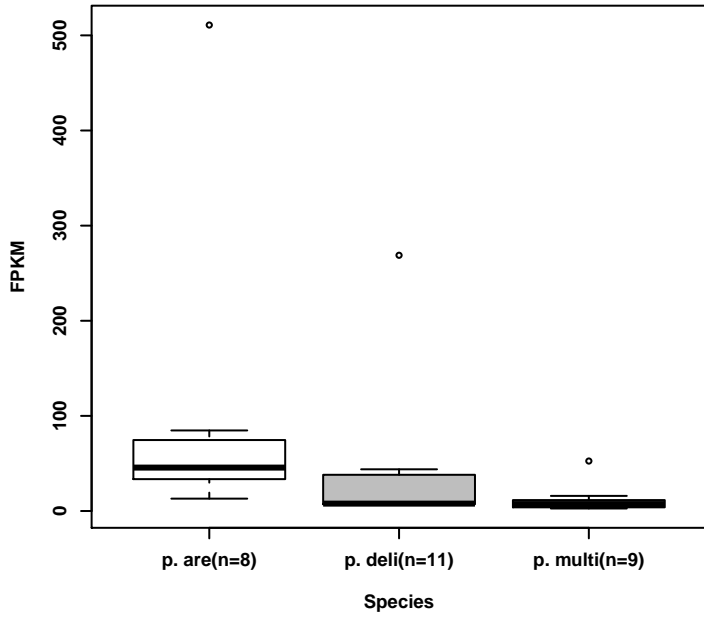
L-serine biosynthesis



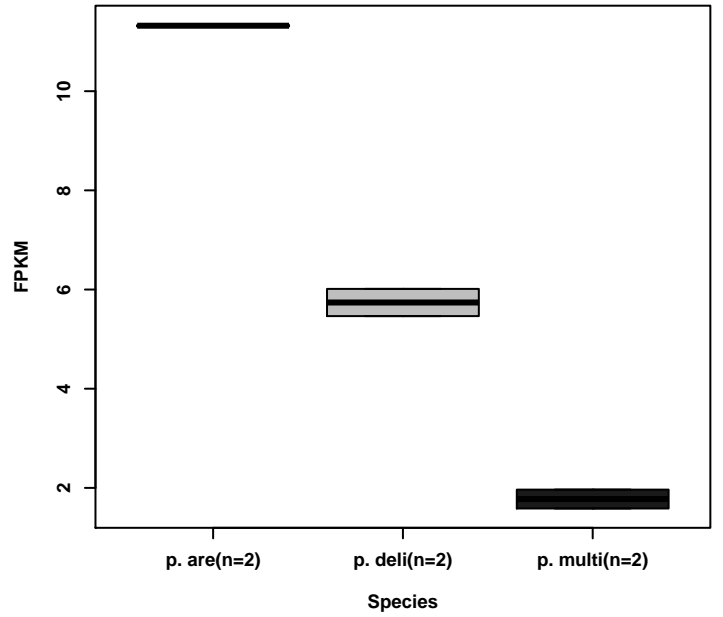
L-threonine biosynthesis



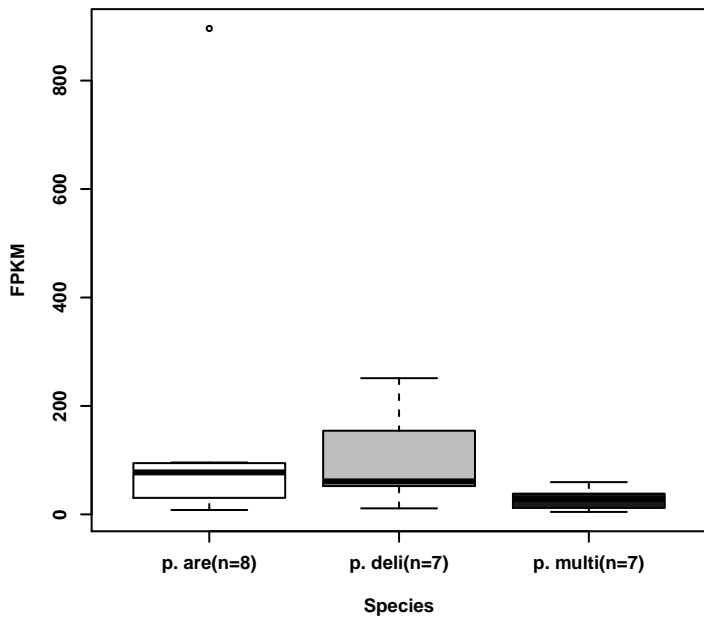
L-tryptophan biosynthesis



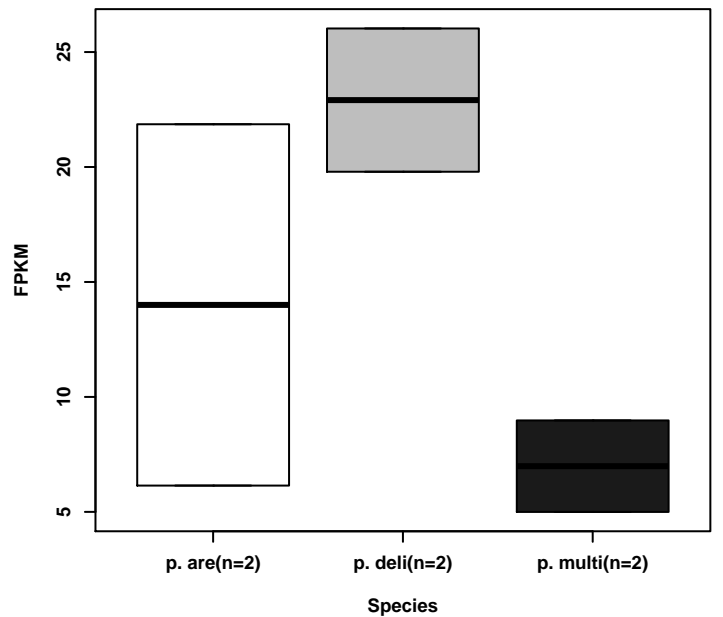
L-tyrosine biosynthesis



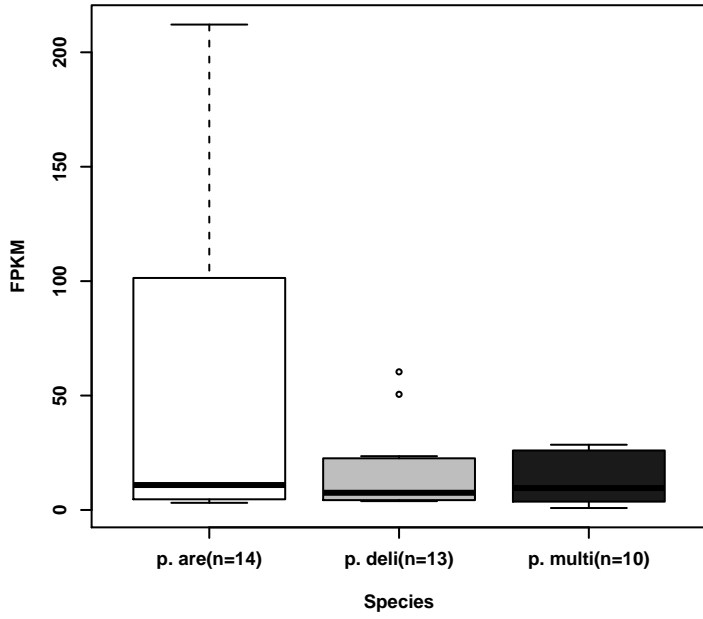
L-valine biosynthesis



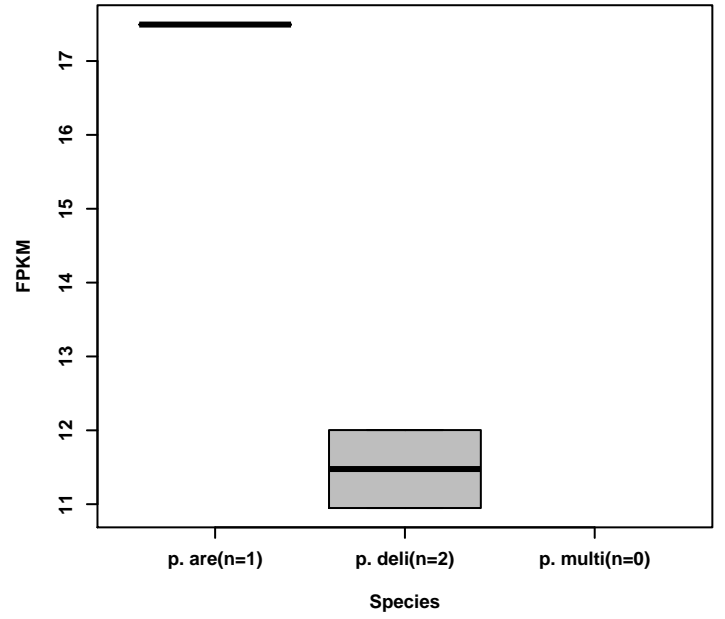
L-valine degradation



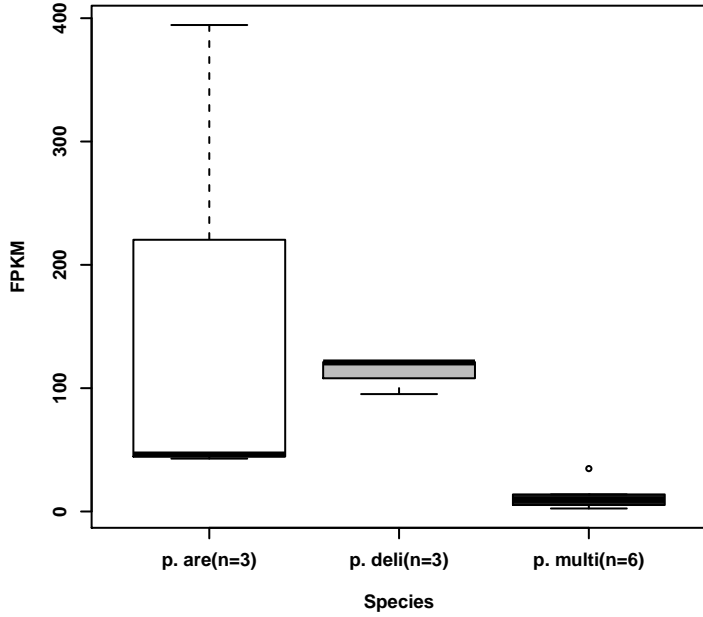
lycopene biosynthesis



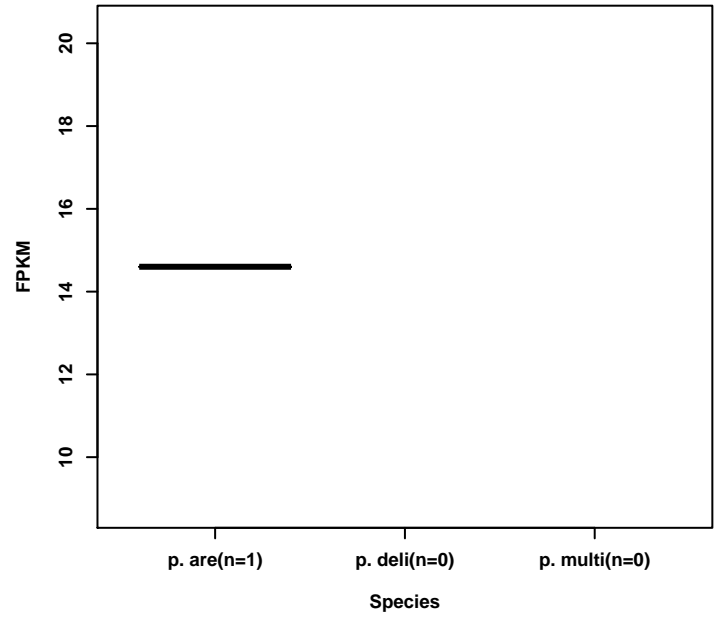
lysine degradation



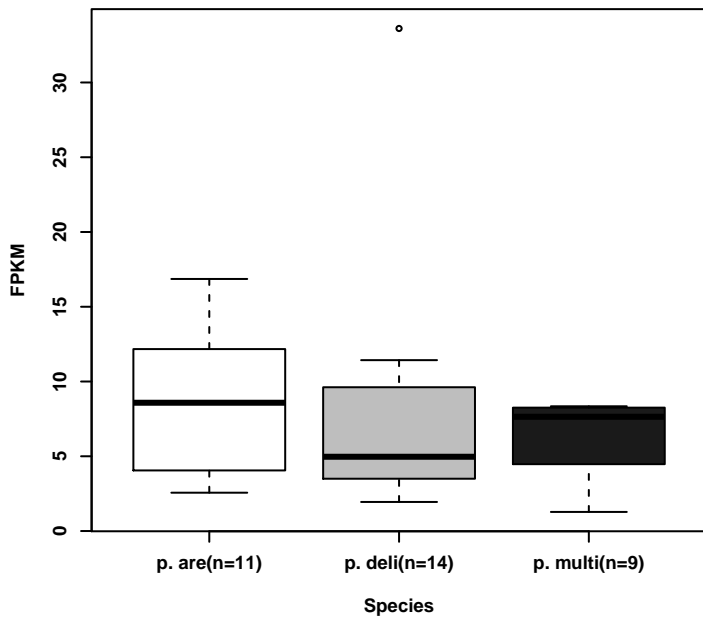
malonyl-CoA biosynthesis



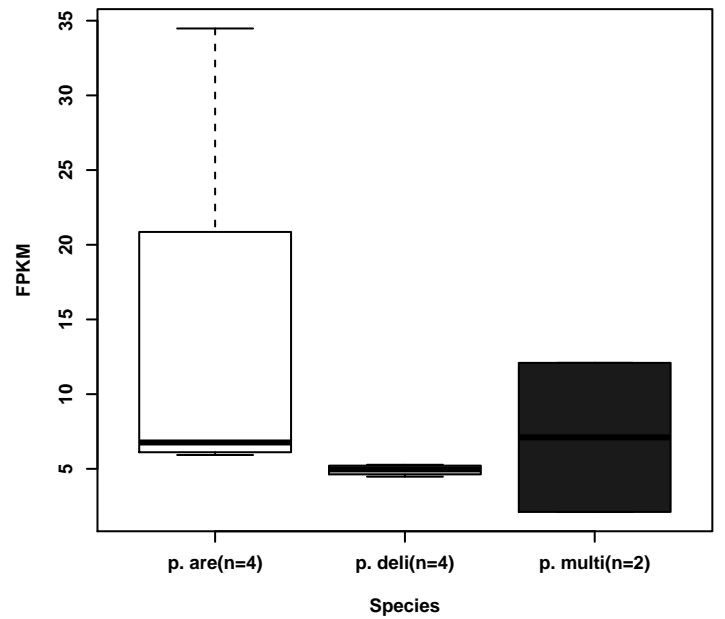
medicarpin biosynthesis



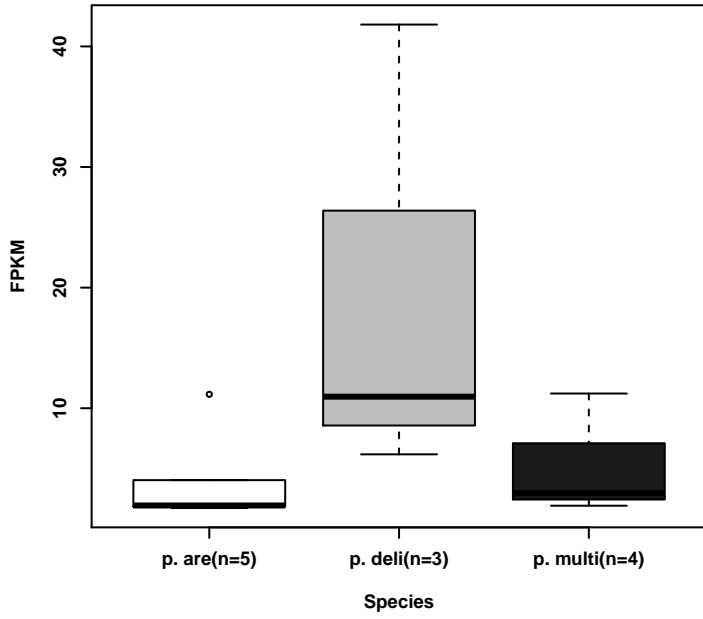
menaquinone biosynthesis



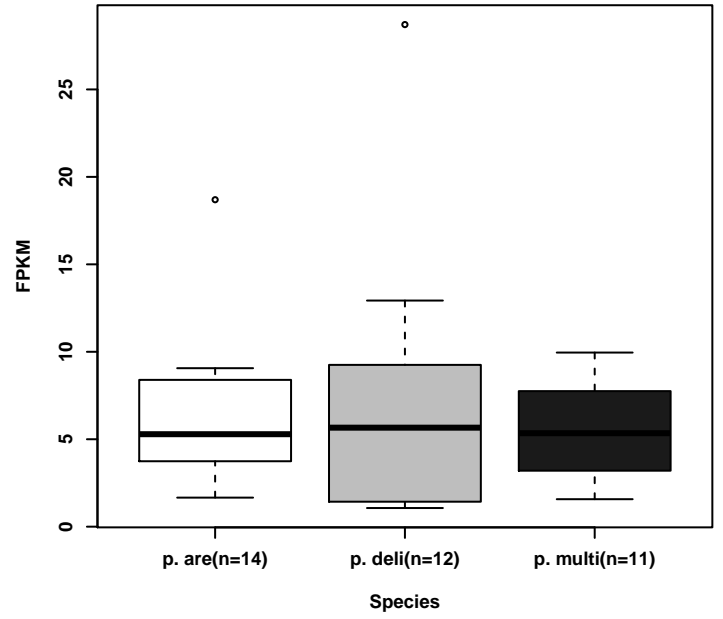
methylglyoxal degradation



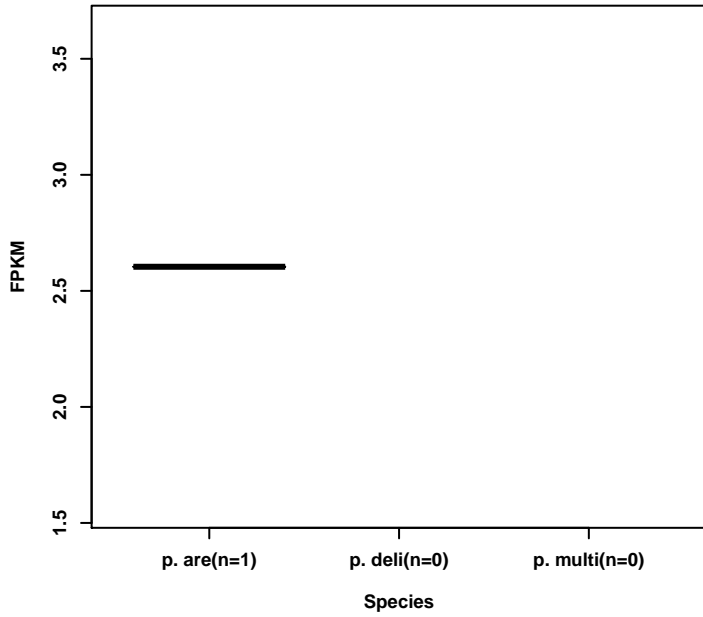
mitochondrial fatty acid beta-oxidation



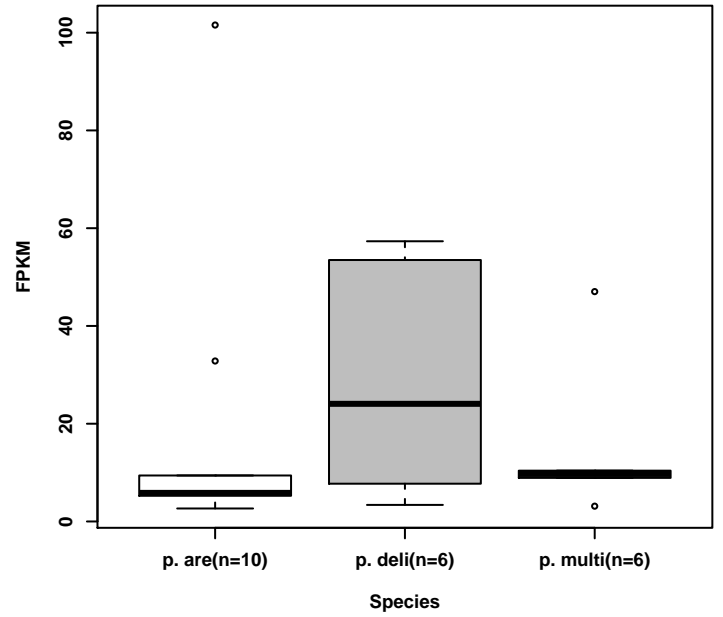
molybdopterin biosynthesis



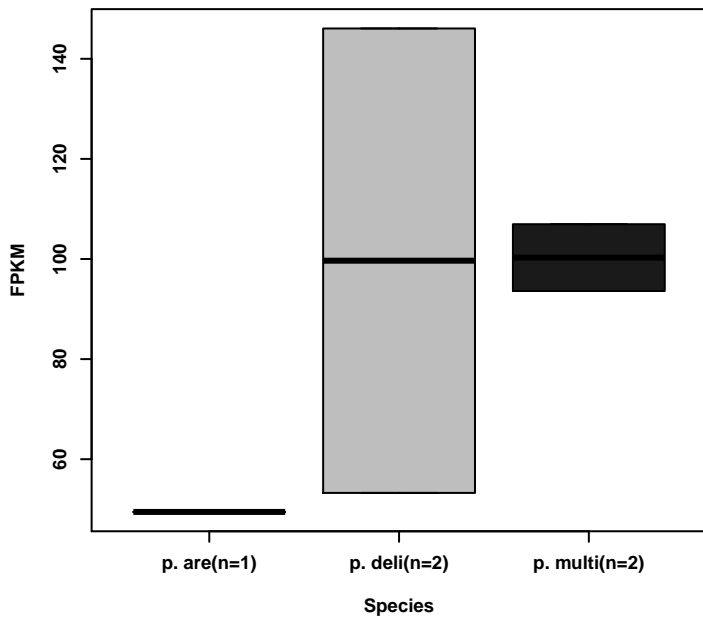
morphine biosynthesis



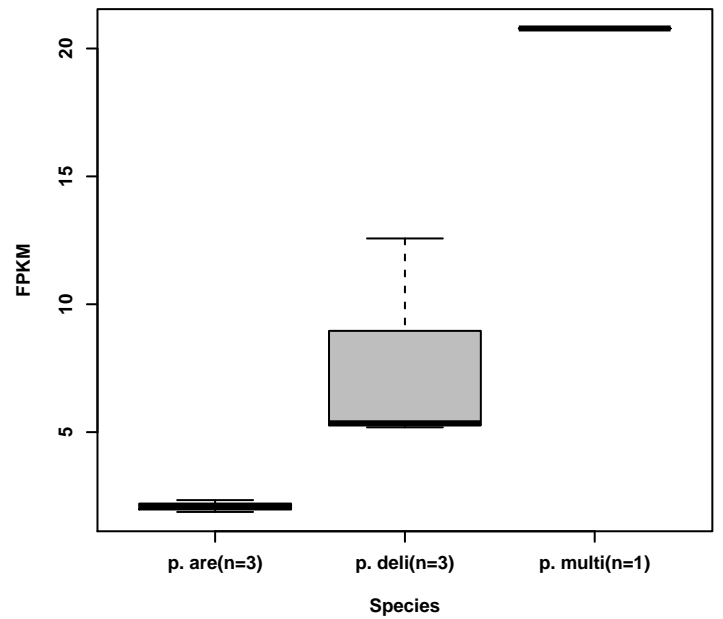
myo-inositol biosynthesis



myo-inositol degradation into acetyl-CoA

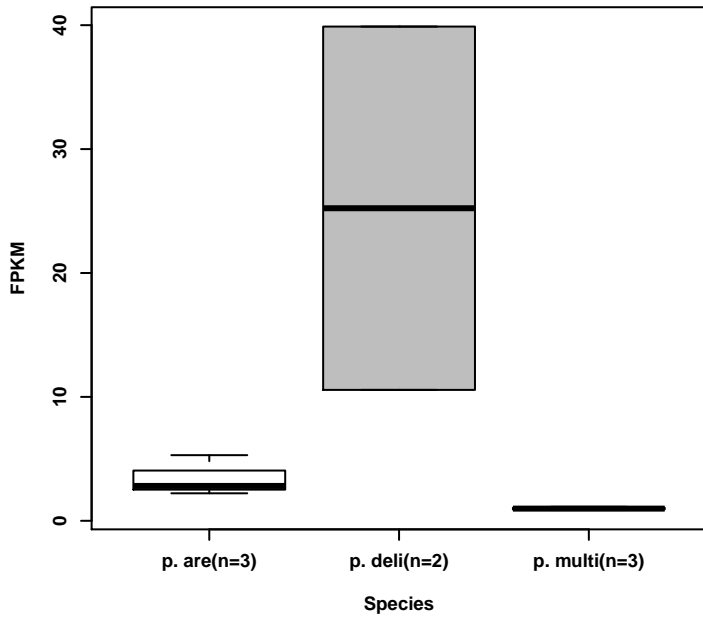


myo-inositol degradation into D-glucuronate

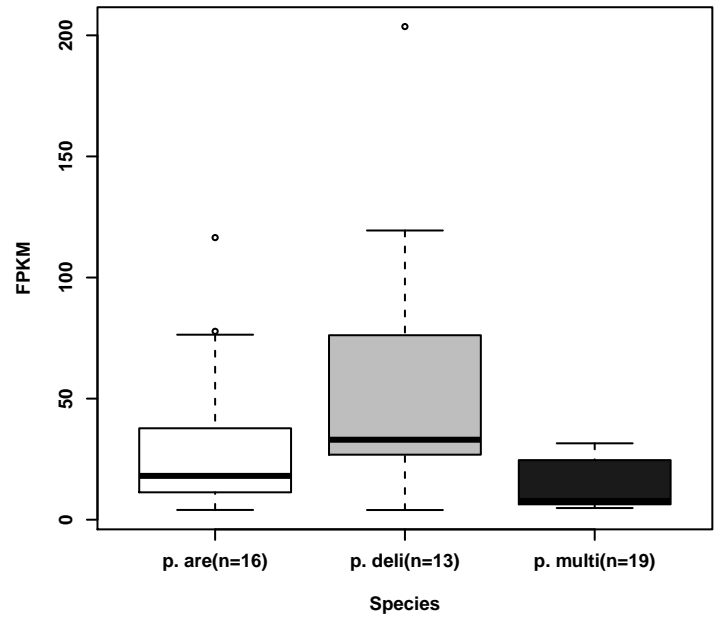




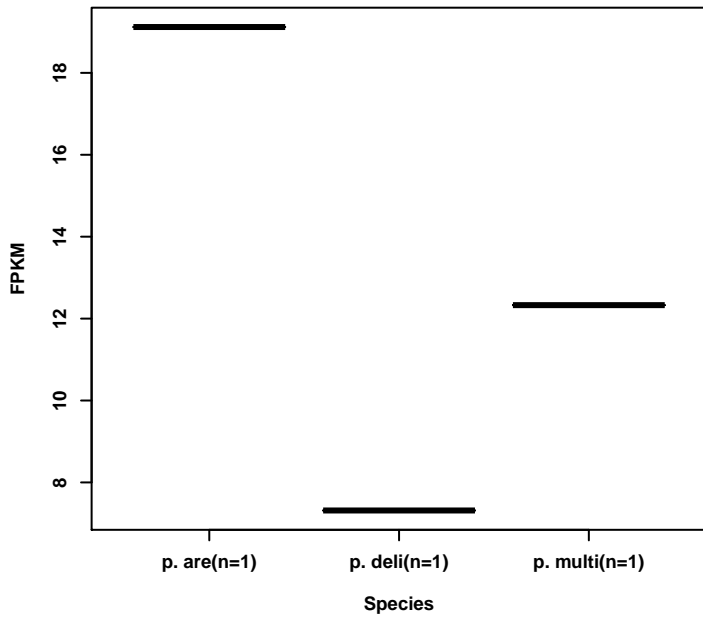
**N(7)-methylguanine-tRNA biosynthesis**



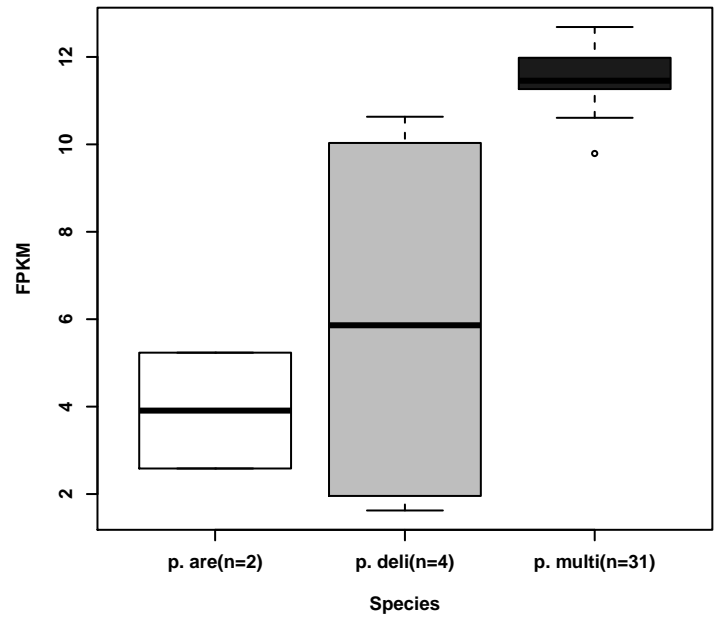
**NAD(+) biosynthesis**



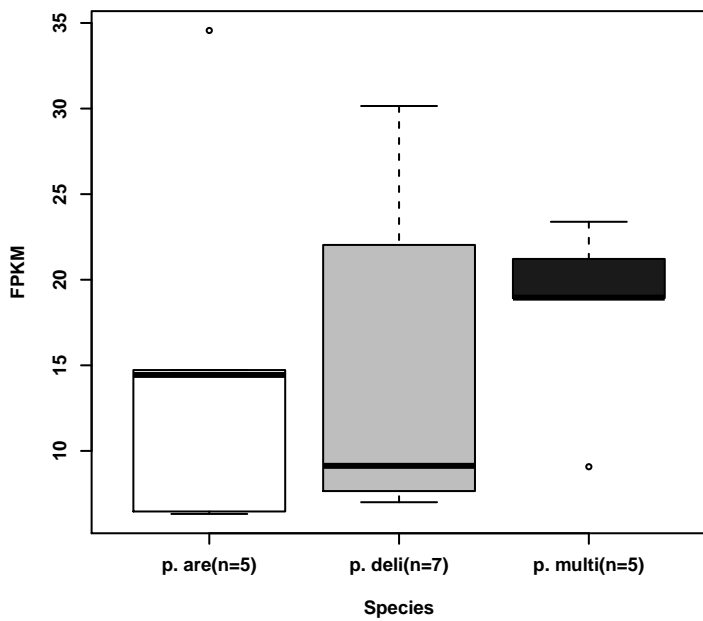
**naphthalene degradation**



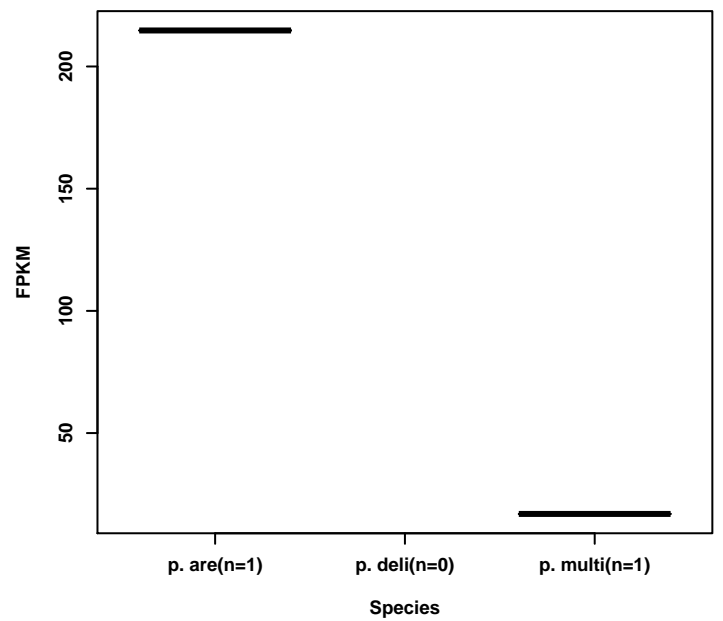
**N-glycan degradation**



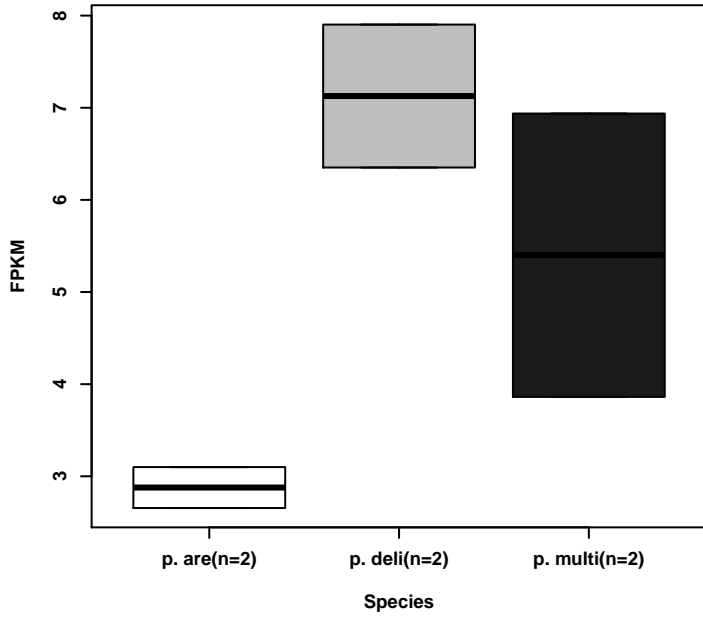
**N-glycan metabolism**



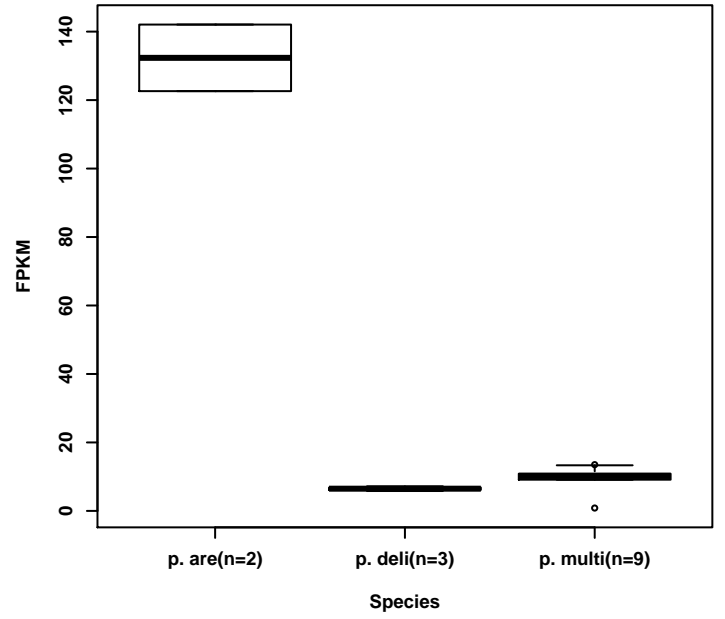
**nicotinate biosynthesis**



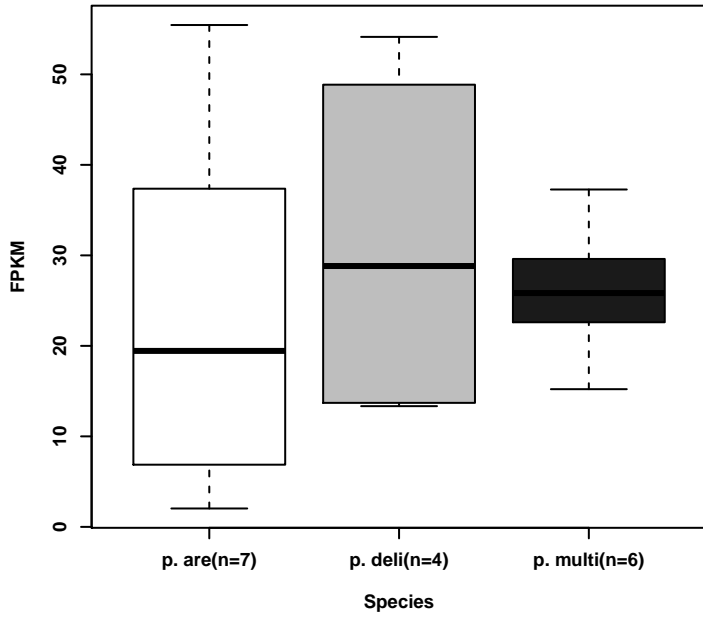
nicotinate degradation



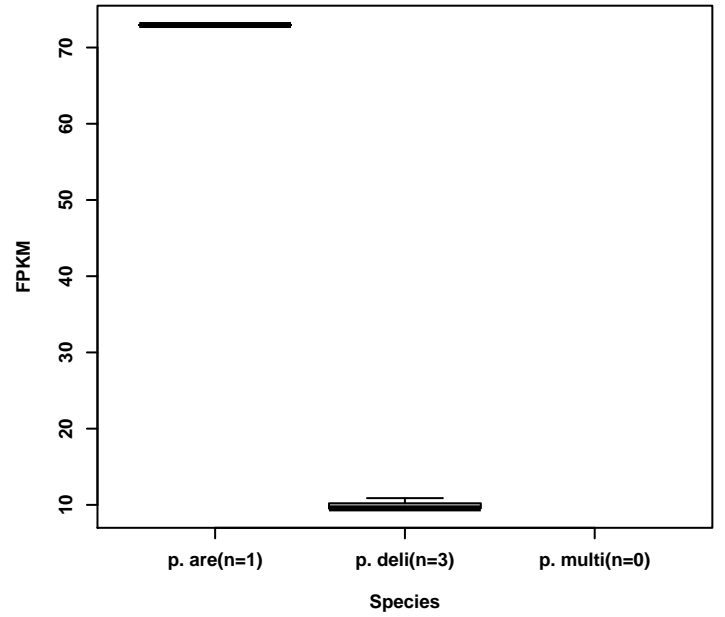
nicotine degradation



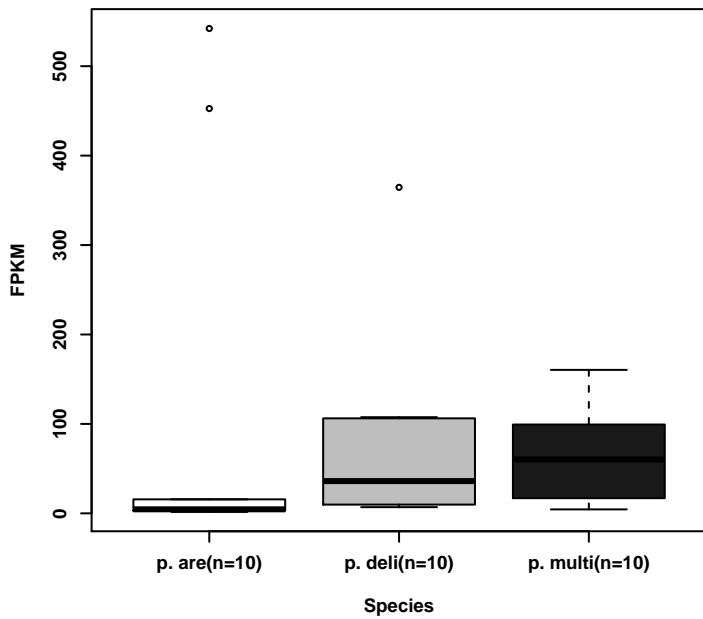
nitrogen metabolism



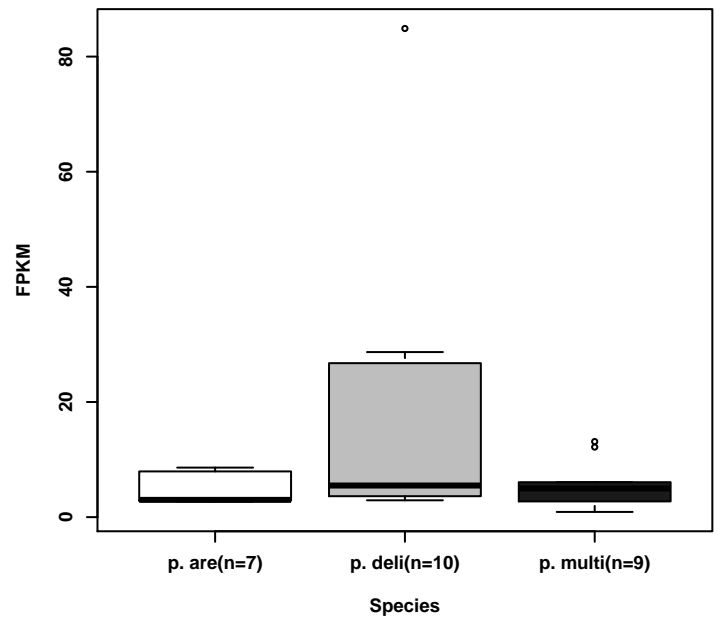
pentachlorophenol degradation



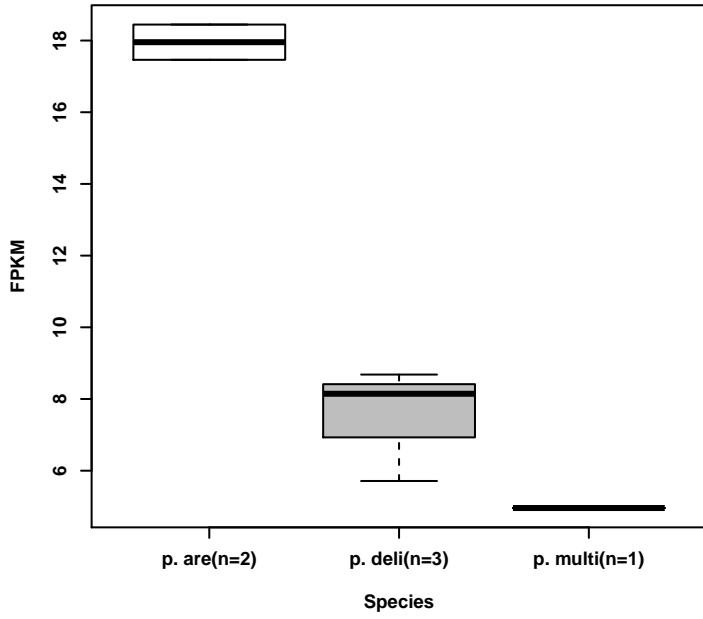
pentose phosphate pathway



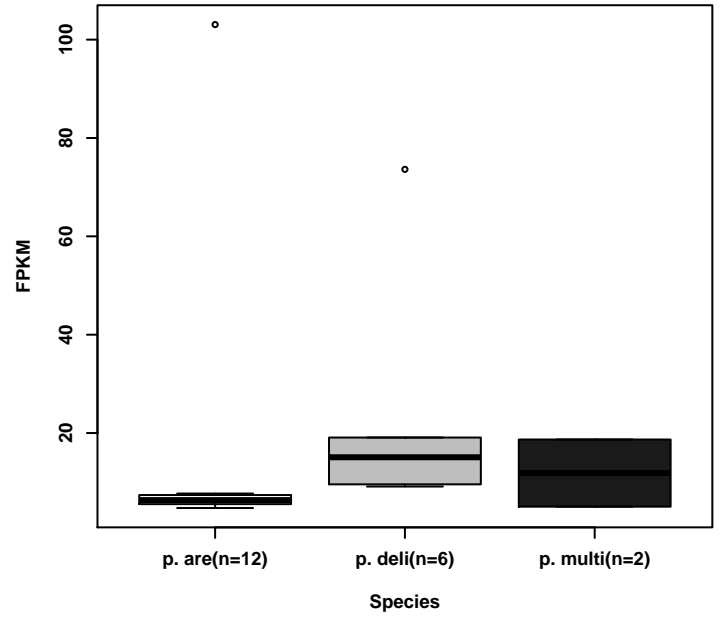
peptidyl-diphthamide biosynthesis



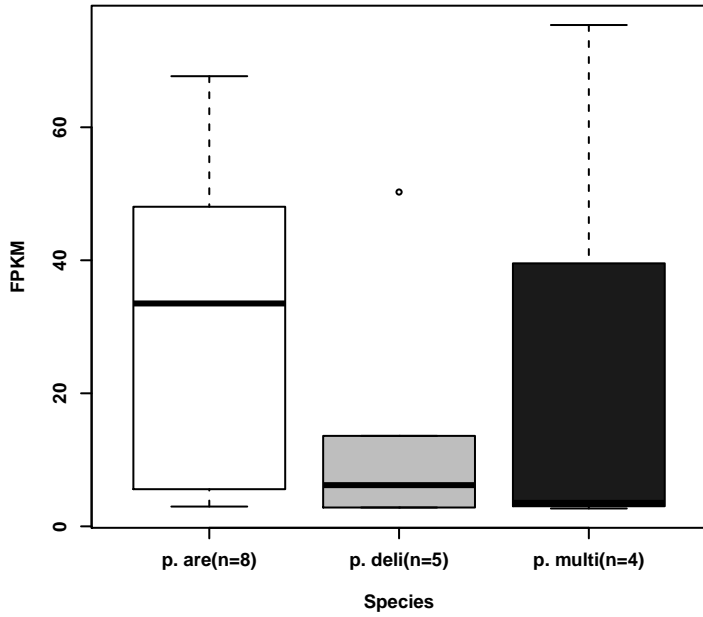
peroxisomal fatty acid beta-oxidation



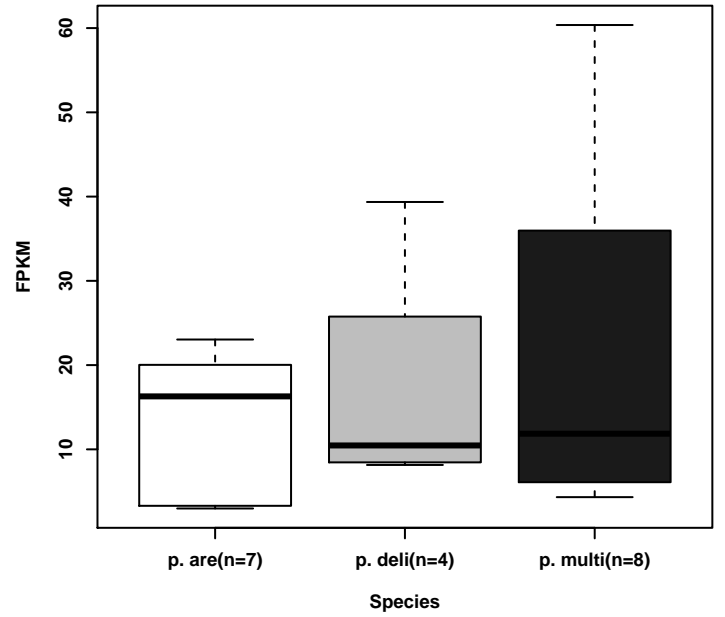
phosphatidylcholine biosynthesis



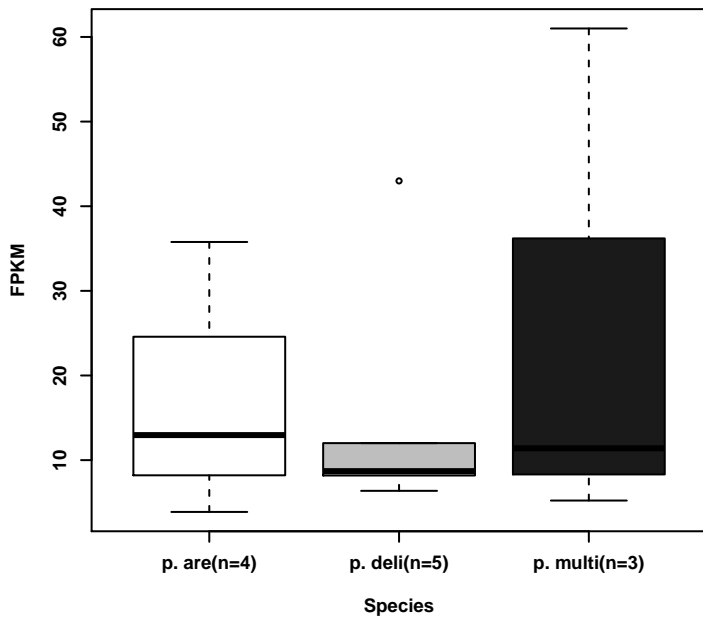
phosphatidylethanolamine biosynthesis



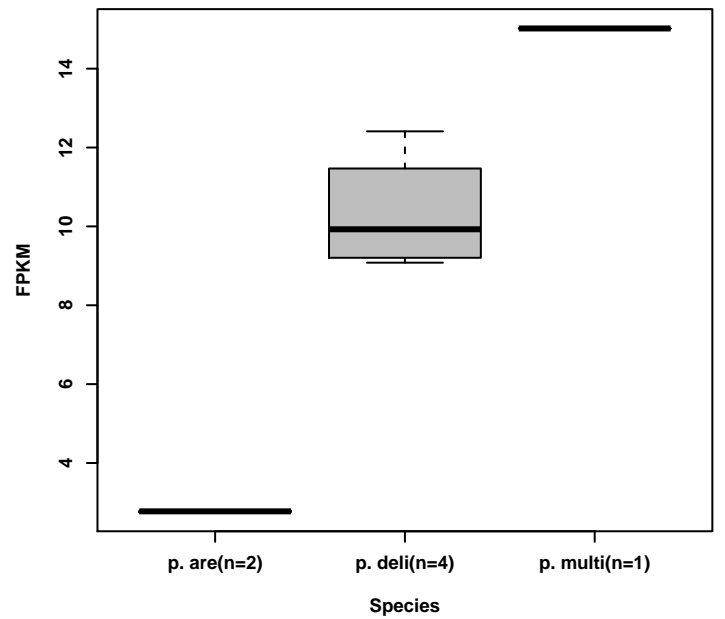
phosphatidylglycerol biosynthesis



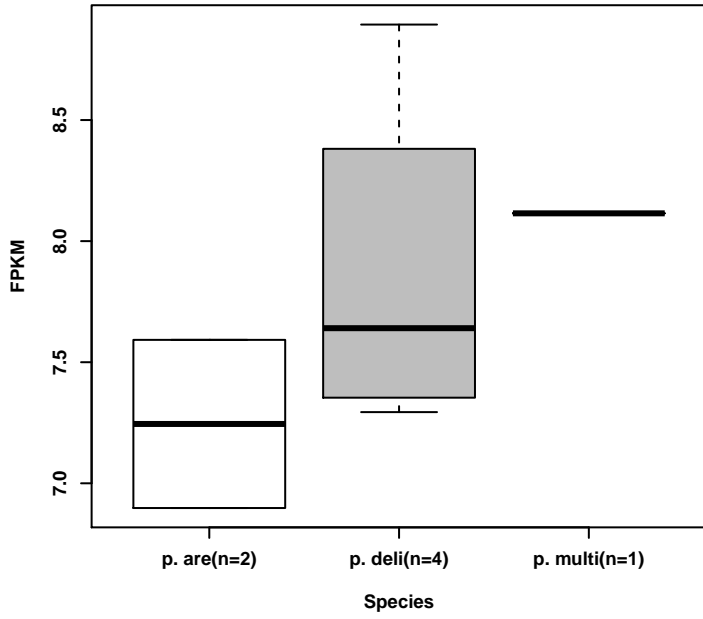
phosphatidylinositol metabolism



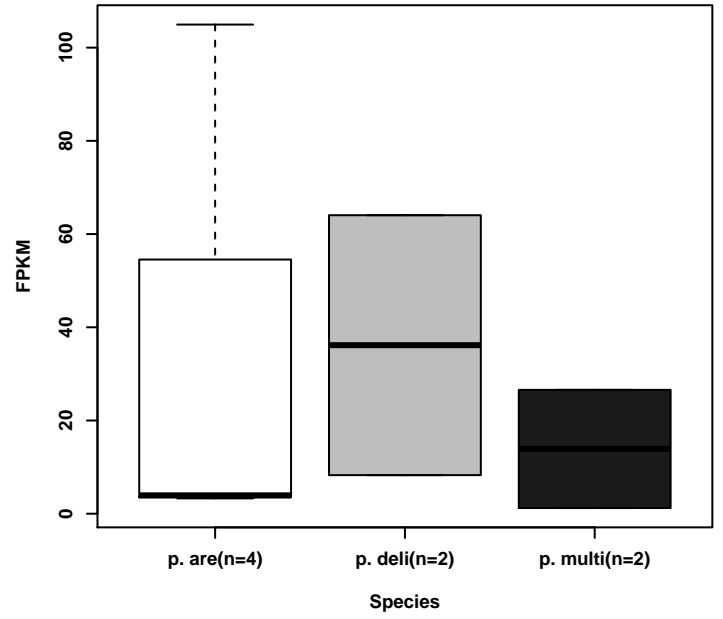
phosphatidylinositol phosphate biosynthesis



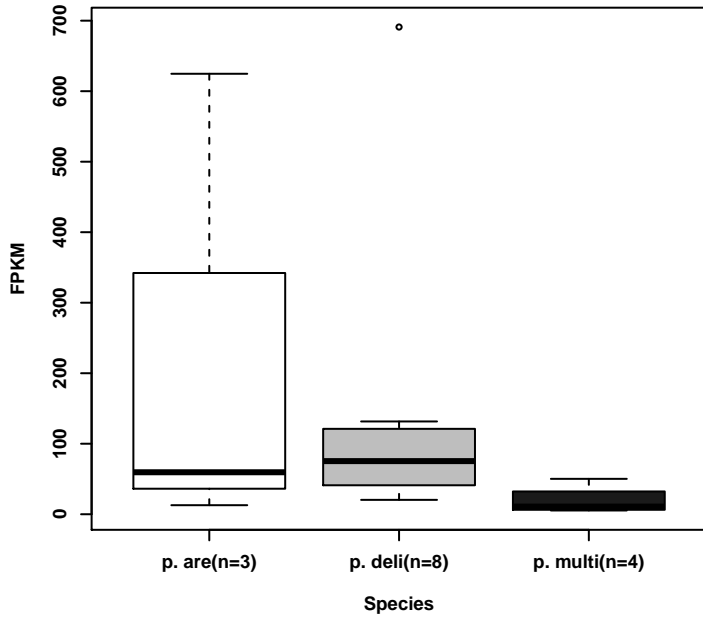
phosphatidylinositol signaling pathway



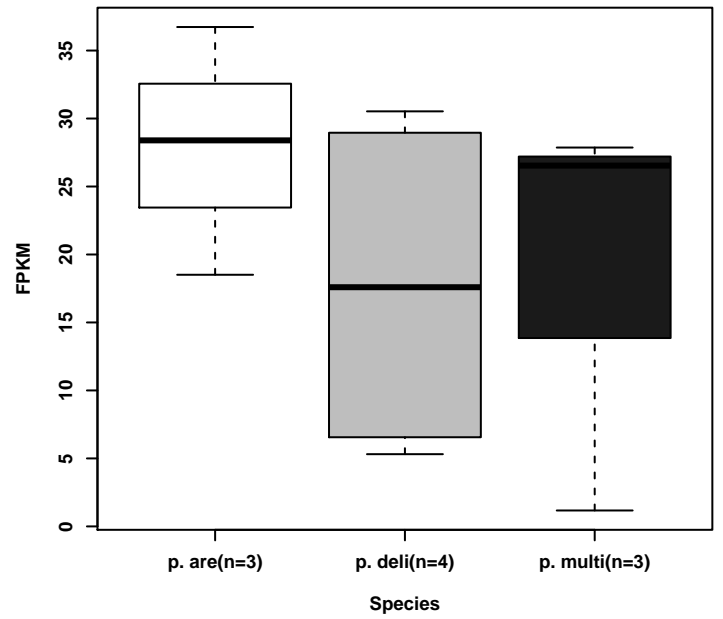
phospholipid metabolism



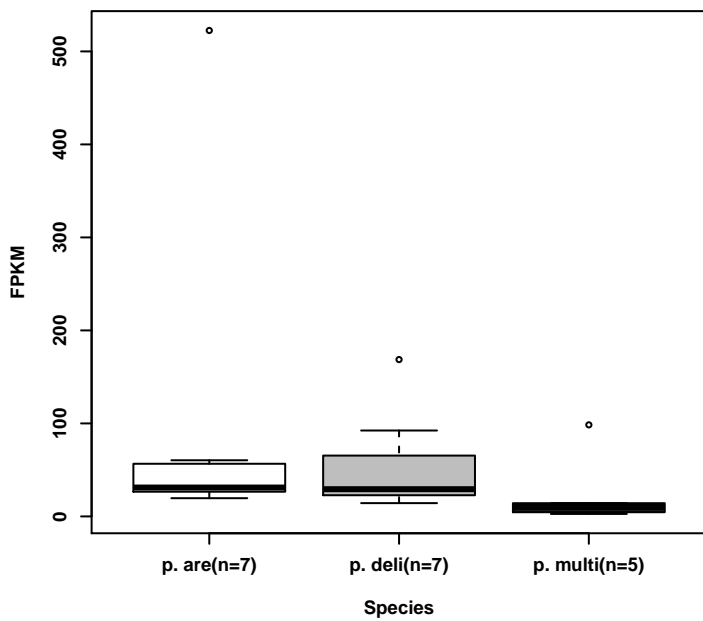
photosynthesis



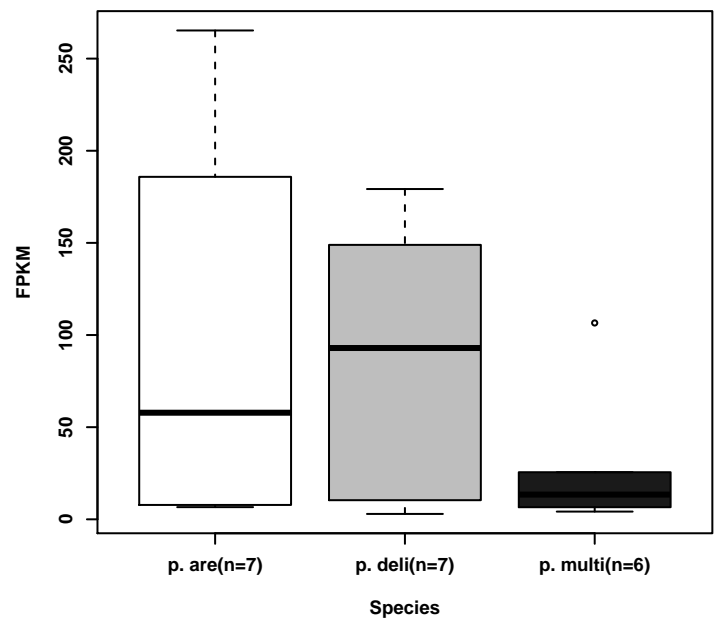
phytoene biosynthesis



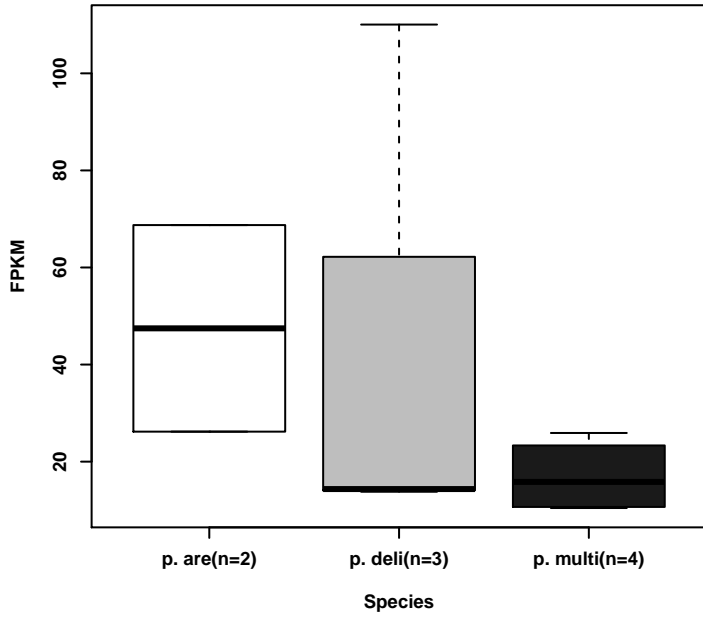
polypeptide chain elongation



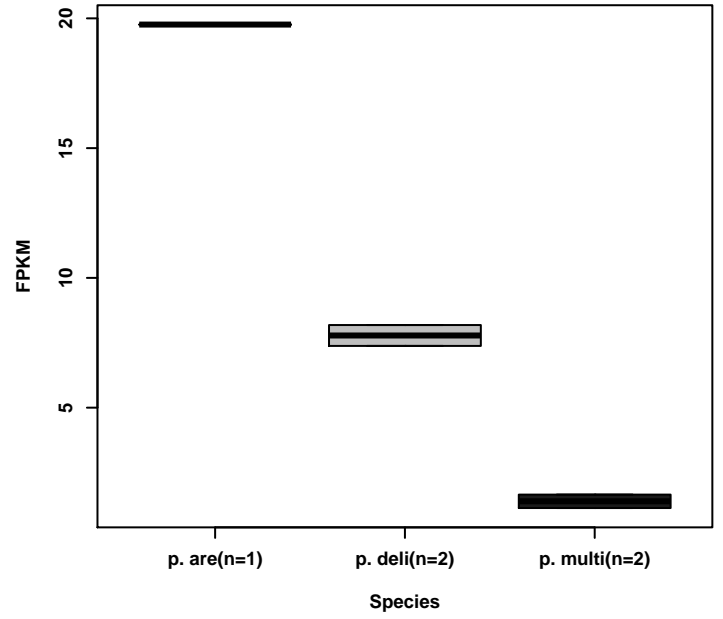
polyunsaturated fatty acid biosynthesis



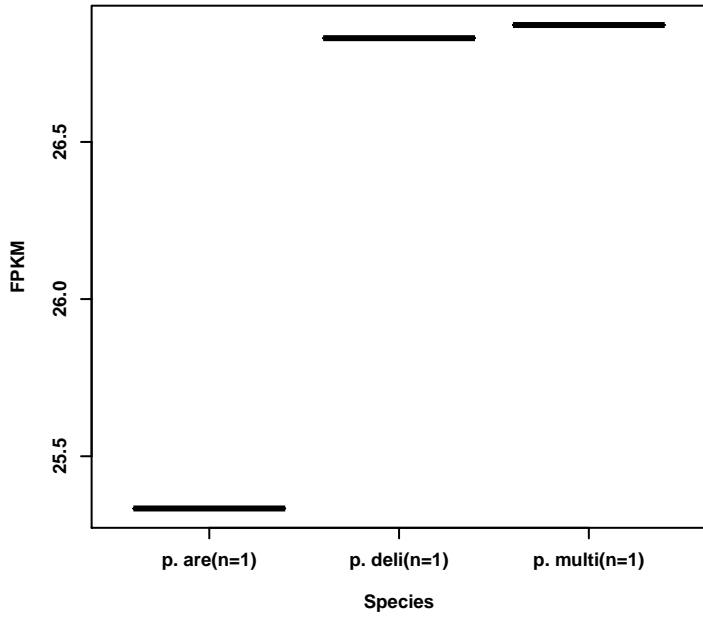
ppGpp biosynthesis



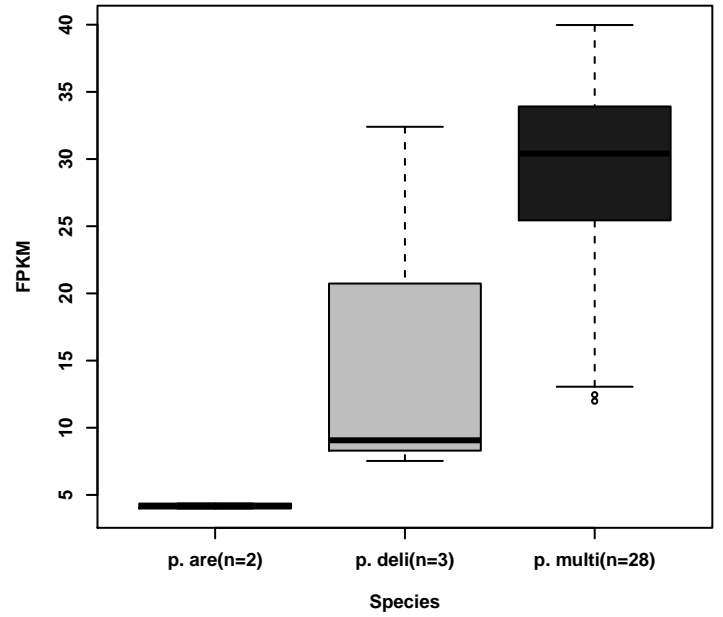
prenylquinone biosynthesis



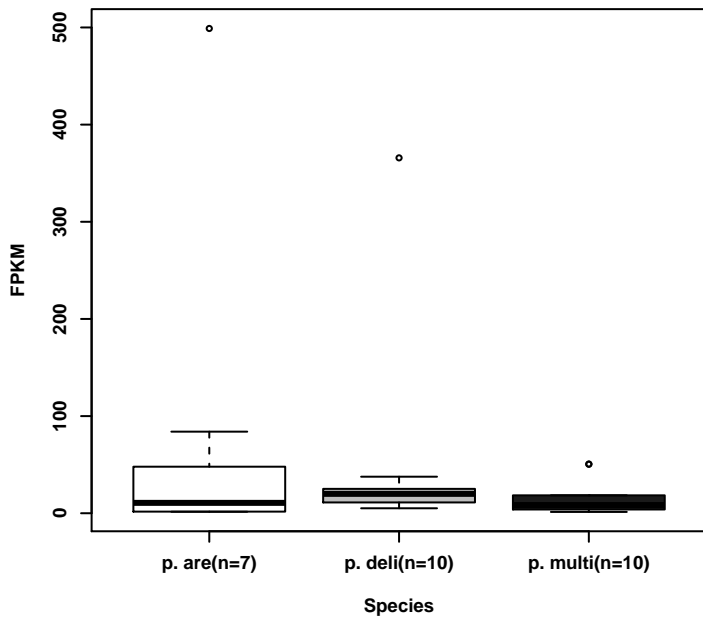
prephenate biosynthesis



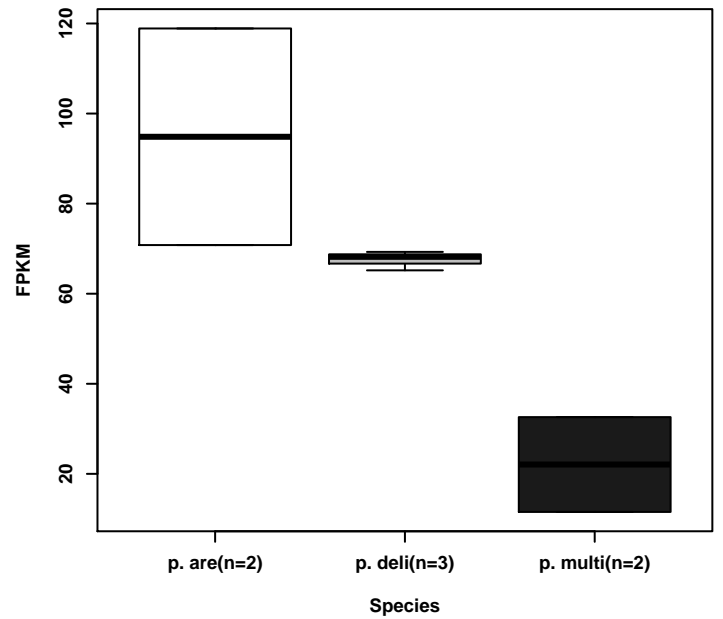
propanoyl-CoA degradation



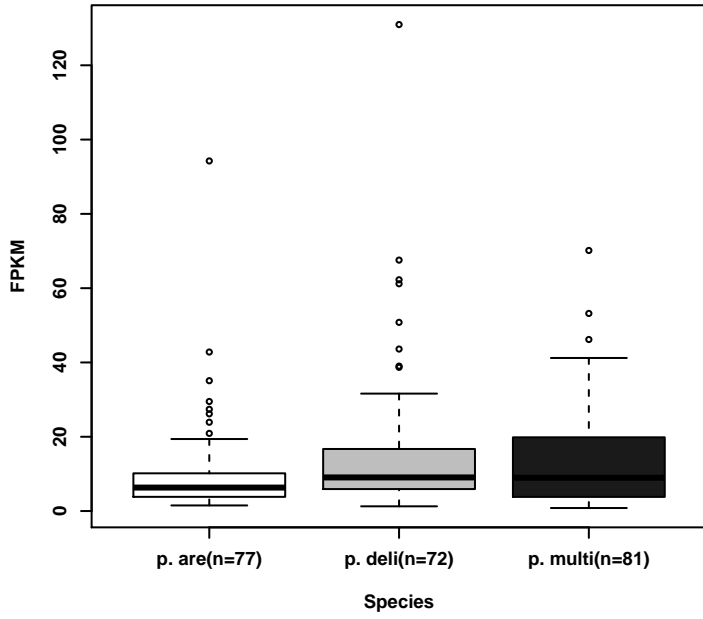
prostaglandin biosynthesis



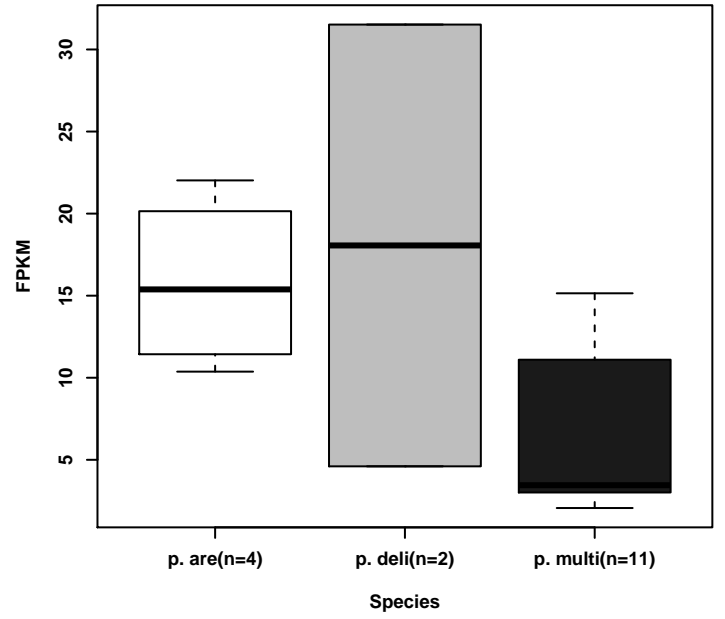
proteasomal ubiquitin-dependent pathway



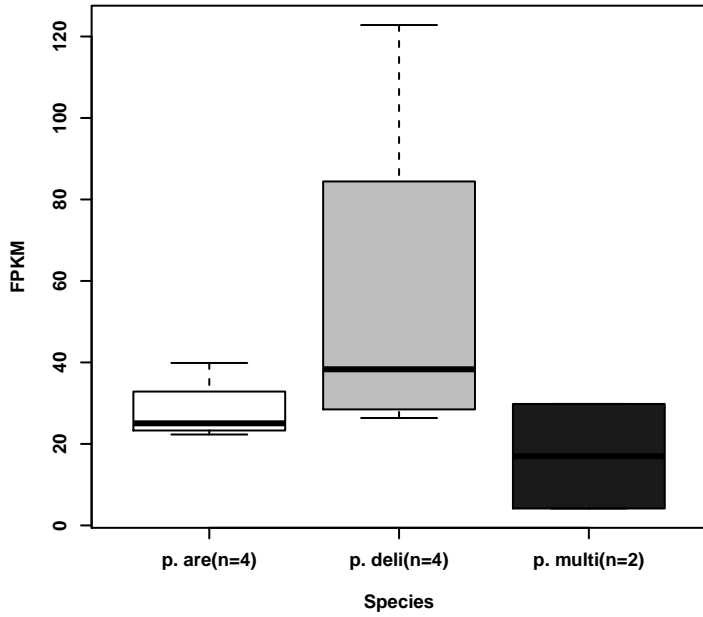
protein glycosylation



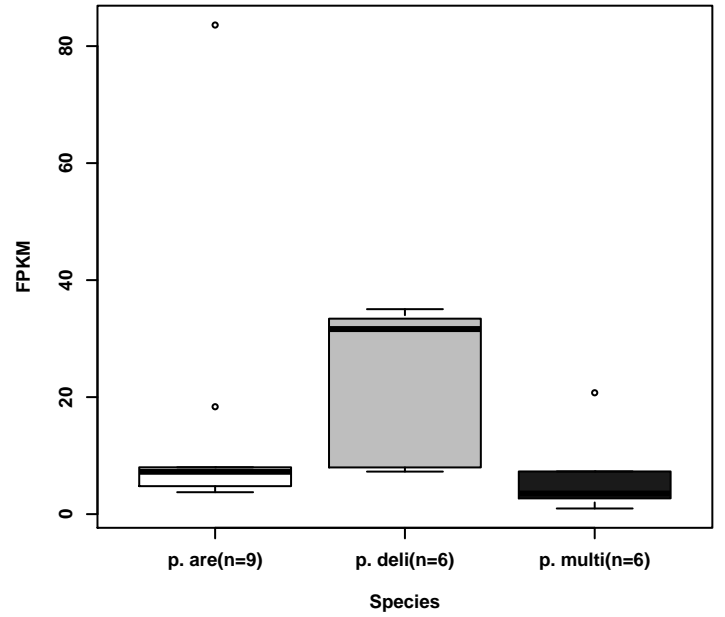
protein lipoylation via endogenous pathway



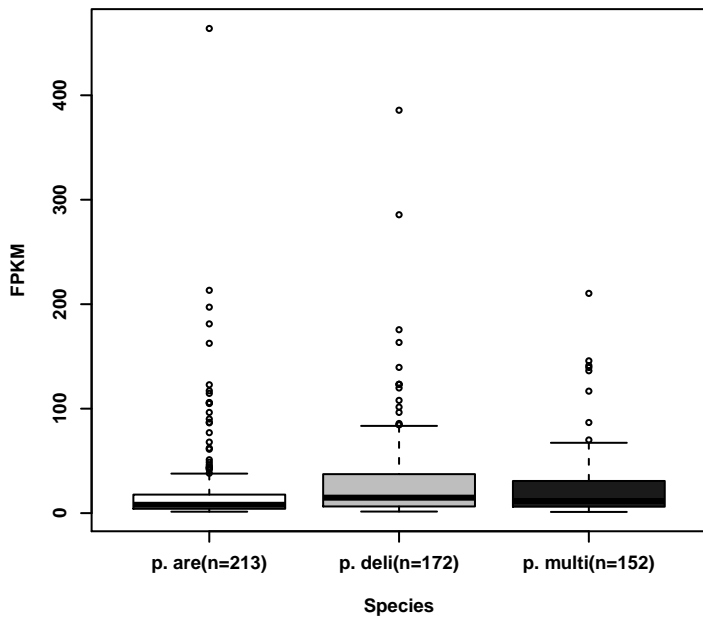
protein neddylation



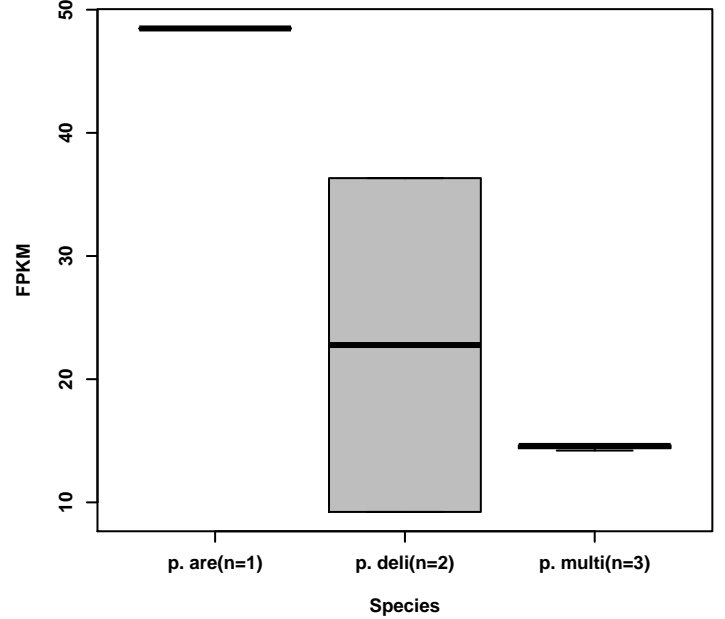
protein sumoylation



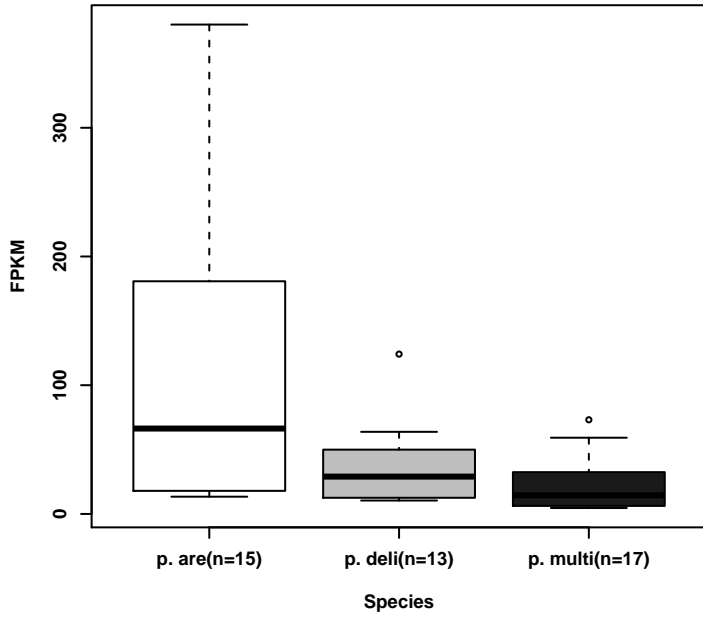
protein ubiquitination



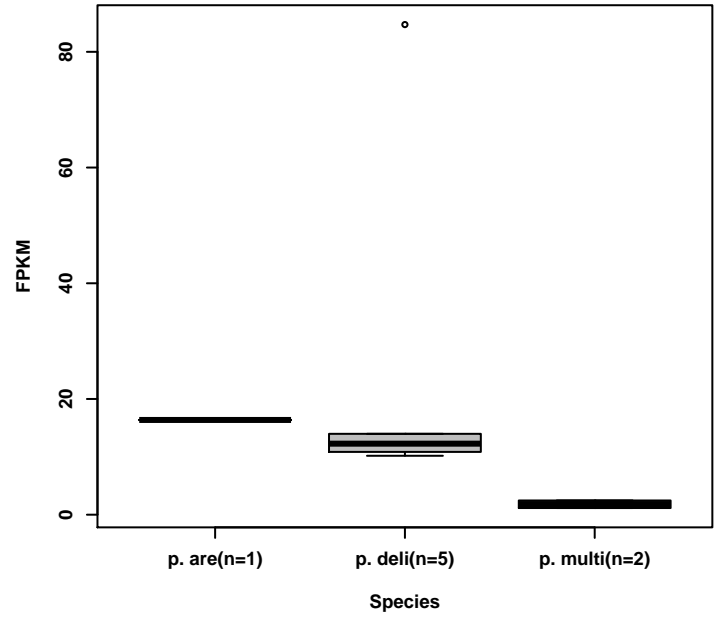
protoheme biosynthesis



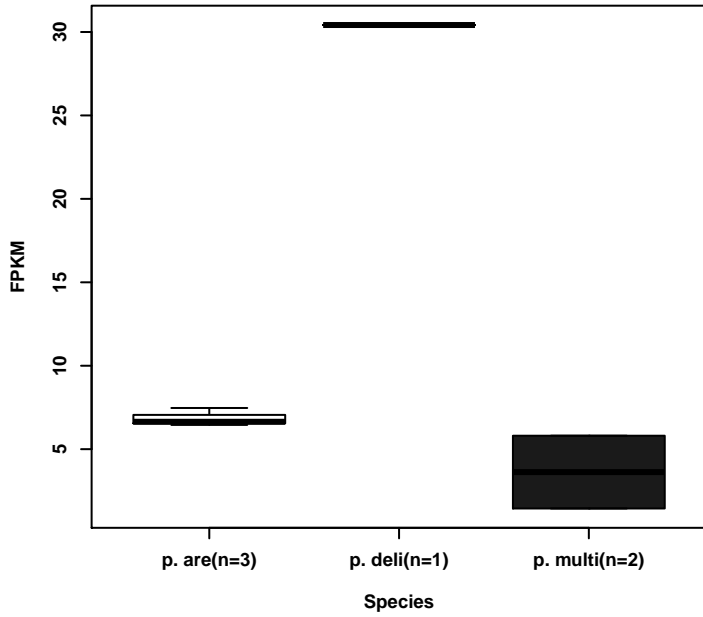
protoporphyrin-IX biosynthesis



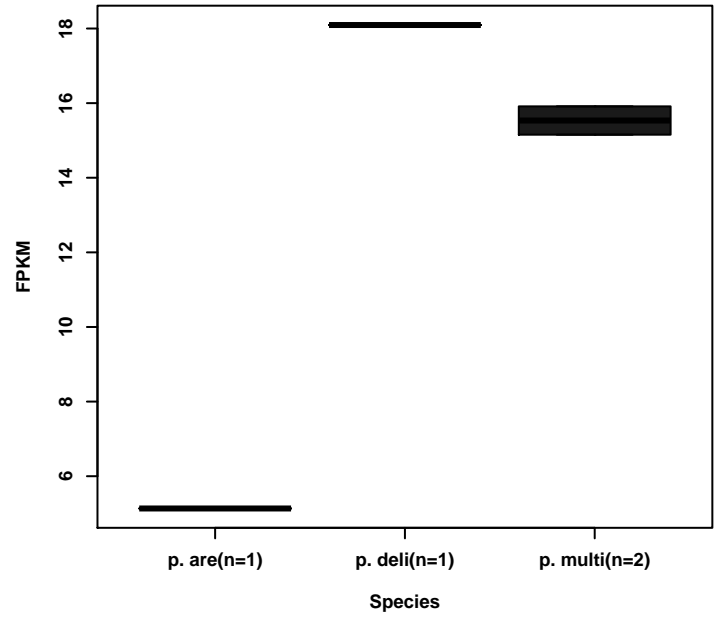
putrescine biosynthesis via agmatine pathway



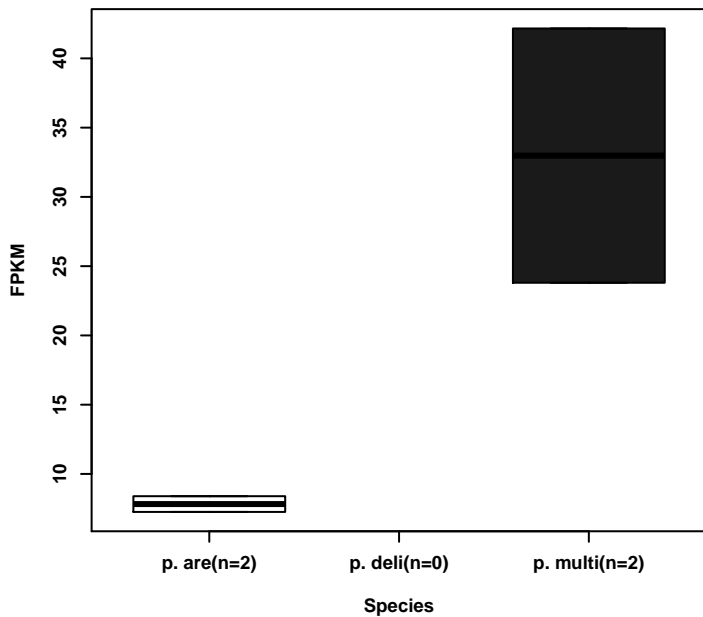
putrescine biosynthesis via L-ornithine pathway



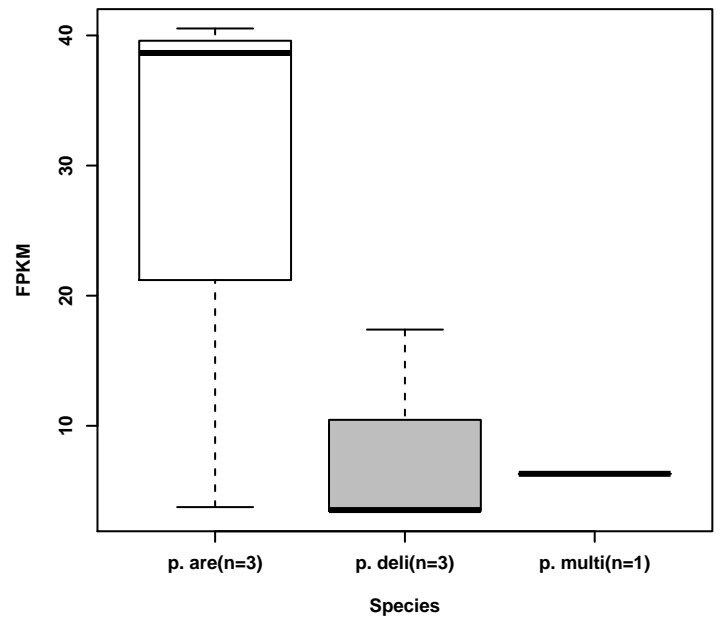
putrescine degradation



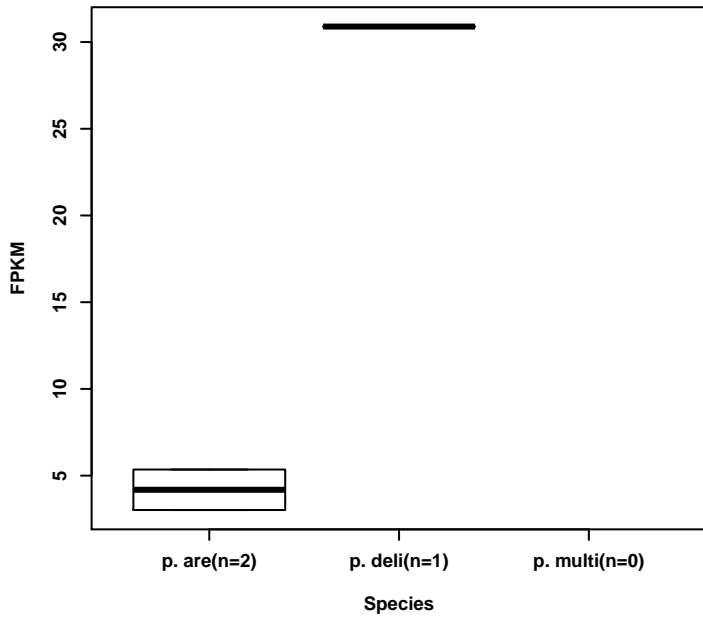
pyridoxal 5'-phosphate biosynthesis



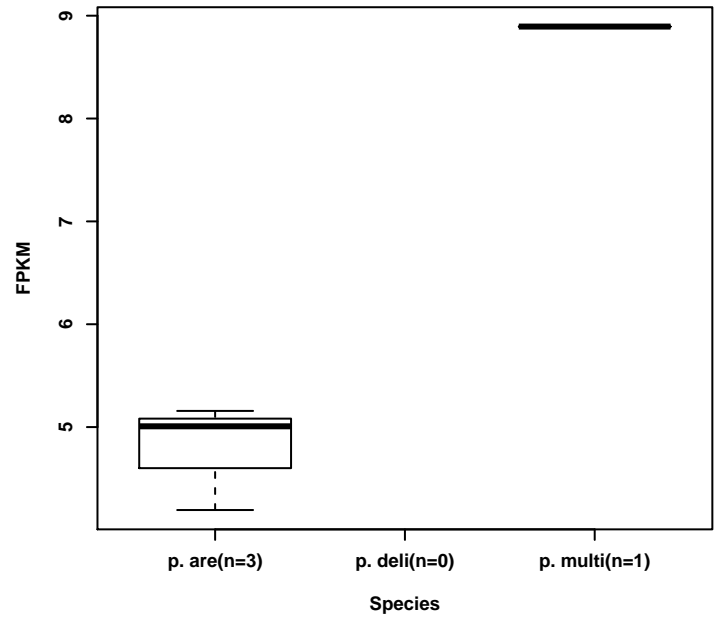
pyridoxine 5'-phosphate biosynthesis



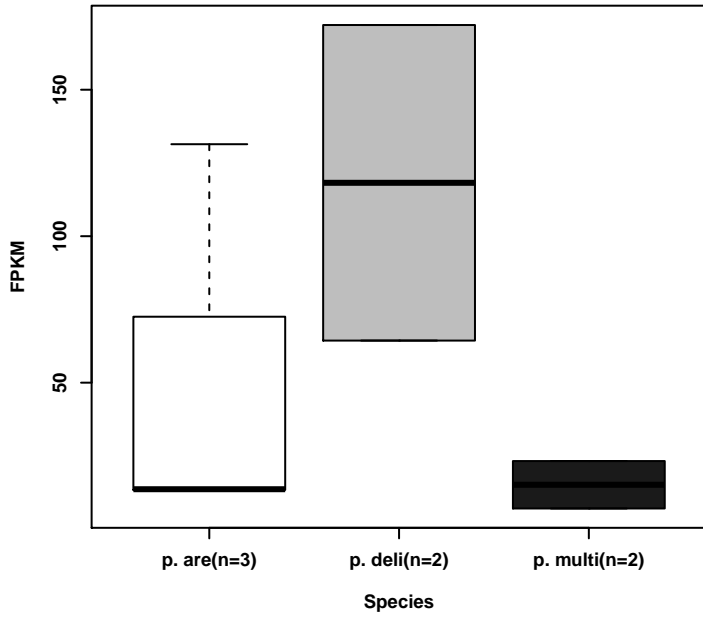
pyruvate fermentation to lactate



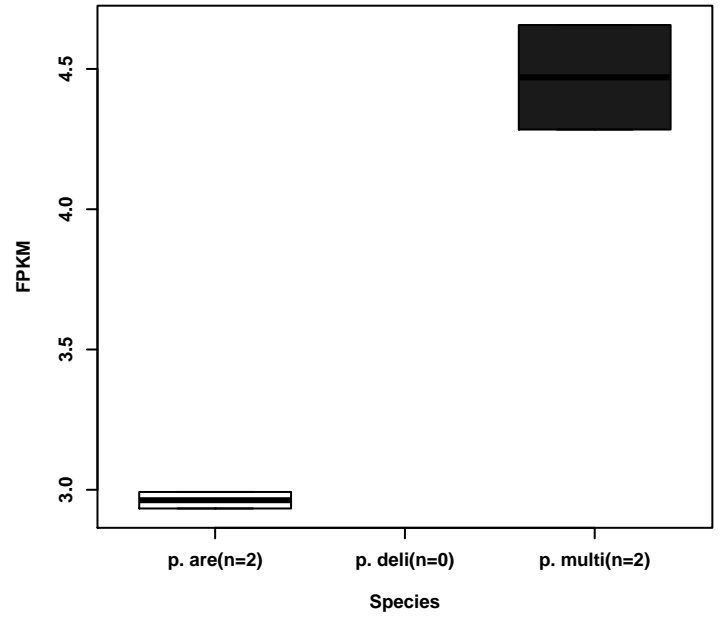
pyruvate metabolism



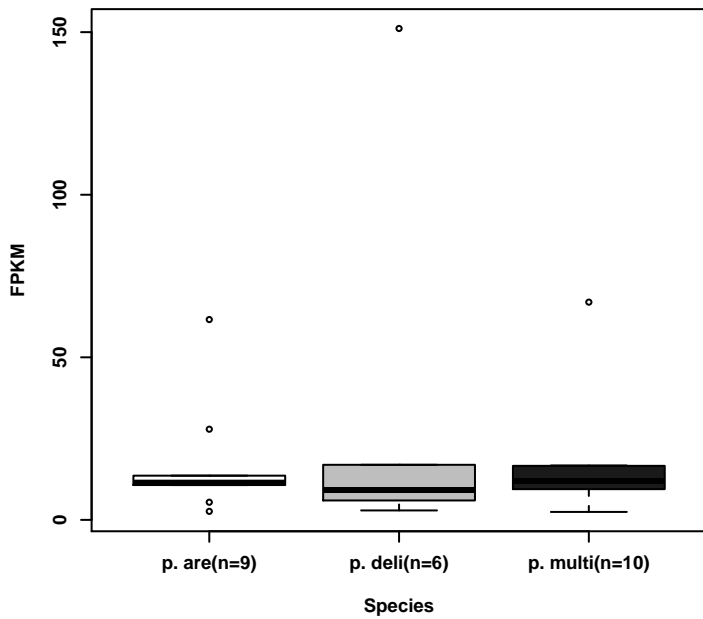
quercetin degradation



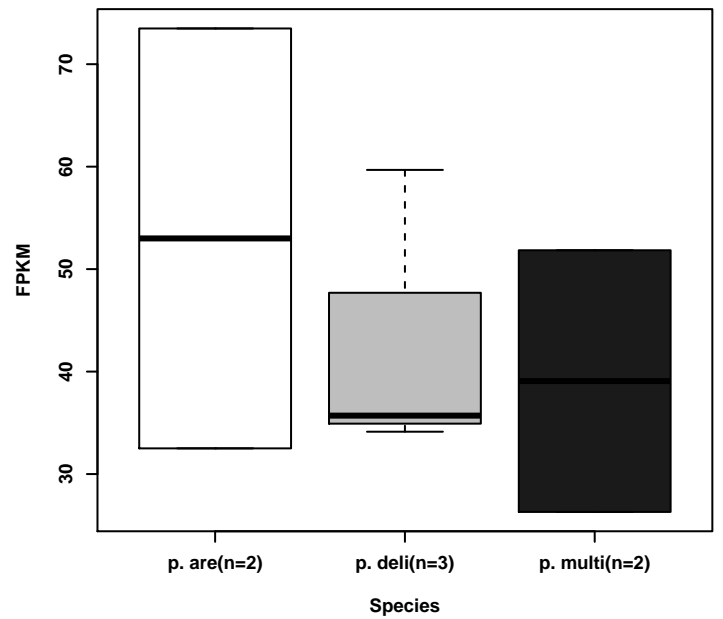
rhamnolipid biosynthesis



riboflavin biosynthesis

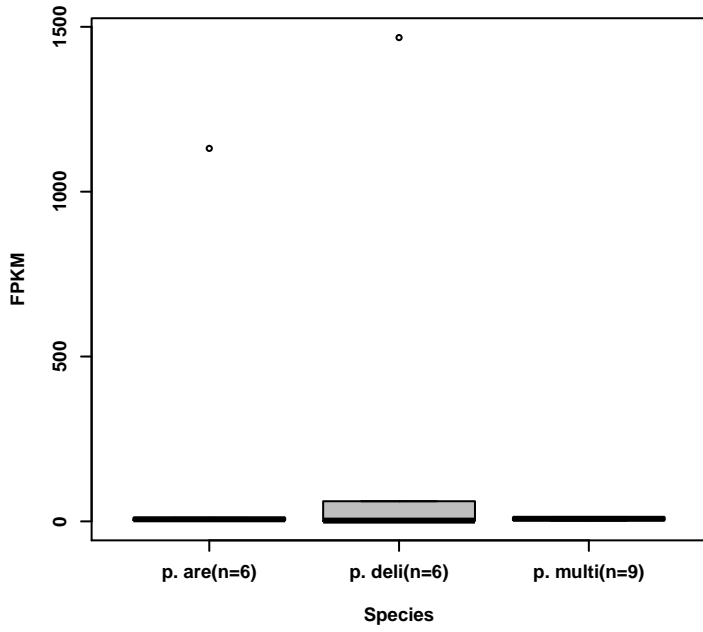


(R)-mevalonate biosynthesis

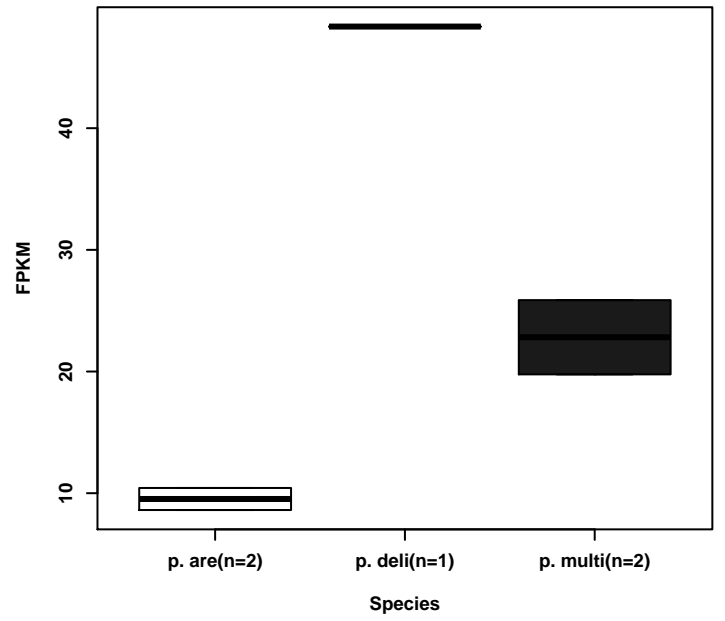




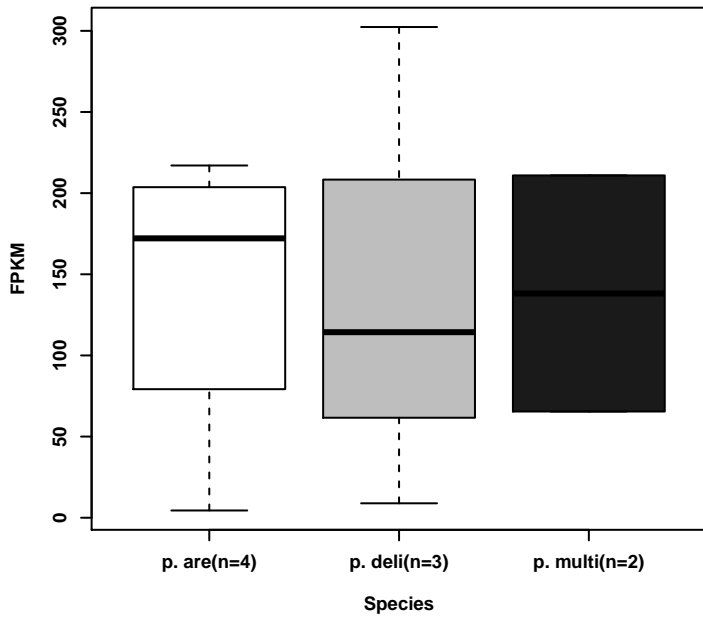
**(R)-pantothenate biosynthesis**



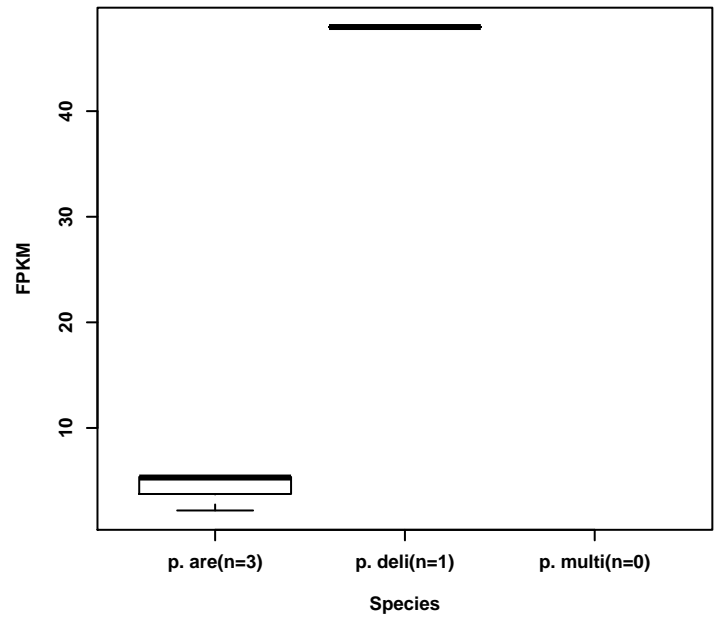
**(S)-3-hydroxy-3-methylglutaryl-CoA degradation**



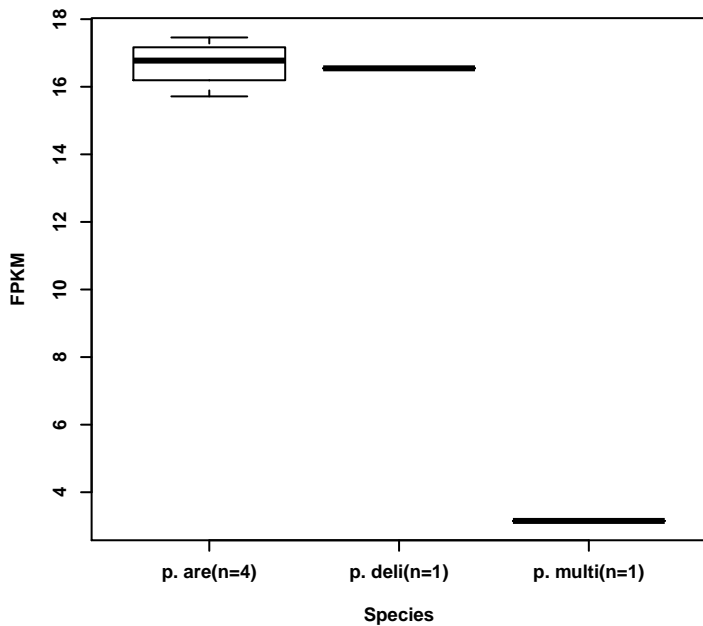
**S-adenosyl-L-methionine biosynthesis**



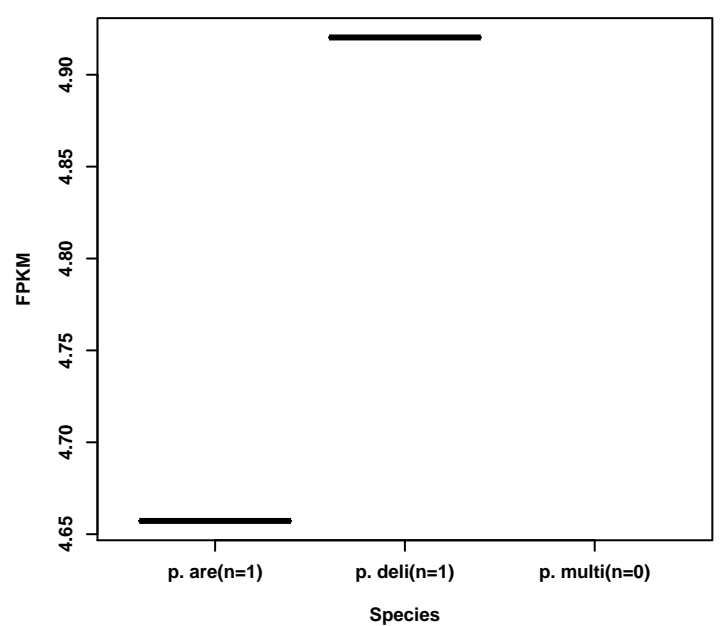
**S-adenosylmethioninamine biosynthesis**



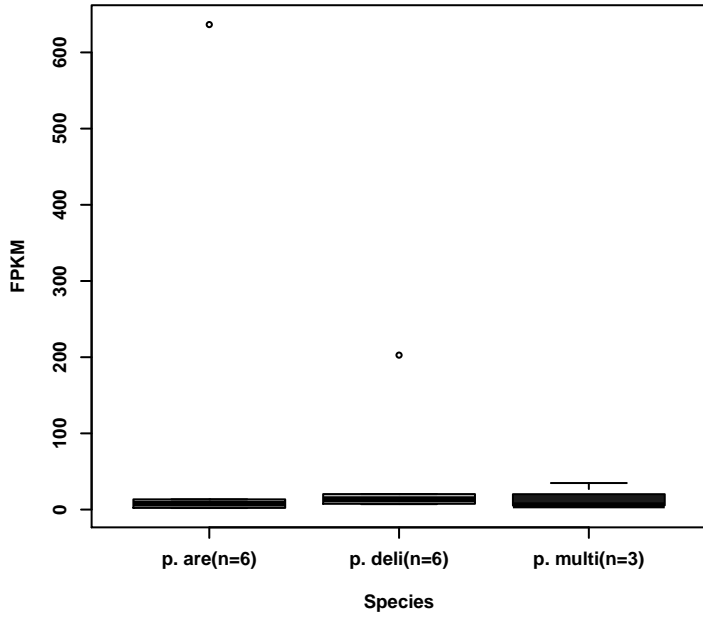
**(S)-allantoin degradation**



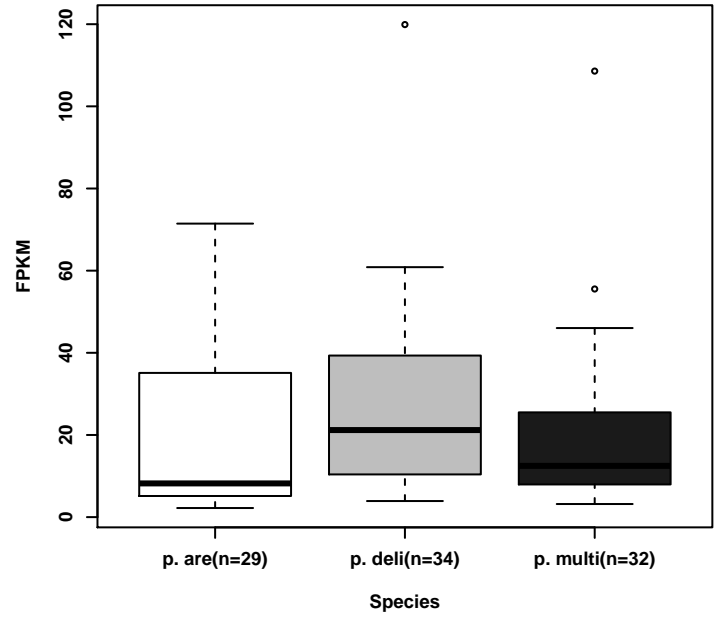
**secologanin biosynthesis**



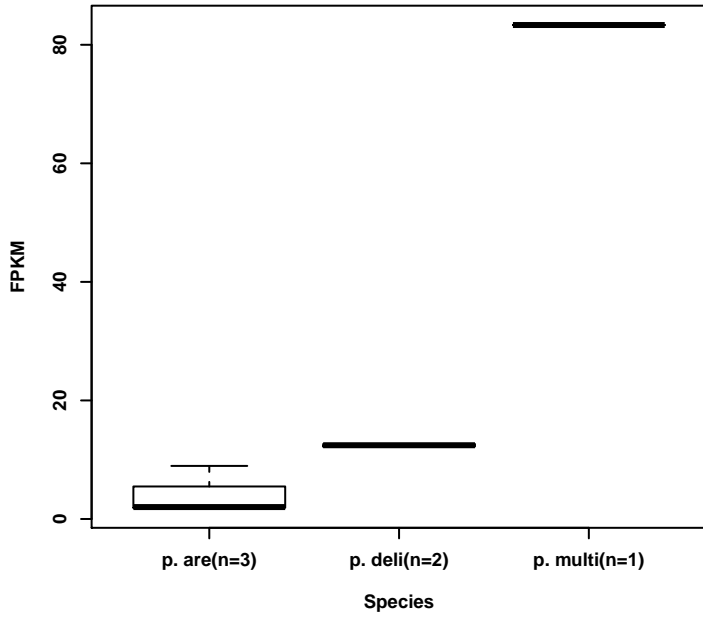
**selenocysteinyl-tRNA(Sec) biosynthesis**



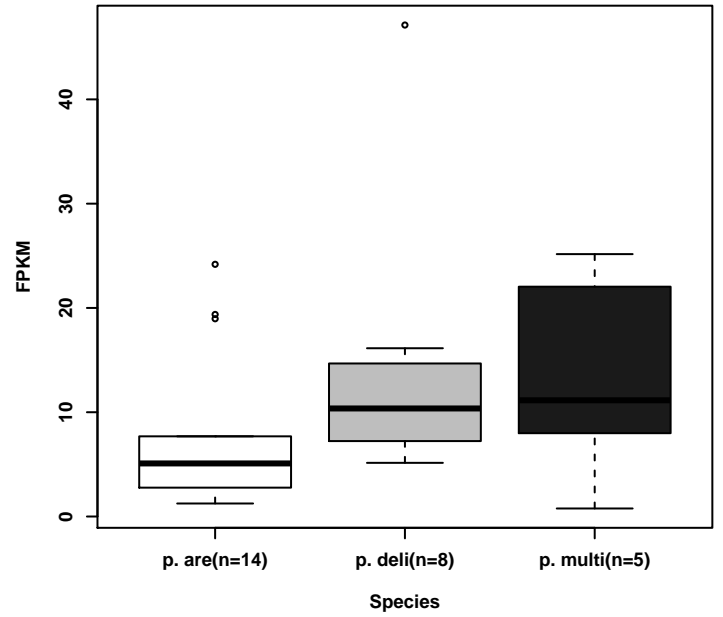
**spermidine biosynthesis**



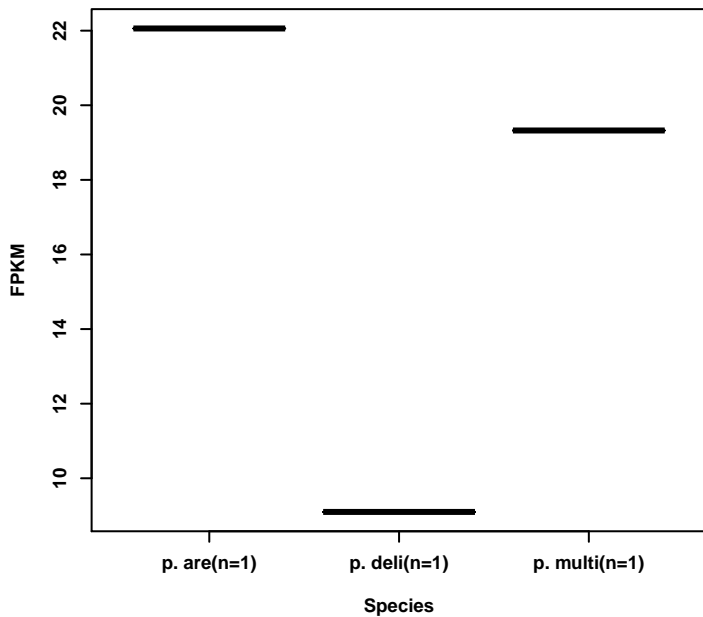
**spheroidene biosynthesis**



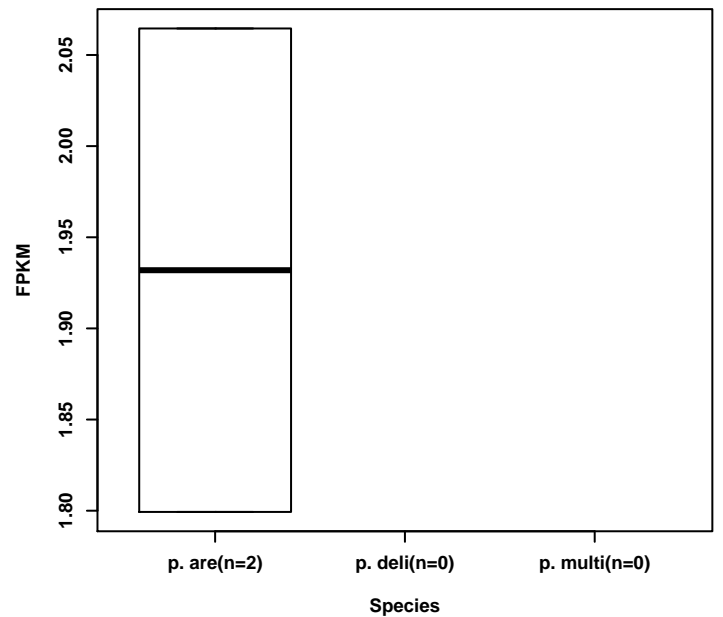
**sphingolipid metabolism**



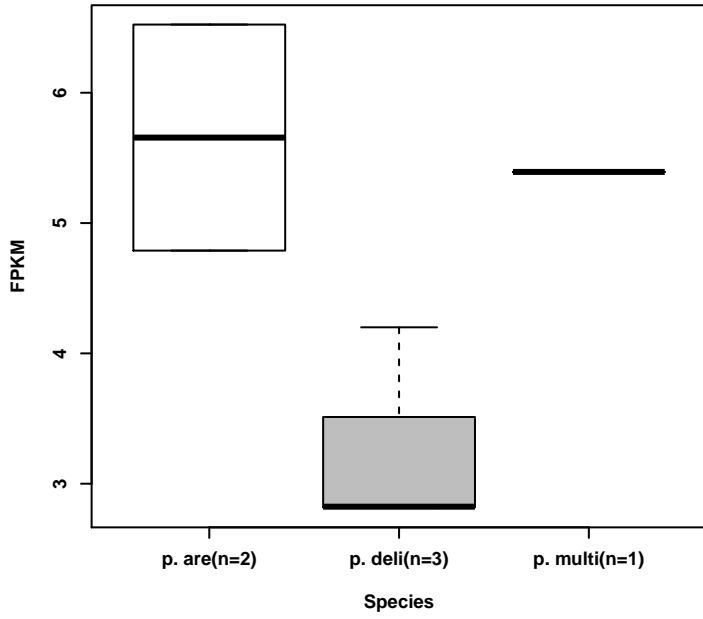
**spore coat polysaccharide biosynthesis**



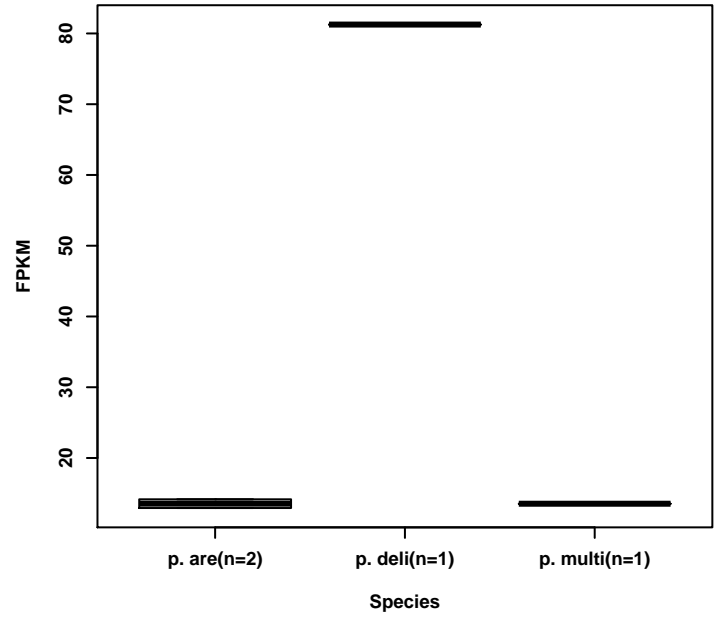
**(S)-reticuline biosynthesis**



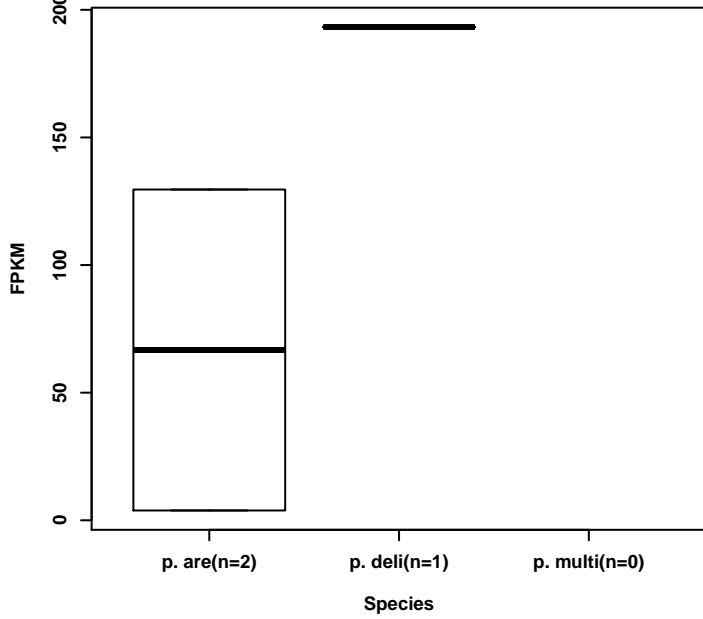
sterigmatocystin biosynthesis



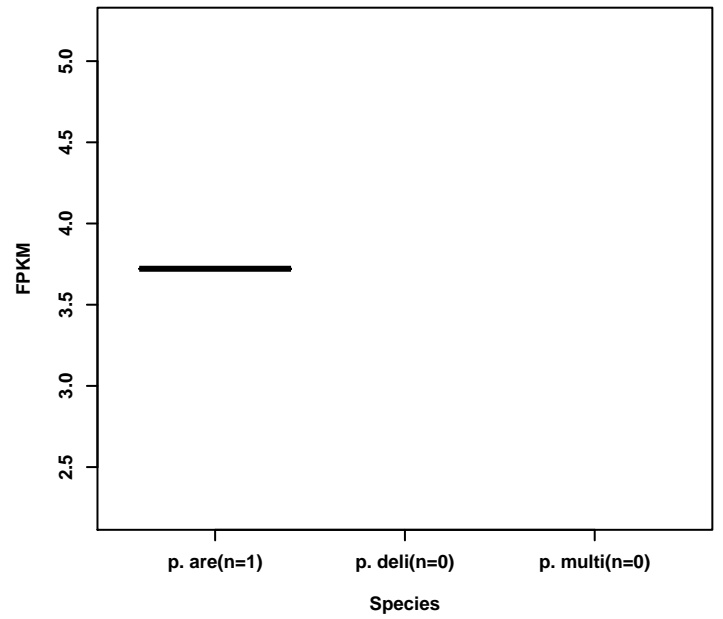
steroid biosynthesis



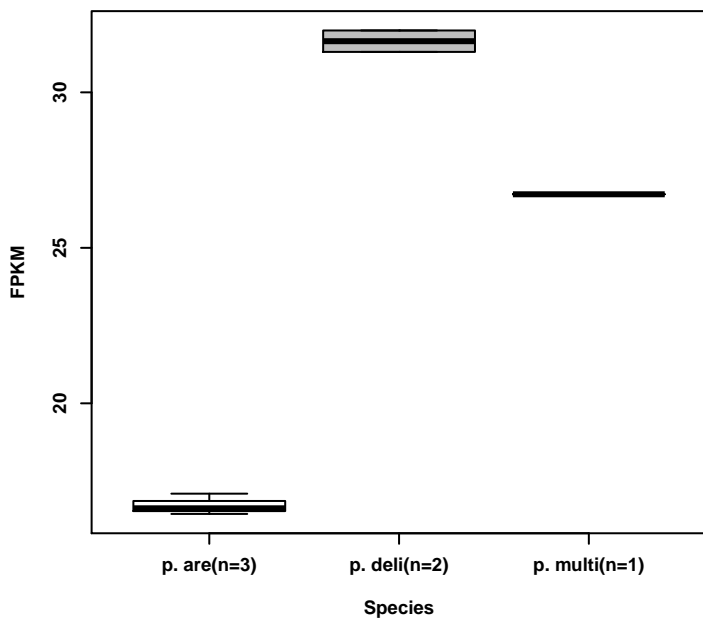
sterol biosynthesis



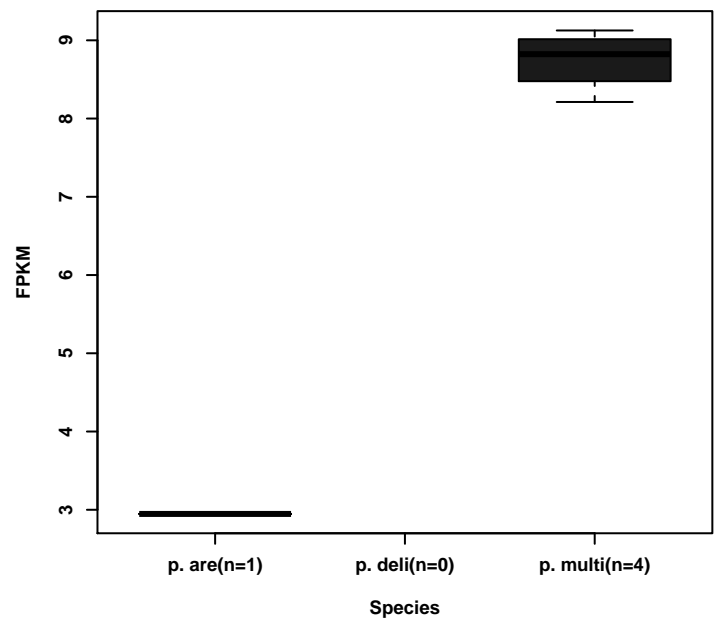
sulfatase oxidation



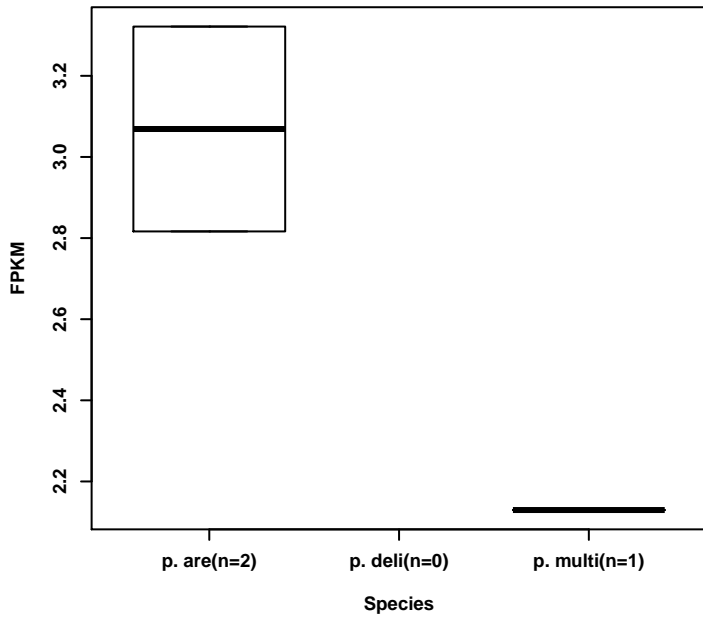
sulfate assimilation



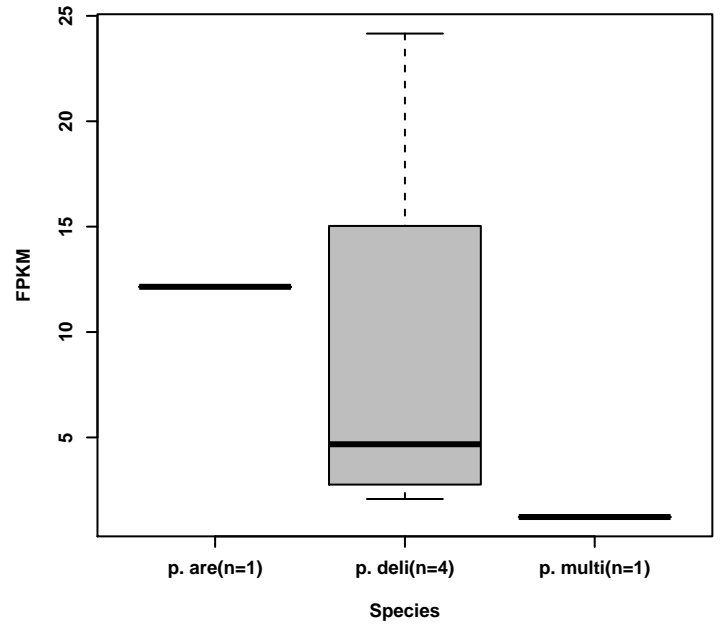
surfactin biosynthesis



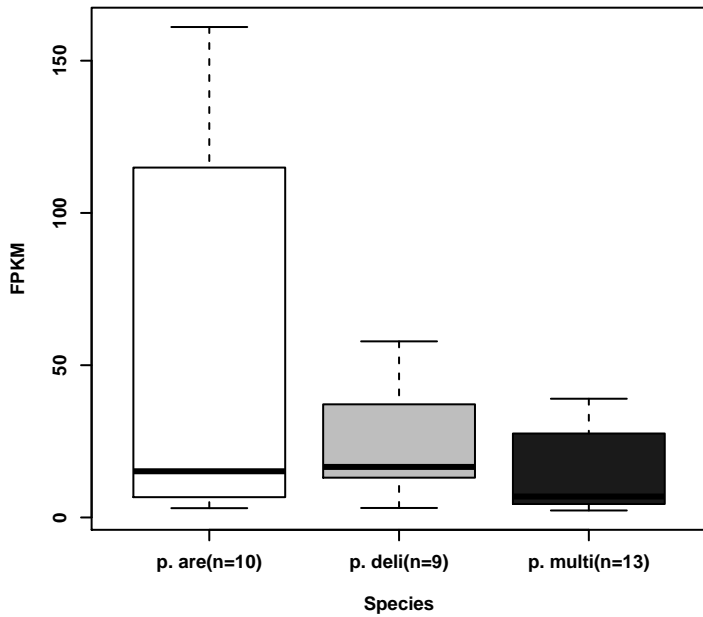
tetracenomycin C biosynthesis



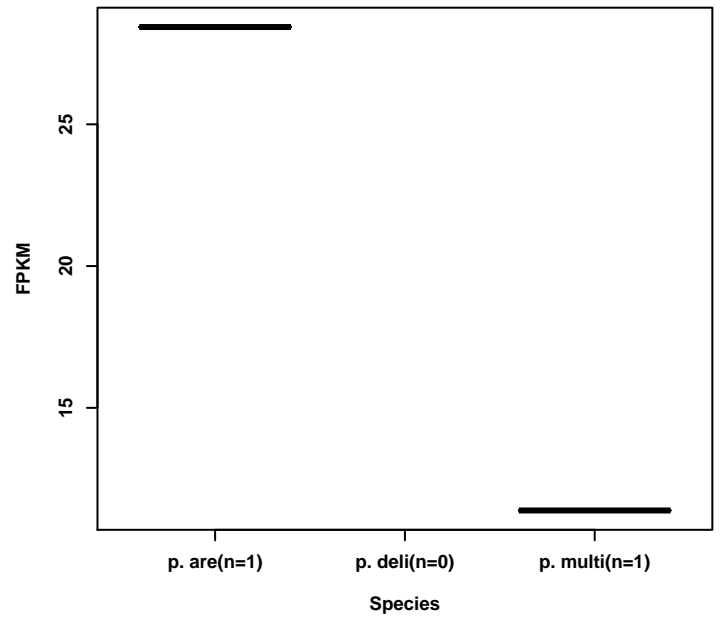
tetrahydrofolate biosynthesis



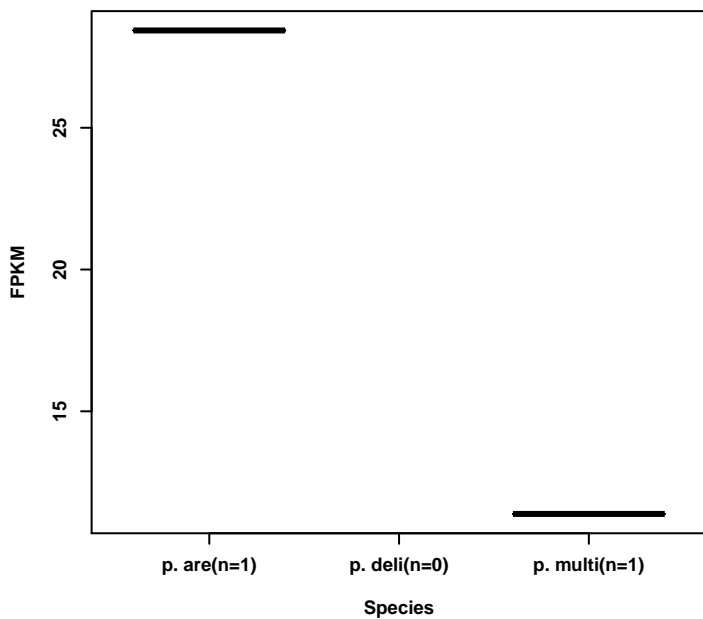
tetrahydrofolate interconversion



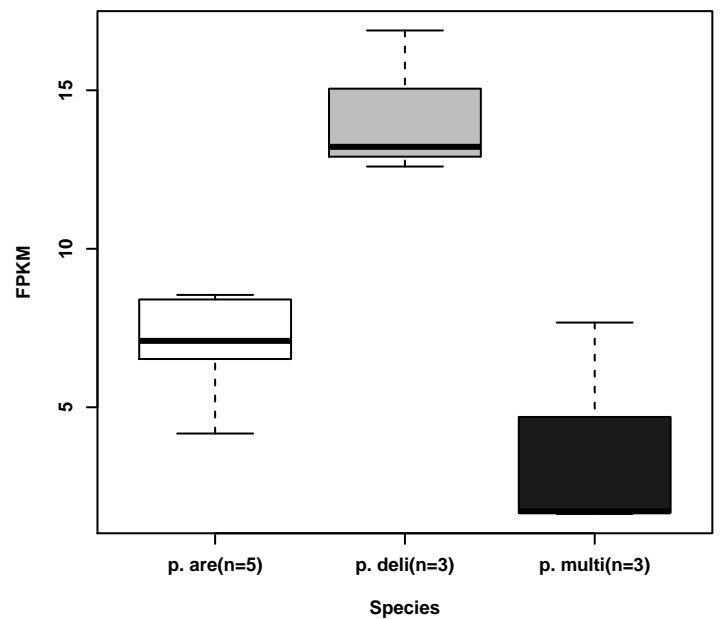
thiamine degradation



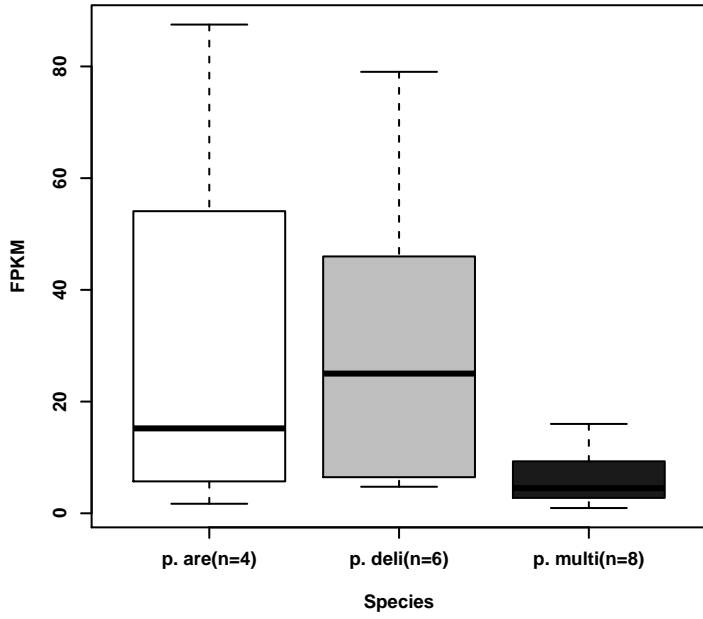
thiamine degradation [regulation]



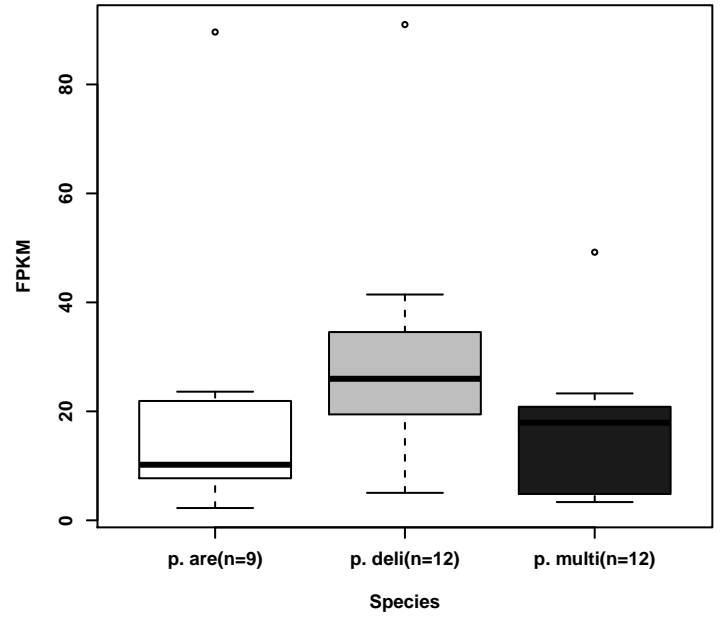
thiamine diphosphate biosynthesis



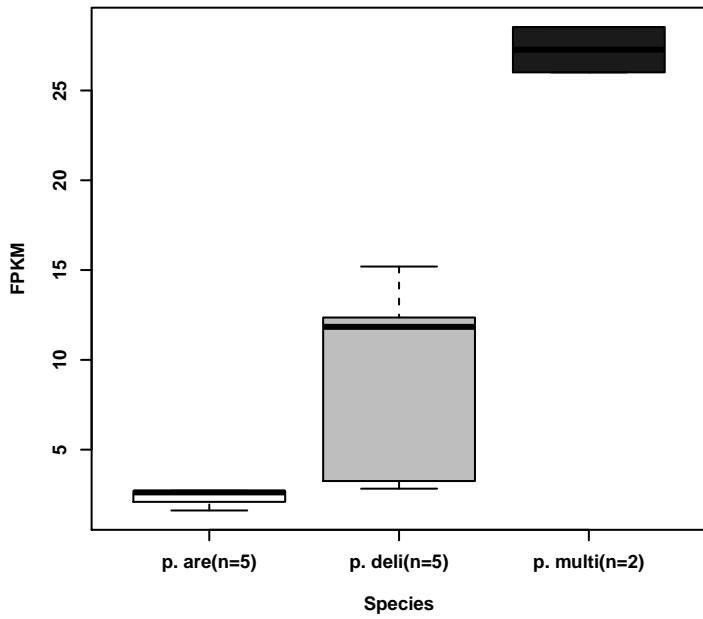
**tocopherol biosynthesis**



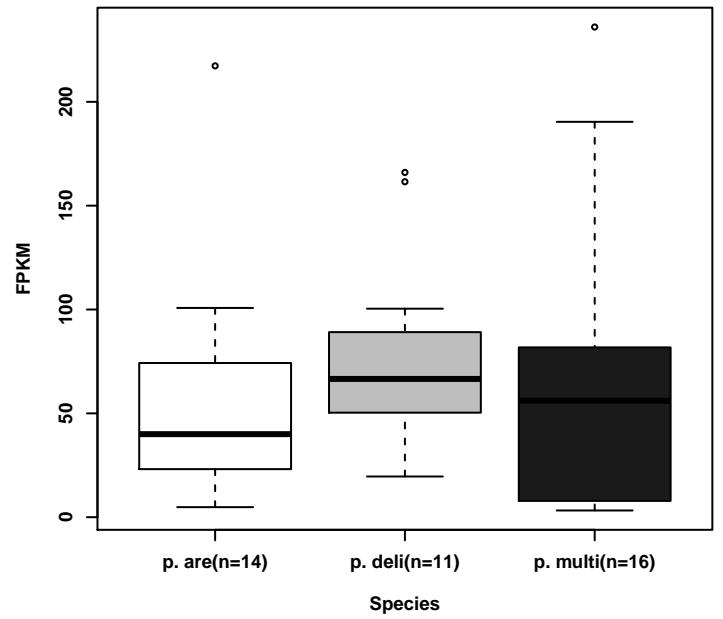
**triacylglycerol biosynthesis**



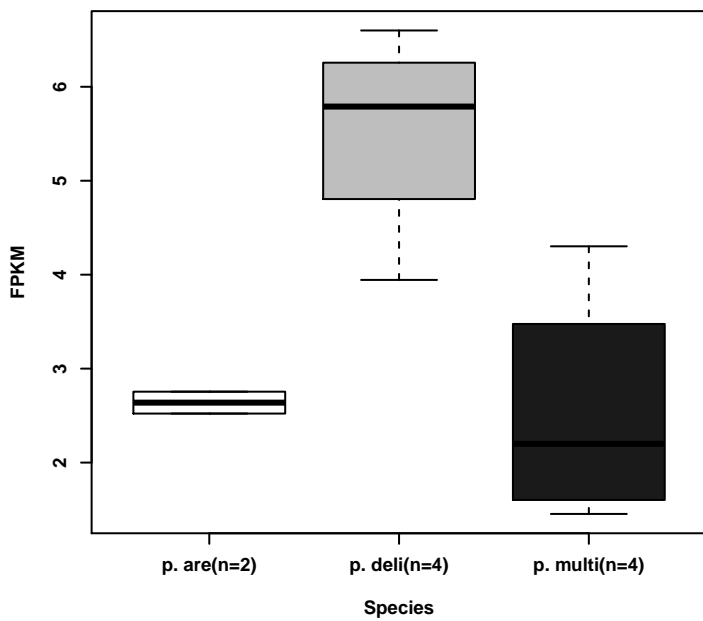
**triacylglycerol degradation**



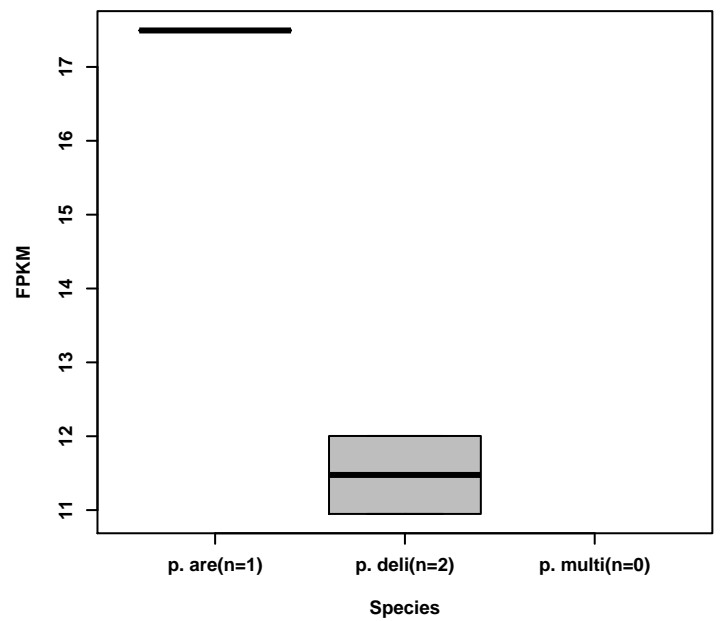
**tricarboxylic acid cycle**

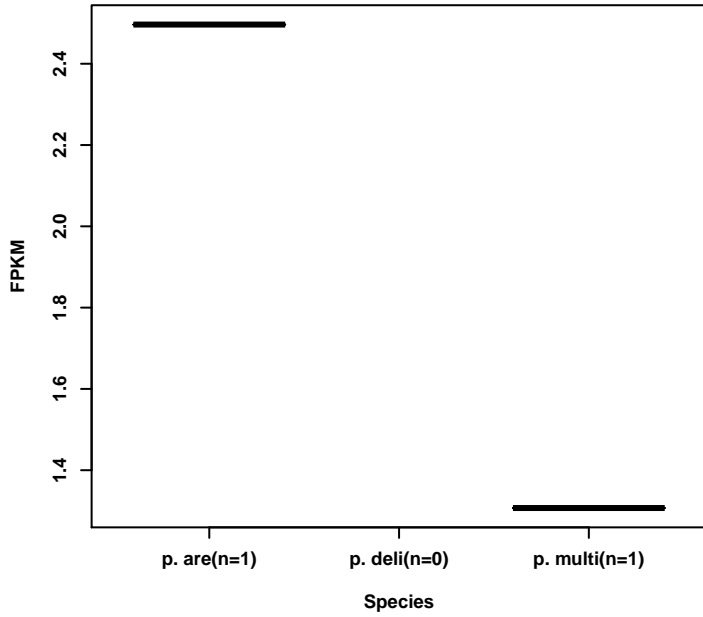
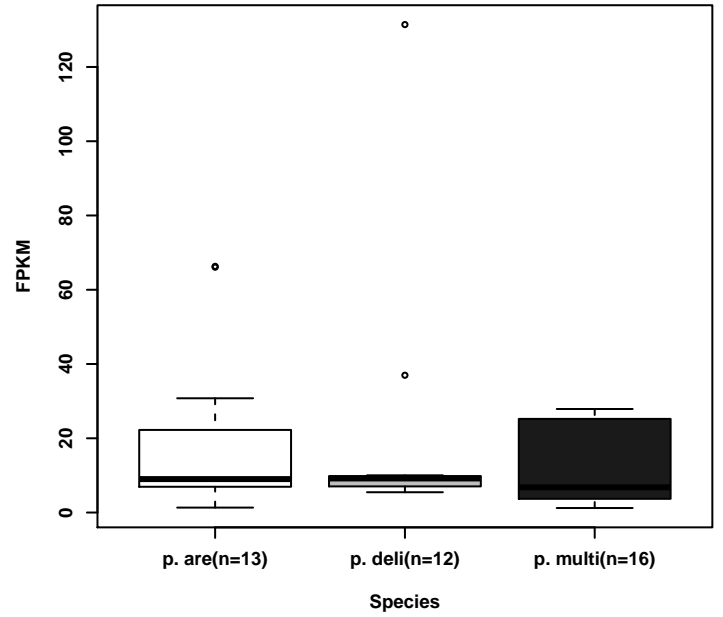
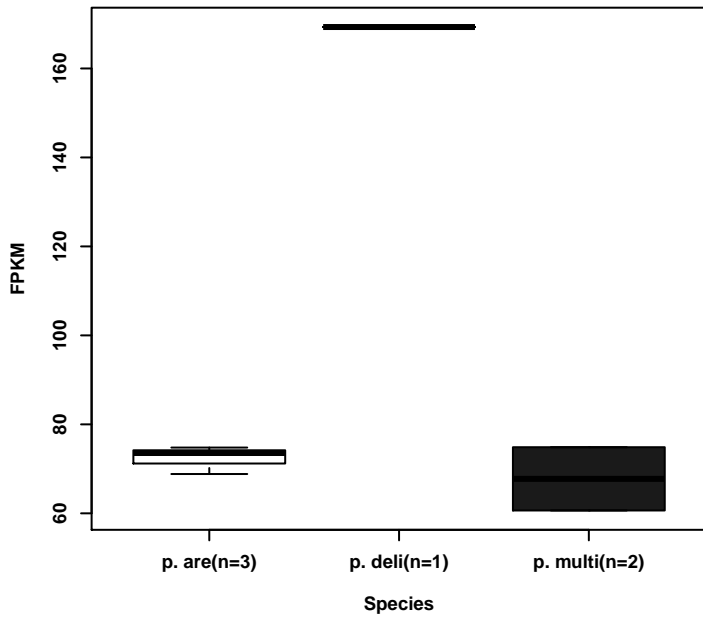
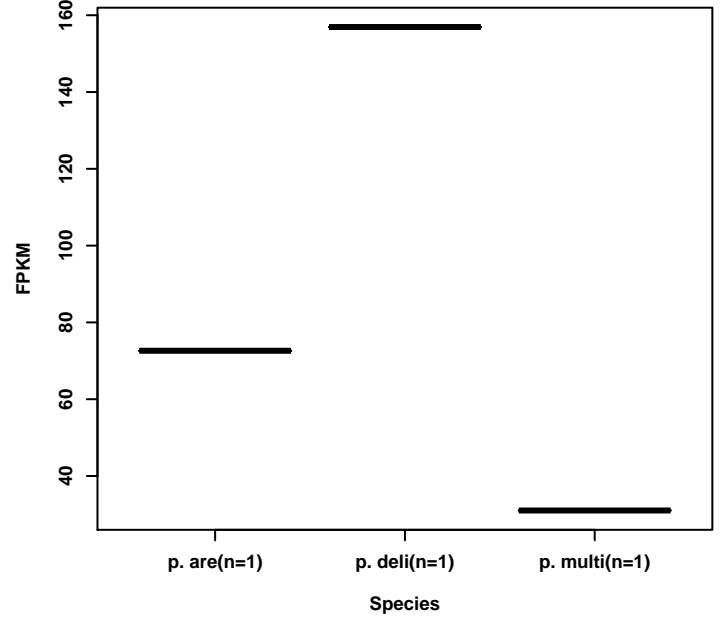
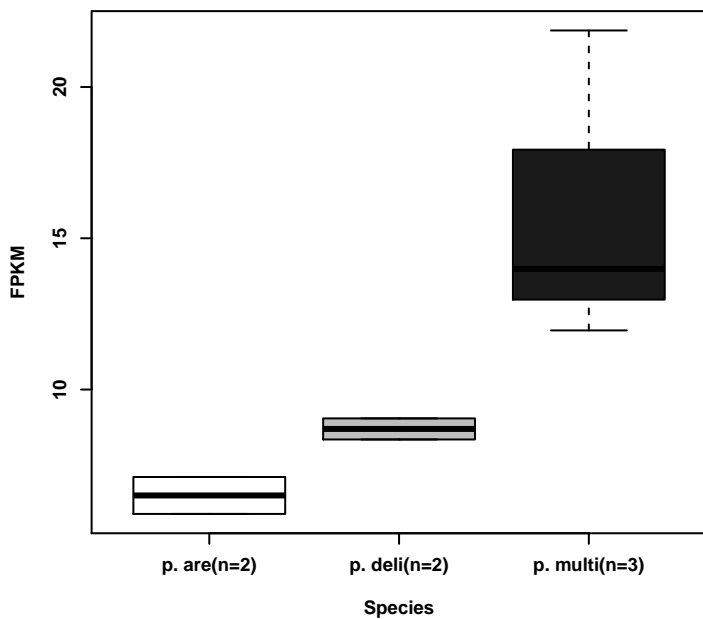
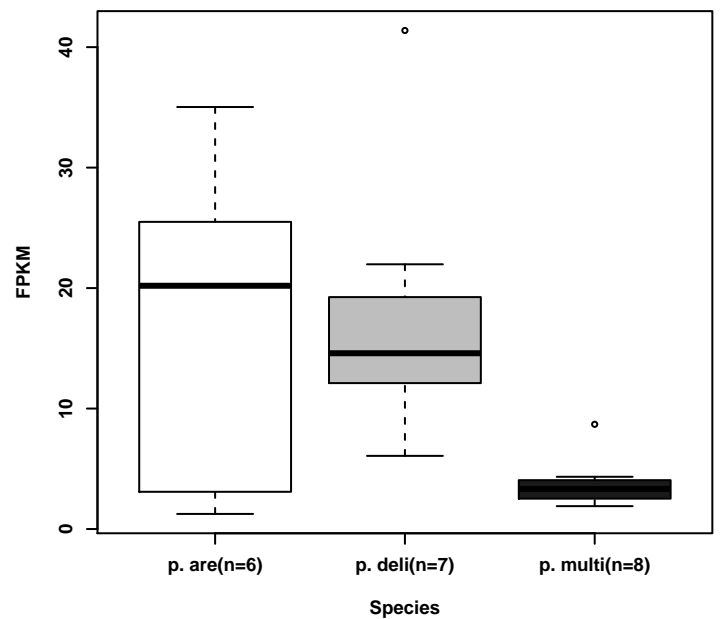


**tRNA-queuosine biosynthesis**

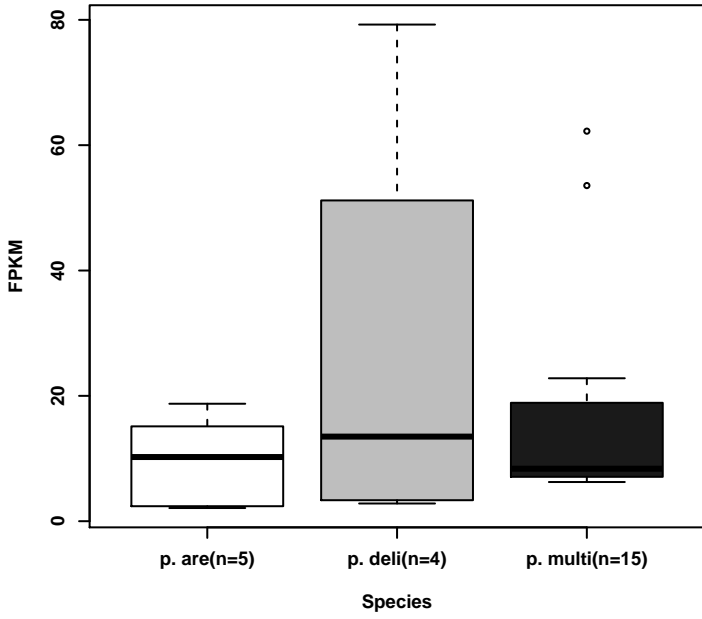


**tryptophan metabolism**

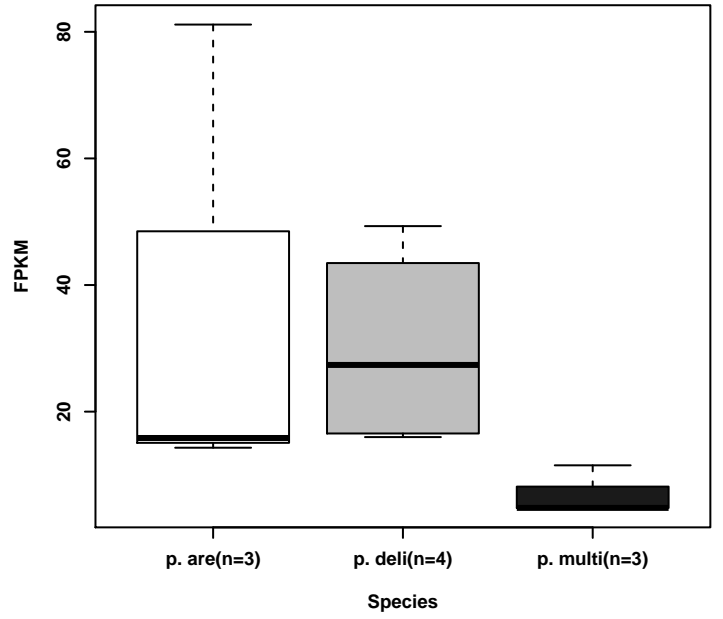


**tyrocidine biosynthesis****ubiquinone biosynthesis****UDP-alpha-D-glucuronate biosynthesis****UDP-alpha-D-xylose biosynthesis****UDP-N-acetyl-alpha-D-glucosamine biosynthesis****UMP biosynthesis via de novo pathway**

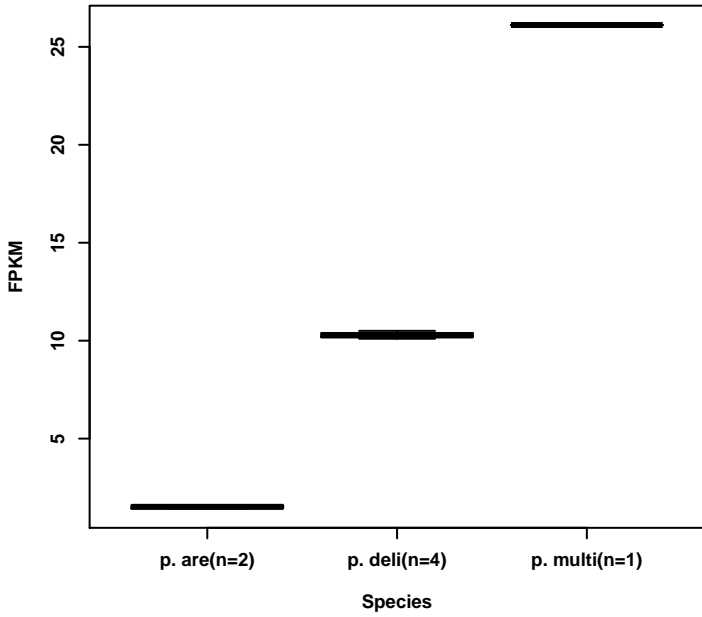
UMP biosynthesis via salvage pathway



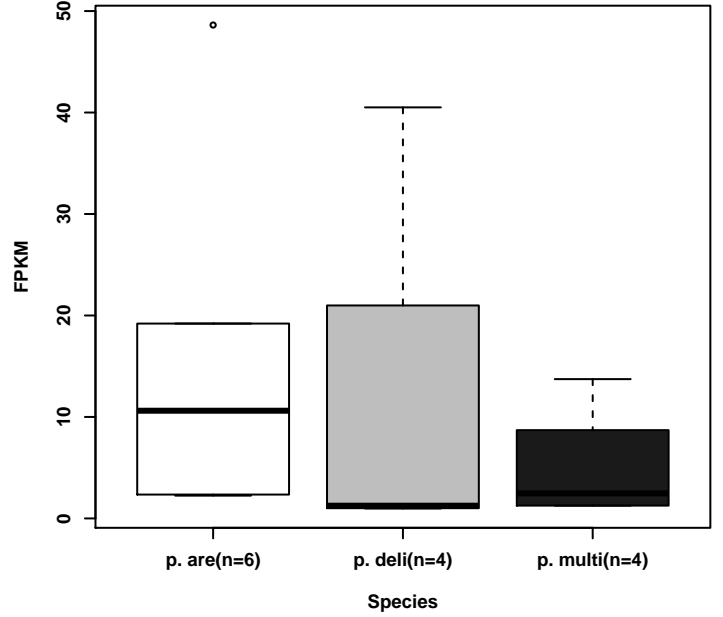
urea cycle



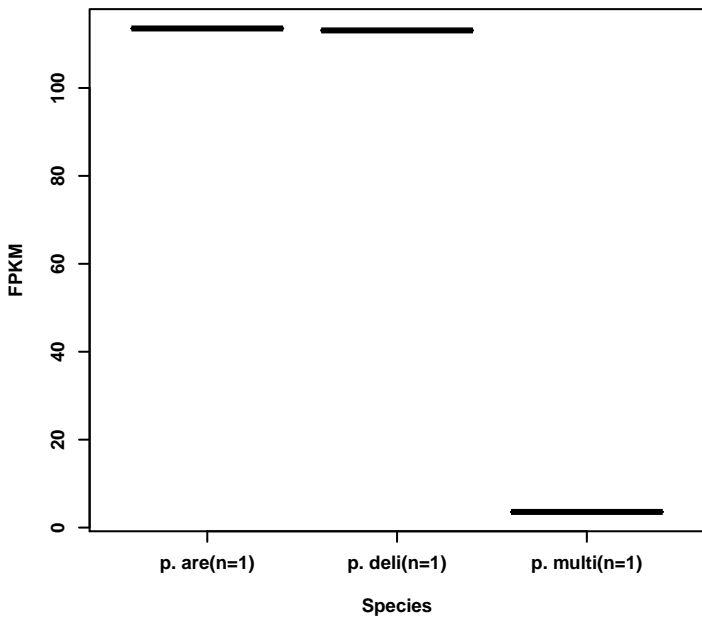
urea degradation



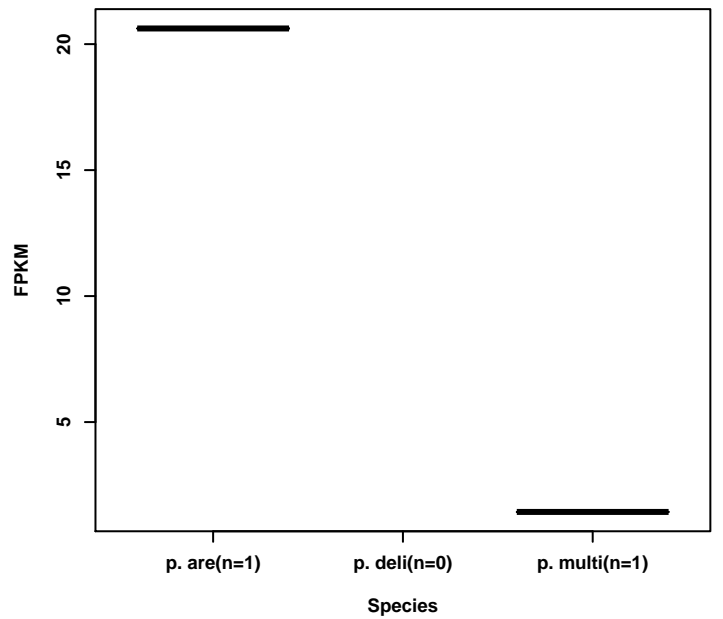
wybutosine-tRNA(Phe) biosynthesis



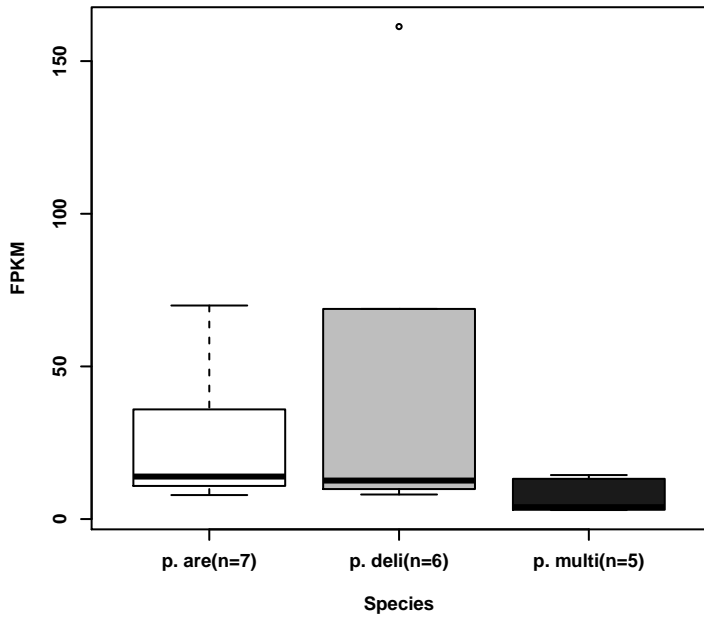
XMP biosynthesis via de novo pathway



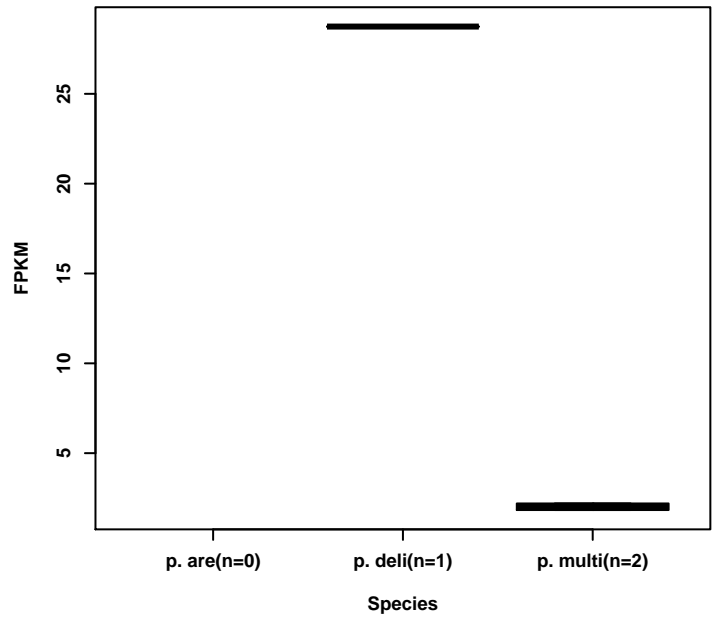
XMP biosynthesis via salvage pathway



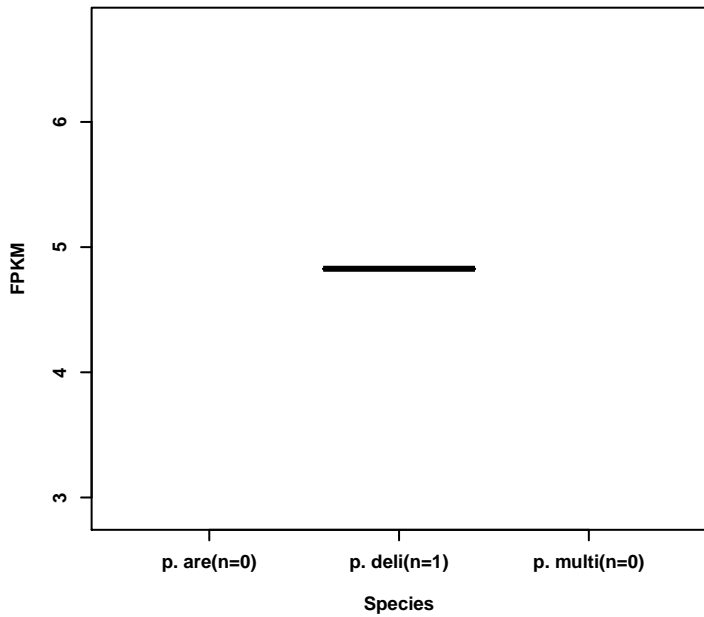
**zymosterol biosynthesis**



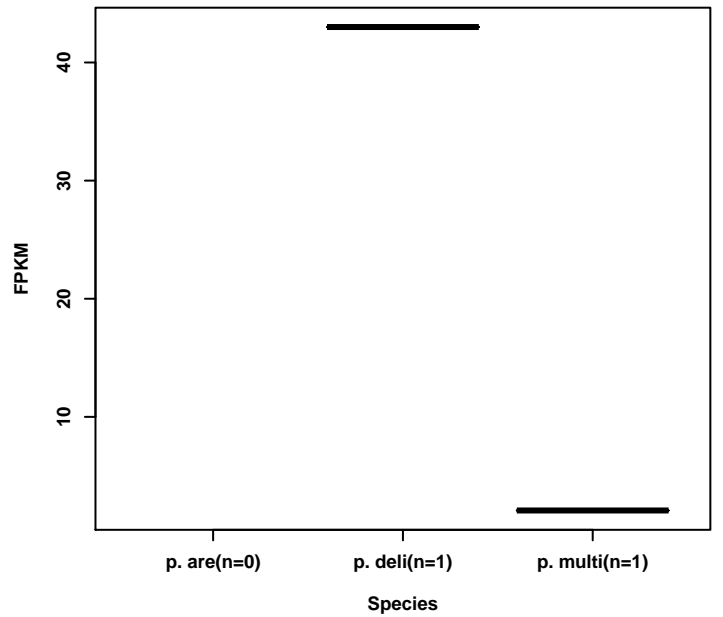
**7,8-dihydroneopterin triphosphate biosynthesis**



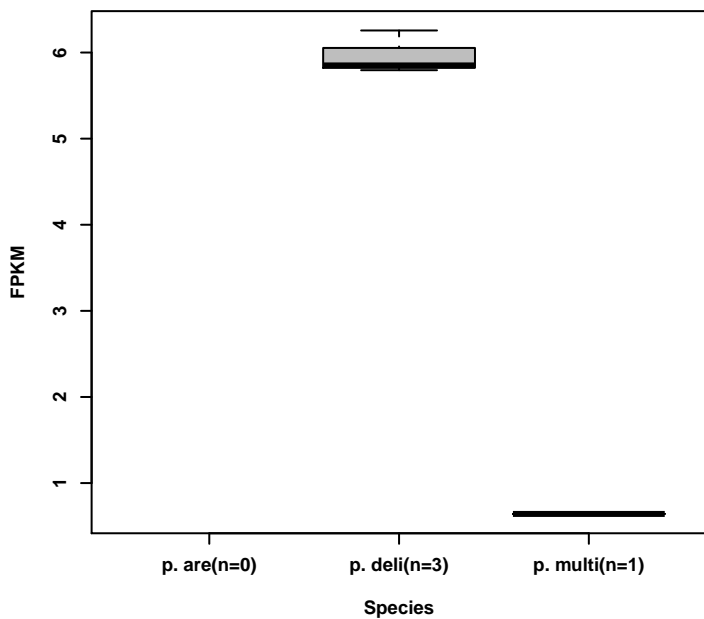
**bacillibactin biosynthesis**



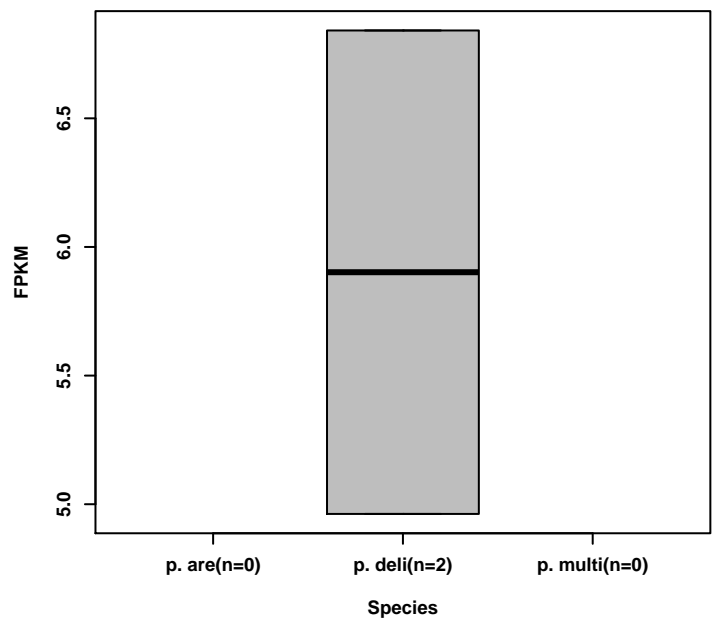
**brassinosteroid biosynthesis**



**candididin biosynthesis**

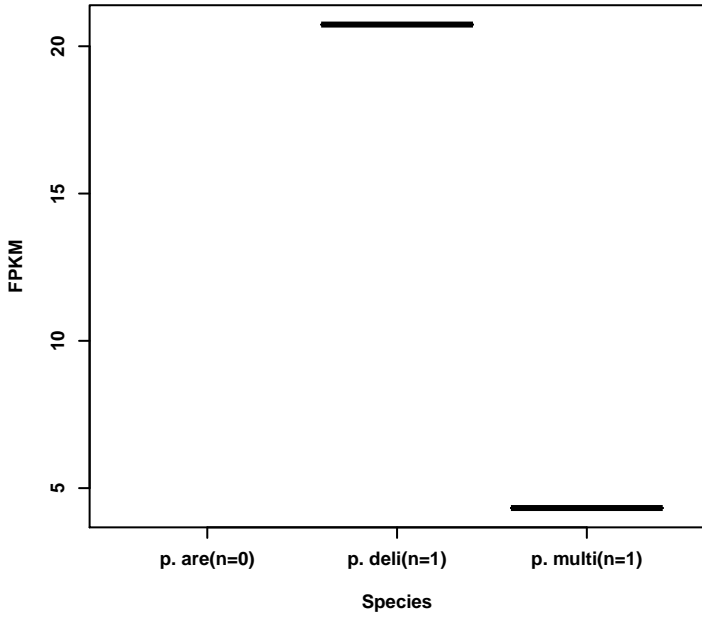


**capsule polysaccharide biosynthesis**

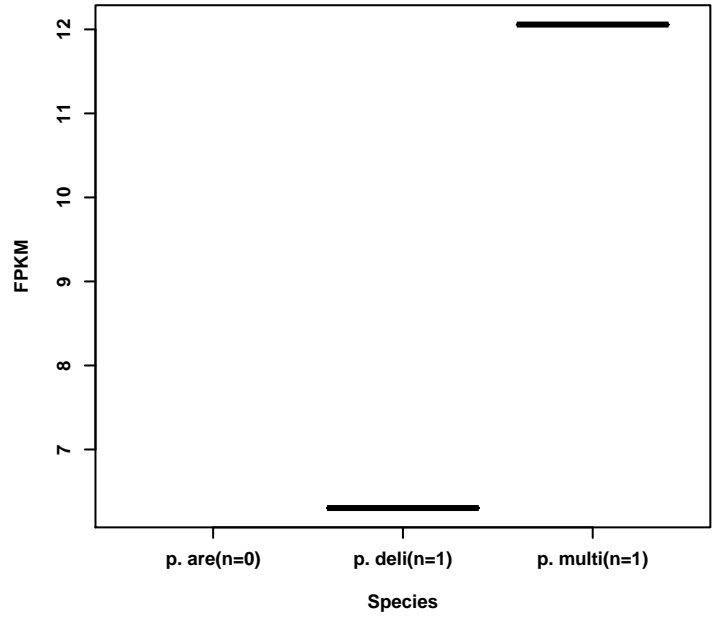




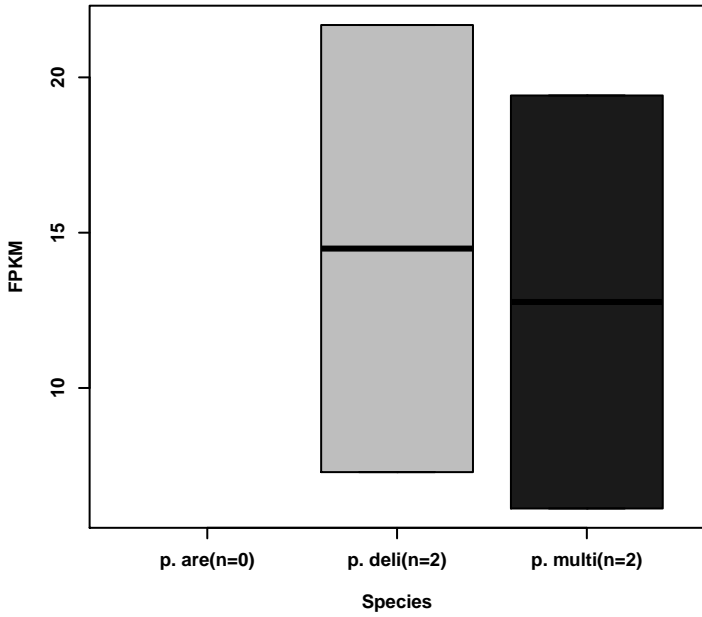
chondroitin sulfate biosynthesis



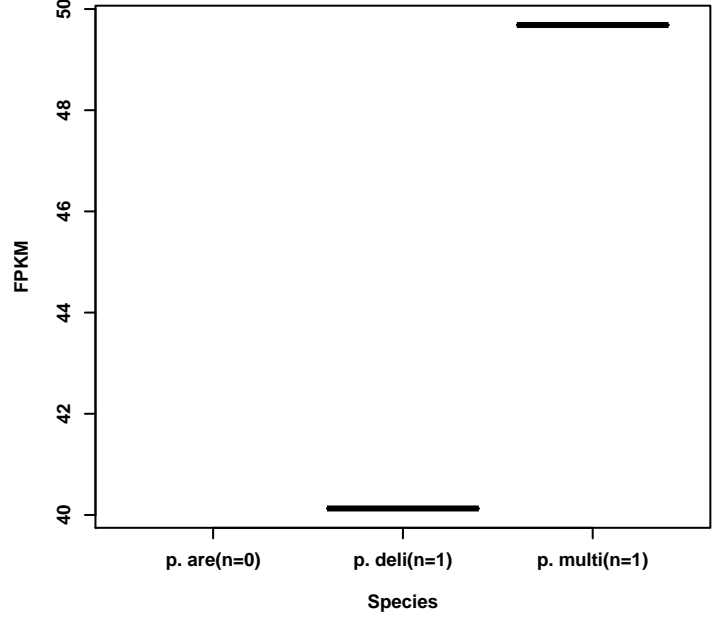
D-galactarate degradation



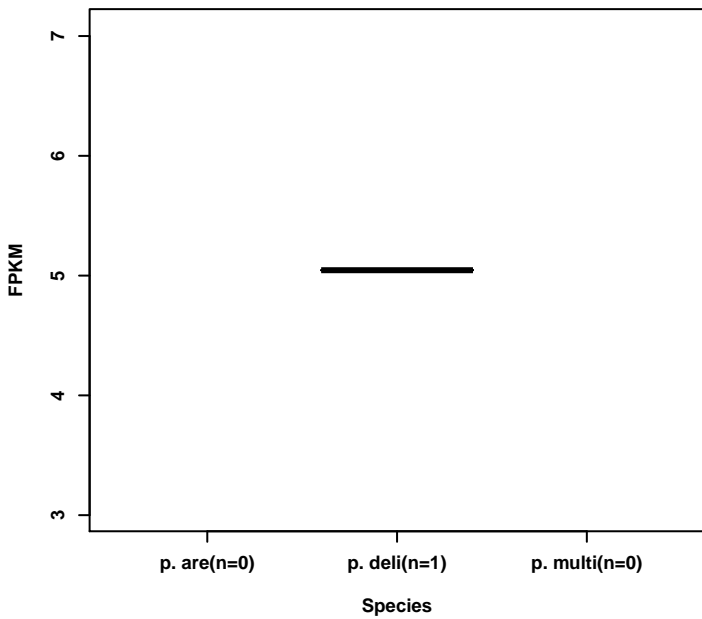
D-xylose degradation



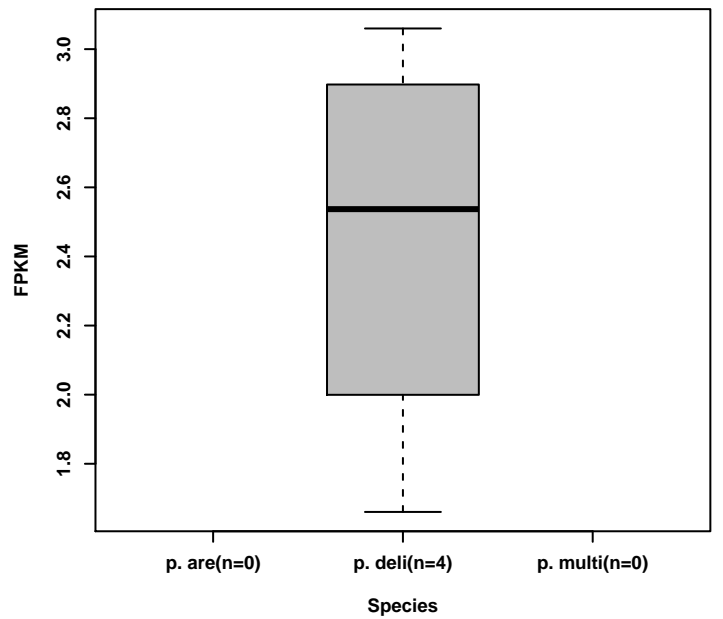
ergothioneine biosynthesis



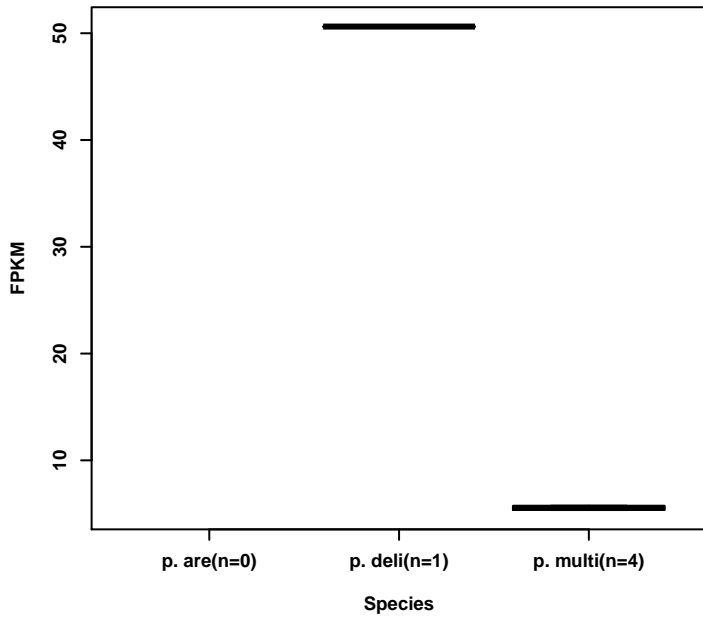
granaticin biosynthesis



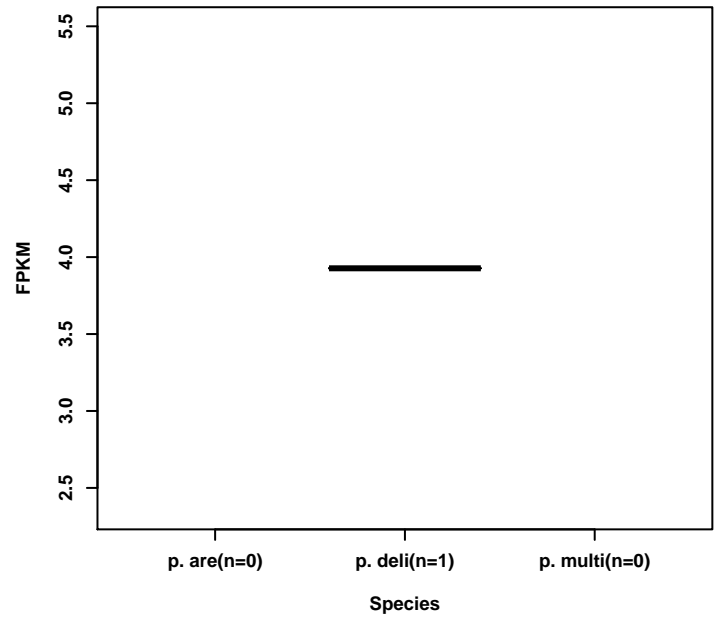
leukotriene A4 biosynthesis



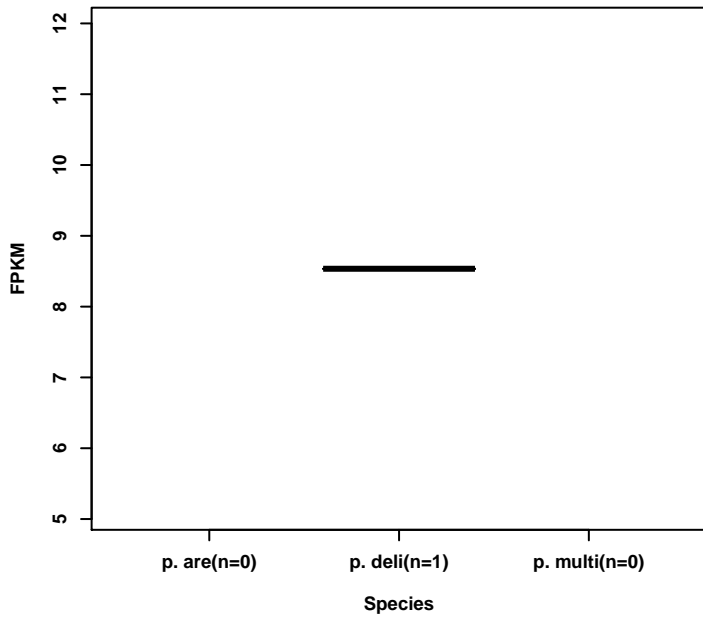
L-proline degradation into L-glutamate



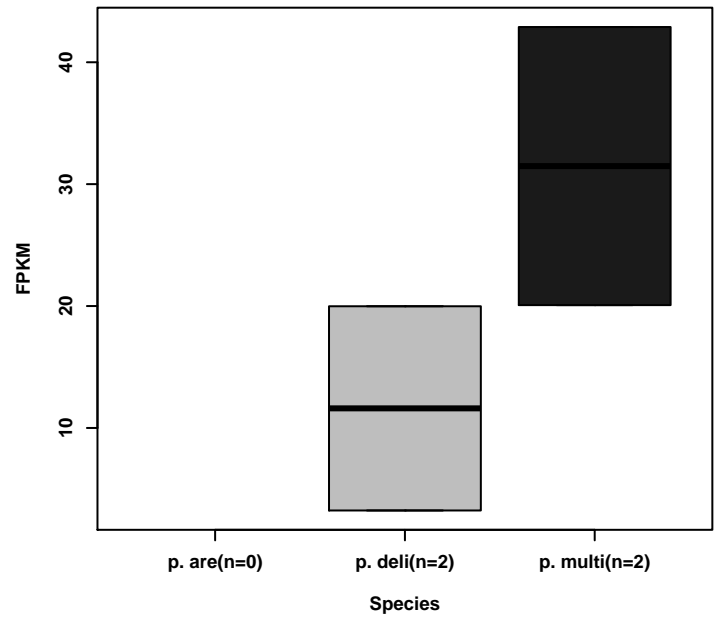
L-tryptophan degradation via kynurenine pathway



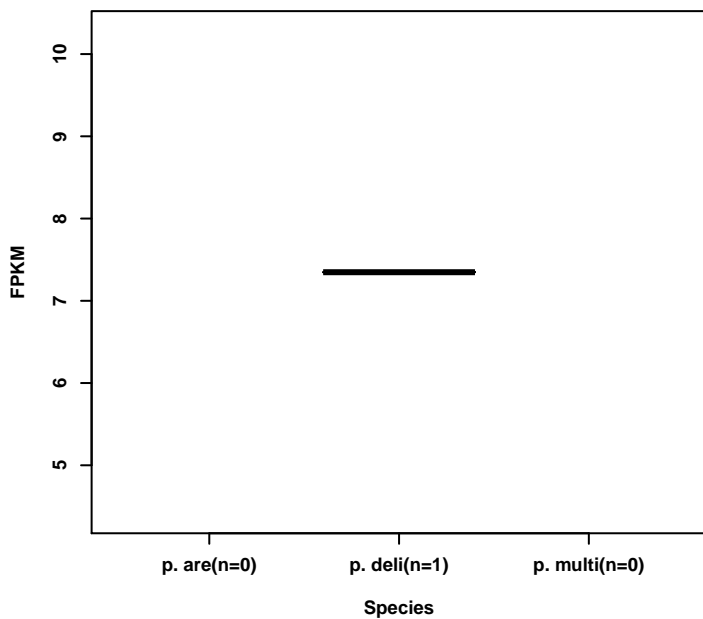
mycinamicin biosynthesis



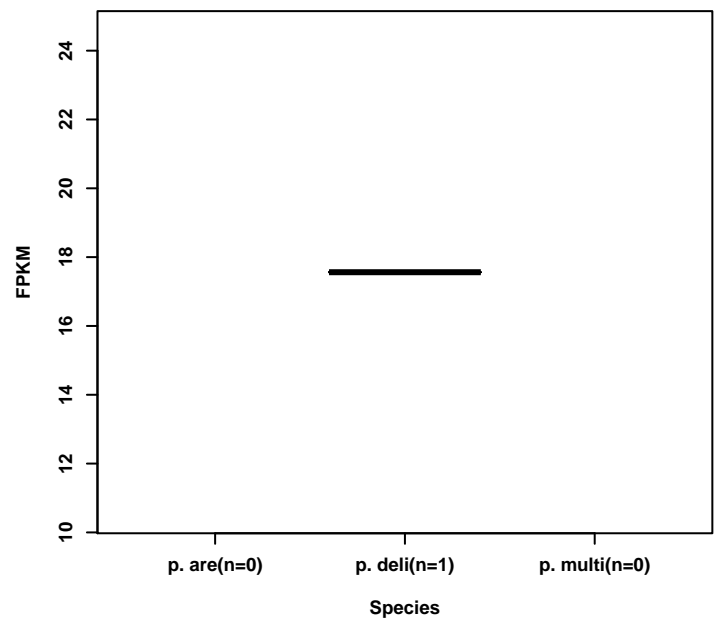
nitrate reduction (assimilation)



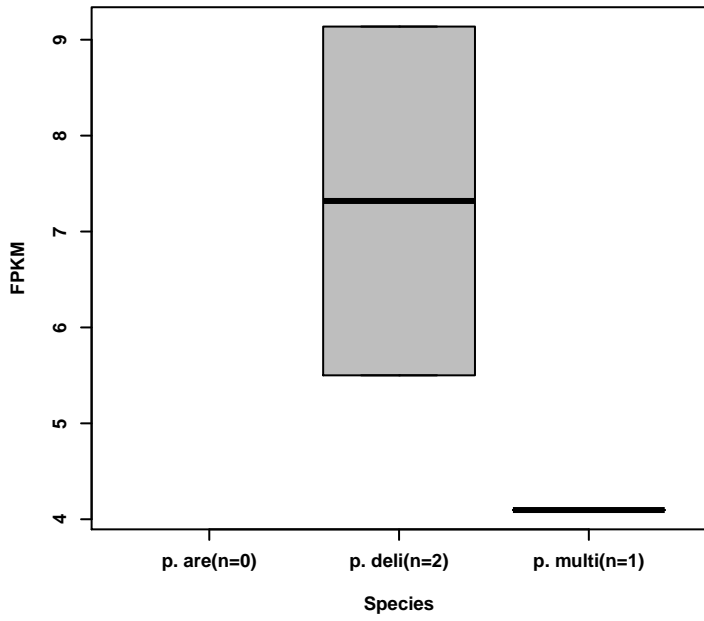
oxylipin biosynthesis



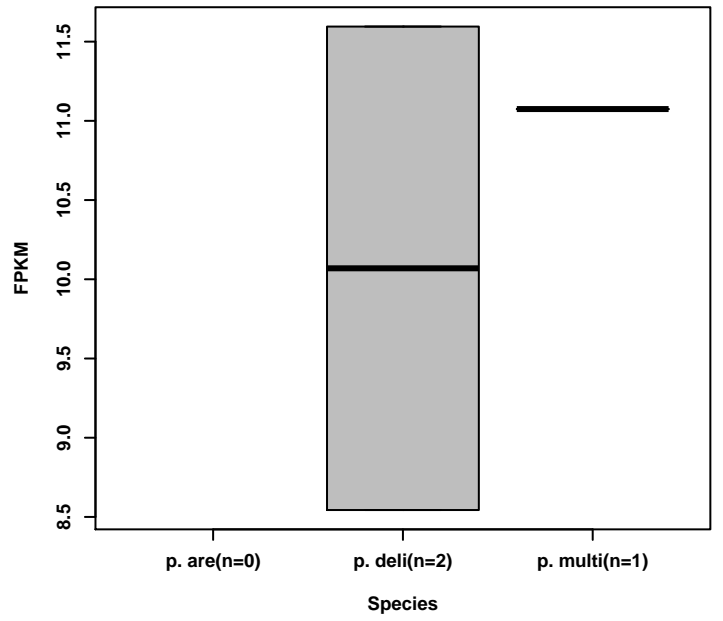
spermine degradation



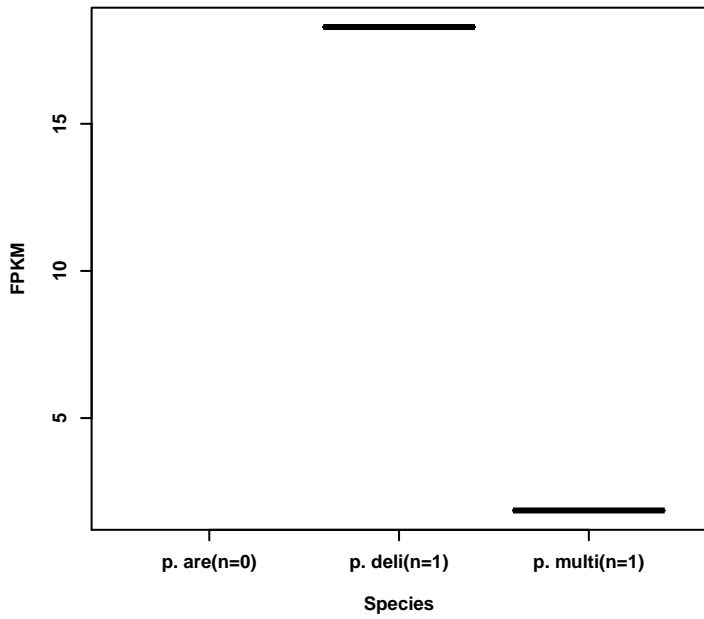
**streptomycin biosynthesis**



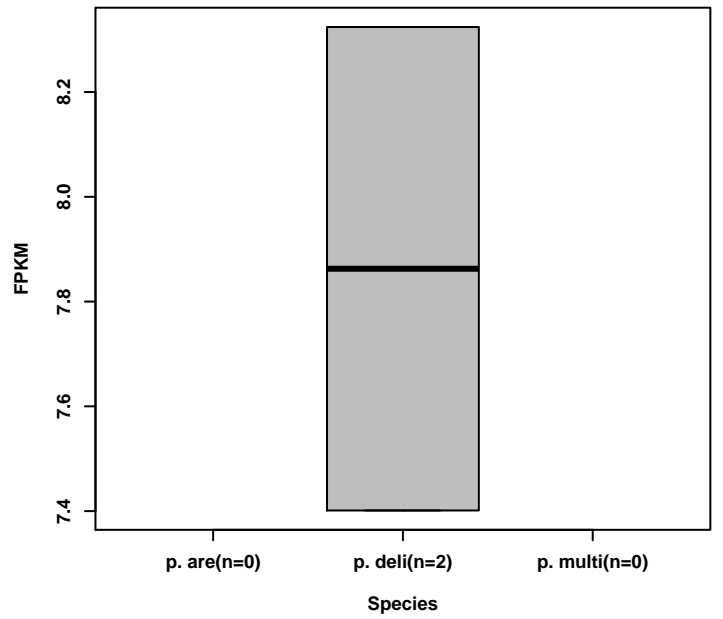
**taurine biosynthesis**



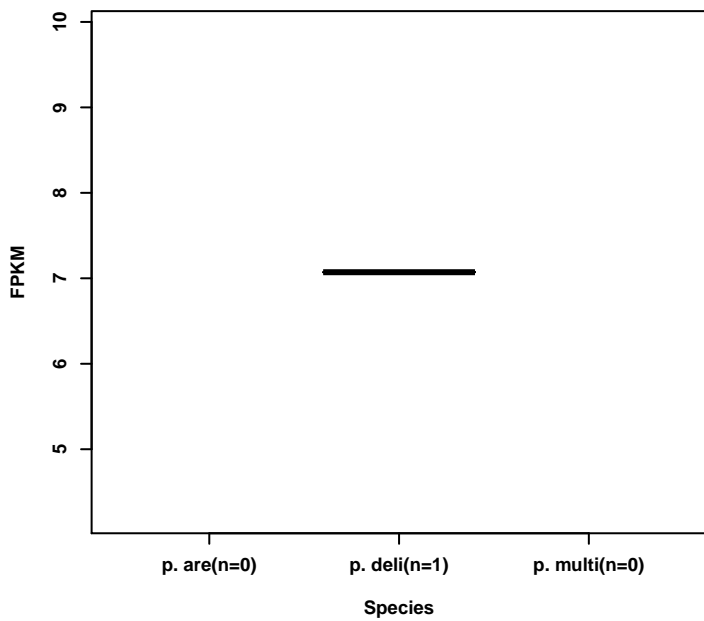
**tetrahydrofolylpolyglutamate biosynthesis**



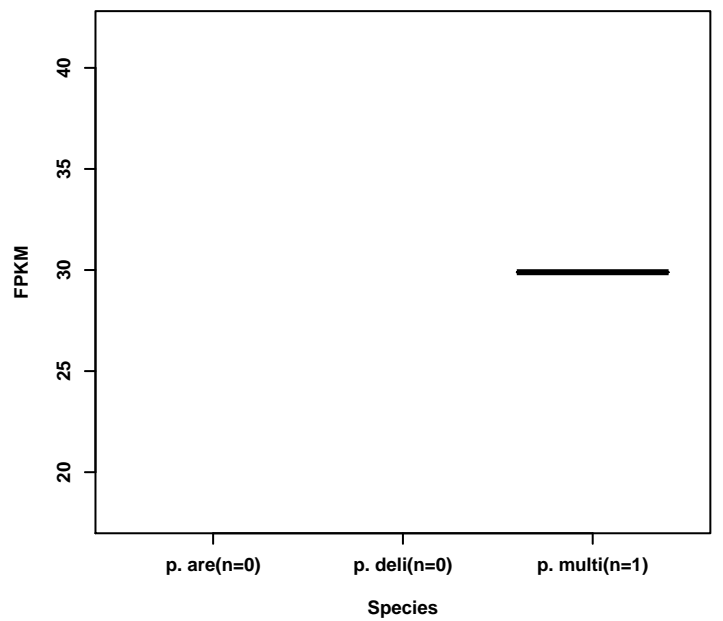
**thyroid hormone biosynthesis**



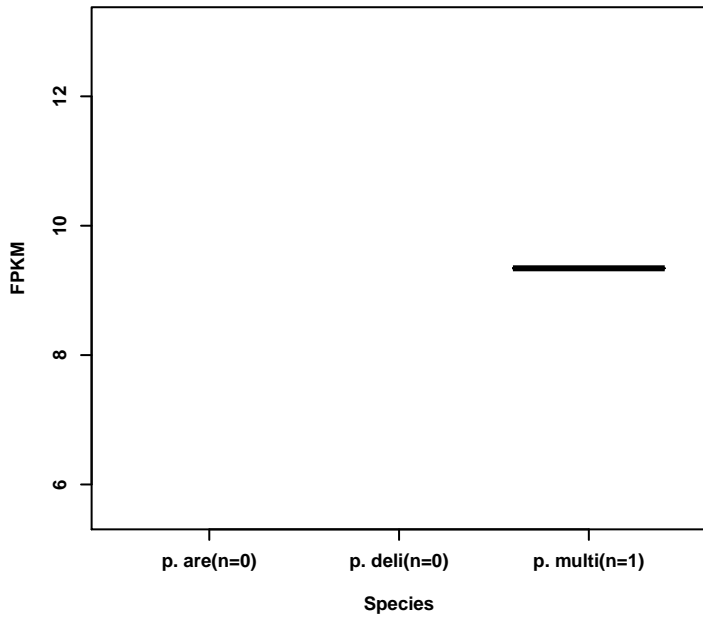
**xylan degradation**



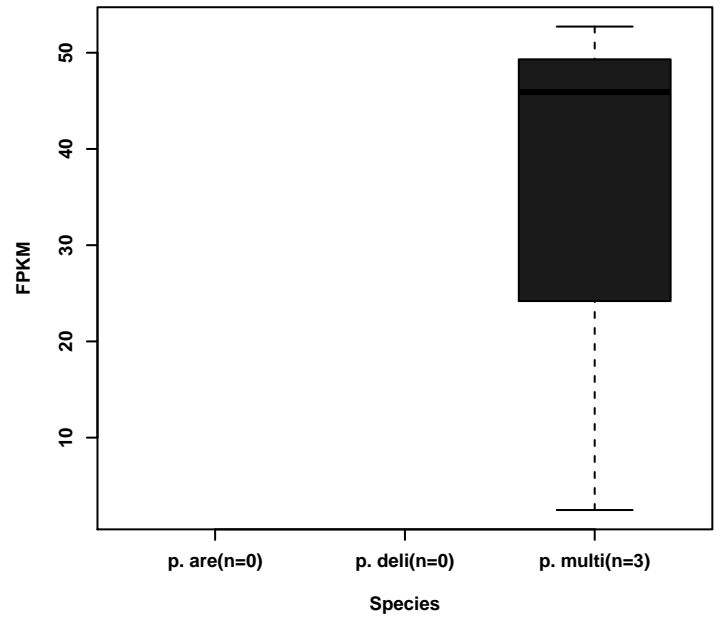
**2-dehydro-3-deoxy-D-gluconate degradation**



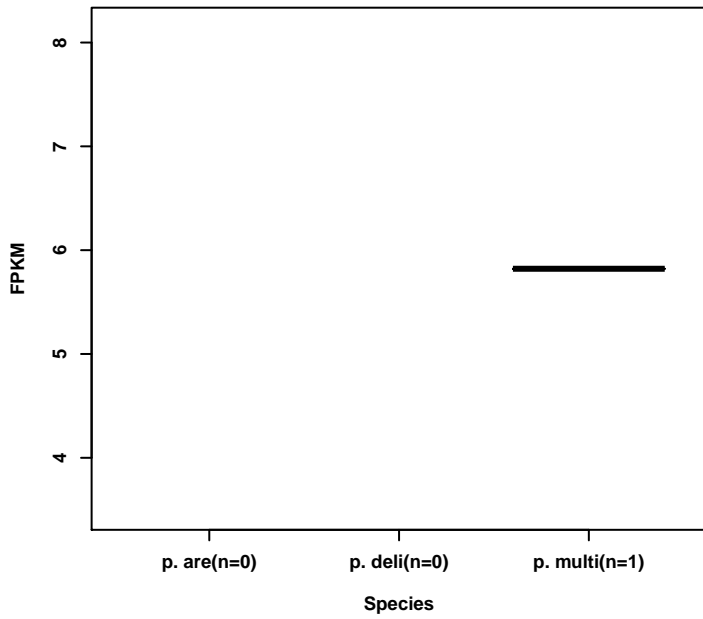
3',5'-cyclic GMP degradation



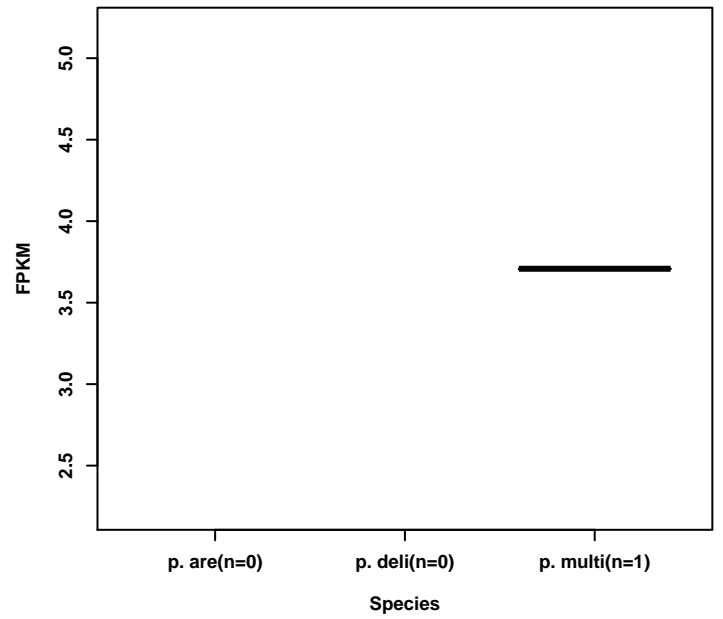
ADP-L-glycero-beta-D-manno-heptose biosynthesis



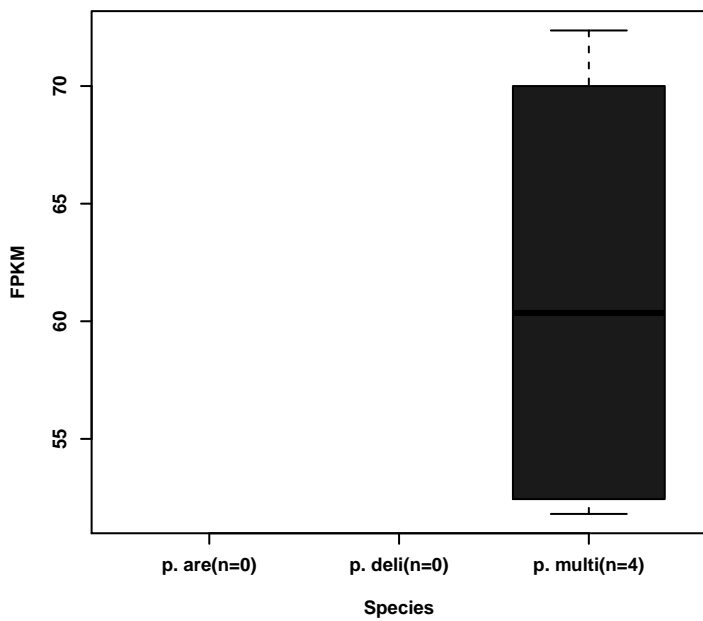
alkane degradation



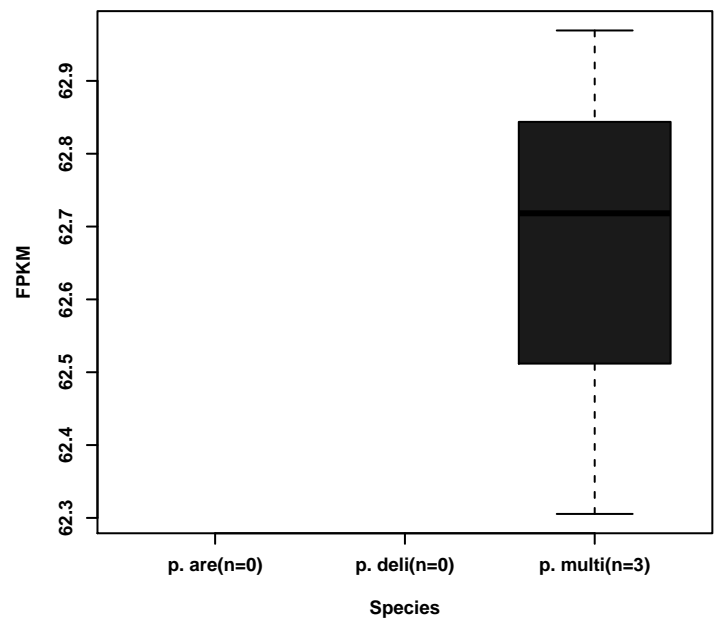
auxin biosynthesis



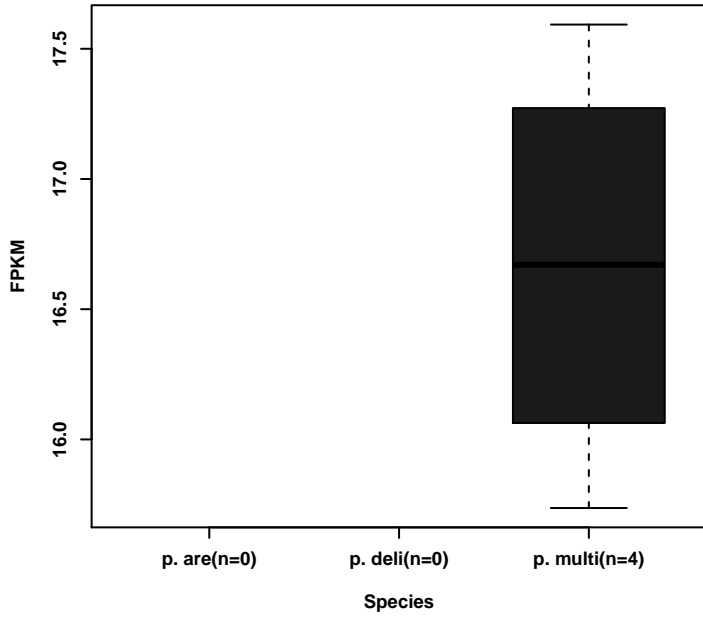
bacteriochlorophyll biosynthesis (light-independent)



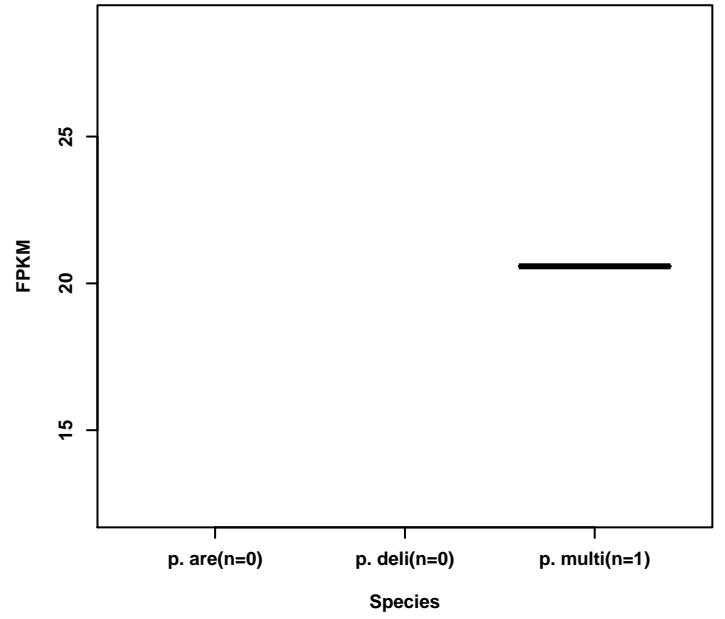
beta-ketoadipate pathway



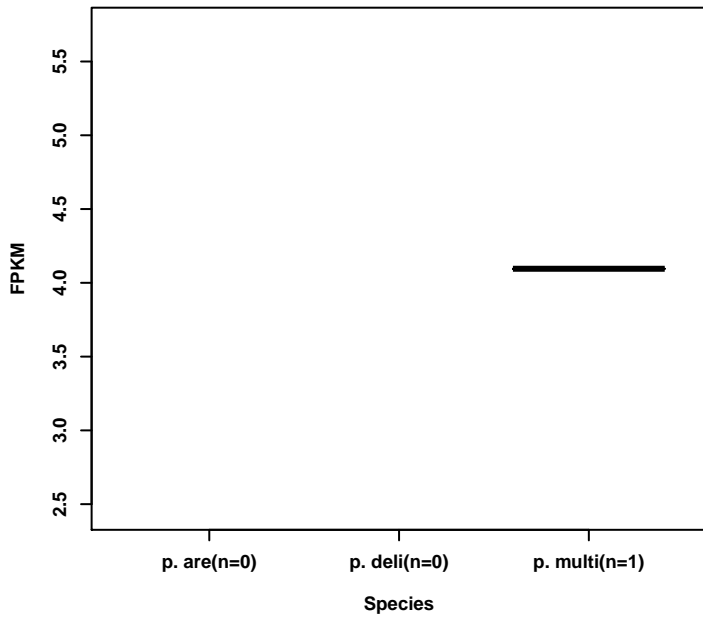
clavulanate biosynthesis



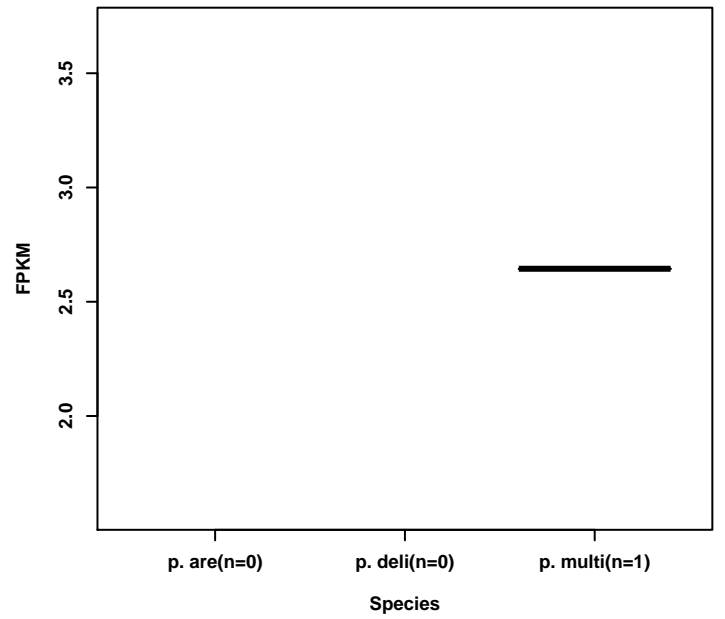
codeine degradation



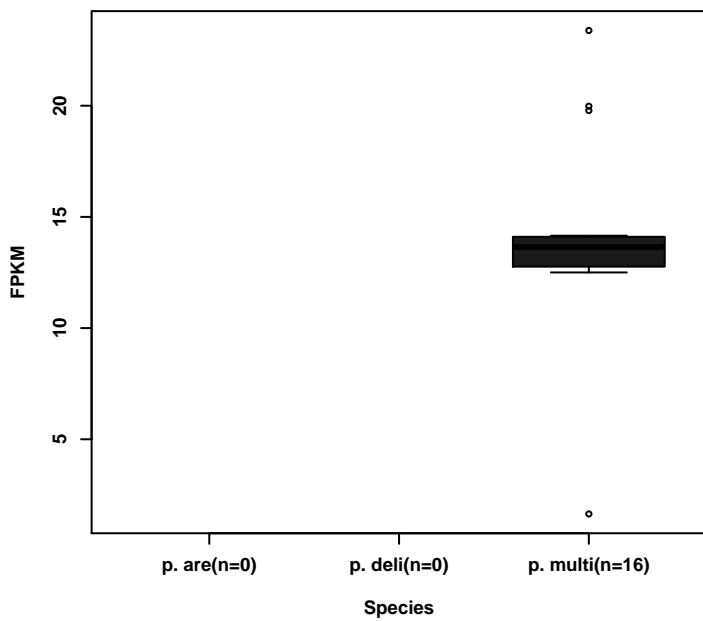
D-alanine degradation



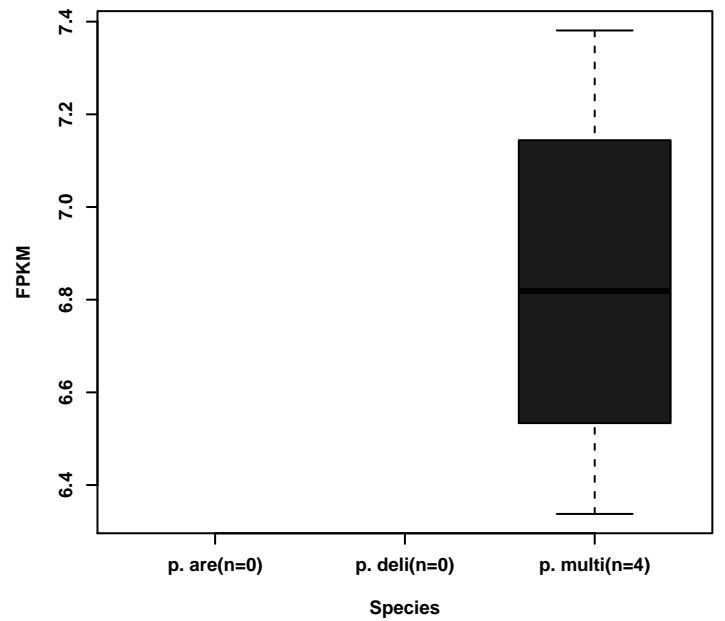
D-sorbitol biosynthesis



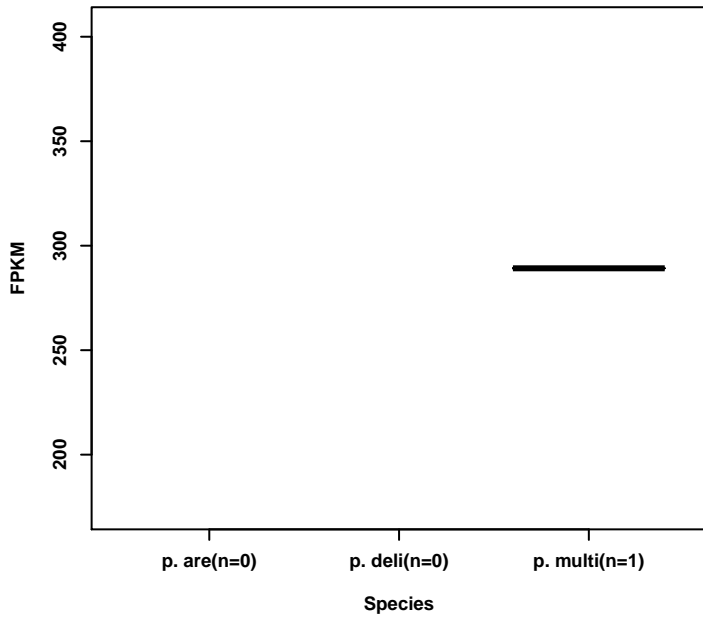
ergosterol biosynthesis



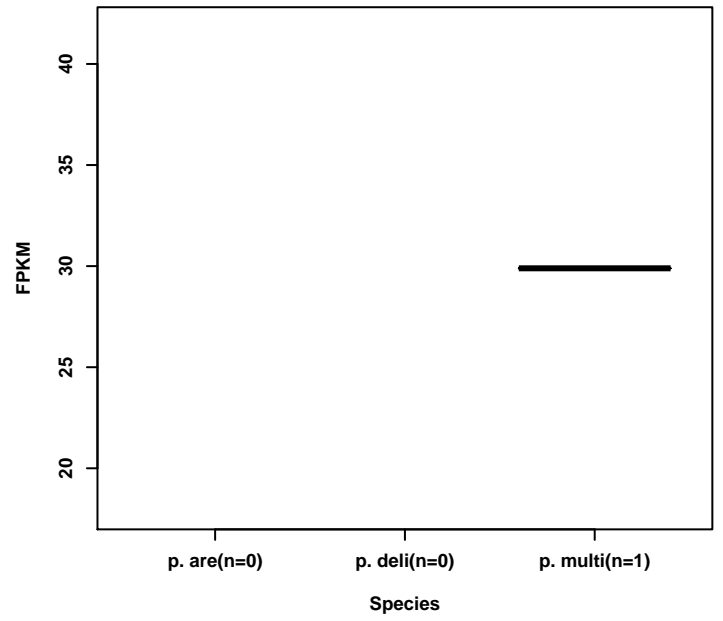
estrogen biosynthesis



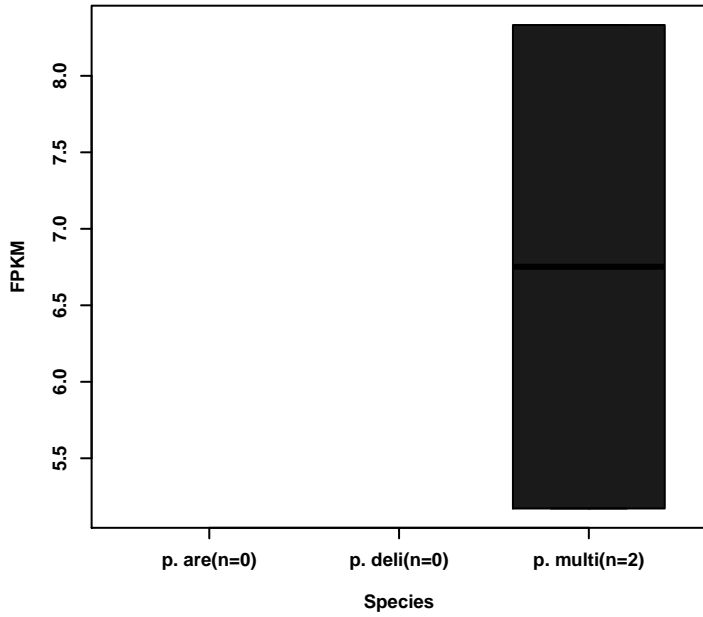
ethanol degradation



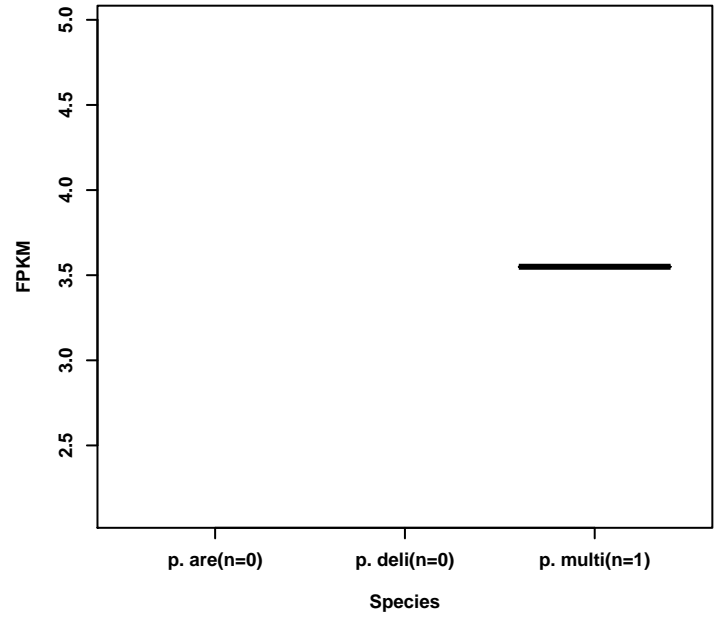
glyoxylate and dicarboxylate metabolism



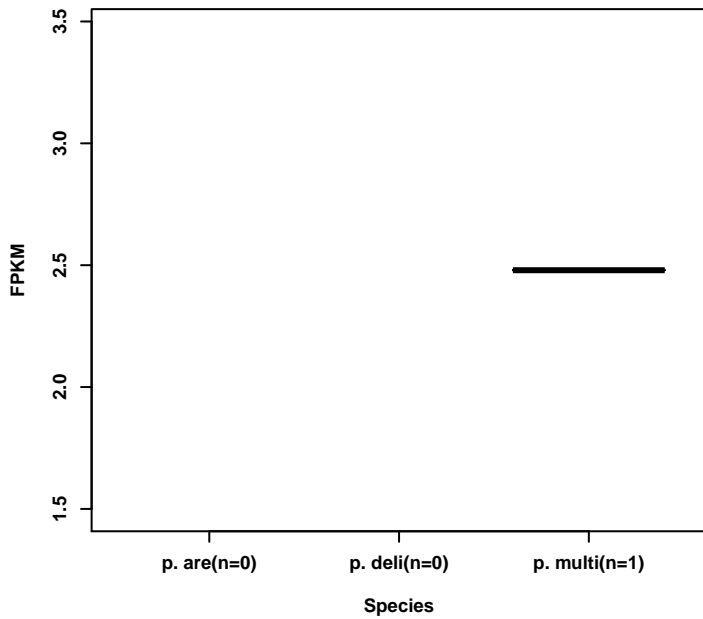
L-arabanan degradation



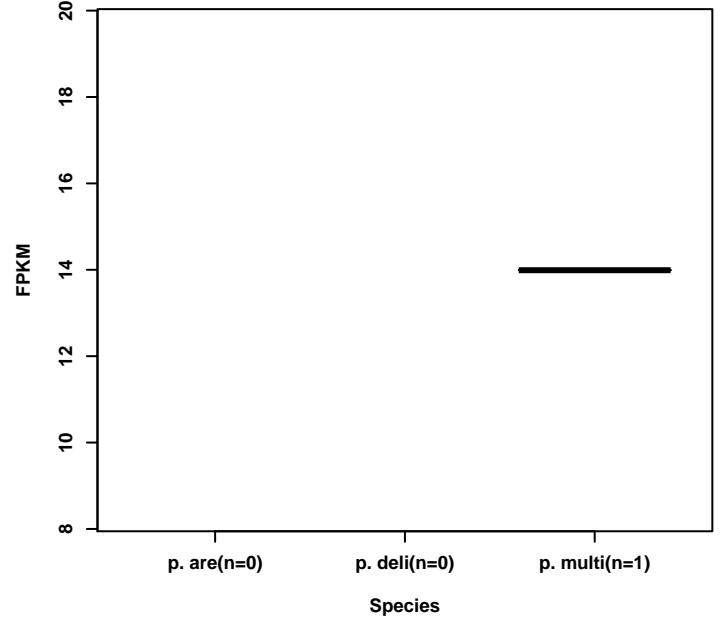
L-lysine biosynthesis via AAA pathway



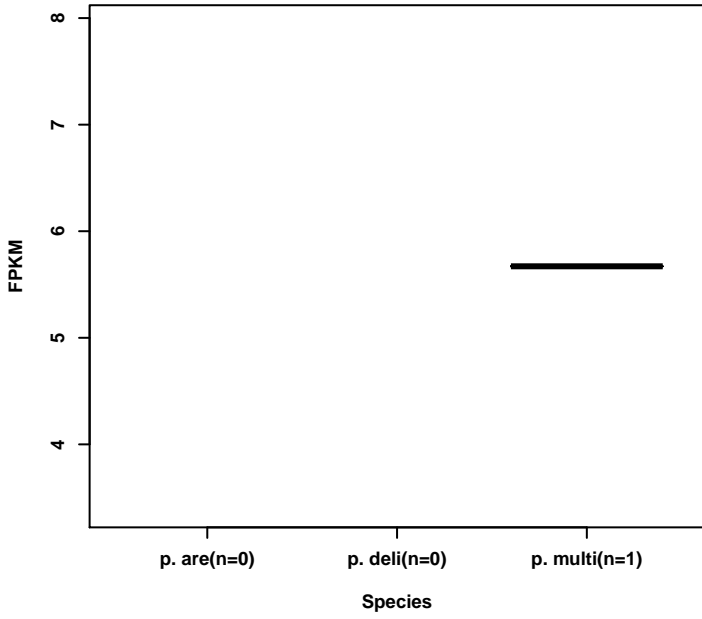
LPS core biosynthesis



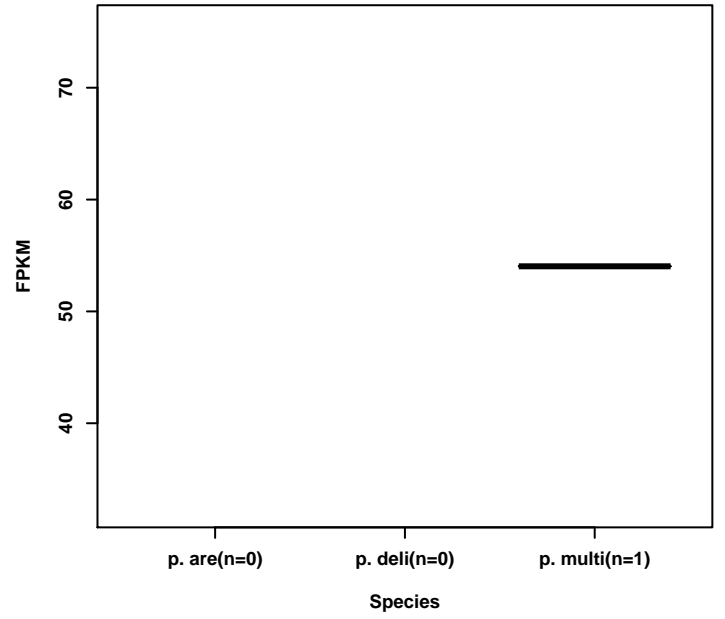
LPS lipid A biosynthesis



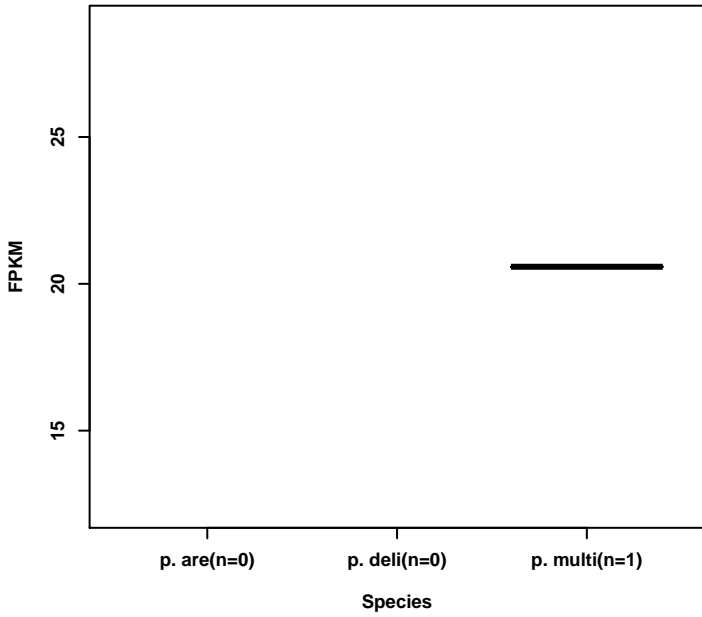
L-threonine degradation via oxydo-reductase pathway



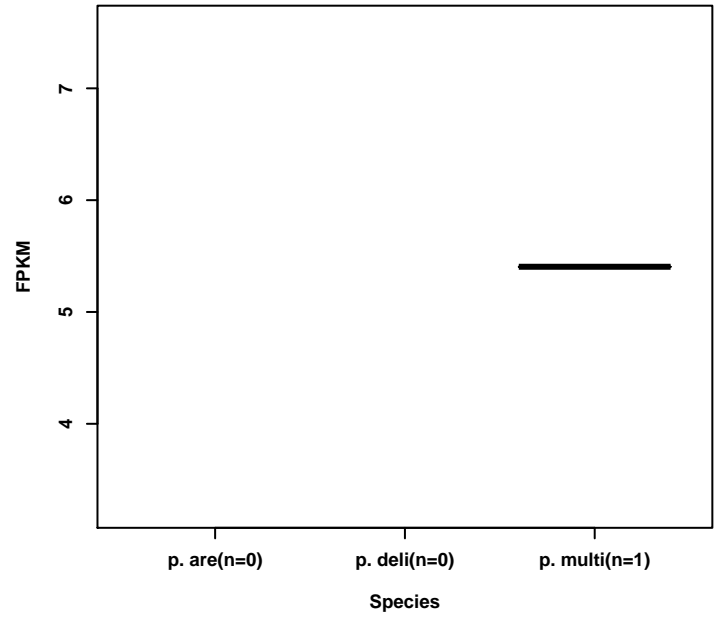
L-threonine degradation via propanoate pathway



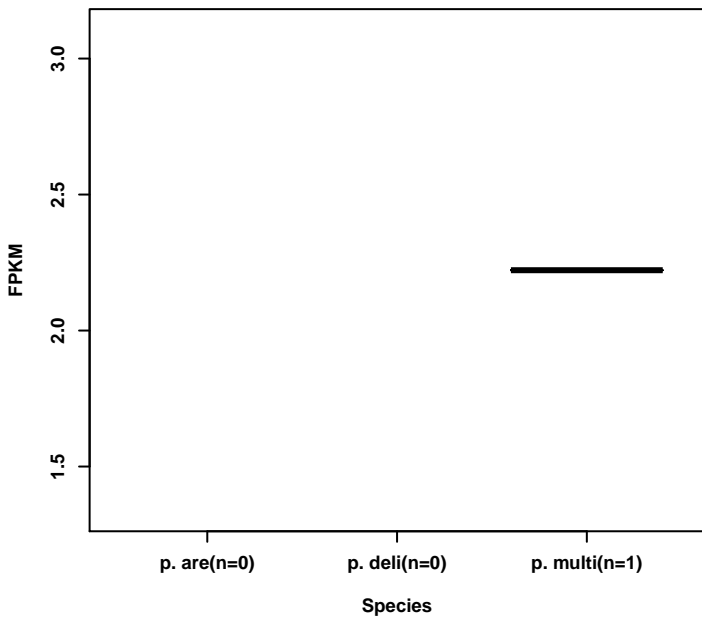
morphine degradation



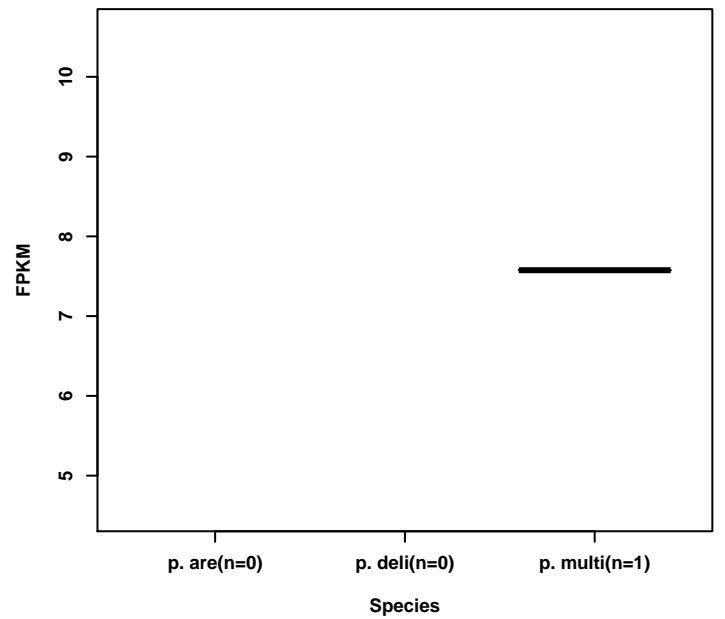
phenylpropanoid biosynthesis



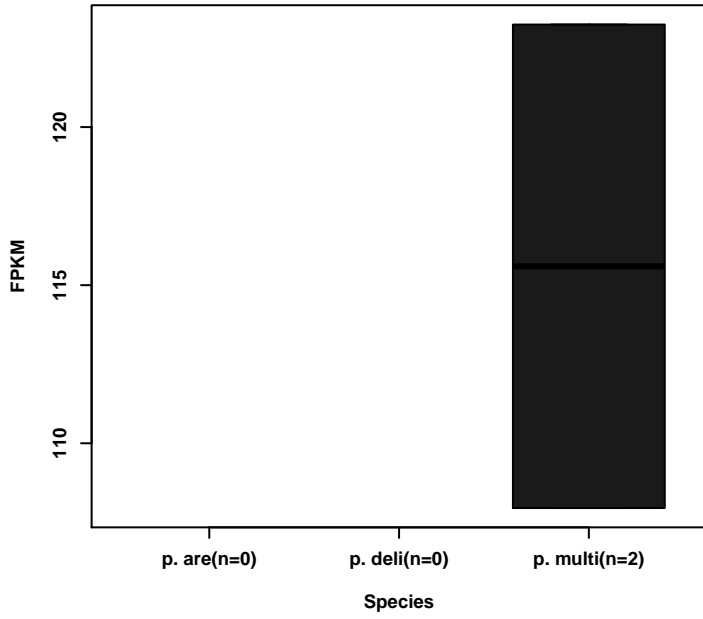
photorespiration



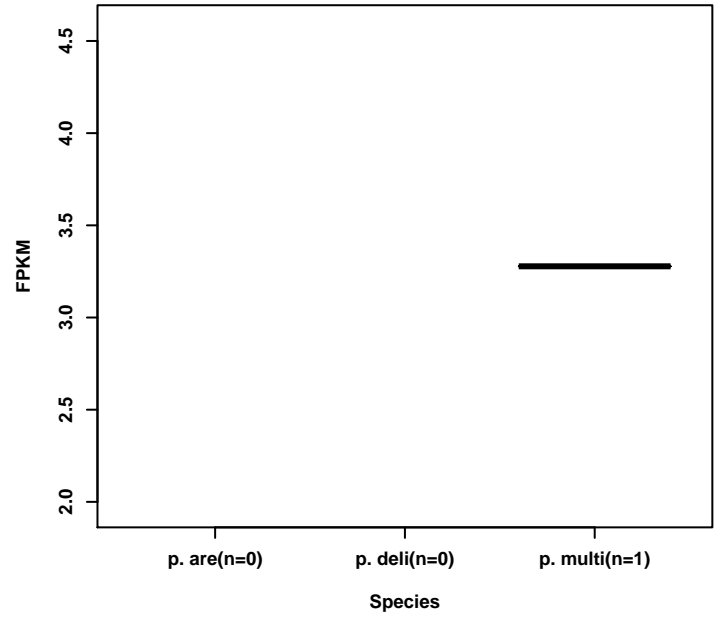
proteasomal pup-dependent pathway



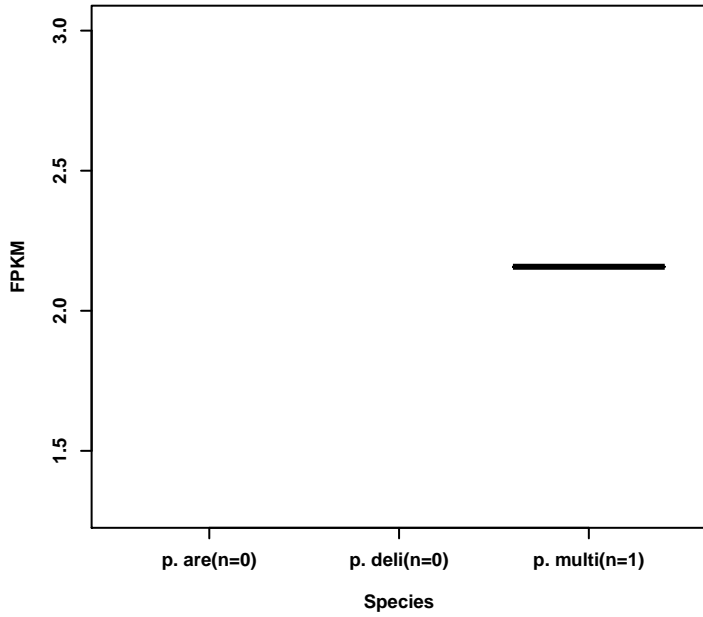
purine nucleoside salvage



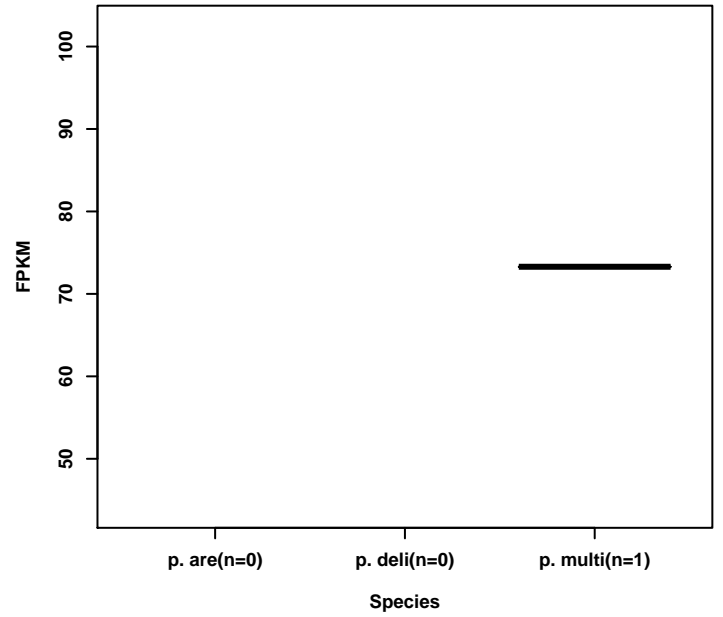
retinol metabolism



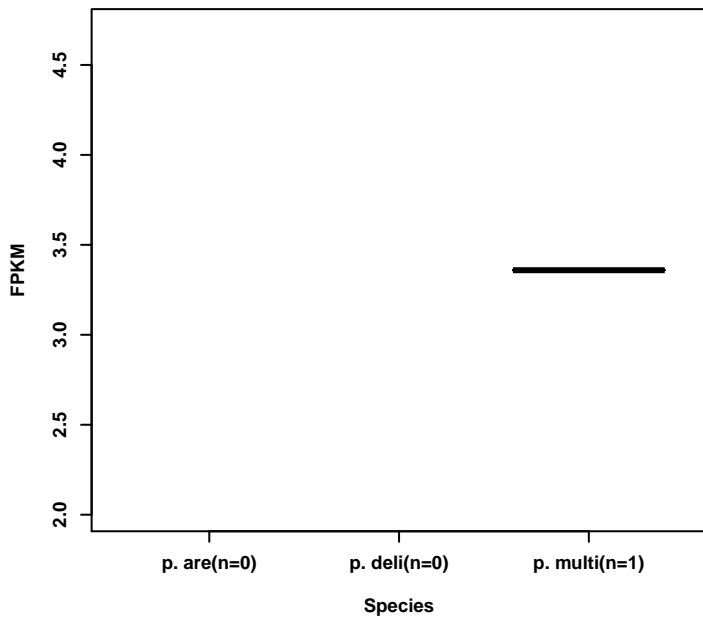
sulfur metabolism



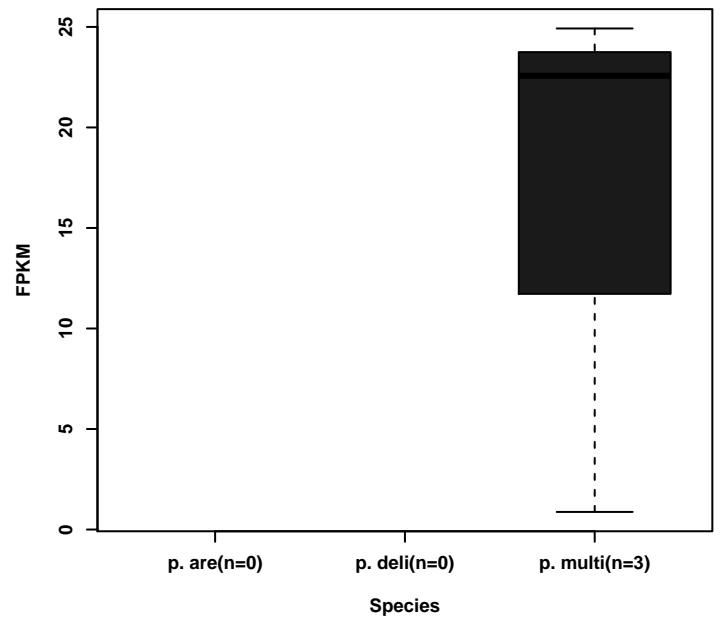
tropane alkaloid biosynthesis



ubiquinone biosynthesis [regulation]

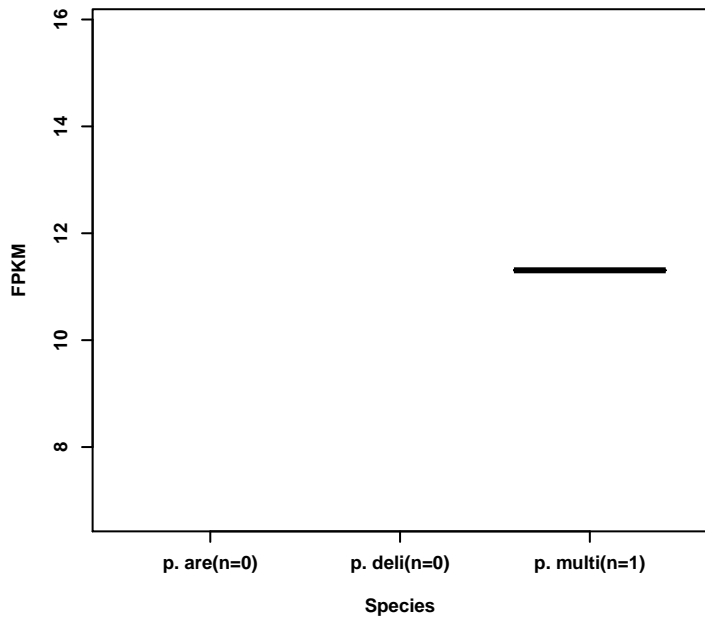


urate degradation





### violacein biosynthesis

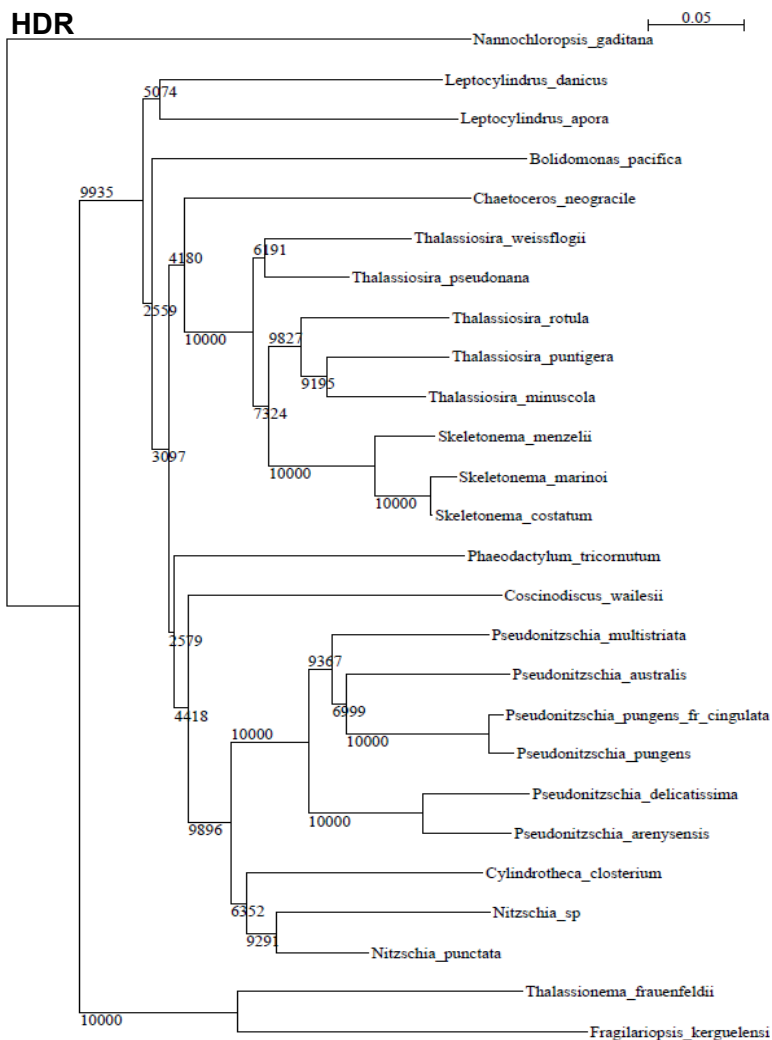
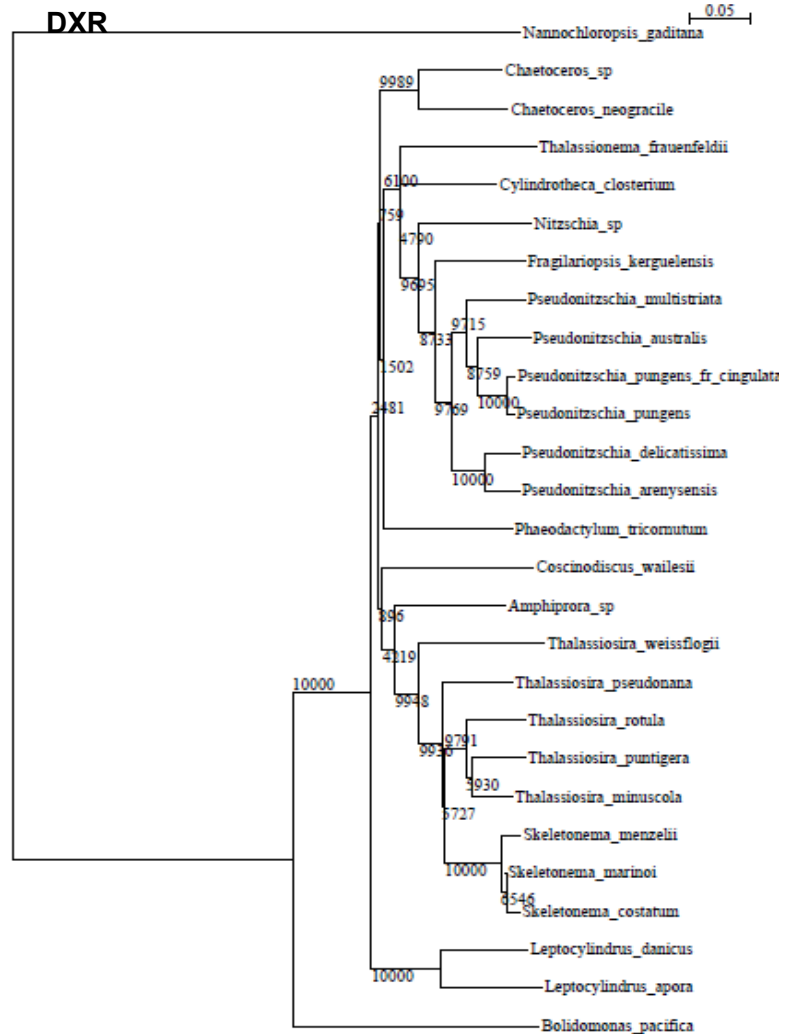
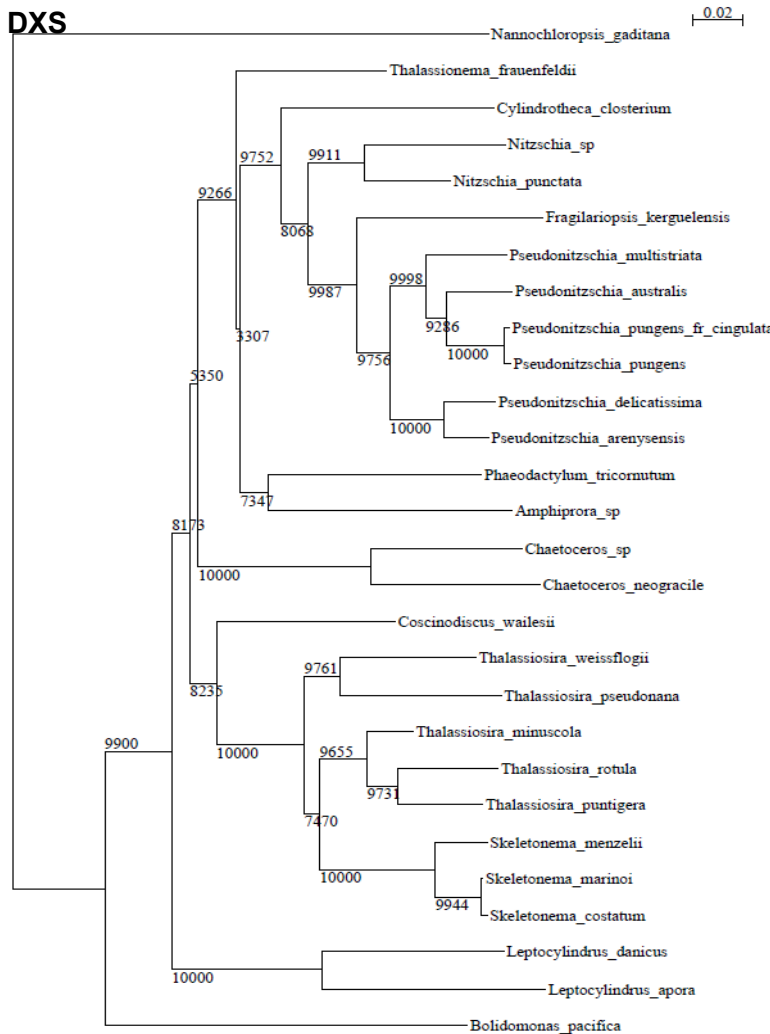


**Supplementary Figure S4. Neighbour joining tree of selected enzymes belonging to the isoprenoids pathways.**

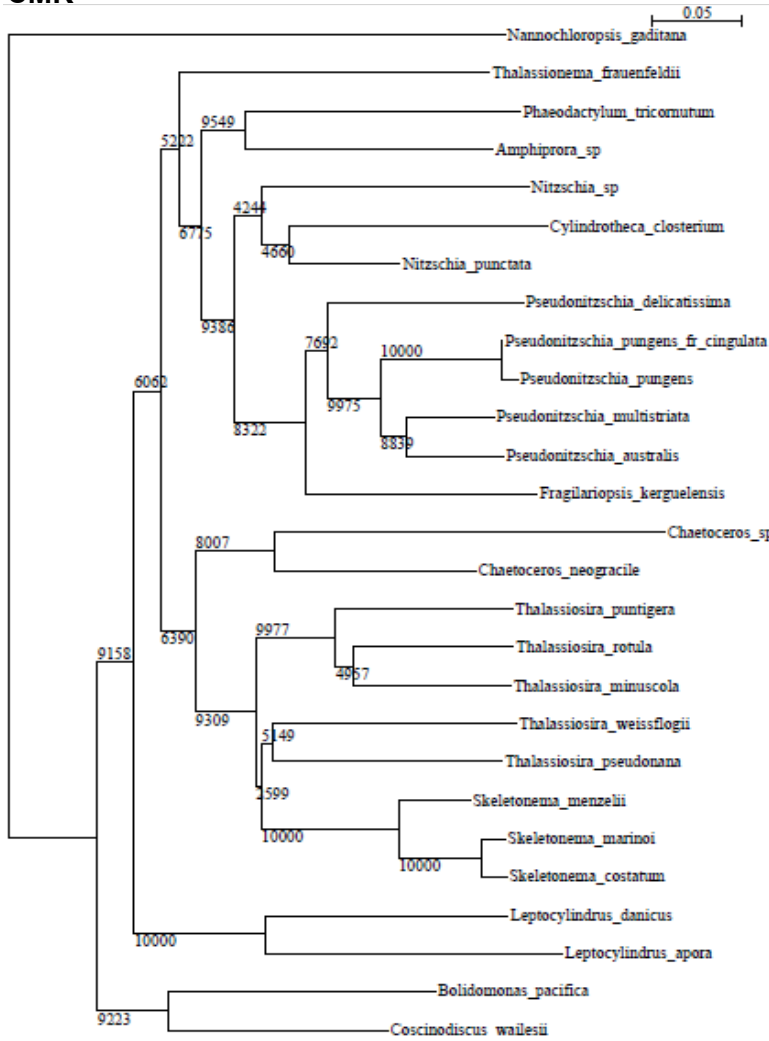
ClustalX neighbor joining tree phylogeny of isoprenoid enzymes including centric and pennate diatom sequences. HMGR = 3-Hydroxy-3-methylglutaryl-CoA reductase; MK = mevalonate kinase; MPDC = diphospho-MVA decarboxylase; DXS = 1-Deoxy-D-xylulose 5-phosphate synthase; DXR = 1-Deoxy-D-xylulose 5-phosphate reductoisomerase; CMK = 4-(Cytidine 5-diphospho)-2-C-methyl-D-erythritol kinase; HDR = 4-Hydroxy-3-methylbut-2-enyl diphosphate reductase; IPPI = isopentenyl diphosphate delta-isomerase.

The bootstrap values at each node represent the numbers of trees having the same node during the tree construction process.

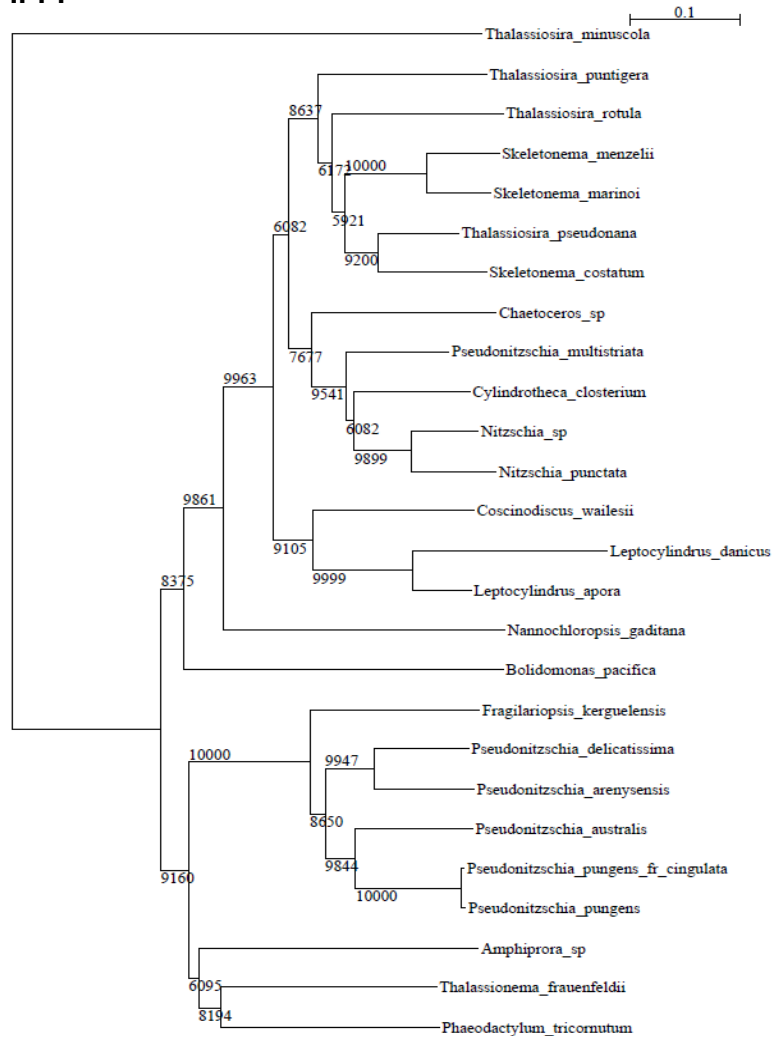




CMK



IPPI



**Supplementary Table S1. Transcriptome size and statistics of selected diatom species sequenced within the MMETSP.**

<b>SPECIES</b>	<b>Total Number of Sequences</b>	<b>N50</b>	<b>Average length</b>	<b>Median length</b>	<b>Maximum length</b>	<b>% GC</b>
<i>Fragilariopsis kerguelensis</i> L26-C5 -Fe_LL	56173	1754	1035	680	16027	41
<i>Fragilariopsis kerguelensis</i> L26-C5 -Fe_SL	48729	1786	1078	744	16511	41
<i>Fragilariopsis kerguelensis</i> L26-C5 Fe_LL	51006	1818	1078	717	16160	41
<i>Fragilariopsis kerguelensis</i> L26-C5 Fe_SL	50930	1927	1109	709	23605	41
<i>Leptocylindrus danicus</i> var. <i>apora</i> B651	17776	1982	1478	1234	16580	40
<i>Leptocylindrus danicus</i> var. <i>danicus</i> B650	30346	1969	1245	893	29076	44
<i>Nitzschia punctata</i> CCMP561 Exponential	24774	1982	1456	1244	9581	46
<i>Nitzschia punctata</i> CCMP561 N-limited	24323	1799	1270	1005	10214	46
<i>Nitzschia punctata</i> CCMP561 Nocodazole	20545	1728	1265	1011	9835	46
<i>Nitzschia punctata</i> CCMP561 Si-limited	20590	1682	1248	1021	8377	46
<i>Pseudo-nitzschia australis</i> Glutammate as N source	18267	2219	1367	1140	8451	47
<i>Pseudo-nitzschia australis</i> NO3 as N source	25559	2213	1546	1371	9261	48
<i>Pseudo-nitzschia australis</i> Urea as N source	24075	2093	1515	1388	8730	47
<i>Pseudo-nitzschia australis</i> NH4 as N source	25365	2186	1524	1359	9246	47
<i>Pseudo-nitzschia fraudulenta</i> Silica limited 390ppm	54003	1697	924	503	10672	56
<i>Pseudo-nitzschia fraudulenta</i> 390ppm_Replete	15456	1940	1217	977	9952	56
<i>Pseudo-nitzschia fraudulenta</i> Silica limited 750ppm	16902	1800	1133	924	7371	49
<i>Pseudo-nitzschia fraudulenta</i> 750ppm_Replete	18640	1868	1056	705	11216	49
<i>Pseudo-nitzschia heimii</i>	16361	2163	1315	1029	13808	51
<i>Pseudo-nitzschia pungens</i> cf. <i>cingulata</i>	17200	2698	1510	1041	20706	49
<i>Pseudo-nitzschia pungens</i> cf. <i>pungens</i>	20080	2192	1587	1307	17488	49
<i>Skeletonema marinoi</i> FE60	19052	2029	1409	1107	13123	45
<i>Skeletonema marinoi</i> FE7	21683	1894	1343	1105	11734	45
<i>Skeletonema marinoi</i> SM1012Hels-07	30397	2475	1688	1311	15814	45
<i>Skeletonema marinoi</i> SM1012Den-03	31328	2519	1664	1269	14168	45
<i>Thalassiosira punctigera</i> C2	32778	1732	942	492	14941	55
<i>Thalassiosira</i> sp. NH16	35814	2192	1587	1307	17488	52
<i>Thalassiosira</i> sp. Th. Sp. FW	12781	2666	1695	1275	27914	47
<i>Thalassiosira weissflogii</i> CCMP1336 20nM Fe	18480	2208	1409	1111	12742	46
<i>Thalassiosira weissflogii</i> CCMP1336 20nM Fe	18662	2250	1428	1119	11100	46
<i>Thalassiosira weissflogii</i> CCMP1336 60 nMFe	18623	2242	1430	1133	13644	46
<i>Thalassiosira weissflogii</i> CCMP1336 60 nMFe	18624	2321	1474	1148	18032	46

**Supplementary Table S8. List of the C4 photosynthetic carbon cycle genes<sup>1</sup> and their corresponding transcript IDs in the three *Pseudo-nitzschia*.**

GENES	<i>Pseudo-nitzschia arenysensis</i>	<i>Pseudo-nitzschia delicatissima</i>	<i>Pseudo-nitzschia multistriata</i>
Carbonic anhydrase	MMETSP0329-20121206 10 MMETSP0329-20121206 11309 MMETSP0329-20121206 11950 MMETSP0329-20121206 16679 MMETSP0329-20121206 19581 MMETSP0329-20121206 21379 MMETSP0329-20121206 21593 MMETSP0329-20121206 22047 MMETSP0329-20121206 2449 MMETSP0329-20121206 6511 MMETSP0329-20121206 7080 MMETSP0329-20121206 9814 MMETSP0329-20121206 9817	MMETSP0327-20121206 12916 MMETSP0327-20121206 13138 MMETSP0327-20121206 13750 MMETSP0327-20121206 19895 MMETSP0327-20121206 2135 MMETSP0327-20121206 7716 MMETSP0327-20121206 89 MMETSP0327-20121206 9960	Comp10350_c0_seq2 comp10602_c0_seq2 comp10604_c0_seq1 comp12090_c0_seq1 comp13851_c0_seq1/2 comp3866_c0_seq1 comp3951_c0_seq1 comp4639_c0_seq1 comp4674_c0_seq1 comp8539_c0_seq1
Malate dehydrogenase (MDH)	MMETSP0329-20121206 3423 MMETSP0329-20121206 8111	MMETSP0327-20121206 13132 MMETSP0327-20121206 2460 MMETSP0327-20121206 8089	Comp13442_c0_seq1/2/5 comp3147_c0_seq1
NAD Malic Enzyme	MMETSP0329-20121206 5779	MMETSP0327-20121206 7012	comp14698_c0_seq1-4
Phosphoenolpyruvate carboxylase (PEPC)	MMETSP0329-20121206 13428 MMETSP0329-20121206 4035 MMETSP0329-20121206 4516	MMETSP0327-20121206 10283 MMETSP0327-20121206 1569 MMETSP0327-20121206 18670 MMETSP0327-20121206 2331	Comp13217_c0_seq1 comp5966_c0_seq1
Phosphoenolpyruvate carboxykinase (PEPCK)	MMETSP0329-20121206 3602 MMETSP0329-20121206 4400	MMETSP0327-20121206 12843 MMETSP0327-20121206 1562 MMETSP0327-20121206 19812	
Pyruvate kinase (PK)	MMETSP0329-20121206 12757 MMETSP0329-20121206 22082 MMETSP0329-20121206 6571 MMETSP0329-20121206 7148 MMETSP0329-20121206 8630	MMETSP0327-20121206 13875 MMETSP0327-20121206 10290 MMETSP0327-20121206 1193 MMETSP0327-20121206 12468 MMETSP0327-20121206 19729 MMETSP0327-20121206 7774 MMETSP0327-20121206 8994	Comp10607_c0_seq1-3 comp13968_co_seq1 Comp15562_c0_seq3 Comp7347_c0_seq1-3 Comp9010_c0_seq1 Comp9501_c0_seq1-8
Pyruvate-phosphate dikinase (PPDK)	MMETSP0329-20121206 17567 MMETSP0329-20121206 4315 MMETSP0329-20121206 13199	MMETSP0327-20121206 1616 MMETSP0327-20121206 9199	Comp4924_c0_seq1

Transcripts IDs were retrieved with a name search in the annotation tables.

<sup>1</sup> Kroth, P.G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M.S., Stanley, M.S., Kaplan, A., Caron, L., Weber, T., et al. (2008). A Model for Carbohydrate Metabolism in the Diatom *Phaeodactylum tricornutum* Deduced from Comparative Whole Genome Analysis. PLoS ONE 3.

**Supplementary Table S9. List of the Urea cycle genes and their corresponding transcript IDs in the three *Pseudo-nitzschia*.**

GENES	<i>Pseudo-nitzschia arenysensis</i>	<i>Pseudo-nitzschia delicatissima</i>	<i>Pseudo-nitzschia multistriata</i>
Ornithine cyclodeaminase	MMETSP0329-20121206 16601 MMETSP0329-20121206 12770	MMETSP0327-20121206 9866	comp2508_c0_seq1 comp10073_c0_seq2
Agmatine ureohydrolase	MMETSP0329-20121206 21557 MMETSP0329-20121206 6021	MMETSP0327-20121206 14705 MMETSP0327-20121206 232	comp4538_c0_seq1 comp4538_c0_seq2
Urease	MMETSP0329-20121206 20423 MMETSP0329-20121206 3502 MMETSP0329-20121206 22452 MMETSP0329-20121206 4831	MMETSP0327-20121206 18997 MMETSP0327-20121206 10411 MMETSP0327-20121206 8851 MMETSP0327-20121206 1463	comp14288_c0_seq2
Aspartate aminotransferase	MMETSP0329-20121206 13842 MMETSP0329-20121206 14937 MMETSP0329-20121206 4801 MMETSP0329-20121206 7116	MMETSP0327-20121206 4731 MMETSP0327-20121206 5235 MMETSP0327-20121206 9520	comp11127_c0_seq1-5 comp12679_c0_seq1 comp12679_c0_seq3 comp14098_c0_seq1 comp7127_c0_seq1 comp8502_c0_seq1 comp9942_c0_seq2
ODC2 - ornithine decarboxylase 2	MMETSP0329-20121206 5745 MMETSP0329-20121206 19848 MMETSP0329-20121206 11971 MMETSP0329-20121206 6946	MMETSP0327-20121206 11700 MMETSP0327-20121206 12906 MMETSP0327-20121206 4293	comp12949_c0_seq1 comp8700_c0_seq1
Urea carrier	MMETSP0329-20121206 11970 MMETSP0329-20121206 4419 MMETSP0329-20121206 4956	MMETSP0327-20121206 13582 MMETSP0327-20121206 17340 MMETSP0327-20121206 596	comp11844_c0_seq1 comp15168_co_seq1
Arginase	MMETSP0329-20121206 21557 MMETSP0329-20121206 6021	MMETSP0327-20121206 14705 MMETSP0327-20121206 232	comp4538_c0_seq1 comp4538_c0_seq2
Fumarate lyase Argininosuccinate lyase	MMETSP0329-20121206 21824 MMETSP0329-20121206 2578 MMETSP0329-20121206 14971 MMETSP0329-20121206 5583 MMETSP0329-20121206 6329	MMETSP0327-20121206 10118 MMETSP0327-20121206 10669 MMETSP0327-20121206 17196 MMETSP0327-20121206 17458 MMETSP0327-20121206 7445 MMETSP0327-20121206 7743	comp8093_c0_seq1 comp12477_c0_seq1 comp4933_c0_seq1 comp13109_c0_seq1
Allantoicase	MMETSP0329-20121206 23359 MMETSP0329-20121206 14584 MMETSP0329-20121206 23318 MMETSP0329-20121206 22834	MMETSP0327-20121206 2270 MMETSP0327-20121206 6752	comp7156_c0_seq1
Ornithine transcarbamylase	MMETSP0329-20121206 14902 MMETSP0329-20121206 15636	MMETSP0327-20121206 12919 MMETSP0327-20121206 19499 MMETSP0327-20121206 6375	comp8342_c0_seq1 comp5549_c0_seq1
Argininosuccinate synthase	MMETSP0329-20121206 14126 MMETSP0329-20121206 9021	MMETSP0327-20121206 15742 MMETSP0327-20121206 15160 MMETSP0327-20121206 11703	comp2518_c0_seq1
Urea-proton symporter DUR3	MMETSP0329-20121206 4419 MMETSP0329-20121206 4956 MMETSP0329-20121206 11970	MMETSP0327-20121206 13582	comp15168_c0_seq1
CPS III- carbamoyl-phosphate synthase (mitochondrial precursor)	MMETSP0329-20121206 10336 MMETSP0329-20121206 6833 MMETSP0329-20121206 7908 MMETSP0329-20121206 22067	MMETSP0327-20121206 6375 MMETSP0327-20121206 5850 MMETSP0327-20121206 11066 MMETSP0327-20121206 4753	comp15368_c0_seq1 comp12152_c0_seq1

Transcripts IDs were retrieved with a tblastn search in each *Pseudo-nitzschia* transcriptome using *Phaeodactylum tricorutum* proteins as a query. *P. tricorutum* protein IDs were taken from Levitan et al., 2015<sup>1</sup>.

<sup>1</sup> Levitan, O., Dinamarca, J., Zelzion, E., Lun, D.S., Guerra, L.T., Kim, M.K., Kim, J., Van Mooy, B.A.S., Bhattacharya, D., and Falkowski, P.G. (2015). Remodeling of intermediate metabolism in the diatom *Phaeodactylum tricorutum* under nitrogen stress. Proc. Natl. Acad. Sci. U.S.A. 112, 412–417.



## Supplementary Note

>Pseudo-nitzschia\_multistriata\_NOS

MGIFSKHCCGGDTLNQVIHHQPEHKKDEADNEETCNRTDLWPHTVTSRMERSESSLEFSRRVMTSNYRSQRD  
ASTPQRSSNSQSSIKSTCSSVESTESDLGHLKEKAIEEALTASYLDVRNVESRHRSESSPTEQSLNTSMH  
SKRMILKRDSERGFQDKKHFYESLVASTEDVLEKINEGWGIKGSFTMEPAMYGAPLSKVPLEQKHAVLISQ  
SWNKVLAFRSMFAEALISRWRLLASVEYIEEQALREKNPVPNSRRHFPSFHQQLWQVGFSPDVQALNLI  
ESNNDPIAKCFETRATDLQVLVVSMTAVRGLCPHSQTIQRESYRPIDGSADSDPDISQVFFHRKDCTTFEDFCF  
LFARYWVKPHHWLTLCRAFLWAMKSQNPYCIDDEKDDLDRPADSVHARFFFTGMIILPILLEANLRRANYMRKG  
VFEELKVCISNDRLGNKFDITIGVEMFKKLFETFPEIKDHFCEENDVEDMSYGLFEMIYVLVQGSGRDLANPQSN  
LRKILREQSMKWVELGLSTHVVYLIGREILEGMETIFGDSVNHSKEKLRNAAFALWKHSIQFVLHPILLQQNL  
QAEALKFYNDVAAELKWSTSTKQNRLEFVNMEILATGEYTHTQEELEIGARLAWRNSAKCIGRIAWNTLVVRD  
RRHCTQPSEIFNEVCEHMKLATGGASIQSVMTVFRPQRHNELEFGTRFWSSQFIRYCGYKDRDSEEIMGDPACV  
DLTSYLIEKKLWTPPERRTAFDILPLVLKIPGNDIPFIYQLPKSVTHEVHLEHPEFPEVKDLGYKWAAPPAIS  
NFMNMLGGIKYSAAPFNGWVFSIEVARNLLERYHVTEPLAIAMNMLSDKFLAQKVSAELESAIIHSFEKSDF  
TIVDPDTVGHSFITHCKRERSAGRECPAQWSWIGgptnPVWHLEMRDFRVVPQFEYCCDPWRVLDIDGVSGTK  
AIDTTGHILSEDDFSQKVSMPVLIAYGSETGKAEAVANSLARTLKFRCRPTVLTNLNGALEYNQELGKFSHILI  
ICSTFGAGEPPLNAKKFANAELAGKVTGVKYAVLALGSSLYPDFCKAGKIDSQMEAGANSIMNIVCADA  
AKGNQSDIYNWSIFLKKLLLPNSLLVELESSGIPVEDTKAAEYRIRWLAKDEYDEEYSKHRLRTGTMKCRKNQDL  
FEGSDPLRSTRHIELEQIDKTDIAYETGDHLSVVPINSMDMVSRFCNCFAPELEKAAESSGFHRLKLGKASRKK  
FLQKTSARMNCTPSVLWQTHQPFYIESVEDQGTTPCNAKHLTNNNLADVLQFSLDLSLSSSESYLVEFLQMLNS  
KLEEASLTLCASDFQTMAKDAVQQCNEEGDSGMINNIKATYPTIVHFLEKFNDFCKPLETRKNPLLGLADV  
LVIMPRLGPRQYSISSSALTSPNKVSIITVGVVKFETS RDVLVEGVC SHDLAGLEPGQMINASVVKSPFRAPRS  
IKSPLIMIGAGTGLAPFMGFLRDRSLAMKKAGASRRGMFGKCHLFFGCRSEGEMLYKDQLLEWESSGVAKLHI  
AYSREPGIHCNRYVHDAMEDRGAEALFALMTSEPKTHLYICGDVNMAKACSEKCSILLQEHGSMKINATR  
FM SNMRVSNRFQLDVGWGTREVEIVDIDFSYHGNANTRLIKEASRRTIMREFNEKY

**IDs of the transcriptomes in the MMETSP used for the Isoprenoids analysis.**

<b>SPECIES</b>	<b>MMETSP ID</b>
<i>Amphiprora sp</i>	MMETSP 0724
<i>Bolidomonas pacifica</i>	MMETSP 0785
<i>Chaetoceros neogracile</i>	MMETSP 1336
<i>Chaetoceros sp</i>	MMETSP 0200
<i>Coscinodiscus wailesii</i>	MMETSP 1066
<i>Cylindrotheca closterium</i>	MMETSP 0017
<i>Fragilariopsis kerguelensis</i>	MMETSP 0733
<i>Leptocylindrus apora</i>	MMETSP 0322
<i>Leptocylindrus danicus</i>	MMETSP 0321
<i>Nitzschia punctata</i>	MMETSP 0744
<i>Nitzschia sp</i>	MMETSP 0014
<i>Pseudo-nitzschia australis</i>	MMETSP 0139
<i>Pseudo-nitzschia pungens</i>	MMETSP 1061
<i>Pseudo-nitzschia pungens fr cingulata</i>	MMETSP 1060
<i>Skeletonema costatum</i>	MMETSP 0013
<i>Skeletonema marinoi</i>	MMETSP 1039
<i>Skeletonema menzelii</i>	MMETSP 0603
<i>Thalassionema frauenfeldii</i>	MMETSP 0786
<i>Thalassiosira minuscola</i>	MMETSP 0737
<i>Thalassiosira puntigera</i>	MMETSP 1067
<i>Thalassiosira rotula</i>	MMETSP 0403
<i>Thalassiosira weissflogii</i>	MMETSP 0878

## Isoprenoids protein sequences used for phylogenetic trees.

### HMGR protein sequences

#### >*Amphiprora\_sp*/HMGR

PFHVHLKNVKRAMSSAMTLGMQIDAVINELSNLSSTQQYAIIVIGVGLVATLLLGSADH  
PGLEMTTAPPKTAVSPSKPVKTDKNSQPEPRWYIFRMINYAMMVAFVASVADVFVWNY  
AAYA VDSAVLLRFLIGWSACLIYFFGFFGVSVFVHDVPPVDEGLAPSPPQPEKIPSAKM  
VSSSRNQTVTIPKSMKKKAVEAPCPKDVPCSDPSSFATSSKKAPPASNLTELSDDQIA  
NLVLTGQMKDHELEKRLDCHRAVDVRRLVIREKLAKVGNPSSLQELPSQPSLDYSRV  
HGANCEIVVGYVPLPVGMVGPLRINDMTVYVPMATTEGCLVASANRGCKAISVTGAN  
AMVIRDGITRAPCIRMP SAMEAARLKLWCDEPTNFAAMKAAFETTSFGKCLKSCSPTV  
AGRNVYLRLVCFSGDAMGMNMVSKGSLAVIDLLRAEFPELKLIALSGNMCTDKKAAA  
TNWLEGRGKSVVTECVIPKEVVNKTLLKTTVEAMVETNINKNLIGSAMAGAIGGFNAH  
AANMVTA VFLATGQDPAQNVESNCITLLEETEDGDLHMSCTMPSLEVGTVGGGTTLF  
AQAACLQAMGCKGGDRQNP GANAQRLALITATAVMAGELSLIAALAANTLVQAHMT  
HNRKK

#### >*Bolidomonas\_pacifica*/HMGR

MGEGSPSLEPSTSATNLKGLDSLQLHDTQINNGVRPTMHHDNHMGNHEDKVDKLDKE  
DNTSSARPVSSSISPSISPSISSSISSSISSSISSSISIRPSTNSPTNSIAPKIDIEKLVEIAKTINSND  
NKYTIRDVNALLGGGQETIETALLRRYVSGLQSPSSLPYTGGGVVDYDSISGRNCDNCV  
GYVSVVPGIAGPVKYSFIDLTDGGVKEDKDLISHIPIATTEGGLIASVSRGCKALRDSCT  
VTTCVTKDAMTRCPALEFKTVKDAVAFKVYVEGDGIGVVKESFEDTTKYGKLERVEC  
KVVGRMVYFRVEAKTGEAMGMNMVSKGTENVLRVLKQDWEFNVITLSGNGCSDKK  
PSATNFVKGRGKSVTAEAIKREVVERTLKSNPENMKRTNIYKNQVGS AVAGNLGGM  
NAHAANVVAGVFIATGQDPAHTVEGSMCTTVMDVTPEGDLVSVTLPSICVGTIGGGT  
ELPPQKASIKSMGCEGTGGSKRLAGLIG

#### >*Cylindrotheca\_closterium*/HMGR

MASAA YQPTTAPTLGMRLDAVLSQVDELSSTQLYGLIVAATVTL CVFLLGTGSNTDVE  
FAATNSKMKTGTNDDGLKKPNASTNNTASTTRQPRWHIFKWVNYLASA AFLSSVGMF  
CMNASQYLH HESKGVLVQFLVGWSVFLMYFFGFFGISLIHEDIPKDNQDDQGIEQPRN  
TTRYAAGSSSTTAAAKQALHTAAPSAPVCS DPASFKASSSSVPSNLKELPDEDIAN  
LVLMNKLKDHELEKRLDPFRAVTVRRLVVNQKLSTLLTDGKTNPVTNVLEKLPSTPS  
LDYSRVFGANCEIVVGYVPLPVGLVGPLSLNDET VYVPMATTEGCLVASTNRGAKAIT  
QGGGAQARIVRDGITRAPCVRMASAMEAADLKVWCEEPQNF AVLKQAFESTTSFGKL  
QACNPTVAGKNVYLRLVCFSGDAMGMNMVSKGSLAVIETLQAKFPSLQLVALSGNM  
CTDKKAAATNWLHGRGKSVVVEAIIPKDVVRGTLKTTVDALIFTNTNKNLIGSAMAGS  
IGGFNAHASNIVTA VFLATGQDPAQXVXSSNCITXMEPTDDGDLWISCTMPSIEVGTVG  
GGTSLPAQSACLEAIGCKGGGATPGENAQKLATVVAATMAGELSLLAALAANTLVQ  
AHMQHNRKPAPKK

#### >*Fragilariopsis\_kerguelensis*/HMGR

MASTEVMPTLGMRLDAMISHIDRLPSSQLYGVIVAATVGLCVVLLGTGNSNLELQQQQ  
RQRQQQMMNVDDLKKPINNRIANVGGKQPKWHIFKWVNYLAVGAF LWSVYTFCSNA  
SQYLH HESQGVLVQFLVGWSVFLLYFFGFFGISLIHEDIPKEEDET TMTVSKNIVSRKP  
YNASTNSTTTMMKKKNSIHPPASCAPVCS DPSTSFKSSGASLSSNSVNLEELADEEIANL  
VLNNKVKDHELEKRLDPFRAVXVRRIAFNQKIASVLS DNTNKQNNINNTANVLDKLP  
TPSLDYSRVYGANCEIVVGYVPLPVGLVGPLTINDET VYIPMATTEGCLVASSNRGAKA  
ITQGS GAKARIVRDGXTRAPCIRMRSAMEADLKLWCEEPSNFLILKQAFESTTNFGKL  
KECNPTVAGKNVYLRLVCFSGDAMGMNMVSKGSLAVIETLQNEFPSLQLVALSGNMC  
TDKKAATNWLNGRGKSVVVEAVIPREVVEKTLKTTVKALVHTNINKNLIGSAMAGA

IGGFNAHASNIVTAIFLATGQDPAQNVESNCITLIEETDDNDLWISCTMPXIEVGTVGG  
GTSLEAQSACLEAIGCKGGGATPGENARKLATVVAATMAGELSLLAALAANTLVQA  
HMHNRKPNASK

>*Leptocylindrus\_apora*/HMGR

MEGITSSPPTVGGQLDAFLEGLQNLSTHLYLVAFATCLAVSFVVLNSGQGRYLDVKG  
KDHQKAASQKKDLIVANDMPQKASATGKKS VVSGGREPRWYIFKIMNYVLATSFLVS  
VVHFLNSEEYIMSNDHIKTLGAVWTLFVLYFWGFFGVSVFVDTDEHLEEEESPSSPSSS  
HKVAAPENVSTVLPVPPXPVHPPAPFAPVCS DPSTQKASSMKT PGDISSLSNEELANLV  
LENKLDHKLLETENPTRA VVRRLVFEK KLGSLGKKGALNDLPYEHDLSYERVFGA  
NCEIVVGYVPLPVGMVGPCTLNGESVYIPMATTEGCLVASTNRGCKAISAGGGAVSTL  
LRDGITRAPCLRFESAAEAAALAIWVEETHNFLKKNAFESTTSFGKLLSAKPTVAGKN  
CYLRLKCFSGDAMGMNMVSKGSLAVVELLRQNFPTLKLVALSGNMCTDKKAAATN  
WIEGRGKSVVVEATIPKR VVRTVLKTTVKAIVD TNVQKNLVGSAMSGTLGGFNAHAS  
NIVTAVFLATGQDPAQNVESNCITLMEETKEGDLWISCTMPSIEVGTVGGGTS LPAQ  
AACLKAIGVKG GGDIPGSNARKLAHVAVATMAGELSLLAALAANTLVQA HMQHNR  
KPATSK

>*Leptocylindrus\_danicus*/HMGR

MAQTADPMTIGTRLDALIASGAEYLNNASQAQVCATL FVSSVAFS FALLNCGKGAHGS  
SLDPTL KPLYKNTMEPIKTKVKSPDGSREPRWYIFKMLNYTAVATFSTSVLHFIYSDV  
YMNDAQMMKLMGAWTLFVLYFFGFFGVSVFVDTDDHLEGSSPAEDEISEMTADQASP  
SAVVAVAKQMPVKAAAPIHPPAPSHPVCS DKNLTKDAL TSSSATPAVKKSSSLDLQS  
MTNEELADLVL TNKMKDHQLETKLNPTRA VPIRRLVFEK KLASLGHAKSLDELPEHS  
LSYERVFGANCEIVVGYVPLPVGMVGPCTLNGESVYIPMATTEGCLVASTNRGCKAIT  
AGGGAVSTLLRDGITRAPCLRFESAAEAAALALWAQEPHNFAKLKA AFESTTSFGKLL  
SATPTVAGKNCYLRLKCFSGDAMGMNMVSKGSLAVVDLLRQHFP TLKLVALSGNMC  
TDKKPAAINWIEGRGKSVVVEATIPKDIVRTVLKTTVKAIVD TNLQKNLIGSAMSGSVG  
GYNAHASNIVTAVFLATGQDPAQNVESNCITIMEETDDGDLWISCTMPSIEVGTVGGG  
TSLPAQAACLKAIGVKG GGDIPGGNARKLAHVAVATMAGELSLLAALAANTLVQA H  
MQHNRKPATPAKK

>*Nitzschia\_punctata*/HMGR

MAATFDTMPTTIGMKLDAMIAEIDNLSSTQLYGVIVAATVVL CVVLLGTGHSNLDLQH  
SNNNDPLLKKQPAVAPSGNIKQPRWHIFKWINYLAVAAFLWSVCTFCLNASQYLHHEE  
SQGVLVKFLLGWSVFLLYFFGFFGVSLIHEDIPKEEAGAASSAANRLSSVSQSNKTKTV  
ETSSKNKALHGAACTPVCSDPSSFKVSSSVPSNIKELADEDVADLV LKNKVKDHELE  
KRLDPFRAVTVRRMTANRKLASVLPQNKPNVLDKLPATPSLDYSKVHGANCEIVVGY  
VPLPVGLVGPLSLNNETVYVPMATTEGCLVASTNRGAKAITQGGGAQARIVRDGITRA  
PCVRMESA MEAADLKVWCEQPENFARLKQAFEGTTSFGKLQACHPTVAGKNVYLRL  
VCFSGDAMGMNMVSKGSLAVIETLQKEFPSLQLVALSGNMCTDKKAAATNWLQGRG  
KSVVVEATIPKDVVRTTLKTTVAALVHTNMHKNLIGSAMAGSLGGFNAHASNIVTAV  
FLATGQDPAQNVESNCITLMEETDEGDLWISCTMPSIEVGTVGGGTSLEAQAACLEAI  
GCKGGGATPGENAKKLATVVAATMAGELSLLAALAANTLVQA HMTHNRKSNKK

>*Nitzschia\_sp*/HMGR

MASAA YQPTTAPTLGMRLDAVLSQVDELSSTQLYGLIVAATVTL CVVLLGTGSNTDVE  
FAATNSKMKTGTNDDGLKKPNASTNNTASTTRQPRWHIFKWNYLASAFLSSVGMF  
CMNASQYLHHESKGVLVQFLVGWSVFLMYFFGFFGISLIHEDIPKDNQDDQGIEQPRN  
TTRYAAGSSSTTAAAKKQALHTAAPSAPVCS DPASFKASSSSSVPSNLKELPDEDIAN  
LVLMNKLKDHELEKRLDPFRAVTVRRLVVNQKLSTLLTDGKTNPVTNVLEKLPSTPS  
LDYSRVFGANCEIVVGYVPLPVGLVGPLSLNDET VYVPMATTEGCLVASTNRGAKAIT  
QGGGAQARIVRDGITRAPCVRMASAMEAADLKVWCEEPQNF AVLKQAFESTTSFGKL  
QACNPTVAGKNVYLRLVCFSGDAMGMNMVSKGSLAVIETLQAKFPSLQLVALSGNM  
CTDKKAAATNWLHGRGKSVVVEAIPKDVVRGTLKTTVDALIFTNTNKNLIGSAMAGS

IGGFNAHASNIVTAVFLATGQDPAQXVXSSNCITXMEPTDDGDLWISCTMPSIEVGTVG  
GGTSLPAQSACLEAIGCKGGGATPGENAQKLATVVAAATMAGELSLLAALAANTLVQ  
AHMQHNRKPAPKK

>*Phaeodactylum\_tricornutum*/HMGR

MRIHQYL YVPDRVSFVVRSTSSRPTFRVPVTPSHPPCLLVRSDSISTKTSASNTLSARHPPA  
PSAPVCSDPSSFTPIHTSKTNITLDNA AICQLVL TNQIKDHELEKRLDAHRAVQVRRLV  
VAHKLDTLEHINAHALDNL PSEPSLDYTRVHGANCEIVVGYVPLPVGLVGPLTVNGET  
VYVPMATTEGCLVASTNRGAKAITAGGGATAVLLRDGITRAPCVRMPSSAAQAHLKL  
WCETPQHFSTLKRAFESTTSFGKLLKCAPTVAGRNVYLRLTCFSGDAMGMNMVSKGS  
LAVIETLQEFPELVVALSGNMCTDKKAAATNWLEGRGKSIVVEATIPKDVVTNTLK  
TTVHSMVQTNLHKNLIGSAMAGALGGFNAHASNIVTAVFLATGQDPAQNVESNCITL  
LEETEEGDLWISCTMPSIEVGTGGTSLPAQAACLQAMGVRGGGATPGAHAQKLAQI  
VASATLAGELSLLAALAANTLVQAHMQRKPAK

>*Pseudonitzschia\_arenysensis*/HMGR

MAETMTPMPTIGMQLDAMIAQIDALPSTQLYGLIVAATVGLCVVVLGTGNSNLEMEQ  
HDLKKPKMVPVDGGKQPNWQIFKVINYIAVA AFLYSVFMFCSNASKYLH HESQGVLA  
QFLVGWSVFLMYFFGFFGVSLIHEDIPSEVDSVPSQPKSVAPTKPVHPPAACAPVCSDPS  
SFKAKKSCVPENIKELGDEEIADLV LANKVKDHMLEKLLDPFRAVTVRRIACNRKLS  
VHGNTSNVLDKLPSEHALDYSRVYGANCEIVVGYVPLPVGLVGPLTINDESFYVPMAT  
TEGCLVASSNRGAKAICQGGGAKARIVRDGITRAPCLRMNSAMEAADLKIWCEKPAN  
FAILKKAFESTTSFGKLI ECNPTVAGKNVYLRLVCFSGDAMGMNMVSKGSLAVIEKLQ  
EYFPSCQLVALSGNMCTDKKAAATNWLHGRGKSVVVECIIPKEVVRTTLKTTVAALV  
HTNVNKNLIGSAMAGAIGGFNAHASNIVTAIFLATGQDPAQNVESNCITLMEKQDNG  
DLWMCCTMPSIEVGTGGTSLPAQAACLEAIGCKGGGATPGANAKQLATVVAAAT  
MAGELSLLAALAANTLVQAHAHNRKPAACK

>*Pseudonitzschia\_australis*/HMGR

MPTIGMRLDAMVAQVDKVPSTQLYGLIVAATVGLCVVLLGTGNSNLEIQLEQNCNND  
LRKPKMVTVDAGKQPKWHIFKWINFVAVA AFLCSVFTFCSNASQYLHDESQGVLVQF  
LVGWSVFLMYFFGFFGV SFIHEDIPREEVDVIPSKPQNKSLSTNNISHNKNKNSISNKK  
SVHPPAACAPVCSDPSSFRSSKVASLPDDLKELGDEEIANLVLSNKKIKDHMLEKLLDPF  
RAVTVRRIACNQKLASVHGRDQQNGNVLDKLPSEPSLDYSRVHGANCEIVVGYIPLV  
GLVGPLTINGESVYVPMATTEGCLVASSNRGAKAITQGGGAKARIVRDGITRAPCLRM  
DSAMEAADLKIWCEQPSNFAVLKKAFESTTSFGKLECNPTVAGKNVYLRLVCFSGD  
AMGMNMISKGSLAVIETLQDKFPCKLVVALSGNMCTDKKAAATNWLLGRGKSVVVE  
CVIPKEVVRTTLKTTVAALVHTNINKNLIGSAMAGAIGGFNAHASNIVTAVFLATGQDP  
AQNVESNCITLMEEQDNGDLWICCTMPSIEVGTGGTSLPAQAACLEAIGCKGGGS  
TPGANAQKLATVVAAATMAGELSLLAALAANTLVQAHMVHNRKPAPKK

>*Pseudonitzschia\_delicatissima*/HMGR

MAETMTPMPTIGMQLDAMIAQIDALPSTQLYGLIVAATVGLCVVVLGTGNSNLEMEQ  
HDLKKPKMVPVDGGKQPNWQIFKVINYIAVA AFLYSVFMFCSNASKYLH HESQGVLA  
QFLVGWSVFLMYFFGFFGVSLIHEDIPSEVDSVPSQPKSVAPTKPVHPPAACAPVCSDPS  
SFKAKKSCVPENIKELGDEEIADLV LANKVKDHMLEKLLDPFRAVTVRRIACNRKLS  
VHGNTSNVLDKLPSEHALDYSRVYGANCEIVVGYVPLPVGLVGPLTINDESFYVPMAT  
TEGCLVASSNRGAKAICQGGGAKARIVRDGITRAPCLRMNSAMEAADLKIWCEKPAN  
FAILKKAFESTTSFGKLI ECNPTVAGKNVYLRLVCFSGDAMGMNMVSKGSLAVIEKLQ  
EYFPSCQLVALSGNMCTDKKAAATNWLHGRGKSVVVECIIPKEVVRTTLKTTVAALV  
HTNVNKNLIGSAMAGAIGGFNAHASNIVTAIFLATGQDPAQNVESNCITLMEKQDNG  
DLWMCCTMPSIEVGTGGTSLPAQAACLEAIGCKGGGATPGANAKQLATVVAAAT  
MAGELSLLAALAANTLVQAHAHNRKPAACK

>*Pseudonitzschia\_multistriata*/HMGR

MAQTMGTVPTIGMRLDAMVAQIDQLPSTQLYGLIVAATVGLCTVLLGTGNSNLEKQL  
QQNDDLKPKASAAPVGVDRGKQPKWHLFKWINFVAVATFLYSVFLFCSNATQYLH  
HESQGVLVQFLVGWSVFLMYFFGFFGVFSIHDDIPREAEVIPSQPSQPASGASKKTV  
HPPAACAPVCSDPSSFKSKASPPDNLKELPDEEIAIDLVLNKNVKDHMLEKLLDPFRAVT  
VRIACNRKLASVLGRDQANSNVLEKLPSEHSLDYSRVYGANCEIVVGYVPLPVGLVG  
PLTLNGESVYVPMATTEGCLVASSNRGAKAITQGGGAKARIVRDGITRAPCLRMGSAM  
EAADLKIWCEQPENFAVLKKAFFSTTSFGKLEKNPTVAGKNVYLRLVCFSGDAMGM  
NMISKGSLAVIETLQSYFPSCQLIALSGNMCTDKKAAATNWLHGRGKSVIVECFIPKEV  
VRTTLKTTVAALVHTNLNKNLIGSAMAGAIGGFNAHASNIVTAVFLATGQDPAQNVES  
SNCITLMEENENGDWMCCTMPSIEVGTVGGGTSPLPAQAACLEAIGCKGGGSTPGANA  
KKLATVVAAATMAGELSLAALAANTLVQAHMVHNRKPAPKK

**>*Pseudonitzschia pungens*/HMGR**

MAQTMGMTMATIGTRIDNVIAQFDQLPSTQLYGLIVAATVGLCVVLLGTGNSNFELHHQ  
RQKQLQQNDDLKPKMAAMPGRQPRWHIFKWINFVAVGAFLCSVFIFCSNASRYL  
HHESQGVLVQFLVGWSVFLMYFFGFFGVSLIHDDIPREEEESLPSKPQTKSVPTNTSAPK  
KTVHPPAACTPVCSDPSSFKSSKPSVPENLKELSDDEIAALVLGNKVKDHMLEKLLDPF  
RAVTVRRIACNQKLASVLGREDKSNVLEKLPSEPSLDYSRVFGANCEIVVGYVPLPVGL  
VGPLTINDES VYVPMATTEGCLVASSNRGAKAITQGGGARAKIVRDGITRAPCLRMNT  
AMEAADLKIWCEKPNFAILKQAFESTTSFGKLEKNPTVAGKNVYLRLVCFSGDAM  
GMNMVSKGSLAVIETLQKYFPTCQLVALSGNMCTDKKAAATNWL YGRGKSVVVECV  
IPKEVVRTTLKTTVSALVHTNLNKNLIGSAMAGAIGGFNAHASNIVTAVFLATGQDPA  
QNVESNCITLMEEEVNGDLWMCCTMPSIEVGTVGGGTSPLPAQAACLEAIGCKGGGV  
TPGANAKKLATVVAAATMAGELSLAALAANTLVQAHMAHNRKPASKK

**>*Pseudonitzschia pungens fr. cingulata*/HMGR**

MAQTMGMTMATIGMRIDNVIAQFDQLPSTQLYGLIVAATVGLCVVLLGTGNSNFELDH  
QRQKQLQQNDDLKPKMAAMPGRQPRWHIFKWINFVTVGAFLCSVFIFCSNASRY  
LHHESQGVLVQFLVGWCVFLMYFFGFFGVSLIHDDIPREEEESLPSKPQTKSVPTNTSAP  
KKT VHPPAACTPVCSDPSSFKSSKPSVPENLKELSDDEIAALVLGNKVKDHMLEKLLDP  
FRAVTVRRIACNQKLASVLGREDKSNVLEKLPSEPSLDYSRVFGANCEIVVGYVPLPVG  
LVGPLTINDES VYVPMATTEGCLVASSNRGAKAITQGGGARAKIVRDGITRAPCLRMN  
TAMEAADLKIWCEKPNFAILKQAFESTTSFGKLEKNPTVAGKNVYLRLVCFSGDA  
MGNMVSKGSLAVIETLQKYFPTCQLVALSGNMCTDKKAAATNWL YGRGKSVVVE  
CVIPKEVVRTTLKTTVSALVHTNLNKNLIGSAMAGAIGGFNAHASNIVTAVFLATGQD  
PAQNVESNCITLMEEEVNGDLWMCCTMPSIEVGTVGGGTSPLPAQAACLEAIGCKGGG  
VTPGANAKKLATVVAAATMAGELSLLSALAANTLVQAHMAHNRKPVSKK

**>*Skeletonema costatum*/HMGR**

MTFSEPTTIGMRIDAVIASLESPLHTNPEILYASLIGASLMFSFLVLNSGGSSMALPDGGD  
ADSLNKPKKISNAKPVSSSDEPQPKWHILRLTNYVVTFGFLLSVLKFNATTYLDDST  
SLLQFLIWSISMICYFFGFFGISFIELDELESEPQNHQHQMLRKPMPTKESAEPVVKVVK  
MHPPASSAPVCTDPTSFKKPSIPSNIKELSNGEIASLVLQDKVKDHQLEKLLDPHRAV  
EVRRKVDNQLDSLGRGGALAEPLHKHLDYKRVLGANCEIVVGYIPIPVGMAGPITL  
NGESVYIPMATTEGCLVASTNRGCKAITQGGGAVSSILRDGITRAPCVRLPSAKEASEVF  
LWIERPENFQKLKEVFESTTSFGKLEARPTVAGKNVYIRLRCFAGDAMGMNMISKGS  
LAVIECLRQVFPNISLVALSGNMCTDKKAAAINWIEGRGKSVVIEATIPKVVVRSTLKTS  
VPAIVEANLNKNLIGSAMAGVVGGFNAHASNNVTAVFLATGQDPAQNVESANCITLM  
EETPEGDLWISCTMPSIEVGTVGGGTGLPAQAACLKAIGVKGGGANPGDNAKQLAHV  
VAAATMAGELSLMAALASNSLVAAHMTHNRKPASK

**>*Skeletonema marinoi*/HMGR**

MTFSEPTTIGMRIDAVIASLESPLHTNPEILYASLIGASLMFSFLVLNSGGSSMALPDGGX  
ADSLNKPKKISNAKPVSSSDEPQPKWHILRLTNYVVTFGFLLSVLKFNATTYLDDST  
SLLQFLIWSISMICYFFGFFGISFIELDELESEPQNHQHQMLRKPMPTKESAEPVVKVVK

MHPPASSAPVCTDPTSFKKPSIPSNIKELSNGEIASLVLQDKVKDHQLEKLLDPHRAV  
EVRRLKVDNQLDSLGRGGALAEPLHKHDLKYKRVLGANCEIVVGYIPIVGMAGPITL  
NGESVYIPMATTEGCLVASTNRGCKAITQSGAVSSILRDGITRAPCVRLPSAKEASEVFL  
LWIERPENFQKLKEVFESTTSFGKLLERPTVAGKNVYIRLRCFAGDAMGMNMISKGS  
LAVIECLRQVFPNISLVALSGNMCTDKKAAAINWIEGRGKSVVIEATIPKGVVVRSTLKTS  
VPAIVEANVNKNLIGSAMAGVVGGFNAHASNNVTAVFLATGQDPAQNVESANCITLM  
EETPEGDLWISCTMPSIEVGTVGGGTGLPAQAACLKAIGVKGGGENPGDNAKQLAHV  
VAAATMAGELSLMAALASNSLVAAHMTNHRKPASK

>*Skeletonema\_menzelii*/HMGR

MTSEATTIGMRIDAVIASLESPLHTNPEILYASVIGASLLFSFLVLNSGGSSMALPDGGD  
ADSLNPKKISKAKLVSPSDEPQPKWHILRLTNYIVTFGFLSVLKFNATTYLDDSTSL  
LQFVIIWSISMCYFFGFFGISFIELDELENEPQNHQHQLRKPMPTKESADPPVKVVKMH  
PPASSAPVCIDPTSFKKPSIPSNIKELSNSEIASLVLQDKVKDHQLEKLLDPHRAVEVR  
RLKVVDVQLDSLGRGGALAEPLHKHDLKYKRVLGANCEIVVGYVPIVGMAGPITLNG  
ESIYIPMATTEGCLVASTNRGCKAISQSGAVSSILRDGITRAPCVRLPSAKEASEVHLW  
IESPENFQKLKEVFESTTSFGKLLERPSVAGKNVYIRLRCFAGDAMGMNMISKGSLAV  
IECLRQRFPNLSLVALSGNMCTDKKAAAINWIEGRGKSVVIEATIPKGVVVRSTLKTSVP  
AIVEANLNKNLIGSAMAGVVGGFNAHASNNVTAVFLATGQDPAQNVESANCITLMEE  
TPEGDLWISCTMPSIEVGTVGGGTGLPAQAACLKAIGVKGGGENPGDNAKQLAHVVA  
AATMAGELSLMAALASNSLVAAHMTNHRKPASK

>*Thalassionema\_fruenfeldii*/HMGR

MSSADTAMTWGMRLDVFLSQMDSTMVSDNLSSTQLYGLIVLVTGVISFSILGFGPKDE  
LPAAVTNNNSNNNNXNNKKKIVGSNYNNKEPKWFVLRYLNYVAVAAAFVASVAMFG  
WNAQAILKDTSSLLKFLMGWCVFLCYFLGFFGISFVDPDDIVESSNTSPNATTTTTTPA  
AAASKPKKVLPAPSAPVCSDPVSKTRSKLGDINHLSDEEVVALVLNNQLKDHQLES  
KLNPDRAVVVRRQIVAQKLATLGCGLDNLPCQPSLDYSKVHGANCEVVVGYIPIVGM  
MVGPLTLNGESVYIPMATTEGCLVASTQRGAKAISSGAGATCTILKDGITRACLRLPS  
AQQAAQLAKWCMDENNFKLKKAFESTTSFGKLLKAQPTVAGRNVYLRLVCFSGDA  
MGMNMVSKGSLAAIDLLRSHFPQVELIALSGNMCTDKKAAATNWLEGRGKSVVVEA  
TISKQVVETLKTVAIVDTNLQKNLVGSAMAGTLGGYNAHASNIVTAVFLATGQDP  
AQNVESSQCITLLEETKEGDLWMSCTMPALEVGTVGGGTSLKAQRACLECMGVAGGS  
REEPGKNAQKLATMVAATLAGELSLLAALAANTLVAAHMKHNRK

>*Thalassiosira\_minuscola*/HMGR

MPSSPTAPPPPPTLGMQIDAVIASLESISTPQLYGTVIAISLAFS FYLLNSGGGSSSPALQM  
MTNDTKEQQDADADSLRKPAAAAATSSSSSESSSEDRPEPRWHILRFTNYVAALGFL  
VSVLQFASNASAYLNDNSLLQFLTVWSLFFVYFVGGFFGISFVELDDFVDGAQISSVA  
RQRAQQQQQQQQQQSQNMTVAASQAKTPGRKSHKTDGSDSPPIKVVNVHPPATSAP  
VCSDPKSFKKPSAASIPPNLKELPNSEIASLVLQDKIKDHQLEKLLDPHRAVQVRRLKFD  
ARLSALGNGGALDELPHRHLDLEYKRVLGANCEIVVGYVPIVGMAGPVTLNGESVYIP  
MATTEGCLVASTNRGCKAISQSGASSTILKDGITRAPCVRLPSAKEAAEVALWIETPE  
NFATLKEAFESTTSFGKLLSARPTVAGKNVYVRLRCFSGDAMGMNMISKGSLAVIECL  
RQIFPALSLVALSGNMCTDKKAAAMNWIEGRGKSVVIEATIPADVVRSTLKTSVTAIVE  
ANLNKNLIGSAMAGTVGGFNAHAANNVTAVFLATGQDPAQNVESSNCITLMEETPEG  
DLWMSCTMPSIEVGTVGGGTGLPAQAACLRAIGAKGGGENPGDNARRLAHVVAAT  
MAGELSLMAALASNSLVAAHMAHNRKPASK

>*Thalassiosira\_pseudonana*/HMGR

MNHLLSNYQPSTSQIYSLVIVLSVAFSFRLLNSGDDLGSKLSSTVEQAKRGEKGTSSK  
YAKANNKNNKQSYDFPQPKWHILKLTNYVVTTFLLSIITFLSNASVYLNNTALMTF  
LGVWSGLLCYFFGFFGISFVELDDL VVTDADGQQRQQGVVTQPQKQASCKSRKVESP  
NDPPVKVISLHPPASSTPVCSDPKSKSPTNTGIPTNVKDLSNEEIATLVLQDKIKDHQLE  
KLLDPHRAVAVRRLKFDALLSSLGKTTGEDDKKGGVLSLPEHDLKYKRVLGANCEI

VVGYIPIVGVIPVTLNGESVYIPMATTEGCLVASTNRGCKAITQGS GARSTILRDGIT  
RAPCVRLPSAHEAAQVHLWIEEADNFAKLKEAFESTTSFGK LIEASPTVAGRNVYIRLR  
CFSGDAMGMNMISKGSLAVIECLREQFPQLSLVALSGNMCTDKKAAAMNWIEGRGKS  
VVIEATIPKDVVRSTLKTSPVAIVEANVNKNLIGSAMAGTVGGFN AHAANNVTAVFLA  
TGQDPAQNVESNCITLMEVSPEGDLWISCTMPSIEVGT VGGGTGLSAQSACLRAIGVK  
GGGENPGDNARQLAHVVACATMAGELSLMAALASNSLVAAHMQHNRKPASK

**>Thalassiosira\_puntigera/HMGR**

XDTSELKGSKSSNRSPIXPSQNSESMGDATSAPPTVGMQIDAAIAQLE NVSTTQLYAI VI  
AASLAFS FYLLNSGGGSSALTIMNEKDGDADSLRKPEAR KRKDRPRASTSSDEPEPRW  
HILRITNYVVAAGFLLSVLQFASNASTYLNDSTSL LQFLSVWSIFLCYFFGFFGISFIELD  
DFVDQQPSQQQQPSHKLKKAESISDPDPVKV VNVHPPSSSAPVCTDPTSFKKPPSNVP  
GNLKDLPNSEIASFVLQDKIKDHQLEKLLDPHRA VEVRLKFD AKLDSLGRGGALADL  
PHKHDL DYKRVLGANCEIVVGYLPIPVGLAGPITLNGESVYIPMATTEGCLVASTNRGC  
KAISQGS GASSTILKDGITRAPCVRLPSAKEAAK VALWIGTPENFLTLKKA FESTTSFGK  
LLDATPTVAGRNVYIRIRCFSGDAMGMNMISKGSLAVIETLKKVFPDLSLLALSGNMC  
TDKKAALNWIEGRGKSVVIEATIPKDVVRSTLKT SVQAIVEANVNKNLIGSAMAGTV  
GGFN AHAANNVTAVFIATGQDPAQNVESNCITLMEETSEGD LWISCTMPSIEVGT V  
GGTGLPAQSACLKVIGVKGGGENPGDNARQLAHV VAAATMAGELSLMAALASNSLV  
AAHMSHNRKPTSK

**>Thalassiosira\_rotula/HMGR**

MVVATSAPPTVGMQIDAAIASLETSTTQLXATVIAVSLXFSFYLLNSGGESSAL TMMN  
ESSDSL RKPKXDRKRMKTQQQSQREETDEPEPRWHILRITNYV VATGFTLSVLQFASN  
ASTYLNDSTSL LQFLMVWSLFLCYFFGFFGISFIELD DDFVDQQPSQQQPSQQTQVAPSR  
QLKKAESSDPPVKV VNIHPPASSAPVCTDPTSFKK PSSKIPSNL KLPNSEIASLVLQDKI  
KDHQLEKLLDPHRA VEVRLKFD AKLDSLGC GGALTELP HKHALDYKRVLGANCEIV  
VGYIPIVGMAGPITLNGESVYIPMATTEGCLVASTNRGCKAISQGS GASSTILKDGITR  
APCVRLPSAREAAEVALWIGKTXNFLSLKEAFESTTSFGKLL DATPTVAGKNVYIRLRC  
FSGDAMGMNMISKGSLAVIDALRQIFPRLSLLALSGNMCTDKKAAAMNWIEGRGKSV  
VIEATIPQDVVRSTLKT SVRAITEANVNKNLIGSAMAGTVGGFN AHAANNVTAVFLAT  
GQDPAQNVESNCITLMEETPEGDLWISCTMPSIEVGT VGGGTGLPAQSACLKAIGVKG  
GGENPGDNAKQLAHV VAAATMAGELSLMAALASNSLVAAHMAHNRKPASK

**>Thalassiosira\_weissflogii/HMGR**

MAHPSSYERVSSPPPTIGKHIDAVIAKIDSLPSSYVYTLAIAASLTFSFYLLNSGGGSNSM  
QMEPHGDVDRASAIKKPKQKSRRSLDEPQPKWHILKLTNYVATSGFLLSVLKFVSNAS  
MYLNDSTSL LQFLTIWSIFLCYFFGFFGISFIELD DDFEQESH PQHQAPVSAINVSHAKHCN  
AVKSGSLEPPVKVINVHPHSTSAPVCSDPASFRKPSSTIPGNLKDMSNEEIASLVLDDKI  
KDHQLEKLLDPHRA VEVRLKFD AQLQSLGRGGALAE LPHKHDL DYKRVLGANCEIV  
VGYVPIVGMAGPITLNGESVYIPMATTEGCLVASTNRGCKAITQGS GAFSTIVRDGITR  
APLVRLPSAREAAEVLLWIEDENNFKILKEAFESTTSFGK LIEARPTVAGRNVYIRLRCF  
SGDAMGMNMISKGSLAVIECLRRKFPQLSLAALSGNMCTDKKAAAINWIEGRGKSVVI  
EATIPKEVVRSTLKTSPVAIVEANTNKNLIGSAMAATVGGFN AHAANNVTAVFLATGQ  
DPAQNVESNCITLMEETSDGDLWISCTMPSIEVGT VGGGTGLPAQSACLKAIGVKGG  
GDSPGDNSKKLAHV VAAATMAGELSLMAALASNSLVAAHMAHNRKPVSK

**MK protein sequences**

**>Amphiprora\_sp/MK**

MSENQDSAAKRQRHERIGESFLKRTRAFGKLILFGEHFV VYKVPALVGAVSAYTDCEV  
EFLEQANGDGD LQIIDNRPAVPNYKVKKAE EGKEAIGLV LKHFGLDPKESPMKLTFGG  
DLCAVSGIGASAAQVVSLSRALS LALPKGMTEEEINAAGYE GEKGYHGTSGIDNTAA  
TFGGVLRFRQTDGDPFVKKKIAEPIRIVYASTGITASTTTV VGDVKAKKEANEDWFDG



LLKKYVELVGKGEKALDEGDLTTLGKLMDENHALCQELTVSCKELDDLVDAAARAAG  
AIGAKMSGTGRGGLMLALTPTAELQDKVASALEKSAPQVWKTTF

>*Bolidomonas pacifica*/MK

XLGSEGESIIHGAPSGIDNSVSTFGGCLRMRIEGEDGEIRMVKDLMKMEATEILITNTF  
VPKSTKDLVAGVKRRWEEARKVYDPIFLAMSEISEAFLKGGGLGLERVNELVRLAQGLL  
RTIGVTHPAIEMVVEKSRAYGVETKLTGAGGGGGCVITLLG

>*Cylindrotheca closterium*/MK

MSDVSKTQAYGKLLIFGEHFVVYHVPALVGAVAASTTCELEFVQGGEAVEIVDNRPA  
VPNYKVKKAEEGRQAIDLVLKHFVNAEKQPKDTNLKLTFGGDLCCVSGIGASAAQTV  
SLSRAIDQVLKLESTEEINQGGYEKEKGYHGTSPSIDNTAATYGGLLKFRQRTDTPIFT  
KKDIKEPIRIVYASTGITASTTEVVGDVRAKKEADPAWFDNLLEQYKTLVKKAEDAMD  
QGDIPLLGSLMNENHALCQELTVSCDELQDLVIAARSAGAIGAKMSGTGRGGLMLCLT  
PTEEIQNKVFEALEQRAPQVW

>*Fragilariopsis kerguelensis*/MK

KLRTNNIMTEEIETSKRTSPPTNEKEDVDVTKKXKVDDEKNKELSTLAKTHAYGKLLIF  
GEHFVVYKVPALVGAVAAAYTDCSIEVVVDVDESNDVKTNEGDESKANNDTGIVEIVDN  
RPAVPNYKTKKAEEGKKAVDLVLNHFKIFGKGGKAKLTFGGNLCCASGIGASAAQVV  
ALARAIKQTFPQFATLTDDEINAAGYEKEKGYHGTSPSIDNTAATFGGLLKFRTEGEP  
IFEKKLKEPIRIVYASTGITSSTTEVVGDVRAKKEADPAWFDALLKKYCDIVSKGEKV  
LEDSNINELGKLMDENHKILQDLTVSCDELDTLVVAAKKAIGAIGAKMSGTGRGGLMI  
AIVGCEETQNAVADALDKIAPQVWKTSTFQ

>*Leptocylindrus apora*/MK

XVSCFGGAVYYQKEAATGSVKAQFDQLTIAKMDIILTNTFVPRSTKALVAGVGQRKENF  
KFAPYVITAIGNISDQFRDIITSNTQDPSVISDLIQTNQLLQSLGVSHETLNEICRITQKM  
DISSKLTGAGGGGCALTYIEPDCKYTKREVMDMIEQSGHGFKCLESSVGGFGVLFVDF  
AX

>*Leptocylindrus danicus*/MK

XTTTVAASAPGKLVILFGEHAVVHSQPAIAASLDDLRIYAVIRHEEVTSKGMLTVNMPNF  
KKGLVFRAPTASCLYETNVTRDVTQDVSPPSDEVQKIHKFIQKAGASHTCNSALASD  
QVEALTPLFFMMNGIIPALLNAKSSVSIYVRSEDLPLGAGLGSSAAFCVACSAALYRLR  
EKTGIMGVAAVDRNIKDNTRPEQASLDIINAWAFASEGVIHGNPSGLDNTVSCYGGA  
VYYQKDAETGNLRAFHELVPPLDMILTNTFIPRSTKALVAGVHKVKEDFEFTVDVFA  
AIGRISDQFRKALDPGDRGYRDTAKISQLIRTNQNLLKAIGVSHKALDEICHITEQMGV  
ASKLTGAGGGGCTFTYIPDCEYTRDEIREKIEESPHGFKCLSSLVGGTGVRYMNPDTIP  
IVDLAKMRNQTTDNSLPWDRTLQYILALGSLTALGFVAVGNMKKK

>*Nitzschia sp*/MK

MPPSADDTTTTTTKRPPEEETVTTDEGEETAkkakTEGEETEATNADAAATATDTAEDD  
CDTKGFGKLLIFGEHFVVYKVPAYVGAVAAAYTDCEMELLEDSKDKTDAPTLDIVDNR  
PAVPNYKVKKADEAKEAISLVLKHFQIDRSVKLVFGGNTAVSGIGASAAQVVAMA  
RAXYEKEKGYHGTSPSIDNTAATFGGLLKFRTEGDPIFTKKSLEPIKIVYASTGITSS  
TTTVVGDVRAAXNSNXTEGDPFTTKKSLKEPIKIVYASTGITSSTTTVVGDVRAKKEAEP  
AWFDLLAKYVALAEKGEKALEGDITTLGQLMDENHGLLQQLTVSCKELDDLVDAA  
AKKAGAIGAKMSGTGRGGLMLALTSPSEVQNAVAXALEKAGAPQVWKTSTFQ

>*Phaeodactylum tricornutum*/MK

MAGTPLINSRVCASAPGKAILFGEHAVVYGEPAVAAALDDLRIYAVIRHEEVTSKGMLTVNMPNF  
VIRVVMPLDPTPVDFALPVSLSFTSMESTNTPPTPDDATVLAQLLHTADPTLDDFSVQAL  
TPVLYLCNQILLKPLRFRSESNPSADGGYELWVRSQDLPVGAGLGSSAAFGVACAAALI  
QYRQNLTAHAGTAAGTTSTNTPIQYGGPDSATLSEIDRYAYYSEILLHGTSPSIDNAVS  
AHGGAIIFTKDVASTNGTVKMEHLTPSDDLTLVYTHVPRSTKTLVAGVRHFYQRHV  
GFVTLILEAMGAIARDFEQAIRDRTNAHHGDHASDSHDFGERVLTMVRTNQYLLQAV  
GVSHPSLDHVCAVVHEHFRDYGAAKLTGAGGGGCAFLVWKPGLAADVLATQRQLQ

YALTTTLTPSPPYRYQCLASVVGEGVLFLPAGDFPWDATTPSSATPSRSATNVGRLVS  
ALTVSALVATTWMVWRSIGTKAR

>*Pseudonitzschia\_arenysensis*/MK

MTESTSPTS VSKRIPEEEV ADETAKKAKVETSADSDSTLEKTHAYGKLILFGEHFVYK  
VPALVGAVAAAYTDCKIEVADVDEKDAANEKETGVVEIVDNRPAPVNYKTKKAEEGK  
KAIDLVLNHFKIFGKGKKATLTFGGNLCCASGIGASAAQVVALARAVKETMPEFSKLT  
DDEINAAGYEGEKGYHGTSPGIDNTAATFGGLLKFQRTDGEPIFEKKS MKEPIRIVYAS  
TGITSSTTEVVGDVKSKEADPAWFDDLMKKYCDVATSGEKAIEEGDMTKLGELLNE  
NHKLLQDLTVSCKELDTLVDAARAAGAIGAKMSGTGRGGLMLALPTTEESQNAVAEA  
LEKIAPQVWKTSFQ

>*Pseudonitzschia\_australis*/MK

MAPSIEETTPPSADAETSEAPVTEDASKRPSEEEAVADTAKKAKVETETETNTDAKTD  
SADADVDAERLAKTHAYGKLILFGEHFVYKVPALVGAVAAAYTDCSIKVEDAGDSAD  
DSETGIVEVIDNRPAVPDYKTKKAEEGKKAIDLVLNHFKIFGKGKKATLTFGGNLCCAS  
GIGASAAQVVALSRAVKETLPEFSKLTNDEINTAGYEGEKGYHGTSPGIDNTAATFGGL  
LKFERTDGDPIFEKKTIKEPIRIVYASTGITSSTTEVVGDVRAKKEADPAWFDDLLKKYC  
ELAAKGEKALEDGDNTELGKLLNENHKLQDLTVSCKELDTLVDAARAAGAIGAKMS  
GTGRGGLMLALPTTEIQNAVAEAELEKVAPQVWKTSFQ

>*Pseudonitzschia\_delicatissima*/MK

MTETSSPTS VQDTSKRAPEEEVAAETAKKAKVDTSAESESKLEKTHAYGKLILFGEHFV  
VYKVPALVGAVAAAYTDCSIEVAXVXEKDAANEKETGVVEIVDNRPAPVNYKTKKAE  
GKKKAIDLVLNHFKIFGKGKKATLTFGGNLCCASGIGASAAQVVALARAVKETLPEFSK  
LTEDEINAAGYEGEKGYHGTSPGIDNTAATFGGLLKFQRTDGEPIFEKKAIKEPIRIVYA  
STGITSSTTEVVGDVRAKKEADPAWFDDLMKKYCEVATKGEKAIEEGDMTKLGELLN  
ENHKLQDLTVSCKELDTLVDAARKAGAIGAKMSGTGRGGLMLALPTTEESQNAVAE  
ALEKIAPQVWKTSFQ

>*Pseudonitzschia\_multistriata*/MK

MAPSTEDTTPEATVAETTPAAPAEETA AKRPSEEEAENS AKKAKVDDSAAPAVDADAE  
RLAKTHAYGKLILFGEHFVYKVPALVGAVAAAYTDCSIKLEDAPADSKDESPTGIVEVI  
DNRPAPVNYKTKKAEEGKKAIDLVLNHFKIFGKGKKATLTFGGNLTCASGIGASAAQV  
VALARAVKETLPEFAKLT DDEVNAAGYEGEKGYHGTSPGIDNTAATFGGLLKFQRTD  
ADPIFEKKTIKEPIRIVYASTGITSSTTEVVGDV KAKKEADPAWFDDLLKKYCEVATKG  
EKALEDGDTALLGSLLENHKLQDLTVSCKELDTLVDAARAAGAVGAKMSGTGRG  
GLMLALPTTEIQNAVAEAELEKIAPQVWKTSFQ

>*Pseudonitzschia\_pungens*/MK

MAPSTDENPAEPAVAETPATDQIAKRPSEDEPD PETAKKAKVATTESDSSSTADAERLA  
KTHAYGKLILFGEHFVYKVPALVGAVAAAYTDCSIEVEDAPSSDSTDDQETGIVEVIDN  
RPAVPNYKTKKAEEGKKAIDLVLNHFKIFGKGKKATLTFGGTLCCASGIGASAAQVVA  
LARAVKETLPERFGKLT EDEINTAGYEGEKGYHGTSPGIDNTAATFGGLLKFQRTDSDP  
IFEKKTIQEPIRIVYASTGITSSTTEVVGDVRAKKEADPAWFDDLLQKYVELAKKGETAI  
EKGDIPELGKLLDENHKLQDLTVSCKELDDLVAARAAGAVGAKMSGTGRGGLML  
ALPTTEIQNAVAEAELEKIAPQVWKTSFQ

>*Pseudonitzschia\_pungens\_fr\_cingulata*/MK

MAPSTDETPAEPAVAETPATDQTA KRPSEDEPD PETAKKAKVATESTAESTAEAKATA  
KAKAERLAKTHAYGKLILFGEHFVYKVPALVGAVAAAYTDCSIEVEDAPADSTDDKE  
TGIVEVIDNRPAVPNYKTKKAEEGKKAIDLVLNHFKIFGRGKKATLTFGGNLCCASGIG  
ASAAQVVALARAVKETLPERFGKLT EDEINTAGYEGEKGYHGTSPGIDNTAATFGGLL  
KFRRTDSDPIFEKKTIKEPIRIVYASTGITSSTTEVVGDVRAKKEADPAWFDDLLRKYVE  
LAEKGETAIEKGDIPELGKLLDENHKLQDLTVSCKELDDLVAARAAGAIGAKMSGT  
GRGGLMLALPTTEIQNAVAEAELEKIAPQVWKTSFQ

>*Skeletonema\_costatum*/MK

XTRAFGKLILFGEHFVVYKVPALVGAVSAYTDCKLEYTTEPGLEVIDNRPAVPLYKEK  
KMDEGIEAINLTLTHLGVDTKKSGVKLTFGGDLCCASGIGASAAQVVSLARAVNIADS  
RSLSEDEINAAGYEGEKGYHGTSPGIDNTAATFGLLRFQRTEGAPIFDKKTLPVSP  
IRIVYASTGITASTTEVVGDVREKKEADEAWFADLIGKYTKLCEEGQKAVEEGDMDKLGKL  
MNLNHEYLQELTVSCKELDDL VVXARAAGAVGAKMSGTGRGGLMIALTPTEEQNEV  
GSALEKXAPQVWKTTF

>*Skeletonema\_marinoi*/MK

XFPVDQTRAFGKLILFGEHFVVYKVPALVGAVSAYTDCKLEYTTEPGLEVIDNRPAVPL  
YKEKKMDEGIEAINLTLTHLGVDTKKSGVKLTFGGDLCCASGIGASAAQVVSLARAVN  
IADSRSLSEDEINAAGYEGEKGYHGTSPGIDNTAATFGLLRFQRTEGAPIFDKKTLPVSP  
IRIVYASTGITASTTEVVGDVREKKEAXEAWFADLIGKYTKLCEEGQKAVEEGDMDKLGKL  
MNLNHEYLQELTVSCKELDDL VVAARAAGAVGAKMSGTGRGGLMIALTPTEEQNEV  
NEVYSALEKLAPQVWKTTF

>*Skeletonema\_menzelii*/MK

MTSFPVDQTHAFGKLILFGEHFVVYKVPALVGAVSAYTNCKLEYTAEPGLEVIDNRPA  
VPFYKEKKMEEGIEAIDLTLTHLGVDTKKRGVKLTFGGDLCCASGIGASAAQVVSLAR  
AVNIADSRSLSEDEINAAGYEGEKGYHGTSPGIDNTAATFGLLRFQRTEGAPIFDKKT  
LATPIRIVYASTGITASTTEVVSDVRAKKEADEAWFADLIGKYTDLCEQGQKAVEEGD  
MDKLGKLMNLNHEYLQELTVSCKELDDL VVAARAAGAVGAKMSGTGRGGLMIALTP  
TEEQNEVASALEKLAPQVWKTTF

>*Thalassionema\_frauenfeldii*/MK

XLIYLLNILTSEYLHGLEITVRSNFPMGAGLGSSAAFSVALTAALLRLKLALSEKPSLP  
PLGRPAENHLDTINTYAFYAETLLHGNPSGIDNSVSTYGGGICYTKDLVNPGKGSMERF  
RMPPLSVMLVNTNVPKSTKALVASVRTMYEEHPEVVGCILKCIGEITQEFRNLVEKKET  
KNQKVLTLVNTNQHMLRSVGVSHESLDHVCHVTETISEGQAATKLTGAGGGGCAMIV  
LQTNECNNSVSIKLTQKIQALEGGWHRWNWSSLGGS

>*Thalassiosira\_minuscola*/MK

MVNNFPTQRTRAFGKLILFGEHFVVYKVPALVGAVSAYTDCELEFLDKPGLEVIDNRPA  
VPL YKVKKKDEGDEAINLTLTHLGIDPKERGVKLTFGGDLCCASGIGASAAQVVSLA  
RAVNIADSRSLSEEEINAAGYEGEKGYHGTSPGIDNTAATFGLLRFQRTEGAPIFEKKS  
FPSAIRIVYASTGITASTTEVVGDVRSKKEADEAWFASLLGKYTSLVGEAEQVVEEGDL  
NKLGLMDQNHALLQELTVSCKELDDL VDAARAAGAVGAKMSGTGRGGLMIALTP  
EEVQNSVAAALEKLAPQVWKTTF

>*Thalassiosira\_pseudonana*/MK

MADDFPIHQTHAYGKLILFGEHFVVYKVPALVGAVSAYTDCKVEYLNPEPGLEVVDNR  
PAVPLYKEQKMEEGMEAINLTLKHLGVDTPKARGVRLTFGGDLCCASGIGASAAQVVSLA  
LARAFNVVDGRSMTEDEINAAGYEGEKGYHGTSPGIDNTAATFGLLRFQRTEGAPIFEKKS  
DKKSLKSPIRIVYASTGITASTTKVVGDVRAKKEADEAWFANLMEQYKVLVEDGQTA  
VEAGDLTTLGKLMQNHVLLQELTVSCKELDDL VAAAREAGALGAKMSGTGRGGLM  
IALTPTEEQSAVADALGELAPQVWKTTF

>*Thalassiosira\_puntigera*/MK

MSDFPTDQTRAFGKLILFGEHFVVYKVPALVGAVSAYTDCKLEYLDEPGLEVIDNRPA  
VPLYKEKKKDEGEEAINLTLAHLGVDTPKARGVRLTFGGDLCCASGIGASAAQVVSLA  
AVNFADSRSLSEDEINAAGYEGEKGYHGTSPGIDNTAATFGLLRFQRTEGAPIFEKKS  
FPSAIRIVYASTGITASTTKVVGDVRSKKEADEAWFADIMGKYTTLVEEAEQIVEAGNL  
IKLGKLMNQNHSLQELTVSCKELDDL VDAARAAGAVGAKMSGTGRGGLMIALTPTE  
DIQNT

>*Thalassiosira\_rotula*/MK

MDDEFPTDQTRAFGKLILFGEHFVVYKVPALVGAVSAYTDCKLEYLDKPGLEVIDXRP  
AVPRYKTKKHECDKSINLTLTHLGVDTPKARGVKLTFGGDLCCASGIGASAAQVVSLA  
RAVNAADSRSLSEDEINAAGYEGEKGSHTGTPSGIDNTAATFGLLRFQRTEGAPIFEKKS

SFPSAIRIVYASTGITASTTEVVGDVRSKKEADEAWFSGLMGKYTTLVEEAEQVVVGG  
DLEKLGKLMQNHALLQELTVSCKELDDLVDAAARAAGAMGAKMSGTGRGGLMIALT  
PTEDIQNNVATALEKIAPQVWKTTF

>*Thalassiosira\_weissflogii*/MK

XRAFGKLILFGEHFVYKVPALVGAVSAYTDCKLEYVDDKPGLEVIDNRPAVPHYKET  
KSAEGLEAINLTLSHLGVYDYNRGRVCLTFGGDLCCASGIGASAAQVVSLLARAVNIADS  
KSLSEDEINAAGYEKGYHGTSGIDNTAATFGGLLRFQRTDDAPIFEKKTLASPIRIV  
YASTGITASTTKVVDVRAKKEADEAWFANLIDQYKILVEEGQNAIEKGDLPGLGKLM  
DKNHELLQELTVSCKELDDLVAARHTGALGAKMSGTGRGGLMIALTPTEAIQNAVA  
DALEKIAPQVWKTTF

## MPDC protein sequences

>*Amphiprora\_sp*/MPDC

MTTPPQRMVMCTVTAPTNIAMVIKYWGKASVEHNTPTNSSLSVTLDQSDLRAVTTVATS  
PDFDQDCLWLNKLEEGAQNSKRFRACVDGVKALATDKVDGETRAVVVSKAEWQE  
WKVHVSSYNTFPTAAGLASSAAGYAALVAALTELYQAQETFDHQLSTIARQSGS  
SLYGGFVAWRQGSAAENDWKDSLAPVAPVEHWPEIRAVILVTNAQKKHTSSTTGMQ  
NSVRTSPFLKYRGEAVVPQRMHDMERAWLAKDFDAFAKLTMQDSNQFHSTCLDTYP  
PIFYLNDRDIIRLIHAYNDFYATTMNDGGGSGSNGATSLRVAYTFDAGPNAVLYTTA  
AHVQEVQLVQFFFPEPSLKETNGQTTLPYVRGMLAVEETTTTPLDEQLVQYCQEAL  
LLPTSGKVQYMYVTQSGPGPQVLLPCSLASDVKTDGDDQKMEINLDPPTGLNTYVKS

>*Bolidomonas\_pacifica*/MPDC

XLSVVFCAEEGYEGELSTIARQSGSAAAGAFREVSLMGSKEDGSDSMAEQICNENHWP  
EMRAVILVISDLKKDTSSTEGMGTSVATSELLGFRARKGGIVDKRMEIEKAYKERDFE  
TFGRITMQDSNQFHATCLDTYPIFYMNDTSKSVIKIVHAYNKWKGRVAAAAYTFDAGP  
NAVYICLEDVMDEVLALMLQYYPGKMF

>*Coscinodiscus\_wailesii*/MPDC

XSTQDEHNGAPFAVIHMETVSAPTNIACIKYWGKASIALNTPINSSVSITLSQSDLRAV  
TTASASTLFSNDRLWLNKKEEDISKSRFRACVVGVRALATDKICPTTGNVLVKKEDW  
NGMHVHVSSYNNFPTAAGLASSAAGYAALVAALAKLFNAVETYPGELSTIARQSGS  
ACRSLYGGFVAWRMGVVKEDGTDIAEQIADENHWPEIRALILVVSDDKKKGTSTAG  
MGTSVATSELLTYRAKEIVPERMTIEKAFLERDFETFGRTMEDSNQFHATCLDTYPI  
FYMNDVSRSIIRIVHVYNDWSGRICAAAYTFDAGPNAVLYTLKEYAVELCALMLKFYPG  
EGQDYINNEEIAKARDYCLDEELVA AVLQNGLAPQTGDVKMIYYTRSGPGPMSLGFE  
NALIDPKSGLNTYEP

>*Cylindrotheca\_closterium*/MPDC

SGSACRSLYGGFVAWRKGGKEDWSDSMAQQVADEHHWPELRALILVVSDDAKKDT  
STAGMSTSVATSTLLQHRIHSVVPDRMKEIETAFSGKDFQAFGKVTMRDSNQFHAVCM  
DTYPIFYMNDXSRMVVRIVHAYNAWAGEIRAAAYTFDAGPNAVLYTLQYVVEVGA  
XMKHYFPETGALSDYINNAEYTEKVQKFSLSXELLEATDKTGRVPASGDVKQFYTKS  
GPGPQTLSMEESNLDPKTGLNTYQP

>*Fragilariopsis\_kerguelensis*/MPDC

XSTERCYQVDLSKXNQSKIKSMSNTVENNNHAPVWMATVSAPTNIACIKYWGKADT  
KWNTPTNDSVSVTLQDGLRAVTTAAASSKFEKDRWLWLNKKEEGADDCTRFACV  
DGVKVLAAADPKWKDLKVHISSYNTFPTAAGLASSAAGYAALTAALVELMDAKETY  
KGQLTTIARQSGSACRSLYGGFVAWKAGSPEVSTEDGTPWADSIQIADENHWPEL  
RAVIMVVSDDAKKDTSTAGMSTSVATSTLLQHRIAKVVPERMVAIEKAFLGKDFESFG  
ELTMRDSNQFHAVCADTYPIFYMNDVSRIVRIAHAYNKWAGETRVAITYDAGPNA  
VLYTIEKYSVEVGALMTHYFPSIDADGFFNDSIYANDVTTFSLSPELLAATDKTGRKSH  
AGDVKMMYHTKSGPGPQVLSNEKDGNDLKTGLNIYTPK

>*Leptocylindrus\_apora*/MPDC

MSEIFSATVSAPTNIACIKYWGKASIPLNTPINDSVSVTLHQKDLKAVTTASCSASFKEK  
KLWLNIGIPVSLEGKRFYSCLNGVKALAKEKGFDMKVHVSSYNTFPTAAGLASSA  
AGYAALTFAFKELFDAKEDYEGQLSTIARQGSACSRLYGGFVAWRKGNDPYNNGE  
DSMAVQIAPESHWPEIRALILVVSDKKKDTSSSTKGMSTSVATSKLLAHRADSIVAKRM  
EDIEKAYLRKDFESFGKITMMDSNQFHATCLDTPPIFYMNDISKGVILQIHKYNEFCGE  
IRAAITFDAGPNAVLYTLEKYQIEILALVLHCYPCSDKSAFVSNSKLAEEARAFKLSDE  
LVDLVAVNHASVGDVKKVYCTESGPGPLRLDEKESLIDLVTGDNLYNA

>*Leptocylindrus\_danicus*/MPDC

XHVTVSAPTNIACVKEYWGKASIPLNTPINDSVSITLQQRDLKAVTTASCSTAFETDELW  
LNGAKVSLDGNRRFYNCLNGVRALGKKKLEGKVHVSSYNTFPTAAGLASSAAGYAA  
LVYAFKELFDAVEEYEGQLSTIARQGSACSRLYGGFVAWRKGSDDPYTEGKDSMAV  
QIAPETHWPEIRALILVVSDAKKDTSSSTKGMSTSVATSKLLAHRASQIVPERMEQIEKAF  
LAKDFEAFGRITMMDSNQFHATCLDTPPIFYMNDISKSIISLVHXYNDLYGEIRAAITF  
DAGPNAVLYTLEKYQVEILAFMLKCFSSIGDDFVSNSELAQASECELKPELLEFVQV  
AEEGDVKKIYCTESGPGPQRLDDGLCLIDLETGLNKYSP

>*Nitzschia\_punctata*/MPDC

AKRKDSAADAVPWSDSMAVQIADENHWPEIRALILVVSDAKKDTSSSTAGMSTSVATS  
TLLQHRIQTVVPERMEGIEKAFXKDSNQFHAVCMDDTPPIFYMNDISRMVVRIVHAYN  
QWAGEIRAAITFDAGPNAVLYTLDKYAIEV GALMLHYFPAPTAQNGTSYINKSDYAK  
EIEEYKLDPELLAATEKTGRTPAAGDVKMMYYTKSGPAPQVLSMEEANLDPTTGLNT  
YHP

>*Nitzschia\_sp*/MPDC

MSSTSAADDGTPIYMATVTAPTNIAVIKYWGKASAEWNTPTNDSVSVTLDQSDLKAV  
TTVAASSKFQXPSLVXLNGVEEVDVYKNKRFRACVDGVKALAADKDKWSGLKVHV  
SSYNTFPTAAGLASSAAGYAAALVAALQLMDAREVYPGQLTTIARQGSACSRLY  
GFVAWRKGTPEGGNLSSRNTPWDNSLAEQVADEHHWPELRALIMVVSDAKKDTSSST  
AGMSTSVATSTLLQHRIPNVVPARMEAIEASFKAADFDFKFEITMKDSNQFHAVCMDDT  
YPIFYMNDVSRMIVRIVHAYNTWSNEIRAAITFDAGPNAVLYTLDKYAVEV GALML  
HYFPSSSSDDGGYINNPTYAESIKSFKLDPSLLEATDKTGRSPSSGDVKMFYYTKSGPGP  
QVLGSEHSNLDPKTGLNIYQP

>*Phaeodactylum\_tricornutum*/MPDC

MTTPATHPIFLATVSAPTNIAVVKYWGKADEHYNTPINSSCSVTLHQDDLRAVTTVAV  
SKDFVQDRLWLNGVEVPHAATSRFRACVDGVLALAPDKYHTDDDNNNKTVIAIAQH  
EWPTLHVHVSSYNTFPTAAGLASSAAGYAAALVAALVQLTGATETFPGEFSTLARQGS  
SACSRLYGGGLVAWHAGTAEQWRDSRAEQLADEASWPALRAVIAVVSDAQKDTAST  
AGMQASVKTSPLLAFFRAAHVVPQRMQELTQAWRRRDFPVFGKITMQDSNQFHATCL  
DTPPIFYMNDVSRQIIRIVTAYNDYAGEIRAAITLDAGPNVLYVLEPHRPVLAALLR  
HFFPASGLEEQNDEVLDPALVHAAEATGRVPRDGDVRRHYVTRPGPGPRVLDNDADG  
TIDPHTGLNR YQPSG

>*Pseudonitzschia\_arenysensis*/MPDC

MSTADAPIYMATVSAPTNIACIKYWGKADAKWNTPTNDSVSVTLDQTDLRAVTTAAA  
SSKFEKDRLWLNGEEVVGASESKRFRACVDGVKLLAADPAKWKDLKIHISYNTFPTA  
AGLASSAAGYAAALVAALAEALMDAKETYKGLTTLARQGSACSRLYGGFVAWRAG  
TPEGRSDDDIPWSDSMAVQIADAEHWPEIRAVIMVVSDAKKDTSSSTAGMSTSVATSTL  
LPHRISTVVPDRMKAIEAFLAKDFAAFGELTMRDSNQFHSVCADTXPPIFYMNDVSR  
QIVRIAHAYNKWAGETRVAITYDAGPNAVLYTLDKFVIEV GALMTHYFPPSGGVEGFC  
NDAFTXDIAAFELDVLLXATDKCGRTPASGDVKMYYHTKSGPGPQVLDNDKDGXL  
DPVTGLNRYTR

>*Pseudonitzschia\_australis*/MPDC

MTTEGEGAPIYMATVSAPTNIACIKYWGKADAKFNTPTNDSVSVTLTDQSDLRAVTTAA  
ASSVFEKDRLWLNNGEEEEGASESKRFRACVDGVKLLAADPSKWKDLKIHSSYNTFPT  
AAGLASSAAGYAALVAALAQLMDAKESYQGQLTTLARQSGSACRSLYGGFVAWRA  
GTPEGRXDDDIPWSDSMAVQIADENHWPELRAVIMVVSADAKKDTSSSTAGMSTSVATS  
TLLPHRIAKVPERMVSIEKAFLAKDFEAFGELTMRDSNQFHSCADTYPPIFYMNDVS  
RQIVRIVHAYNKWAGETRAAYTYDAGPNAVLYTLEKHSVEVGALMAQYFPPSGGVDG  
FYNDASYASSVQDFALDPSLVEAADKCGRTPASGDVKMVYHTKSGPAPQVLDNEKDG  
NLDPKTGLNRYTKPN

>*Pseudonitzschia delicatissima*/MPDC

MAQDAPIYMATVSAPTNIACIKYWGKADAKWNTPTNDSVSVTLTDQSDLRAVTTAAAS  
ARFEKDRLWLNNGEEEEGASESRRFRACVDGVKLLAADPAKWKDLKIHVSSYNTFPTA  
AGLASSAAGYAALVAALAEMLDAKEEYKQQLTTLARQSGSACRSLYGGFVAWRAG  
TPEGRSEDDVAWSDSMAVQIADENHWPEIRAVIMVVSADAKKDTSSSTAGMSTSVATST  
LLPHRISTVVPDRMKAIEAAFLAKDFAAFGELTMRDSNQFHSCADTYPPIFYMNDVS  
RQIVRIAHAAYNKWAGETRVAYTYDAGPNAVLYTLTKYSIEVGALMTHYFPSTGDAGF  
FNDAAFQAQIEAA FELDAVLLAATDKCGRTPASGDVKMVYHTKSGPGPQVLDNDKDG  
NIDPTTGLNRYTRN

>*Pseudonitzschia multistriata*/MPDC

MTDAPVYMATVSAPTNIACIKYWGKADAKFNTPTNDSVSVTLTDQSDLRAVTTAAASP  
KFEKDRLWLNNGEEEEGASTSKRFRACVDGVKLLAADPEKWKDLKVHVSSYNTFPTAA  
GLASSAAGYAALVAALAEMLDAKEAYPGQLTTLARQSGSACRSLYGGFVAWRAGT  
PEGRSDDKEPTPWSDSMAVQIADETHWPELRAVIMVVSADAKKDTSSSTAGMSTSVATST  
LLPHRIAKVPERMVAIEKAFLARDFAAFGELTMRDSNQFHSCADTYPPIFYMNDVS  
RQIVRIVHAYNQWAGETRVAYTYDAGPNAVLYTLEKHSVEVGALMAHYFPPGEASVE  
SFYNDASFASDLGGFSLDPSLLAATDKCGRTPASGDVKMVYHTKAGPGPQVLDNETD  
GNLDPVTGLNRYTKPN

>*Pseudonitzschia pungens*/MPDC

MATEGEDSPIYMATVSAPTNIACIKYWGKADAKFNTPTNDSVSVTLTDQSDLRAVTTAA  
ASKRFESDRLWLNNGTEEEGAASSKRFRACVDGVKLLAADPAKWKDLKVHVSSYNTFP  
TAAGLASSAAGYAALVAALAEMLDATEAYPGQLTTLARQSGSACRSLYGGFVAWR  
AGTPEGRSDDEIPWSDSLAVQIADETHWPELRAVILVVSADAKKDTSSSTAGMSTSVATST  
LLPHRISTVPERMEAIEKAFLSKDFETFGELTMRDSNQFHSCADTYPPIFYMNDVSR  
QIVRIVHAYNKWAGETRAAYTYDAGPNAVLYTLEKHSVEVGALMAHYFPPSGGADGF  
YNDAGYEASVGGFVVDPSLLEATDKCGRTPASGDVKMVYHTKSGPGPQVLDNDKDG  
NLDPVTGLNRYTKPN

>*Pseudonitzschia pungens fr. cingulata*/MPDC

MATEGEDSPIYMATVSAPTNIACIKYWGKADAKFNTPTNDSVSVTLTDQSDLRAVTTAA  
ASKRFESDRLWLNNGTEEEGAASSKRFRACVDGVKLLVADPAKWKDLKVHVSSYNTFP  
TAAGLASSAAGYAALVAALAEMLDATEAYPGQLTTLARQSGSACRSLYGGFVAWR  
AGTPEGRSDDEIPWSDSLAVQIADETHWPELRAVILVVSADAKKDTSSSTAGMSTSVATST  
LLPHRISTVPERMEAIEKAFLSKDFETFGELTMRDSNQFHSCADTYPPIFYMNDVSR  
QIVRIVHAYNKWAGETRAAYTYDAGPNAVLYTLEKHSVEVGALMAHYFPPSGGADGF  
YNDAGYEASVGGFVVDPSLLEATDKCGRTPASGDVKMVYHTRSGPGPQVLDNDKDG  
NIDPVTGLNRYTKPN

>*Skeletonema costatum*/MPDC

XPTNIACIKYWGKADSHYNTPIINDSLSLTLDQSDLRAVTSASASLAFDKDRLWLNNGSEE  
XGAASSRRFRACVDGVRALATEKIXPDTKEVIVSKDDWKKMHVYIASYNTFPTAAGL  
ASSAAGYAALVAALVELFGAKESYPGEFTAIARQSGSACRSLYGGFVAWRAGGEKS  
DWSDSIAEQVADEAHWGEVRAVICVVSADAKKETSSTAGMSTSVATSELLAFRAKEVV  
SKRMKIIIEAXKAKDFEAFGKVTMMDSNQFHATCLDTYPPIFYMNDVSRIVQLVHRY

NDWAGEIRAAAYTFDAGPNAVLYTLDKHVIELLALVLKYYPCESEDLMGDYVNXKXM  
VSKALAFELDSGLLQALEKSCXIRDIGDVKMVYCTKGGPGPQVLDAX

>*Skeletonema\_marinoi*/MPDC

XSAPTNIACIKYWGKADSHYNTPIINDSLSLTLTLDQSDLRAVTSASASLAFDKDRLWLN  
GSEEVGAASSRRFRACVDGVRALATEKIDPDTKEVIVSKDDWKKMHVYIASYNTFPTAA  
GLASSAAGYAALVAALVELFGAKESYPGEFTAIARQGGSGSACRSLYGGFVAWRAGGE  
KSDWSDSIAEQVADEAHWGEVRAVICVVSADAKKETSSTAGMSTSVATSELLAFRAKE  
VVSKRMKIIIEAIKAKDFEAFGKVTMMDSNQFHATCLDTYPPIFYMNDVSRIVQLVH  
RYNDWAGEIRAAAYTFDAGPNAVLYTLDKHVIELLALVLKYYPCESEDLMGDYVNNK  
DMVSKALAFELDSGLLQALEKSCNIRDIGDVKMVYCTKGGPGPQVLDASSEVFEFKVE

>*Skeletonema\_menzelii*/MPDC

PKADGDTNPTIHKATVSAPTNIACIKYWGKADSHYNTPIINDSLSLTLTLDQSDLRAVTSAS  
ASLAFNRDRLWLNGLSEEVDAASSKRFRACVDGVRALATEKIDPDTKEIIVSKDDWKK  
MHVYIASYNTFPTAAGLASSAAGYAALVAALVELFGAKESYPGEFTAIARQGGSGSACR  
SLYGGFVAWRAGGEKSDWSDSIAEQVADETHWKEVRAVICVVSADAKKETSSTAGMST  
SVATSELLAFRAKEVVSKRMKLIIEAIKAKDFESFGKVTMMDSNQFHATCLDTYPPIFY  
MNDVSRIVQLIHRYNWAGEIRAAAYTFDAGPNAVLYTLDKHVIELLALVLKYYPCE  
KDLMGDYVNNKDMVTKALAFELDSLLSLEKSCSVRDIGDVKMVYCTRGGPGPQAL  
DASSEVDFDKVE

>*Thalassiosira\_minuscola*/MPDC

XKSADNPEPIHTATVSAPTNIACIKYWGKANSHYNTPIINDSLSLTLTLDQSDLRAVTTASA  
SLSFDRDRLWLNGLDEVLDASKAKRFRACVDGVRALATDKLDPDTNQVIVSIDDWKN  
MHVHIASNTFPTAAGLASSAAGYAALVAALVQLFGAKESYPGEFTAIARQGGSGSACR  
SLYGGFVAWRAGGTKDDWTDSDIADQVADELHWEEMRAVILVVSADAKKETSSTAGMS  
NSVATSELLAHRAKEVVPKRMKIIIEAIKSKDFESFGKVTMMDSNQFHATCLDTYPPIF  
YMNDVSRIVQMIHKYNEWAGEIRAAAYTFDAGPNAVLYTLDKYVVELLALVLKYYP  
AQSDLVGEYVNNNSAMSKAALEFELDPSLTEAIEKTCKVHEYGDVKMVYCTKGGPGP  
KILDASCEIFDFKIE

>*Thalassiosira\_pseudonana*/MPDC

IHTATVSAPTNIACIKYWGKASSKYNTPIINSSLSLTLTLDQSDLRAVTTAAASTSFTKDRL  
WLNGLSEEANAFTSKRFRACIDGLRALATDKVDPTTNEVIVSKSQWQSMHVHVASYNT  
FPTAAGLASSAAGYAALVASLVELYNAKESYPGEFTAIARQGGSGSACRSLYGGFVAWR  
AGGMKEDWSDSIAEQVADEMHWKEMRAVILVVSADAKKETSSTVGMETSVATSELLA  
HRAKEIVPKRMKIIIEDAIQAKDFEAFGKVTMMDSNQFHATCLDTYPPIFYMNDVSRV  
QMVTRYNEWAGEIRAAAYTFDAGPNAVLYTLDKYVVELLALVLKHYPQSPELLGEYV

>*Thalassiosira\_puntigera*/MPDC

XFEKDRLWLNGLSEVLDASKAKRFRACVDGVRALATDKIDPDTNKVIVSKDDWKRMH  
VHIASNTFPTAAGLASSAAGYAALVAALVRLYGAEESYPGEFTAIARQGGSGSACRSL  
YGGFVAWRSGGAKDDWSDSVAEQVADESHWEEVRAVILVVSADAKKETSSTAGMSNS  
VATSELLAFRAREVVPKRMGIIIEKAIKAKDFEAFGKVTMMDSNQFHATCLDTYPPIFY  
MNDVSRVSVQMVHRYNEWAGDIRAAAYTFDAGPNAVLYTLDKHVVELLALVLKYYP  
AQSEDLVGDYVNNNSVLSKQALEFELEPSLTEAVEKTGGYASTVT

>*Thalassiosira\_rotula*/MPDC

XNPAIHKATVSAPTNIACIKYWGKADSHYNTPIINDSLSLTLTLDQSDLRAVTTASASLSFE  
KDRLWLNGLSEVLDAKAKRFRACVDGVRALATDKIDPSTNQVIVSKDDWKMHMVHI  
ASYNTFPTAAGLASSAAGYAALVAALVRLYGAKESYPGEFTAIARQGGSGSACRSLYGG  
FVAWRAGGAKEDWSDSIAEQVADESHWEEVRAVILVVSXAKKETSSTAGMSNSVATS  
ELLGFRAKEVVPKRMKTIEAAXAKDFESFGMVTMMDSNQFHATCLDTYPPIFYMND  
VSRVSIQMVHRYNEWAGEIRAAAYTFDAGPNAVLYTLDKHIVELLALVLKYYPQSQD  
LVGDYVNNNEVLSKEALEFELDSSLTEAIEKTGRVHEYGDVKMVYCTKGGPGPMTLDA  
SSXIFDFKIE

**>Thalassiosira\_weissflogii/MPDC**

XKVVVALSKVDGDDPIIHTATVSAPTNIACIKYWGKANKHYNTPINSSLSLTLDQSDLR  
AVTTATASPSFSDRLWLNQEVDSACTSKRFRACVDGVVALATDKVDSSSGKILVSK  
DDWKKMHVHIASYNTPFTAAGLASSAAGYAALVAALVELYGAKESYPGEFTAIARQG  
SGSACRSLYGGFVAWRQGGEEKEDWSDSIAEQVADESHWPEMRAVILVVSDAKKETSS  
TDGMSTSVATSELLAFRAKEIVSKRMKTIEEAIKAKDFEAFGKVTMMDSNQFHATCLD  
TYPPIFYMNDVSRSIQMVHKYNSWAGGIRAA YTFDAGPNAVLYTLEEYVVELLALVL  
KYYPEVGEDYVNNVNLKNEALEFQLNPSLIECVEQSSRIHETGDVKMVYCTKGGTGPQ  
VLDAVSSVIDFKIE

**HDR protein sequences**

**>Bolidomonas\_pacifica/HDR**

MKLVSAAALILLYASPLASSFAFSLSSSTTSISTSSSLRRSPFISRSSLHANPLDVRDKGVETK  
KDKRRSYMAKDTYYKNGFKGVREEVKVNMEQQFKSETVGELKSSDHVLFDRDNVEVH  
LAKDFGFCWGVERSIA MAYEARTHFPDRKLHITNELIHNPQVNGKLQEMDVQFVPKT  
GGEDGGKDFSAVGDGDVVILPAFGASIEEMALFDSKNVEVVDTTCPWVSKVWNTVD  
MHQRRGLTSVIHGKYAHEETVATASFCETYIIVKDMKEAEYVADYIVNGGDREEFLKK  
FENAMSEGFDPDLDLKLGLANQTTMYKKETRAIGQLFQRAQMSKYGPIDVKEHYME  
FDTICDATQERQDAIHDLVNNKDELGLDFILVGGWDSSTQHLLEIPLHAGLRAFHN  
EGGCIGADNTITHRAMDGTIVKEEFMGKGLVKMGVTS GASTPDRAVQDALDRIFLLKK  
VNPEYSSSQLQESILRADP

**>Chaetoceros\_neogracile/HDR**

MKFTSAALLLSAVTTASAFAPNLNLSTRYTSISAPLQMSETTETEEEEAKISKKQQRLLFM  
KNDKFFRNGFKEVRKDVSTMEGQFKSNVDEMKENKFMEREGVKVHLASDFGFC  
WGVERSIALAYEAVKHFPDRKVHITNELIHNPVNDNLHAMDVNFIQKTDDGKDFTV  
VGENDVVILPAFGASFEEMQFFDQKNVDVVDTTCPWVSKVWNTVDTHQKKGLTSIIH  
GKYEHEETVATVSFCEDYICVKNMKEAEMVADYITNGGDKA AFMNHFKHAHSDGFD  
PDTMLDCLGLANQTTMYKKETRAIGQMFQKTMMEKFGPVAETHYMEFDTICDATQ  
ERQDAVHDLVVKSEDELGLDFILVGGWDSSTAHALLEIPQKAGVRSFHINRAECITAD  
NTITHRTMSGEIVTESLILDTTKDVVMGVTS GASTPDAAVQDSLSSIFMLKKLSAKE

**>Coscinodiscus\_wailesii/HDR**

ILFRELSFGTTFANWTPPDVNTPVDTFAPSLVTASSNLLHLRHNTPTRIMKFNPTLLL  
LLLTSSPVTNAFVVITPSPIQLSSSSSPPTTTTRISMSTTADAADSSTTRKSKKQDRLRW  
MKEPSFHRRGFKDQRDGAEKRMNDSFKSDIVQEMRDSNFKIEREGVTVYLAKDFGFC  
WGVERSIALSYEAVDHFPGKNIHLTNELIHNPVNDLKNMNVQFIEKLEGGKKRFEA  
VEDGDVVILPAFGASYEEMDYLDKKNVQVVDTTCPWVSKVWNTVDTHQKNGLTSIIH  
GKYGHEETVATTSFCEDYICVKDYKEAEMVADYIVNGGDKEAFLEYFSKAVSKGFDP  
DKMLNKVGLANQTTMYKKETRAIGQMLQKAMMKKYGPVEADKHYSEFDTICDATQ  
ERQDAVAELVENASTLDLDFILVGGWDSSTAHALLEIPHKAGVRSFHIDRAERINADN  
TITHRTVEGEIVTEPFLDLSKDVVMGVTS GASTPDSAVQASLGSIFLLKEVHDA

**>Cylindrotheca\_closterium/HDR**

MKISGAAVFAIAIASVEAFSPSVQNRNGATALFSTAEESKPRTKKEGRLGFFKSENFH  
RKGFKVVRDNVEDTIKDQYQSSLVDDLKSSNYVIEKDNVYVYLAKDFGFCWGVERSI  
ALAYEAVKHYPDKNLHITNELIHNPVNDRLTEMNVKLIIEKLGEGKKDFDVIEDGDVV  
ILPAFGASYEEMDFLDKKNVEIVDTTCPWVSKVWNTVDAHQRKGLTSIIHGKYAHEET  
VATTSFCEDYICVKDMKEAEMVADYMINGGNREEFLEYFKNAVSEGFDPDKMLEKVG  
LANQTTMYKKETKSIGQLFQKTIMKKFGPSKVDEHYFEFDTICDATQERQDAVTELVE  
NKNDYGLDFILVGGWDSSTAHALLEIPHHQGIRSFHINRAECISADNTITHRTVDGEIV  
TEPLLLDLKDVVLGVTS GASTPDAAVQDALSQIFLLKMKMQGAAKEESASKSD

**>Fragilariopsis\_kerguelensis/HDR**



SIMTTSHVVPVATSFQPIRRYGCSLRSASASSLSSSIIDKDTTTTDDVVIEKKKQHENKK  
KDDTTTTTKEKKKPKVYIKNEREQIARKKDFYRGAGVFKDVKRDVTATMRAQFDSE  
LMNAMKENPNYICIEKEGVEFFLAKTQFDSELMNAMKENPNYICIEKEGVEFFLAKDHG  
FCWGVERSINLAYS AVETFPNAKVHITNELIHNPMVNDRLHLKNVNFQKLERTPEEEE  
NENEKSDCCSTKKKNNAILDANPNAKDFSAIQEGDVVMLPAFGATLDEMKELDANK  
VKVVDTTCPWVAKVWNTVHTHQVRGLTSIIHGKYAHEETMATKSMAETYLCVKNID  
EAQYVVSDFILQELQQSDADKDRFESQEEQEAHQQQMTAAFMLKFRDACS AHFDPYVH  
LRKIGLANQTTMYKKETRAIGQLLQKTMMHTFGPELLNDHYEFDTICDATQVRQDA  
VDELCTMTLDETHPITLDFILVVGGFDSNTAHLLEIPHMRKVPSYHINTAECIKATNTIT  
HRTVDGVVIEDEFPLTPAMLWENEEQEQEGSTNTKKTCKLLRVGVTSGASTPDKEVQ  
DALGRIMMINKLKEEEEEHKEYE

**>Leptocylindrus\_ apora/HDR**

MKFSTFAILTLIQTTTVAAYTRSSTSLRMSSTDSSTSSPTTTK KIERNIMKAPSFFRAGF  
KEQREEVEGIMQDQFKSNLVDDLRSNEYVIERDGVKVLAKDFGFCWGVERSIALAY  
EAVKQFPEKNVHITNELIHNPEVNDHLKDMKVNFDKTEEGGKDFEVVGDKDVVILPA  
FGASLEEMKYFDK KDVTVVDTTCPWVSKVWNTVDTHXKKGLTSIIHGKYAHEETIAT  
TSFCEDYICVKDMKEAEYLCDYMLNGGDKEEFMKNKFAKAMSEGFDPDTMLEKVGIA  
NQTTMYKKETRAIGQLLQKTMMMEKHGPQNVNEHYMEFDTIXDATQERQDAVAELTE  
NSEKLDLDFILVIGGWSSNTAHLLEIPMNAGVRSFHINRADCITADNTITHRLMDGTIK  
TEEFLLNKREDGSVKMGVTS GASTPDAAVQESLGQIFLLNKVMFGKDA

**>Leptocylindrus\_ danicus/HDR**

SNAFVVPGRNGRAFSTKVG MSTTEKADEAKISKKKKRIDIMKNEGFFRNGFKEVRED  
AEQTMNQKFSELVDSLRAEEFKMERDGVTVYLAKDYGFCWGVERSIALAYEAVNH  
YPDRKVYITNELIHNPEVNDNLSAMKVNFIKKEDEGSKKFEVVQDNDVLLPAFGASL  
EEMQYFDEKNVDIVDTTCPWVSKVWNTVDKHQQNGLTSVIHGKYAHEETIATTSFCE  
DFICVKNMDEAEYVSNYILNGGDKEEFMKKFEKAVSEGFPDMLSKLGLANQTTMY  
KKETRAIGKLLQRTMMEKFGPAELNERYMEFDTICDATQERQDAVSELTVNAESLGLD  
FILVVGWSSNTAHLLEIPHMAGVRSFHINRADCIGADNTITHRSVEGEIVTEKFLIDA  
DKKDVVMGVTS GASTPDAAVQEALSQIFVLNAV LKTKQ

**>Nannochloropsis\_ gaditana/HDR**

MSPRSMPVLLVAVLAVLCTASVTA FHTPSSPSAPSIWGAKRGAASAAALTTRRQAAAD  
GEAVENVLSPKQERRRIQSPNFHRQGFKAESQMVSEFTSQLINEMKDSTYTFQ  
KGDLTIKLAKSFGFCWGVERAVAMA YEARAHFPEQRIHITNEIHNPGVNERLADMNID  
FLPVDEGKKQFEHIEKDDVVILPAFGATI QEMQLLDKKG VQIVDTTCPWVAKVWTAVE  
KHRKLESTSVIHGKWNHEETIATASFADV IIVKDIHEAEYVCNYIVNGGDKAEMK  
KNAVSEGFPD VDLK VGLANQTTMYKRETRAIGKLF EKTMMHKYGP AEVKNHYAE  
FDTICDATQERQDAITDLMEEK LDFILVGGFDSNTAHLKEIPMQHIPSFHIDRAERIS  
ADGSIEHREIDGSIVKTANFLPDGPLTIGVTS GASTPDAYLQEAIERLFLRLTKGQEAPQ  
KDTVNA

**>Nitzschia\_ punctata/HDR**

KEERL NFMKSEQFHRRGFKEVRDRVEKDIQTQYQSELVNDLKTSNYMIENKDANVKV  
YLAKDFGFCWGVERSIALAYEAVQHYPDRKLHITNELIHNPEVNDRLTEMKVN LIEKV  
GEGQKDFSTIEGDVVILPAFGASYEEMEMLDKKNVEVVDTTCPWVSKVWNTVDMH  
QRKGLTSVIHGKXGHEETIATTSFCEDYICVKNMKEAEMVADYIVNGGDKEEFLKYFS  
KAVSKGFDPDTMLKKVGLANQTTMYKKETKAIGQLFQKAMMKKYGPAEVDQH YME  
FDTICDATQERQDAVSELVENAKEYGLDFILVVGWSSNTAHLLEIPHMAGIRSFHIN  
RAECIGADNTITHRTVDGEIVTEPFLLDIEKDVVMGVTS GASTPDAAVQDSL SQIFLLKQ  
LNKKKEEGQKA

**>Nitzschia\_ sp/HDR**

MKFSQSVMLAVLTATSSVHAFTTTTPAQHQPLAFVGTKASPPTSSSVSVLHMSTTEEE  
ATASTKKQTAGGMSKKKEERLQFMKSDQFHRRGFKEVRDKVESDIEEQFNAKLVEDL

KESNYVLDKTDEVGKVVYLAKDFGFCWGVERSIALAYEAVQHYPDRKLHITNELIHNPEVNDRLTDMKVNLIKSTKGEKNFEKIEEGDVVILPAFGASYEEMDYFDKKNVEVVDTCPWVSKVWNAVDTHQRKGLTSVIHGKYGHEETVATTSFCEDYICVKDMKEAEMVADYIVNGGDKEEFLKYFEKAVSKGFDPDTMLSKVGLANQTTMYKKETNAIGKLLQKTLMKKFGPAKVDEHYEFDTICDATQVRQDAVSELVDNVDEYGLDFILVVGWWDSSNTAHLLEIPHMAGVRSFHINRADCIKADNTITHRTVGGEIVTEKFLDTSKDVVMGVTSGASTPDSVVQDSLSQLFLLKQTQAKEP

**>*Phaeodactylum\_tricornutum*/HDR**

MRFASSAVVFMMAVASTVTAFAQQTAFKPSRISIVVQOSTSATEAEEKRATKKEERLRMMKSDRFHRRGFKEVREGVESNMEDQFQSPIVNSLRTSNFVMDRDGVKVVYLAKDFGFCWGVERSIALAYEAAEHFPDRKLHITNELIHNPEVNENLKAKNVQFIEKLDGDTKNFASVQDGDVVILPAFGASFEEMDYFDKKNVEIVDTCPWVSKVWNTVDKHQKQGLTSVIHGKYAHEETVATTSFCEDFICVKNFKEAEMVANYILNGGDKDAFMKHFENAVSEGFDPAKHLKLVGLANQTTMYKKETRAIGQLFQKTMMKKFGPVKSKEHYMEFDTICDATQERQDAIHDMVESAQKDGLDFILVIGGWDSSNTAHLLEIPVHAGIRAFHINRAECIGADNTITHRTVDGEIVTTQLLEDMDKEVVMGVTSGASTPDAAVQDSLSQIFLLKKMYDESKK

**>*Pseudonitzschia\_arenysensis*/HDR**

MKFSSSSISALALLAANGSDAFAPAIQSPGLASTELFSTKSQSKKNDRLEFMKNPQYHRRGFTDVRPKIEEKMEEDYTSELVEDLKSNNYLVEKDGVKMYLAKDFGFCWGVERSIALAYEAVEHYPDRKLHITNELIHNPEVNDLTEMKVNLIKESTTGAKDFSPVQEGDVVILPAFGASYEEMEMLDKKNVEVVDTCPWVSKVWNAVDKHQKGLTSVIHGKYAHEETLATTSCEDYICVKNMDEAQMVDYMEKGEDNTPEDKEAFLNFFKNAISEGFDPTMLTKLGLANQTTMYKKETKAIGQLFQKAMMKIYGPSKVDEHYMEFDTICDATQERQDAVSELVENAKDLELDFILVVGWWDSSNTAHLLEIPHNEGVRSFHINRADCIGADNTITHRSVEGEIVTETFLTSPEKNVVMGVTSGASTPDGVSVDALGQIFLLKQMLSAKEE

**>*Pseudonitzschia\_australis*/HDR**

MKFSATLSAVLFLGSYANTCNAFSPAITSSQSRTIATSTTTALGATMSKKKEERLEFMKNPQFHRRGFKDVRGKVEETMENEYKADLVEDLKSNNYLLEKDGVKMYLAKDFGFCWGVERSIALAYEAVEHYPDRTLHITNELIHNPEVNDLSAMKVNLIKESTTGKDFSTVEEGDVVILPAFGASYEEMDLFDKKNVEVVDTCPWVSKVWNAVDKHQKGGSTSVIHGKYAHEETLATVSFCDDYICVKDMKEAEMVIAVMEKGADNTAEDKEAFMKHFEKAVSEGFDPTMLDKIGLANQTTMYKKETKAIGQLFQKSIMKIHGPAEVDQHYMEFDTICDATQERQDAVAELVHSASGLGLDFILVVGWWDSSNTAHLLEIPHHAGVRSFHINRADCIGADNTITHRTVEGEIVTEKFLTTLDDDDKDVVLGVTSGASTPDGVSVDALSQIFLLKQMATNKEE

**>*Pseudonitzschia\_delicatissima*/HDR**

MKFSSSSISALVLLASQGSEAFAPAPQGLGIASSTTELFSSTSEKRTQSKKNDRLGFMKNDQYHRRGFADVRPKIEEKMEEDYTSELVEDLKSNNFLVEKDGVKIYLAKDFGFCWGVERSIALAYEAVEHYPDKKLHITNELIHNPEVNDLTEMKVNLIKESTTGSKDFSVVEEGDVVILPAFGASYEEMEMFDKKNVEVVDTCPWVSKVWNAVDKHQKGLTSVIHGKYXXEETLATTSCFEEYICVKNMDEAXMVVEYMQKGKENTPEEKEAFLKFFENATSEGFPDTMLNKVGLANQTTMYKKETKAIGQLFQKAMMKIYGPSAVDEHYMEFDTICDATQERQDAVTELVVKADDLKLDLDFILVVGWWDSSNTAHLLEIPHHAGVRSFHINRADCIGADNTITHRTVDGEIVTEKFLTSPEKNVAMGVTSGASTPDGVSVDALGQIFLLKQMLSAKDE

**>*Pseudonitzschia\_multistriata*/HDR**

MKFSSTALSVVLLLGAGCDAFSPLANANTNTGAKASATALQAAAQGSMSQKKGDRLEFMKNPQYHRRGFKDVREKVEETMEEQYKSDLVEDLKSNNFLMEKEGVKMYLAKDFGFCWGVERSIALAYEAVEHYPDKKLHITNELIHNPEVNDLSTSMNVNLIKESTTGSKDFSPVEEGDVVILPAFGASYEEMEMLDKKNVEVVDTCPWVSKVWNAVDQHQRGGLTSIHGKYAHEETLATTSCEDYICVKDMKEAQMVDYIQKGKENTKEDKEAFMKYFEKA

VSEGFDPDTMLDKVGLANQTTMYKKETKAIGQLFQKAIMKTYGPAKVDEHYMEFDTI  
CDATQERQDAVSELVDNASGLGLDFILVVGWSSNTAHLLEIPHKAGVRSFHINRAD  
CIGADNTITHRTVDGEIVTEEFLTTLDDTDRDVVLGVTSGASTPDGSVQDALSQIFLLKK  
MASEKAE

>*Pseudonitzschia\_pungens*/HDR

MKFSSATVSAFVFLGSTCHAFSPVPRNTNVQTTVLRVSAEQAPEVNTGSMSAKKKDRL  
NFMKNPQFHRRGFKEVRTSVEETMEQEYKSDLVEDLKSNNYLLQKDGVKTYLAKDFG  
FCWGVERSIALAYEAVEQYPDRKLHITNELIHNPEVNDALTNMKVNLIAKESTTGTKD  
FSPVEDGDVVILPAFGASYEEMEMFDKKNVEVVDTTCPWVSKVWNAVDKHQRKGLT  
SIIHGKYAHEETLATTSCFEDYICVKDMKEAQMVDYMEKGADNTPEDKAAFLKYFE  
KAVSEGFDPDTMLGKIGLANQTTMYKKETKAIGQLFQKSIMKIHGPAEVDQHIMEFD  
TICDATQERQDAVSELVENASGLGLDFILVVGWSSNTAHLLEIPHKAGIRSFHINRA  
DCIGADNTITHRTVDGEIVTEKFLTLDEPDREVILGVTSGASTPDGSVQEALGQIFLLK  
KLAAKEE

>*Pseudonitzschia\_pungens\_fr\_cingulata*/HDR

MKFSSATVSAFVFLGSTCHAFSPVPRNTNVQTTVLRSTAEQAPEVNTGSMSAKKKDRL  
SFMKNPQFHRRGFKEVRTSVEETMEEYKSDLVEDLKSNNYLLQKDGVKTYLAKDFG  
FCWGVERSIALAYEAVEHYPDRKLHITNELIHNPEVNDALTNMKVNLIAKESTTGTKD  
FSPVEDGDVVILPAFGASYEEMEMFDKKNVEVVDTTCPWVSKVWNAVDKHQRKGLT  
SIIHGKYGHEETLATTSCFEDYICVKDMKEAQMVDYMEKGADNTPEDKAAFLKYFE  
KAVSEGFDPDTMLGKIGLANQTTMYKKETKAIGQLFQKSIMKIHGPVGVVDQHIMEFD  
TICDATQERQDAVSELVENASGLGIDFILVVGWSSNTAHLLEIPHKAGIRSFHINRAD  
CIGADNTITHRTVDGEIVTEKFLTLDEPDREVILGVTSGASTPDGSV

>*Skeletonema\_costatum*/HDR

MAALAGSAAAFAPSAQSLNRITSSLSMSTEVAEESKKRVTKKEKRLQFMKTETFHRRG  
FKDVREGVEKSMEEQFKSQLVDELTSNDFVLERDGVKVFLLAKDFGFCWGVERSIALA  
YEAVEHFPGKTVHITNELIHNPEVNDKLHDMVDVQFIEKLGEGKKDFSNVNDGDVVILP  
AFGASFEEMSMMNDKNVDIIDTTCPWVSKVWNTVDQHQQKGLTSVIHGKYAHEETV  
ATISFCEDYICVKDMKEAEMVVDYILNGGDKEAFLKYFENAVSKDFDPDTMLDKVGL  
ANQTTMYKKETRAIGQMMQKAMMKKFGPVDKDHYWEFDTICDATQERQDAIHEL  
VEKSDDLGLDFILVVGWSSNTAHLLEIPQKAGVRSFHINKSDCIGADNTITHRTMAG  
EIVTEKFIEDWGKDVVMGVVTSGASTPDKAVQDSLSSIFLLKKVHAGED

>*Skeletonema\_marinoi*/HDR

MKFSTSSVVAIAALAGSAAAFAPSAQSLNRITSSLSMSTEVAEESKKRVTKKEKRLQFM  
KTETFHRRGFKDVREGVEKSMEEQFKSQLVDELTSNDFVLERDGVKVFLLAKDFGFCW  
GVERSIALAYEAVEHFPGKTVHITNELIHNPEVNDKLHDMVDVQFIEKLGEGKKDFSNV  
NDGDVVILPAFGASFEEMSMMNDKNVDIIDTTCPWVSKVWNTVDQHQQKGLTSVIHG  
KYAHEETVATISFCEDYIXVKDMKEAEMVVDYILNGGDKEAFLKYFENAVSKDFDPD  
TMLDKVGLANQTTMYKKETRAIGQMMQKAMXKKFGPVDXKDHYWEFDTICDATQE  
RQDAIHELVEKSDDLGLDFILVVGWSSNTAHLLEIPQKAGVRSFHINKSDCIGADNT  
ITHRTMAGEIVTEKFIEDWGKDVVMGVVTSGASTPDKAVQDSXSSIFLLKXVHAGED

>*Skeletonema\_menzelii*/HDR

MKFSSSVLAIAALAGSATAFAPSARSLNRITSSLSMSTEVAEESKKASKKEQRLKFMK  
TDSFHRRGFKDVREGVEKSMEEQFKSKLVDELTSNDFVLERDGVKVFLLAKDFGFCW  
VERSIALAYEAVEHFPGKTVHITNELIHNPEVNDKLHDMNVQFIEKLGEGKKDFSNV  
DGDVVILPAFGASFEEMQMMNDKNVDIVDTTTPWVSKVWNTVDQHQRKGLTSVIHG  
KYAHEETVATISFCEDYICVKDMKEAEMVVDYILNGGDKEAFLEHFKNVSKDFDPDT  
MLDKVGLANQTTMYKKETRAIGQMFQKAMMKKFGPVDKDHYWEFDTICDATQER  
QDAIHELVENAGDLGLDFILVVGWSSNTAHLLEIPQKAGIRSFHINKSDCIGADNTIT  
HRTMSGEIVTEKFIEDWGDVVVMGVVTSGASTPDKAVQDSLSSIFLLKNVHAADN

>*Thalassionema\_frauenfeldii*/HDR

MLFRNLNLLLLLILNSESSHGFTASPPPSILSTQQRRTTVSTLSASVATDDDDTKKKDDKM  
KYLKFORDAIAKKDNYLRGAGVFKDVKNVTEXMAQQFDSELMNEMKKNPNYMLE  
KDGIEFYLAKDHFVCGWVERINLA YSAVETFPDNKLHITNELIHNPLVNKKLHDKNV  
NFIKDNENNKDFSQIEDGDVVMLPAFGATLDEMKLLDEKGVKVVDTTCPWVSKVW  
NTVHKHQVKGLTSVIHGKYAHEETMATKSMCETYLCVKNLEEA EYLADFILNAKPGD  
GRAEKFMKFAKACSPDFDPHRDLLKIGLANQTTMYKKETRAIGQLLQKTIMKKFGPD  
LINEHYFEFDTICDATQVRQDAIDELCDMHLSDDGPTLDFILVVGGFDSNTAHLLEIPQ  
MRGVRSYHINQADCISAENTIRHRNVKGDIVEGETLLLPKXPGSKLRIGVTS GASTPDK  
EVQDALGRIMMLNNLRENGELKA

**>Thalassiosira\_minuscola/HDR**

MKFTTTSIAAIALVSTGSHAFTPSQHTTRATLRPLSMSSTEAVEEKKPTKKEKRLQFM  
KTPTFHRRGFKEVRESVENNMQE QFESELV GELKSNDFVIEKDGVKVHLAKDFGFCW  
GVERSIALA YEAVEHFPGKTVHITNELIHNPEVNDKLDHMDVQFIEKIGEGKKDFSKVG  
DGDVVILPAFGASFEEMSLFDEKNVDVVDTTCPWVSKVWNTVDQHQRKGLTSVIHGK  
YAHEETVATVSFCEDYICVKDMKEA QMLADYIVNGGDKEAFLKYFEKAVSKGFDPDT  
MLDKVGLANQTTMYKKETRAIGQMMQKAMMKKFGPVD AKDHYLEFDTICDATQER  
QDAIHEL VENSDELGLDFILVVG GWSSNTAHLLEIPQKAGVRSFHINKADCIGADNTI  
THRTMSGEIVTEKFIEDVENKDVVMGVTS GASTPDKAVQDSLSSIFLLKKMNEASASA

**>Thalassiosira\_pseudonana/HDR**

MKSPTFYRQGFKDV RPAVEKSMEDQFKSSLVDELKTNDFVIEKDGVKVYLAKDFGFC  
WVERSIALA YEAVEHFPGKTVHITNELIHNPEVNDKLDHMDNVQFIEKLGEGKKDFSKI  
GEGDVVILPAFGASFEEMTLMNKNVEVVDTTCPWVSKVWNTVDQHQRKGLTSVIH  
GKYGHEETVATVSFCEDYICVKDIKEAEMVADYIINGGDKEKFLKYFEKAVSKGFDPD  
TMLDKVGLANQTTMYKKETRAIGQLMQKAMMKKFGPVNAKDHYLEFDTICDATQER  
QDAISDL VENSDELGLDFILVVG GWSSNTAHLLEIPEKAGVRSFHINKSECIGADNTIT  
HRTVDGEIVTEKFIEDIENKDKKLVMGVTS GASTPDKAVQDSL DQIFMLKKVYSKEE

**>Thalassiosira\_puntigera/HDR**

XTHQQADRLFLIVINSGPSTANSISNSNGRTIKMRFSSSTVMAVAALVSTGSHAFTPSPH  
ARRAAARPPSMSTVEAEAPTEERSKPPTKKEQRLKFMKKDSFYRRGFKEVRDDVERT  
MGEQFESELV GELKSNDFVIEKDGVKVHLAKDFGFCWVERSIALA YEAVKQFPEKTV  
HITNELIHNPEVNDHLHMDNVQFIEKV GEGQKDFSKVGDGDVVILPAFGASFEEMSLF  
DAKNVDIVDTTCPWVSKVWNTVDQHQRKGLTSVIHGKYAHEETVATVSFCEDYICVK  
DMKEAQM VADYIINGGDKEAFLKYFEKAVSEGFDPTMLDKVGLANQTTMYKKETR  
AIGQMMQKAMMKKFGPVD AKDHYLEFDTICDATQERQDAIHEL VENSSTLGLDFILV  
VGGWSSNTAHLLEIPQKAGVRSFHINKADCIGADNTITHTMEGEIVTEKFITDIEDGD  
KEVVMGVTS GASTPDKAVQDSLSSIFLLKKMSSGGD

**>Thalassiosira\_rotula/HDR**

MKFSTTSVMAVAALITGTTAFTPSHHTLQRTTALRMSTVEETSVEESKPVRKKEKRL  
QYMKTDTFHRRGFKEVREGVEKAMSEQFESELVDELKSNFVINKDGVKVHLAKDFG  
FCWVERSIALA YEAVEHFPGKTVHITNELIHNPEVNDKLDHMDVQFIEKIGEGKKDFS  
KVG DGDVVILPAFGASFEEMTLFDSKNVDIVDTTCPWVSKVWNTVDQHQRKGLTSVI  
HGKYAHEETVATVSFCEDYICVKDMKEA QMVADYITNGGDKEAFLKYFENAISEGFD  
PDTMLDKVGLANQTTMYKKETRAIGQLMQKTMITKFGPVNTKEHYLEFDTICDATQE  
RQDAIHDL VENADELKLD FILVIGGWSSNTAHLLEIPQKANVRSFHINKADCIGADNTI  
THRTMSGEIVTDKFIFDIEDKDKEVVMGVTS GASTPDAAVQDSLSSIFLLKKMSSSD

**>Thalassiosira\_weissflogii/HDR**

XSHHPHRTSLSLSQSTEVSQETTPATEKKPTKKEKRLQYMKDPNFYRKGFKDV RPNVE  
KTMETQFKSPLVDELKTSNYVIERDGV RVHLAKDFGFCWVERSIALA YEAVEHFPGR  
TVHITNELIHNPEVNDNLHKMNVQFIEKLGEGKKDFSKVEDGDVVILPAFGASFEEMSL  
MNEKNVEVVDTTCPWVSKVWNTVDQHQRKGLTSVIHGKYGHEETIATVSFCEDYICV  
KDYKEAEMVADYIVNGGDKEAFLKYFEKAVSKGFDPDTMLDKVGLANQTTMYKKET

RAIGQLMQKAMMKKFGPVDKDHYLEFDTICDATQERQDAIHELVENASELDLDFILV  
VGGWSSNTAHLLEIPQKAGIRSFHINKADCIGADNTITHRTMSGDIVTEKFIEDIEDAG  
KTVVMGVTSGASTPDKAVQDSLSSIFMLKKLQSGKEE

## **DXR protein sequences**

### **>*Amphipora*\_sp/DXR**

XPLVLSSAAAFSCNQALAPNNKASSSL SAXFPGSKGTGEGWIDASSPVNTEESLQKTLE  
KSLEGSNFQKRLSLLGSTGSIGTQTLEIVDACPDNFVVDALSAGNNAKLMAEQVMKYK  
PKVASLATPEAAEELKQLLKDAGCSDMPSIVHGEEGILEAATVESADTVVTGIVGAAG  
LKPTIEAIKLLKDIANKETLISGGPVINPLVEEYGVNMLPADSEHSAIFQCLQGVPPG  
GLRRVILTASGGAFRDMKDEL FQCKDEPRLIQAKATTHPNWDMGAKITLDSATMM  
NKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTS CIAQLGWADMRLPLVYSVS  
WPHRLKMPYKPLDLAELATLTFKAPDTEKYPCISLAEAGRTGGTMTAVLNAANEA  
NEMFREDVGLGFLDIPKLI EGAMEEHKDDLKTS DITLDDILSCDAWAREYVVKMTDK  
MKTEAVLL

### **>*Bolidomonas*\_pacifica/DXR**

MTYLLITLLLLITPTTLTFKYRAHRPTLPNQISTPLMTKRLHSTPPNKNWVNPCAPVSTP  
DSLTLTLQKSLQGSTFKKNLSILGSTGSIGTQTLDIVRAKPDNFKVTLS SAGRNDLLVE  
QIVEFQPDIVSVM DVEDIKKLEGMLKGKFKPSKIVGGPSGIEVSRDPSSDTLVTGIVGSA  
GLQPTVEGIKHGKDIALANKETLIAGGPVVVPMVEEYGVNMLPADSEHSAVFQSLQGC  
PPGGLSKVILTASGGAFRDFSKEELFDLCESDPSAVLAKATTHPNWDMGAKITLDSAT  
MCNKGLEVIEAHYLFGAGYDDIDVVVHPQSIVHSMVEMRDTSVVAQLGWADMRLPL  
VYALSWPHRLEMDYRRLNLAEIGSLTFKEPDQGKYPCIGLCYEAGRAGGTMTSVLNA  
ANEMANEIYRKDVGMGFLDIPRLIEGAMEGAREQGLYKTVGVELNDILESDQWARDW  
VGRKVEEGLGGKX

### **>*Chaetoceros*\_neogracile/DXR**

MKLTFGALVALLTVPATCAFTITAPSTKSTSLNAMKSGSKGTGEGWLERPAAPVNSEA  
SARLTLQESLKGSSFQKRLSLLGSTGSIGTQTLEIVDACPDNFVVDALSAGNNAELMAE  
QVMKYQPKIASLSTPEAAAELKDRLIAAGCKNMPEIMFGDDGILAAATYDTADTVVTG  
IVGAAGLKPTIEAIKLGKDIALANKETLISGGPVINPLVREYGVNMLPADSEHSAIFQCL  
QGVPEGLRRVILTASGGAFRDFS KDDLFTACMEDPRAVQAKATTHPNWDMGAKITL  
DSATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMIETQDTS CIAQLGWADMRLP  
LVYSVSWPHRLKMPYRPLDLAELSTLTFKPDMDKYPCIGLAYAAGRAGGTMTAVLN  
AANEANEMFRADIGLGLDIPKLVEGAMEAHKSDLKIDDVVLDDILSCDAWAREYV  
ESTAQSMRNGGKILSV

### **>*Chaetoceros*\_sp/DXR**

MPTSAWSFTVTTTPAPSSRQTCRHAVMKS GSKGTGEGWLERPSAPVNSEAAAARLTLQES  
LKGSSFKKRLSLLGSTGSIGTQTLEIVDACPDNFEVDALSAGNNAKLMAEQVLKYQPK  
VASLATAEAAAELKERLVAAGCQNMPEILYGDEGILAAATIDSADTVVTGIVGAAGLK  
PTIEAIKLLKDIANKETLISGGPVINPLVEEYGVNMLPADSEHSAIFQCLQGVPPGGL  
RRIILTASGGAFRDFS KDELFRLCQEDPRAVQAKATTHPNWDMGAKITLDSATMMNK  
GLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCLAQLGWADMRLPLVYSVSWP  
HRLRMPYKPLDLAELS KLTFARPDVEKYPCIRLAEAGRAGGTMTAVLNAANEANE  
MFRADVGLGFLDIPKLIESAMEAHKNLKLDDISLDDILSSDAWAREYVDSAAQKLLK  
NGDKILSV

### **>*Coscinodiscus*\_walesii/DXR**

MRINPTATLLLLVPLTTTSFHLSSPPHKVTTSLNTFQPPGSQGTGEGWISPASPVNTAAS  
LQKTLS ELLGSTFQKRLSLLGSTGSIGTQTLEIVDARPENFVVDALSAGVNTELMMEQ  
VLKYKPKIATMQTADAAEELKRLTDAGCENMPEIMWGDEGIVAAACVDTADTVVT  
GIVGAAGLKPTIEAIKLLKDIANKETLISGGPVILPLVEEYGINMLPADSEHSAIFQCL

QGVPPGGLRRVILTASGGAFRDLSKDELFDKMCKNDPRAIQAKATTHPNWDMGAKITL  
DSATMMNKGLEVIEAHYLFYGASYDDIDIVIHPQSIHSLVETQDTSQIAQLGWPDMLPL  
VYSISWPHRLSLPYKPLDLAELSTMTFKKPDVEKYPCIRLAYEAGRTGGTMTAVLNAA  
NEAVNEKFRDDVGLGFLDIPKLVEMAMEAHKDDLKTEGVTLDDILSCDAWAREFVNT  
AKIEESQIFV

**>*Cylindrotheca closterium*/DXR**

MKLSIAFSVLMAGQAVAFMPNAPKGVQTSMELVSGSKGTGEGWIDPSAPVNSAESLK  
LTLEKSLEGSNFQKRLSLLGSTGSIGTQTLDIVDACPDNFVVDALSAGTNAELMAEQVL  
KYKPKIATLATPEAAATELKERLKAAGCDQMPEILHGDDGILAASTIDTADTVVTGIVGA  
AGLKPTIEAIKLLKDIANKETLISGGPVINPLVEEYGINMLPADSEHSAIFQCLQGVPE  
GGLRRVILTASGGAFRDLSKDELFDQMCKDDPRAVQAKATTHPNWDMGAKITLDSATM  
MNKGLEVIEAHYLFYGASYDDIDVVIHPQSIVHSMVETQDTSQIAQMGWADMRLPLVYS  
VSWPHRLKMPYRPLDLAEVSKMTFMAPDYEKYPCIRLAYEAGRAGGTMTAVLNAA  
EANEMFREDVGLGFLDIPKLVESAMEAHKDDHKTSDVTLDDILSCDEWARQHVVES  
KTLMNQVLV

**>*Fragilariopsis kerguelensis*/DXR**

MFMLFPNSFSSGVVSLLLRGVILAVVSLMASNAFTIQSTSTASPTNLNLHSGSAGT  
GDGWIEPSAPVNSEASLKLTLQKSLEGTSFQKRLSLLGSTGSIGTQTLDIVDACPDNFV  
DALSAGTNSKLMTDQVLKYKPKIVSMATAEAAQELKLNLEDAGCSYMPIMHGDEGI  
LAAATIDSADTVVTGIVGAAGLKPTIEAIKLLKDIANKETLISGGPVINPLVKEHGIN  
MLPADSEHSAIFQCLQGVPEGLRRVILTASGGAFRDLSKDELFRMCEDDPRAVQAKA  
TTHPNWDMGAKITLDSATMMNKGLEVIEAHYLFYGASYDDIDVVIHPQSIVHSMIETKD  
TSCIAQLGWADMRLPLVYSVSWPHRLTMPYKPLDLTEVSQMTFMKPDHEKYPCIGLA  
YEAGRAGGTMTAVLNAAEVANELFRDDVGLSFLDIPKLVIEGAMEAHKDDHKT KDVT  
LEDILSCDAWARQHVMESXETKKTRFFDIGFNPKEPITLNDLVER

**>*Leptocylindrus apora*/DXR**

MKLTAAATFLFATEATHAFIVQRPSINADANVSLNMVFGKSGSKGTGEGWIDPCAPVN  
SEASMRKTLDEALKGASFKKRLSLLGSTGSIGTQTLQIVDACPENFEIDALSAGTNSAL  
MADQVMQYKPKVAAMATKEAADDLAERLKA YGFDEMPEILFGDEGTVA AATVDSA  
DTVVTGIVGAAGLLPTIEAIKLGKDIALANKETLISGGPVINPLVEKHGVNMLAADSEH  
SAIFQCLQGVPEGALRRIILTASGGAFRDFSKEELFSLCENEPELIQKKATTHPNWDMGA  
KITLDSATMMNKGLEVIEAHYLFYGASYDDIDIVIHPQSIHSMVETKDTSCIAQLGWAD  
MRLPIVYSISWPHRLTMPYKPLDLAEIGSLTFKKPDREKYPCIDLAYAAGNAGGTMTA  
VLNAAEANEANEMFRADVGLGFLDIPKLVIEGAMEECHKADHKTTDVSLLDILACDAWAR  
EYVAKASVKNESKIIV

**>*Leptocylindrus danicus*/DXR**

MKLNIAAFLLAQGASAFVPGTTQKSSANVLNVFGKSGSKGTGEGWIDPCAPVNTEA  
AMKKTLNEALKGATFKKRLSLLGSTGSIGTQTLDIVDACPDNFQIDALSAGTNSALMA  
EQVLKYKPVASLATKEAADDLKERLKAAGCEEMPEIFFGDEGTVA AATVGTADTVV  
TGIVGAAGLLPTIEAIKLGKDIALANKETLISGGPVINPLVEKHGINMLAADSEHSAIFQC  
LQGVPEGLRRIILTASGGAFRDYSKEDLFDLCKNNPEFVQQKATTHPNWDMGAKITL  
DSATMMNKGLEVIEAHYLFYGASYDDIDIVIHPQSIHSMVETKDTSCIAQLGWADMRLP  
IVYSVSWPHRLEMPYEPLDLAKIGSLTFDKPDREKYPCIDLAYAAGNAGGTMTAVLNA  
ANEANEKFRANMGLGFLDIPKLVIEGAMEECHKADHKVGDVTLDDILSCDEWARNYVE  
KESLNTATIFS

**>*Nannochloropsis gaditana*/DXR**

PSFPPSLRRHILHFFRKLHPSGPRVHGHNVLQRGCLLPILLVALHRTLHDFGNVQVAN  
LLLPEDFVRALVCGVEGGRHRSRPARVRQLDTGVLLVVGRLGERSQLHQVEGLV  
GHGDAVRPGDVEDGQTHVRPAELSNHARILGLNQGVDDGLRMDDHIDVVVGGPEQ  
VVCLDYFQPFIIHGGAVDSDLGPHVPVWMLLEGVCHGDLLQILHGPVPEGATRGGEDD

AAEAPGGHALQGLKNGAVLTVGGDDVDAVFLQEGHNGWAAADQSLLVREGDGLAC  
LDRFHRGQ

>*Nitzschia\_sp/DXR*

MRFYAASILALAAMAPEPASAFVGVNSFSLTQQSSSSQQSSMLHSSKMKSGSTGTGE  
GWIDPSAPVDSEASLKLTLKLSLEGAEFKKRLSLLGSTGSIGTQTLDIVDVCPDNFVVD  
ALSAGTNSKLMAEQVLKFKPKVASLATPQAAQELVDLLDDANMPKSERPEIMHGDEG  
ILACATVDSADTVVTGIVGAAGLQPTIEAIKLLKDIANKETLISGGPVINPLVKEHGI  
NMLPADSEHSAIFQCLQGVP EGG LRRVILTASGGAFRDLTKDEL FNMCRDDPRAVQAK  
ATTHPNWDMGAKITLDSATMMNKGLEVIEAHYLFGASYDDIDVVIHPQSIVHSMVETQ  
DTSCIAQLGWADMRLPLVYSVSWPHRLKMPYRPLDLAEVSQMTFMKPDFEKYPCIAL  
AYEAGRAGGTMTAVLNAANEANEMFREDIGLGFLDIPTLIESAMEAHKDDHKVNDV  
TLDDILSCDAWARKFVAEKSETMKSQVFA

>*Phaeodactylum\_tricornutum/DXR*

MRLTELSVSLIVGSAAAFAPTTGISATRRTSASTVRLQLVSGSQGTGEGWIDPSAPVNSK  
ASLEKTLKLSLEGSNFQKRLSLLGSTGSIGTQTL EIVDACPDNFVVDALSAGTNVELMT  
QQVLKYSPKVASMSTPEAAKELKRRLQDEGCKSMPLILSGQEGIVAAATVDTADTVVT  
GIVGAAGLEPTIEAIKMKKDIANKETLISGGPVINPLVEKYGINMLPADSEHSAIFQCL  
QGVPPGGLRRVILTASGGAFRDFSDELFRLCDNPRAVQAKATTHPNWDMGAKITLD  
SATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCIAQLGWADMRLP  
LVYSVSWPHRLKMPYKPLDLAELGSLTFSKPDHAKYPCIGLAYAAGRQGGTMTAVLN  
AANEANEMFRADVGLGFLDIPKLI EGAMEAHKEEFKKGVTLEDILHVDQWAREHV  
RQASEQLDKSKVIVV

>*Pseudonitzschia\_arenysensis/DXR*

NMNFLALALFLSAAPTASGFAPVSSSASATQLDVVSGSAGTGEGWIEPSAPVNSAESL  
KLTLKLSLEGATFQKRLSLLGSTGSIGTQTLDIVDACPDNFVVDALSAGTNSKLMTEQV  
LKYTPKVVS MATADAAAELKSNLESAGMSNMPTIMHGDEGILACSTIDSADTVVTGIV  
GAAGLQPTIEAIKLLKDIANKETLISGGPVINPLVKEHGINMLPADSEHSAIFQCLQG  
VPEGGLRRVILTASGGAFRDL SKDEL FNMCRDDPRAVQAKATTHPNWDMGAKITLDS  
ATMMNKGLEVIEAHYLFGASYDDIDVVIHPQSIVHSMIETKDTSCIAQLGWADMRLPL  
VYSVSWPHRLEMPYKPLDLAEVGMQMTFMKPDHEKYPCIALAYEAGRAGGTMTAVLN  
AANEANELFRKDVGLGFLDIPKLI EGAMEAHKDDYKKTDTVTLDDILSCDQWARDV  
MEKSLTMRDRFIGDIGTNP NKEAIFLTDLSES

>*Pseudonitzschia\_australis/DXR*

MNLSIALSLFLSAAPTAFGFAPISSSTPATTLNLNVFSGSAGTGEGWIEPSAPVNTEASLKL  
TLGKSLEGATFQKRLSLLGSTGSIGTQTLDIVDACPDNFVVDALSAGTNV KLMTDQVL  
KYNPTIVSMATEKAAQEVR SNLEDAGISKMPTVLHGDEGILAAATVDTADTVVTGIVG  
SAGLLPTIEAIKLLKDIANKETLISGGPVINPLVKEHGINMLPADSEHSAIFQCLQGVP  
EGGLRRVILTASGGAFRDLTKDEL FNMCRDDPRAVQAKATTHPNWDMGAKITLDSAT  
MMNKGLEVIEAHYLFGASYDDIDVVIHPQSIIHSMVETKDTSCIAQLGWADMRLPLVY  
SVSWPHRLEMPYRPLDLAEVSQMTFMKPDHEKYPCIALAYEAGRAGGTMTAVLNAA  
NEANEIFRKDVGLGFLEIPKL VESAMEAHKDDHKINDVTLDDILSCDEWARAFVMEK  
SESMKTELIL

>*Pseudonitzschia\_delicatissima/DXR*

MNFSTALALFLAAAPTASGFAPVSSSGSSATHLNVVSGSKGTGEGWIEPSAPVNSEASL  
KLTLKLSLEGATFQKRLSLLGSTGSIGTQTLDIVDACPENFVVDALSAGTNSKLMTEQV  
LKYTPKVVS MATEAAAELKSNLEDAGMANMPTIMHGDEGILACSTIDSADTVVTGIV  
GAAGLQPTIEAIKLLKDIANKETLISGGPVINPLVKEHGINMLPADSEHSAIFQCLQG  
VPQGGLRRVILTASGGAFRDL SKDEL FKMCRDDPRAVQAKATTHPNWDMGAKITLDS  
ATMMNKGLEVIEAHYLFGASYDDIDVVIHPQSIVHSMIETKDTSCIAQMGWADMRLPL  
VYSVSWPHRLEMPYKPLDLAEVSQMTFMKPDHEKYPCIALAYEAGRAGGTMTAVLN

AANEANELFRKDVGLGFLDIPKLIESAMEAHKDDYKKNVDVTLLEDILSCDQWARDV  
MEKSETMKSPVLL

**>*Pseudonitzschia multistriata*/DXR**

MKLSIALSLLLSAAPXTVAFAPVSSTAPSSTKLNVSFSGSAGTGDGWIEPSAPVNSEESLK  
LTLEKSLEGSSFQKXLSLLGSTGSGTQTLDIVDACXDNFVVDALSAGTNSKLLTDQVL  
KYTPKIVSMATEKAAKELKSNLEDAGMKNMPTIMHGEEGILAAATIDTADTVVTGIVG  
SAGLLPTIEAIKLLKDIANKETLISGGPVINPLVKEHGINMLPADSEHSAIFQCLQGVP  
EGGLRRVILTASGGAFRDFSKEELFNMCRDDPRAVQAKATTHPNWDMGAKITLDSAT  
MMNKGLEVIEAHYLFGASYDDIDVVIHPQSIVHSMVETKDTSCIAQLGWADMRLPLVY  
SVSWPHRLEMPYKPLDLAEVSQMTFQKPDHDKYPCIALAYEAGRAGGTMTAVLNAA  
NEAANEIFRKDVGLGFLDIPKLIESAMEAHKDDHKVNDVTLDDILSCDEWAREFVMEK  
SETMKTELLPR

**>*Pseudonitzschia pungens*/DXR**

MNLSIALSLLLSAAPTAFGFAPISSTPSTTTLNVSFSGSAGTGEGWIEASAPVNSEASLKL  
TLEKSLEGSSFQKRLSLLGSTGSGTQTLDIVDACPDNFVVDALSAGTNSKLMTDQVLK  
YNPKIVSMATEKAAQELKSNLEDAGLKNMPTILHGDEGILAAATIDSADTVVTGIVGSA  
GLLPTIEAIKLLKDIANKETLISGGPVINPLVKEHGINMLPADSEHSAIFQCLQGVP  
GLRRVILTASGGAFRDLTKDEL FNMCRDDPRAVQAKATTHPNWDMGAKITLDSATM  
MNKGLEVIEAHYLFGASYDDIDVVIHPQSIVHSMIETKDTSCIAQLGWADMRLPLVYSV  
SWPHRLEMPYKPLDLAEVSQMTFMKPDHEKYPCIALAYEAGRAGGTMTAVLNAA  
AANEIFRKDVGLGFLDIPKLI EIAMEAHKDDLKTEDVTLDDILSCDEWARQFVLEKSET  
MKSELILPR

**>*Pseudonitzschia pungens fr. cingulata*/DXR**

MNLSVALSLLLSAAPTAFGFALISSSTPCTTTLNVSFSGSAGTGEGWIEASAPVNSEASLK  
LTLEKSLEGSSFQKRLSLLGSTGSGTQTLDIVDACPENFVVDALSAGTNSKLMTDQVL  
KYNPKIVSMATEKAAQELKSNLEDAGLKNMPTILHGDEGILAAATIDSADTVVTGIVG  
AAGLLPTIEAIKLLKDIANKETLISGGPVINPLVKEHGINMLPADSEHSAIFQCLQGVP  
EGGLRRVILTASGGAFRDLTKDEL FNMCRDDPRAVQAKATTHPNWDMGAKITLDSAT  
MMNKGLEVIEAHYLFGASYDDIDVVIHPQSIVHSMIETKDTSCIAQLGWADMRLPLVY  
SVSWPHRLEMPYKPLDLAEVSQMTFMKPDHEKYPCIALAYEAGRAGGTMTAVLNAA  
NEAANEIFRKDVGLGFLDIPKLI EIAMEAHKDDLKTEDVTLDDILSCDEWARQFVLEKS  
ETMKSELILPR

**>*Skeletonema costatum*/DXR**

MKLPSIAVGFFLASHQASGFSTNLLSSSSTKQSTSSLNVFSVSGSAGTGEGWIDPSAPV  
NSAASLKKTLDEALKGSTFKKRISILGSTGSGTQTLDIVDACPDNFVVDALSAGNNAK  
LMADQVMKYSPKVASLSTPEAAAELRLLKERGCTDMPEIVHGAEGIVEAATVGSAD  
TVVTGIVGAAGLLPTXEAIKLGKDIALANKETLISGGPVINPLVEKHGINMLPADSEHSA  
IFQCLQGVPPGGLRRVILTASGGAFRDFSDEL FEMCENDPRAVQKKATTHPNWDMG  
AKITLDSATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCIAQLGWA  
DMRLPIVYSVSWPHRLKMPYKPLNLAEIGSLTFKAPDTEKYPCISLAYEAGKKGGTMT  
AVLNAAANEANEMFREDVGLGFLDIPKLI EGAMEAHKEDFKVDDVTLDDILSCDAWA  
REYVAENTKNVVKAPILFT

**>*Skeletonema marinoi*/DXR**

XQASGFSTNLLSSSSTKQSTSSLNVFSVSGSAGTGEGWIDPSAPVNSAASLKKTLDEAL  
KGSTFKKRISILGSTGSGTQTLDIVDACPDNFVVDALSAGNNAKLMADQVMKYSPK  
VASLSTPEAAAELRLLKERGCTDMPEIVHGAEGIVEAATVGSADTVVTGIVGAAGLLPT  
IEAIKLGKDIALANKETLISGGPVINPLVEKHGINMLPADSEHSAIFQCLQGVPPGGLRR  
VILTASGGAFRDFSDEL FEMCENDPRAVQKKATTHPNWDMGAKITLDSATMMNKG  
LEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCIAQLGWADMRLPIVYSVSWPHR  
LKMPYKPLNLAEIGSLTFKAPDTEKYPCISLAYEAGKKGGTMTAVLNAAANEANEMF



REDVGLGFLDIPKLIEGAMEAHKEDFKVNDVTLDDILSCDAWAREYVAENTKNVVKAPILFT

**>*Skeletonema\_menzelii*/DXR**

MKLPSIAVGGFFLASHQASGFSTNLLSSSSITKQSSTSSLNVFSVSGSTGTGEGWIDPSAPV  
NSAASLKKTLDEALKGSTFKKRISILGSTGSIGTQTLDIVDACPDNFVVDALSAGNNAK  
LMADQVMKYSPKVASLSTPEAAAELKTLKRGCTNMPEIVHGAEGIVEAATVGSAD  
TVVTGIVGAAGLLPTIEAIKLGKDIALANKETLISGGPVINPLVEKHGINMLPADSEHSAI  
FQCLQGVPVPPGGLRRVILTASGGAFRDFSKDELFCMCEEDPRAVQKKATTHPNWDMGA  
KITLDSATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCIAQLGWAD  
MRLPIVYSVSWPHRLKMPYKPLNLAEIGSLTFKAPDTEKYPCIALAYEAGKKGGTMTA  
VLNAANEANEMFREDVGLGFLDIPKLIEGAMEAHKEDFKVDDVTLDDILNCDAWAR  
EYVAENTKNVVKAPILST

**>*Thalassionema\_frauenfeldii*/DXR**

MKLSIAAAVVCLCSSSSHAFAPTSTTTTTTTTTTISKQOSTNSQLFNQKFSGSAGTGENWI  
DPSAPVNSKESLELTLAKSLEGAQFQKRLSLLGSTGSIGTQTLDIVDACPDNFVIDALSA  
GTNVELMTEQCLKYQPKVASMATPEAAAQLKANLDGKGFRGQILSGDEGILAAATVD  
TADTVVTGIVGAAGLQPTIEAIKLLKDIALANKETLISGGPVINPLVKEYGINMLPADSE  
HSAIFQCLQGVPVPPGGLRRVILTASGGAFRDLTKEELFHMCRDDPRAVQAKATTHPNWD  
MGAKITLDSATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCLAQL  
GWADMRLPLVYSVSWPHRLKMPYQPLDLAKVSQMTFAKPDFEKYPCIALAYEAGRA  
GGTMTAVLNAANEANELFRQDRGLGFLEIPQLIEQTMKHKEDLKTDDITLDDILECD  
AWARQQVTDQSQSLLNKQYV

**>*Thalassiosira\_minuscola*/DXR**

MKLPSLALSLLLLAHEAASFTPAASTRATTSNLNVFSTPGSKGTGEGWIDPSAPVNTAES  
LQKTLSESLKGSTFQKRISILGSTGSIGTQTLIVDACPENYVVDALSAGNNAELMAEQV  
MKYSPKVASLSTPEAAEKLKSILKEKGCADMPEIVYGDEGIVEAATVGSADTVVTGIV  
GAAGLLPTIEAIKLGKDIALANKETLISGGPVINPLVEKHGINMLPADSEHSAIFQCLQG  
VPPGGLRRVILTASGGAFRDFSKDELFCMCEKDPMAVQKKATTHPNWDMGAKITLDS  
ATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCIAQLGWADMRLPIV  
YSVSWPHRLKMPYKPLDLAELGSLTFKAPDREKYPCIQLAYDAGKMGGTMTAVLNA  
ANEANEMFREDVGLGFLDIPKLIEGAMEAHKEDFKDKDVTLDDILSCDAWAREYVA  
ENTKNVVKAPIPL

**>*Thalassiosira\_pseudonana*/DXR**

MKLSLAFSTILLSQSSAFTTLTSTKSTTTTTSSSLNVFSPPGSAGTGAGWIDPSAPVNTPE  
SLQKTLSESLKGSTFQKRISILGSTGSIGTQTLDIVDACPDNFVVDALSAGNNAELMCEQ  
VLKYKPKVASLGTPEAAKLLKELLVKAGCTDMPEIVHGAEGIVEAATVSTADTVVTGI  
VGAAGLLPTIEAIKLGKDIALANKETLISGGPVINPLVEKHGINMLPADSEHSAIFQCLQ  
GVPPGGLRRVILTASGGAFRDLKDELFCMCEKDPMAVQKKATTHPNWDMGAKITLD  
SATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCIAQLGWADMRLPI  
VYSVSWPHRLKMPYKPLNLAEIGSLTFKAPDREKYPCIRLAYEAGRQGGTMTAVLNA  
ANEANEMFREDVGLGFLDIPKLIEGAMEAHKEDFKVKDVTLDDILSCDAWAREYVA  
EQTKNVAKSALIL

**>*Thalassiosira\_puntigera*/DXR**

MKLPSLALSLLLLAHEASSFAPA VGGPTRAATSLNVFSTPGSLGTGAGWIDPSAPVNSA  
ASLEKTLAESLKGSTFQKRISILGSTGSIGTQTLDIVDACPENFVVDALSAGNNAQLMAE  
QVMKYSPKVASLATPEAADELRAILKELGCADMPEIVHGAEGIVEAATVGSADTVVTG  
IVGAAGLLPTIEAIKLGKDIALANKETLISGGPVINPLVEKHGINMLPADSEHSAIFQCLQ  
GVPPGGLRRVILTASGGAFRDLTKDELFCMCEKDPMAVQKKATTHPNWDMGAKITLD  
SATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMVETQDTSCIAQLGWADMRLP  
IVYSVSWPHRLQMPYKPLNLAEIGSLTFKAPDREKYPCIQLAYDAGKKGGTMTAVLNA

ANEAANEMFREDVGLGFLDIPKLIEGAMEAHKEDFKDKDVTLDDILSCDAWAREYVA  
ENTKNVAKQVLVL

**>Thalassiosira\_rotula/DXR**

SGFLLAHDVSSFAPSIGSTRAESSLNVFSTPGSLGTGEGWIDPSSPVNSKESLEKTLSETL  
KGSTFQKRISILGSTGSIGTQTLDIVDACPENFVVDALSAGNNAELMAEQVMKYSPKVA  
SLSTPEAAEKLRTILKDRGCADMPDIVYGDEGILEAATVGSADTVVTGIVGAAGLLPTI  
EAIKLGKDIALANKETLISGGPVINPLVEKHGINMLPADSEHSAIFQCLQGVPVPPGLRRV  
ILTASGGAFRDLSKDELFRMCEEDSAAVQKKATTHPNWDMGAKITLDSATMMNKGLE  
VIEAHYLFGASYEDIDIVIHPQSIVHSMVETQDTSCIAQLGWADMRLPIVYSVSWPHRL  
KMPYKPLNLAIEIGSLTFKAPDREKYPCIQLAYDAGNKGGMATAVLNAANEANEMFR  
ADVGLGFLDIPKLIXGAMEAHKEDFKDKDVTLDDILSCDAWAREYVAEETKTVTKQA  
LIL

**>Thalassiosira\_weissflogii/DXR**

MKVPSIALSGLLLLHSATAFSSLSLPLNASKQSTLLHSSSSSKPIPGSLGTGAGWIDPSA  
PVNTAASLKKTLAKALEGAAFQKRISILGSTGSIGTQTLDIVDACPDNFVVDALSAGTN  
VRLMAEQCLKYRPKVASLSTPRAAEELKEILTKSGQFSKEDMPNIVYGNIEVEAATVS  
SADTVVTGIVGAAGLLPTIEAIKLGKDIALANKETLISGGPVINPLVEKHGVNMLPADSE  
HSAIFQCLQGVPEGGLRRVILTASGGAFRDMTKEELFRMCEEEKEAVQKKATTHPNW  
DMGAKITLDSATMMNKGLEVIEAHYLFGASYDDIDIVIHPQSIVHSMIETQDTSCIAQLG  
WADMRLPIVYSVSWPHRLKMPYKPLDLAQIGSLTFQKPDTEKYPCISLAYEAGRMGGT  
MTAVLNAANEANEMFREDVGLGFLDIPKLIEGAMEGHRENGFKKEDVTLEDILSCDA  
WAREYVKEKTKNVVKKPVVL

**DXS protein sequences**

**>Amphiprora\_sp/DXS**

MMMIMKVAAGIVAIVGLQGVSSFTTTPRLRTSTTTAAPRTSQLLAEQSGGTNASPPGPPY  
SGPAVKPILDSISFPTDMKRLDMAQLKQLAYELRWETLEAVSKTGGHLSSSLGVNELT  
VALHYVFDMPEDDIIWDVSHQCYPHKILTGRRFKMSGRLRQSGGISGFCKRKESEYDSF  
GAGHSSTSISAAQGM SIGKSMMGKRQNNCIAVIGDGAITGGMAYEAMNSAGYLKNRM  
IVVLNDNGQVSLPTGTPSAGGVVPASQLSTYTSNLLVSKPFQDFRSFAKSFNKLLPEGIQ  
GVNKRVD EYVRGLATGGTLFEELGFYVVGPDGHDLDNLVPILEKLRDSPNNKPVLL  
HLKTTKGYGYPPAEASDRMHGVAKFNLGTGAQFKSKPTAPSLTSVFANALIDAATED  
RRVVGITAAMPGGTGMDIFGRRFPKRTFDVIGIAEQHAVTMAAGMACEGLKPFCAIYST  
FMQRGYDQVVHDVAIQNLPVRMVLDRAGLVGNDGPTHHGVDLSYMGICIPNLIIMAP  
SDEIELKNMIMTQAAYDDGPTVCRYPRGTGYGPETLQDLFGYKLEDGQIPAKGEIIPMG  
KGRVIRRP GASRGKSKKDRVCILTIGTRLHEALVAAKEVEEQDPTLGVTVADARFMKP  
LDVDMVRQLADEHSVMITVEEGSIGGFGDHLHFLTL DGLMDDGDLKFRPMVIPDAL  
FEAGSQHEQYDLAGLNAEHIRGTILRLTKRKKAPQLEEVVEEQVSRD

**>Bolidomonas\_pacifica/DXS**

MKFVNALPTLLLTLTPTTSFLPYSPTITTTTTTKTPTTTKLNNNAGPPYGGPVRKPM  
DRVKSPDDMKGMSIQDLKQLSEELRWEVIEQVSKTGGHLGSSLGVIEMTVLHVFST  
PEDKIVWDVAHQCYPHKILTGRDRMNTLRQMGGLSGFCKMKESPYDAFGAGHSSTS  
ISAIQGYSIKSQLNPTGRNNCIAVIGDGAITGGMAYEAMNSAGYLKNRMIVILNDNGQ  
VSLPTGTASAGGTGPASELSAATSRLLVSKQFQDFRSVAKGVSGLPENVQEFNKKIDE  
YARGIVSGGTLFEELGFYVVGPLDGHLDNFVPILSKLRDSPSNKPVLLHVKTNGNG  
YPPAENASDKMHGVAKFDVATGIQKKGPKGPPSLTSVFANELCNIASTDKRVVGITAA  
MPGGTGMDIFGRRFPDRTYDVIGIAEQHAVTMAAAMAAEGLKPFVCIYSTFLQRGFDQ  
VVHDVAIQNLPVRLILDRAGLVGNDGPTHHGAYDLSYLGICIPNL TICGPSDEVELKNMI  
QTAASHDDGPIVVRYPGAGYGEDKLTFLGYDLPDGVPEKGETLEIGKGRVVRTSM  
RSLPRSKSVAVLSFGTRLHESLVAAAEIEEAAPDLSVTVADARFMKPLDVDLVRSLAST

HSVIVTVEENSIGGFGDSVLHFLSLEGLLDNGELRFRPMVLPDAYFEAGPQFEQYEQAG  
LNAKHIKGTILRLTEKIAVPIAV

>*Chaetoceros\_neogracile*/DXS

RKMKLFNAVAALSLAVPAANAFTTPVLSSKYVAKSALRMSEGGPPYAGPLTKPLLDSE  
FPNDMKDFSIKDLKQLSNELRWEVLEAVSRTGGHLSSSLGVVELTVALHYVFDMPD  
IIWDVAHQCYPHKMLTGRRHKFSTLRQLGGISGFCKRKESEYDSFGAGHSSTSI  
MSIGKSMLNKRTNNCIAVIGDGAITGGMAYEAMNSAGYLQNRMIVVLNDNGQVSL  
GTPSAGGTVPASQLSAYTSNLLVSKPFQDFRAFAKSFNLLPENIKDVNKRFD  
SGGTLFEELGFYVGPIDGHDLNMIPILEKLRDGGSNKPVLLHVKTNGQGYPPAES  
SDKMHGKFDLATGAQKKTATAPSLTSIFADSLIQAATEDRAVVGITAAMPGGTG  
MDIFGRRFPKRTFDVGIAEQHAVTMAAGMACEGLKPFVSIYSTFMQRAYDQVIHD  
VQNL PVRMILDRAGVVGNDGPTHHGCDLAYMACV PNLTIMAPSDEIELRNM MKTCL  
DFNDGPTVLRYPGNGYGAETLTELFGYELENGEIPTIGEPLPIGKGRIRRP  
DKNTRGKSREDRIAILSIGTRLHDSLISASDIEDANPTVGVTVADARFMKPLDEDL  
VDDHSILITIEEGSVGGFGDHLHFLAKDGLLDDGNLKFRPMNLPDELFEAATQHEQ  
MAKLNHPYITELLQGLLKKNMKVPVLEEQTAE

>*Chaetoceros\_sp*/DXS

MKLFHAAAAIAFAVPSAFAFTSPVLNTKYVAGSNTQLAMGPPYAGPTAKPLLD  
RDMKDFSIKDLKQLSHELWEVLEAVSKTGGHLSSSLGVIELTVALHYVFDMPDQII  
WDVAHQCYPHKMLTGRRHFLFGGLRQLGGISGFCKRKESEYDSFGAGHSSTSI  
MSIAKSMLNKRTNNCIAVIGDGAITGGMAYEAMNSAGYLQNRMIVVLNDNGQVSL  
GTPSAGGTVPASRLSAYTSNLLVSKPFQDFRDFAKSFNKLLPENIQDVNKRFD  
SGGTLFEELGFFYVGPIDGHDLNMIPILEKLRDSDSNKPVLLHVKTNGQGYPPAES  
SDKMHGKFDLATGVQYKKKATAPSLTSIFADSLIQAATDDRAIVGITAAMPGGTGM  
DIFGRRFPKRTFDVGIAEQHAVTMAAGMVCEGLKPFVSIYSTFMQRAYDQVIHD  
NLPVRFILDRAGVVGNDGPTHHGCDLAYMGCV PNLTIMAPSDEIELRNMVKTCAD  
DEGPTVLRYPGNGYGAEKLEVFVGYKLENGELPSKGEALEIGKGRIRRP  
ANIRGKSRQSRVAISLGLTRLHDSIAAAEIEAANPSVGVTVADARFMKPLDEDL  
DDNSILITIEEGSIGGFGDHLHFLARNGLLDDGNLKVRPMVLPDELFEAATQQEQYDM  
AKLNHPHITELVNNLLSKNMKVPVLEEKVTEQA

>*Coscinodiscus\_walesii*/DXS

MDIKTLNQLAYELRWEVLESVSKTGGHLSSSLGVIELTIALHNVFSAPEDDIIWDVSHQ  
CYPHKILTGRRRDRFSTLRQLNGISGFCKRKESEYDSFGAGHSSTSI  
KRRNNCIAVIGDGAITGGMAYEAMNSAGYLRSRMIVVLNDNGQVSLPTGTQ  
PASQLSAYTSNLLVSKPFQNFDFAKEINRLLPGDLQDINKRIDEYARGV  
GGTLFEELGFFYVGPIDGHDLNMIPILEKLRDSPSNKPVLLHIKTEKGGYPPAEQASDKMHGK  
KFELGTGVQFKGKASAPSFTSIFANSLIDAATEDRAVVGITAAMPGGTGM  
RTFDVGIAEQHAVTFGAGMACEGLKPFVCIYSTFMQRGYDQVVDVAIQNL  
DRAGLVGNDGPTHHGCDLAYMGAIPNLTIMAPSDEIELRNMVATCAAFDDGPTVLR  
YPRGSGYGPDKLQNLFGYTLNDQVPTKGGQVIEIGKGRIMRRPSASRGREKNDRIAILTI  
GTRLHDALVAADDIESTDPSLGVTVADARFMKPLDVLVRELVDQHGVLITVEEGSIG  
GFGDHLHFLALDGDALDNGDLKFRPMVIPDQYIEAGTQFEQYDIAGLNAGHIKGTALR  
LTSRIKVPVLQEE

>*Cylindrotheca\_closterium*/DXS

MRFSLAVALAALSSVSAFAPSTGPQNVLGKSPLRMAEDNVQGGPPYSGPASKPILDSVI  
DAKGLSHTQPASNINSKMRFSLAVALAALSSVSAFAPSTGPQNVLGKSPLRMAEDNV  
QGGPPYSGPASKPILDSVKFPSDMKKLSMSEVKQLAHELWEVLESVSKTGGHLSSSLG  
VIELTVALHYVFDMPEDDIIWDVAHQCYPHKILTGRRDKFSTLRQLDGISGFCKRKESE  
FDSFGAGHSSTSI  
AQAQMSIGKSLLNKRNKNCVAVIGDGAITGGMAYEAMNSAGYLN  
SRMIVVLNDNGQVSLPTGTPSAGGVVPASQLSTYTSNLLVSKPFQDFRSFAKSFNKLLP  
EGIQDVNKRLEDEYARGIVSGGTLFEELGFYVGPVDGHDVDNLVPILEKLRDSPNDKP

VLLHIKTTKGYGYPPAETASDRMHGVGKFNLTGTGVQVKSksiapsfTSLFANSLIDAAT  
EDRAIVGITAAMPGGTGMDFGRRFPKRTFDVGIAEQHAVCMGAGMACEGLKPFVCIY  
STFMQRGYDQVVHDVAVQNLpVRMILDRAGFVGNdGPTHHGcyDLAYLGCIPNLTIM  
SPSDEIELKNMVMTCAAFDDGPTVLRyPRGTGYGAEKLQSHFGYKLDGEVPAKGEVI  
EIGKGRIIRGPNAMIDNSSTTRGKERKDRVAILSIGTRLDEALVAAQEVEDSDENLSVTV  
ADARFMKPLDVLVRQLADHDGVLITXEEGSIGGFGDHVHLHFLALDGLMDSGNLKFR  
PMVMPDAYFEAATQNEQYDQAGLNSQHikGTILRLTERMkVPVLEEA

>*Fragilariopsis\_kerguelensis/DXS*

MRMRVSITVLLFSAAAITIPNAVlGFAPSSSSTTTTSRRKIPFTQHAAVGDDGSGNPHT  
VSGPPYSGPTVKPILDSVNYPSDMKGLDMRQLKQLANELRWEVLEQVSKTGGHLSSSL  
GVTELTVALHYVFDMPDDDIWDVAHQCYPHKILTGRRDQFPTRLQLGGISGFCKRKE  
SKYDSFGAGHSSTISAAQGMSIAKSVLNKRNNCVAVIGDGAITGGMAYEAMNSAG  
YLQSRMIVVLNDNGQVSLPTGTSPASAGGTVPASQLSTYTSNLLVSQPFQDFRDFAKSFNK  
LLPADLQGVNKRIDEYTRGLVSGGTLFEELGFYyVGPIDGHDLNLIPILEKLRDSSNTK  
PVLLHLKTVKGYGYPPAESASDRMHGVGKFNLTGTAQYKKGKPTSPSFTSIFANALIDA  
AVEDRSVVGITAAMPGGTGMDFGRRFPKRTYDVGIAEQHAVCMGAGMACEGLKPFV  
CIYSTFMQRGYDQVVHDVAVQNLpVRMMLDRGGFVGNdGPTHHGcyDLAYXGCIPN  
LTXMAPSDEIELRNMVKTCADFDDGPTVCRYPRGVGYGADKLQSIFGYQLTNNEIPTK  
GEAVPIGKGRIIRGTGGMVDTSSSSTRGKKRDRVAILSIGTRLHEALVAAQEVEDMY  
DLSVTVADARXMKPLDVLVRQLADENGLITIEEGSIGGXGDHVLHFLSLDGLSDE  
GNIKFRPMVIPDAYFEAATQFQQYEQAAGLNAQHikGTILRLANKVEVPVLEQIEQEVL

>*Leptocylindrus\_apora/DXS*

HKEEEAAAAATILISFPFQINNNIINILLTMKYSSAALLSLACFTQTAyGFAPSSRHHT  
HKSNGNSNALQESPNDIDNGPPYSGPAYKPLLDSSSPADMKRLDVKQLNQLAHEL  
RWVLENVSKTGGHLSSSLGVVEMTVALHYVFDMPEDDIVWDVSHQCYPHKILTGR  
RDRFPTLRQAGGISGFCKRKESEYDSFGAGHSTTSISAVQGMSIAKSM LNKRNNCIA  
VIGDGAITGGMAYEAMNSAGYLKNRMIVILNDNGQVSLPTGTQSAGGTRPASQLST  
YTSKLLVSNEFQSVRDVLKNINQLLPSEALQDVNKRIDEYARGIISGGTLFEELGFYy  
VGPVDGHD MENMVGILETLRDSDSNKPVLLHVKTEKGHGYP PAMAAADKMHG  
VGKFDIATGVQYKPKTKPVAPSLTSIFANALIEEAKEDPTVVAITAAMPGGTGLDIF  
GRRFPKRTFDVGI AEQHAVTMAAGMACEGLKPFCCIYSTFLQRGYDSVVHDIAVQ  
NLpVRIIMDRAGLVGNDGPTHHGcyDLAYLGCIPQLIIMAPSDELELKNMIRTVRNY  
DDGPTCLRYPRGSGYGV EKLKGLFGYSDEEINATEPTVLP MGKGRIVRKSARDAK  
DRANRVCILSLGTRLHESLVA ANDIEATDSVAVTVADARFMKPLDIDLIRELASENS  
VMVTIEEGSIGGFGDHVHLHFLT LDGLMDDGNLKFRPMVLPDQYFEAATQFEQYDLA  
GLNASHIKGTVLRRLA QKIQVPVPQES

>*Leptocylindrus\_danicus/DXS*

TKMRFTSAALLALLPSAYAFVPAKNAQHNTALHMSGTESGPPYSGPKVKPLLD  
SVNFPSDMKRFDIKELKQLAHEL RWEVLENVSKTGGHLSSSLGVVEMTVALHHVFD  
MPEDDIIWDVSHQCYPHKMLTGRKDKFSTLRQAGGISGFCKRKESEYDSFGAGH  
STTSISAAQGM SIAKSM LNKRNNCIAIGDGAITGGMAYEAMNSAGYLKNRMIVIL  
NDNGQVSLPTGTQSAGGTRPASQLSTYTSQLLVSKPFQNV RDVLKNINQLLPSETL  
QDVNRRIDEYTRGLISGGTLFEELGFYyVGPVDGHD MENLVGILETLRDSSTKPV  
LLHVKTEKGHGYP PALA AADKMHGVGKFDLGTGVQFKPSKPPVAPSFTSIFAN  
SLIDEATEDRSIVAITAAMPGGTGLDIFGRRFPKRTFDVGI AEQHAVTMAAGMA  
CEGLKPFCCIYSTFLQRGYDSVVHDIAVQNLpVRIIMDRAGLVGNDGPTHHGcyDL  
AYLGCIPQLKIMSPSDEIELKNMIATCAKY DEGPTCLRYPRGSGYGPDKLEKLF  
GYDMAVPKHGEVLEM GKGRIVRKG GGNKLLLG AKQTDEGEKTTRGKDKKDRV  
CILSFGTRLHESLVA ANEIEEADPDLAVTVADARFMKPLDVLVRELA AENS  
VIVTVEEGSIGGFGDHVHLHFLTLDGLMDNGDLKFRPMVLPDQYFEAASQYEQY  
ETAGLNASHIKGTVLRRLA QKIKVPVPQES

>*Nannochloropsis\_gaditana/DXS*

MGVFSASSSSSLSFASFPSSSSTARSLPTAVLLLTTLTALTSSVTA FHAPSTSITSSAHMTF  
PASSRTSSGGARSPLPIRARPRRTVFMETPLETGAGSSSTSNVPPPGTGPYQGSSFPLLD  
SVRYPHDMKRFDLKDLDKQLAHEL RWDTLHHVSKTGGHLGSSLGVIELTVALHYVFNT  
PDDRIVWDVSHQVYPHKILTGRDRMHSLRQTNGLSGFAKRSESEYDAFGAGHSSTSSIS  
AALGMSV GKLLTNKR VNNCVA VIGDGAITGGMAFEALNNAGYLRSRMVVILNDNGQ  
VSLPTGTHSAGGVVPAGALSSYTSRLLSSKPFQDFRSFAKGLNRLMPSEIQDINAKLDE  
YARGLIQGGTLFEELGFYYIGPVDGHDLDNLVPILENL RDSPSTKPVLLHVKTEKGYGY  
PPAEVASDKYHGVAKFDVSTGRQFKGGNKGAPLSLTTTFANALCEIAAEDRTVVGITA  
AMPGGTGMDIFGKRFPKRTFDV GIAEQHAVTFAAGMAIEGLKPFCCYSTFMQRGYDQ  
VIHDVVIQKLPVRMILDRAGLVGNDGPTHHG SFDLSYLGALP DIVIMAPSDELELMNLL  
ETAYETNDLPSV VRYPRGAGYGLETLKSELGYEGLEELPKRGKAVARGKGRIIRSAV  
VPSTAAPGTCRVALLSVGTRLLD TVHAAKEIEENLPGVSVTVADARFVKPLDKEMVSS  
LAMEHDVLITVEENS VGGFGSFVQQFLLNEGLLDGGKLR LRSMLVLPDRFIEAGPQSDQ  
YDQAGLAARHIVEKVEGLVRGREERMLQQEQQQPQPHQQQQQQTGAGVGTVAEMV  
GVGAGKSSSSSSSSSRQGT SFQVLP

>*Nitzschia punctata*/DXS

XGYLQSRMIVILNDNGQVSLPTGTPSAGGTVPASQLSTYTSNLLVSKPFQDFRDFAKSF  
NKLLPGGIQDVNKRIDEYARGLVSGGTLFEELGFYYVGPVDGHDLDXLIPILEKLRDSP  
NNKPVLLHLKTVKGYGYPPAEAASDKMHGVGKFNLTGTAQYKSKSTAPSFTSIFANSL  
IDAAIDDRSIVGVTAAMPGGTGMDIFGRRFPKRTFDV GIAEQHAVCMGAGMACEGLKP  
FVCIYSTFMQRGYDQVVHDVAIQNLPVRMILDRGGFV GNDGPTHHG CYDLSYMGCI  
NLTIMSPSDEIELKNMVMTC AAYDDGPSVLRYPRTGYGAEKLSLFGYKLEND EIPA  
KGEVIPIGKGRIIRGAGGMVDTSSNNR GKERKNRVA ILSL GTRLHEALVAAQEVEDMH  
EDISVTVADARFMKPLD VDLVRQLADDNGVLITIEEGSIGGFGDHV LHFLSLDGLLDDG  
DLKFRPMVIPDSYFEAATQFEQYDQAGLNARHIKGTILRLTKRMQVPVLEPVEQEA

>*Nitzschia sp*/DXS

MARPGQFSVLLWGLTYLSLQVANVAVLGFTAPSQHRYRQPTTTTSLDAEETTKNG  
HTISGPPYSGPASKPVLDSVQYPHDMNRLSMSSELKQLSHEL RWEVLESVSKTGGHLSSS  
LGVTELTVALHYVFDMPEDDIIWDVAHQCYPHKILTGRRSKMSGLRQSGGISGFCKRK  
ESEYDSFGAGHSSTSSISAAQGMSIAKSTLNKRKNNCVA VIGDGAITGGMAYEAMNSAG  
YLQSRMIVVLNDNGQVSLPTGTPSAGGTVPASQLSTYTSNLLVSKPFQDFRDFAKGFN  
KLLPGGIQDVNKRIDEYARGLVSGGTLFEELGFYYVGPVDGHDLDNLVPILEKLRDSPN  
NKPVLLHLKTVKGYGYPPAEAASDKMHGVGKFNLTGTAQFKSKATAPSFTSIFANALI  
DAAVDDRSIVGITAAMPGGTGMDIFGRRFPKRTFDV GIAEQHAVCMGAGMACEGLKP  
FVCIYSTFMQRGYDQVVHDVAIQNLPVRMMLDRGGFV GNDGPTHHG CYDLAYMGI  
PNLTIMSPSDEIELKNMVMTC AAFDDGPTVLRYPRTGYGAEKLSKFGYNLEND EIP  
AKGEIPIGKGRIIRGDGGMIDTASSTRGKQRKDRVA ILSIGTRLSEALDAAKEVEEEYS  
ISVTVADARYMKPLD VDLVRELADDHGV LITIEEGSIGGFGDHV LHFLSLDGLLDDGNL  
KFRPMVMPDAYFETGTQFEQYEEAGLNARHIKGTILRLCQKMKVPVLEQVEQEA

>*Phaeodactylum tricornutum*/DXS

MRLSSALFLLITPTVA AFAPRASLLVRKTNTFVKAEGTNGGGSGPPFSGPAVKPILDSVK  
YPSDMKNLDMRQLKQLANELRWEVLESVSKTGGHLSSSLGVNELTVALHYVFDMP  
DDIIWDVSHQCYPHKILTGRDRMSTLRQSGGISGFCKRKESEYDSFGAGHSSTSSISVAQ  
GMSIGKSILNKRQNNCIAVIGDGAITGGMAYEAMNSAGYLKNRMIVILNDNGQVSLPT  
GTPSAGGIVPASQLSTYTSNLLVSKPFQDFRMFAKNFNKLLPEGVQDVNKRIDEYARGI  
VAGGTLFEELGFFYVGPVDGHDLDNLIPILEKLRDSPSTKPVLLHVKTTKGYGYPPAEQ  
ASDRMHGVAKFNLTGKQIKSKKPIAPSLTSIFANALIDAATEDRSIVGITAAMPGGTG  
MDIFGRRFPKRTFDV GIAEQHAVTMGAGMACEGLKPFVCIYSTFMQRGYDQVVHDVA  
IQNLPVRMVLDRAGLVGNDGPTHHG CYDLSYMGCI PNLIIMAPSDEVELRNMVATQA  
AFNDGPTVLRYPRTGYGLETLKNTFGYTFEGDEIPTKGEVLPIGKGRIVRRPGGFNQN  
QSSTRGKSRKDRVAVLSL GTRLHEALVAAKEVEDIDPTLGVTVADGRFMKPLD VDLV

RELADEHSVLITIEEGSIGGFGDHVHLHFLSLDGLDDGELKFRPMVIPDTLFEAGTQYEQ  
YEQAGLNSEHIRGTILRLTKRLQVPQLQD

>*Pseudonitzschia\_arenysensis/DXS*

MRLSISLLLA AVIAPTFGFAPTSRHHASTKLYAEDSPKVSGPPYSGPAVKPILDSVNYP  
QDMKGLDMRQLKQLANELRWEVLESVSKTGGHLSSSLGVTELTV ALHYVFDMPPTDDI  
IWDVAHQCYPHKILTGRDRFPTLRQLGGLSGFCKRKESEYDSFGAGHSSTSISAAQGM  
SIAKSVLNKRKNNCVA VIGDGAITGGMA YEAMNSAGYLQSRMIVVLNDNGQVSLPTG  
TPSAGGTVPASQLSTYTSNLLVSKPFQDFRDFAKSFNKLLPEDVQSVNKRIDEYTRGLV  
SGGTLFEELGFYVGPIDGHDLNLIPILEKLRDPSNKPVLLHIKTTKGYGYPPAETAS  
DRMHGVGKFNLTGAQYKKGKPGEPSFTSIFANALIDAAIEDRSIVGITAAMPGGTGMDI  
FGRFPKRTFDVGI AEQHAVCMGAGMACEGLKPFVCIYSTFMQRGYDQVVHDVA VQ  
NLPVRMMLDRGGFVGNDGPTHHG CYDLSYMG CIPNLTIMSPSDEMELRN MVMTCAA  
FDDGPTVCRYPRGVGYGADKLQ NIFGYKLDKNEIPQKGEVIPIGKGRIRGAGGLIDTHS  
SSRGKQRKDRVAILSIGTRLHEALVAAQEVEDEY PDLQVTVADARFMKPLD VDLVRQ  
LADDHGV LITIEEGSIGGFGDHVHLHFLALDGLLDEGNL KFRPMVMPDAYFEAGTQFQQ  
YEQAGLNSRHIKGTILRLAKKIEIPILEDMEVDAGESTIA

>*Pseudonitzschia\_australis/DXS*

MRLSITALLLAAYTTPNAGVFGFAPTSRHHASVLNSRTTTVLNAEGDSDKVSGPPYSGP  
TVKPILDSVNYPHEMKGLDMRQLKQLANELRWEVLESVSKTGGHLSSSLGVTELTV A  
LHYVFDMPDDDIWDVAHQCYPHKILTGRDRFPTLRQLNGISGFCKRKESEYDSFGA  
GHSSTSISAAQGM SIAKSVLNKRKNNCVA VIGDGAITGGMA YEAMNSAGYLQSRMIV  
VLNDNGQVSLPTGTPSAGGTVPASQLSTYTSNLLVSKPFQDFRDFAKSFNKLLPEDVQS  
VNKRIDEYTRGLVSGGTLFEELGFYVGPIDGHDLNLIPILEKLRDPSNKPVLLHIKT  
VKGYGYPPAESASDRMHGVGKFNLTGAQFKGKPGEPSFTSIFANALIDAAIEDRSV V  
GITAAMPGGTGMDIFGRFPKRTFDVGI AEQHAVTMGAGMACEGLKPFVCIYSTFMQR  
GYDQVIHDVA VQNL PVRMMLDRGGFVGNDGPTHHG CYDFS YMG CIPNLTIMAPSDEI  
ELRN MVMTCADFDEGPTVLRYP RGVGYGADKLKNLFGYKLDKGEIPAKGEIIPV GKG  
RIIRGQGGMIENKSSTRGKERKDRVAILSIGTRLHEALVAAQEVEDEHE DLSITVADARF  
MKPLD VDLVRQLADDHGV LITIEEGSIGGFGDHVHLHFLSLDGLLDEGNL KFRPMVIPDN  
YFEAATQFQQYEQAGLNSRHIKGTILRLAKKIEIPLDEIAVDNSPGTIAGVYGSRLTGG  
GFGGCTVTLVAQDNAEDLMAHMESEYKSKTGIDCPCFVTRPARGAHLLSIVD HKPLV  
QE

>*Pseudonitzschia\_delicatissima/DXS*

MRLSISALLLA AAAVPTLGFAPASRSHATTKLFAAEESKVSGPPYSGPAVKPILDSVNYP  
QDMNGLDMRQLKQLANELRWEVLE NVSKTGGHLSSSLGVTELTV ALHYVFDMPPTDD  
IWDVAHQCYPHKILTGRREKFPTLRQLGGISGFCKRKESEYDSFGAGHSSTSISAAQGM  
SIAKSVLNKRKNNCVA VIGDGAITGGMA YEAMNSAGYLQSRMIVVLNDNGQVSLPTG  
TPSAGGTVPASQLSTYTSNLLVSKPFQDFRDFAKSFNKLLPEDVQSVNKRIDEYTRGLV  
SGGTLFEELGFYVGPIDGHDLNLIPILEKLRDPSNKPVLLHIKTVKGYGYPPAETAS  
DRMHGVGKFNLTGAQYKTKGGEPSFTSIFANALIDAAIDDRSIVGITAAMPGGTGMDI  
FGRFPKRTFDVGI AEQHAVCMGAGMACEGLKPFVCIYSTFMQRGYDQVVHDVA VQ  
NLPVRMMLDRGGFVGNDGPTHHG CYDLSYMG CIPNLTIMSPSDEMELRN MVMTCAA  
FDDGPTVCRYPRGVGYGADKLQNTFGYKLENNEIPSKGEVIPIGKGRIRGAGGLIDTHS  
SSRGKQRKDRVAILSIGTRLHEALVAAQEVEDEYDDLQVTVADARFMKPLD VDLVRQ  
LADDHGV LITIEEGSIGGFGDHVHLHFLALDGLLDEGNL KFRPMVMPDAYFEAGTQFQQ  
YEQAGLNSRHIKGTILRLAKKVEIPILEDMEVDAGESTIA

>*Pseudonitzschia\_multistriata/DXS*

MRLSIAALLLAAYTSTNTGVHGFAPTSRYHARAPNTVLNAEGDSDKKVSGPPYSGPTV  
KPILDSIKYPHDMNGLDMRQLKQLANEVRWEVLEQVSKTGGHLSSSLGVTELTV ALH  
YVFDMPPTDDIWDVAHQCYPHKMLTGRDRFPTLRQMNGLSGFCKRKESEYDSFGAG  
HSSTSISAAQGM SIAKSILNKRKNNCVA VIGDGAITGGMA YEAMNSAGYLQSRMIVVL

NDNGQVSLPTGTGPSAGGTVPASQLSTYTSNLLVSKPFQDFRDFAKSFNKLLPEDVQSVN  
KRIDEYTRGLVSGGTLFEELGFYYVGPIDGHDLNLIPILEKLRDSPSNKPVLLHIKTTK  
GYGYPPAEAASDRMHGVGKFNLTGAQFKIKGGEPSFTSIFANALIDAAIEDRSIVGITA  
AMPGGTGMDIFGRRFPKRTFDVGIAEQHAVTMGAGMACEGLKPFVCIYSTFMQRGYD  
QVVHDVA VQNL PVRMMLDRGGFVGN DGPTHG CYDLSYMG CIPNLTIMAPSDEIELR  
NMVMTCADYDEGPTVLRYPYPRGVGYGAEKLNLFYKLEKGEIPTKGEKIPIGKGRIR  
GEGGLIDNKSSTRGKERKDRVA VLSIGTRLHEALVAAQEVEDEHEDLSVTVADARYM  
KPLDVDLVRQLADDHSVLITIEEGSIGGFGDHVLFHFLSLDGLLDEGNLKRFRPMVIPDAY  
FEAATQFQQYEQAGLNARHIKGTILRLAKKIEIPILDEIPVDNSESTIA

>*Pseudonitzschia\_pungens/DXS*

AIXNSSEMRLSITALLLAAYTAPNAVFGFAPSSRRHVSVQSSRANTVLNAESES DKISGP  
PYSGPAVKPILDSVNYPHEMKNLDMRQLKQLANELRWEVLEQVSKTGGHLSLGLVT  
ELTV ALHYVFDMPDDDIIWDVAHQCYPHKILTGRDRFPPTLRQLGGLSGFCRKESEY  
DSFGAGHSSTSISAAQMSIAKSILNKRKNNCVAVIGDGAITGGMAYEAMNSAGYLQS  
RMIVVLNDNGQVSLPTGTGPSAGGTVPASQLSTYTSNLLVSKPFQDFRDFAKSFNKLLPE  
DVQSVNKRIDEYTRGLVSGGTLFEELGFYYVGPIDGHDLNLIPILEKLRDSPNNKPVLL  
HIKTVKGYGYPPAESASDRMHGVGKFNLTGAQFKAKGGEPSFTSIFANALIDAAIEDR  
SVVGITAAMPGGTGMDIFGRRFPKRTFDVGIAEQHAVTMGAGMACEGLKPFVCIYSTF  
MQRGYDQVIHDVA VQNL PVRMMLDRGGFVGN DGPTHG CYDFA YMG CIPNLTIMAP  
SDEIELRNMVMTCADFN EGPTVLRYPYPRGVGYGADKLNLFYKLEKGEIPSKGEIPIG  
KGRIIRGEGMIENTSSSTRGKERKDRVAILSIGTRLHEALVAAQEVEDEYDDL SVTVAD  
ARYMKPLDVDLVRQLADDHSVLITIEEGSIGGFGDHVLFHFLSLDGLLDDGNLKRFRPMV  
MPDAYFEAATQFQQYEQAGLNSRHIKGTILRLAKKIEIPVLDEIPVDNSESTIA

>*Pseudonitzschia\_pungens\_fr\_cingulata/DXS*

MRLSITALLLAAYTAPNAVFGFAPSSRRHVSVQNSRANTVLNAESES DKISGPPYSGPA  
VKPILDSVNYPNEMKNLDMRQLKQLANELRWEVLEQVSKTGGHLSLGLVTELTV AL  
HYVFDMPDDDIIWDVAHQCYPHKILTGRDRFPPTLRQLGGLSGFCRKESEYDSFGAG  
HSSTSISAAQMSIAKSILNKRKNNCVAVIGDGAITGGMAYEAMNSAGYLQSRMIVVL  
NDNGQVSLPTGTGPSAGGTVPASQLSTYTSNLLVSKPFQDFRDFAKSFNKLLPEDVQSVN  
KRIDEYTRGLVSGGTLFEELGFYYVGPIDGHDLNLIPILEKLRDSPNNKPVLLHIKTVK  
GYGYPPAESASDRMHGVGKFNLTGAQFKAKGGEPSFTSIFANALIDAAIEDRSVVGIT  
AAMPGGTGMDIFGRRFPKRTFDVGIAEQHAVTMGAGMACEGLKPFVCIYSTFMQRGY  
DQVIHDVA VQNL PVRMMLDRGGFVGN DGPTHG CYDFA YMG CIPNLTIMAPSDEIEL  
RNMVMTCADFN EGPTVLRYPYPRGVGYGADKLNLFYKLEKGEIPAKGEIPIGKGRIR  
GEGMIENTSSSTRGKERKDRVAILSIGTRLHEALNAAQEVEDEYDDL SVTVADARYMK  
PLDVDLVRQLADDHSVLITIEEGSIGGFGDHVLFHFLSLDGLLDDGNLKRFRPMVMPDAY  
FEAATQFQQYEQAGLNSRHIKGTILRLAKKIEIPVLDEIPVDNSESTIA

>*Skeletonema\_costatum/DXS*

MKFISAVVAAAIASAPVCNAFSPSSGRSPLRTIQQSTQSQKTSTSLDQTDSGPPYSGP  
ASKPILDSVKFSPDMNRLNMSELKQLAHEL RWEVIEAVSKTGGHFSSSLGVIELTTALH  
YVFDMPEDDIIWDVAHQCYPHKMLTGRRDRFPPTLRQLNGMSGFCRKESEYDSFGAG  
HSSTSISAAQMSVAKSTLNKRKNNCIAVIGDGAITGGMAYEAMNAGYLSSRMIVVL  
NDNGQVSLPTGTQSAGGVK PASQLSAYTSNLLVSKPFQDFRDFAKSFNRLLPENVQDV  
NKRIDEYARGVVTGGTLFEELGFFYVGPIDGHDLNMPILEKLRDSDSNKPVLLHIKT  
NKGQGYAPALAAASDRMHGVGKFDX

>*Skeletonema\_marinoi/DXS*

MKFISAVVAAAIASAPVCNAFSPSSGRSPLRTIQQSTQSQKTSTSLDQTDSGPPYSGP  
ASKPILDSVKFSPDMNRLNMSELKQLAHEL RWEVIEAVSKTGGHFSSSLGVIELTTALH  
YVFDMPEDDIIWDVAHQCYPHKMLTGRRDRFPPTLRQLNGMSGFCRKESEYDSFGAG  
HSSTSISAAQMSVAKSTLNKRKNNCIAVIGDGAITGGMAYEAMNAGYLSSRMIVVL  
NDNGQVSLPTGTQSAGGVK PASQLSAYTSNLLVSKPFQDFRDFAKSFNRLLPENVQDV

NKRIDEYARGVVTGGTLFEELGFFYVGPIDGHDLENMVPILEKLRDSDSNKPVLLHIKT  
NKGQGYAPALAASDRMHGVGKFDIATGIQQKGKAAAPSFTSIFANSLIDAATEDRAVV  
GITAAMPGGTGMDIFGRRFPKRTFDVGIAEQHAVTFAAGMACEGLKPFCCYSTFMQR  
GYDQVVHDVAIQNLPVRMILDRAGVVGNDGPTHHGICYDLAYLGCIPNLTIMSPSDEIE  
LRNMVATCAGFDDGPTVRLRYPRGVGYGADKLQNLFGYDLEDGEIPTKGVPLEIGKGR  
IRPPGGVHALGEDIAANGSSDSKLGSGGRGAKREDRVAILSLGTRLSESLIAANEVEEM  
DPDIAVTVADARFMKPLDVDLIRELAKDNGIITVEEGSIGGFGDHVLFHFLSLDGALDN  
GDLKFRPMVLPDAYFEAATQYEQYEMAGLNSHIRGTVLRLAEKVEVPELAELEK

>*Skeletonema\_menzelii/DXS*

MKYISAVVAASAPVCNAPVSPSSGRSPLRTIQQSTESQITSTRLFDATADRPASKPILD  
SVKFPSPDMKRLNMSSELKQLAHELWEVIEAVSKTGGHFSSSLGVIELTTALHYVFDMP  
EDDIIWDVAHQCYPHKMLTGRDRKFPTRLQNLNGISGFCKRKESVYDSFGAGHSSTSISA  
AQMMSVAKSTLNKRRNNCIAVIGDGAITGGMA YEAMNAGYLSSRMIVVLNDNGQV  
SLPTGTQSAGGVKPAQLSAYTSNLLVSKPFQDFRDFAKSFNRLPENVDVKNRIDEY  
ARGVVTGGTLFEELGFFYVGPIDGHDLENMVPILEKLRDSDSNKPVLLHIKTNKGQGY  
PALAASDRMHGVGKFDIATGVQQKGKAAAPSFTSIFANSLIDAATEDRAVVGITAAMP  
GGTGMDIFGRRFPKRTFDVGIAEQHAVTFAAGMACEGLKPFCCYSTFMQRGYDQVV  
HDVAIQNLPVRMILDRAGVVGNDGPTHHGICYDLAYLGCIPNLTIMSPSDEIELRNMVA  
TCAGFDDGPTVRLRYPRGVGYGAEKQLDLFGYDLEDGEIPTKGVLEIGKGRIRPPGGV  
RALGKEITANGSSDSKLGSGGRGAKREDRVAILSLGTRLSESLIAADEVEEMDPDIAVT  
VADARFMKPLDVDLIRELAKDNGIITVEEGSIGGFGDHVLFHFLSLDGALDNGDLKFRP  
MVLPTDYFEAATQYEQYEMAGLNAHHRGTVMLRAQKVEVPELAELEK

>*Thalassionema\_frauenfeldii/DXS*

MSGLRQSGGISGFCKRKESKYDSFGAGHSSTSISAAQMMSVAKSMLNKRNNNCIAVIG  
DGAITGGMA YEAMNSAGYLRSMIVVLNDNGQVSLPTGTSPASQLSTYTSN  
LLVSKPFQDFRSFAKSFNKLPEEIQGLNKRIDEYARGIVSGGTLFEELGFFYVGPVDGH  
DLENMVPILEKLRDSDPKPVLLHLKTIKGYGYPPAEEASDKMHGVAKFNLGTGAQIK  
PKAGNPSMTSIFANSLIDCATEDRAIVGITAAMPGGTGMDIFGRRFPKRTYDVGIAEQH  
AVTMAAGMACEGLKPFCCYSTFMQRGYDQVVHDVAIQNLPVRMILDRAGLVGNDG  
PTHHGICYDLAYMGCIPNLTIMAPSDEIELKNMVSTCAAFDEGPTVRLRYPRGTGYGADK  
LQSLFGYELEGGEIPAKGEVLPKGRIVRPAGGFNKNQINTRGKERKDRVAILSLGTRL  
HESLKAANEIEEQDDKLGVTVADARFMKPLDVDLIREL VDDHSLVITVEEGSIGGFGDH  
VLHFITLDGLMDDGDLKFRPMVLPDVLFEAATQTEQYEQAGLNAQHIGTILRLTKRV  
NVPVLEEA

>*Thalassiosira\_minuscola/DXS*

MKELEQLAHELWEVIDAVSQTGGHFSSSLGVIELTVLHYVFDMPEDDIIWDVAHQ  
YPHKMLTGRDRKFPTRLQNLGGISGFCKRKESEYDSFGAGHSSTSISAAQMMSVGSML  
NKRNNNCIAVIGDGAITGGMA YEAMNAGYLSSRMIVVLNDNGQVSLPTGTQTAGGV  
KPAQLSAYTSNLLVSKPFQDFRDFAKSLNRLFPENVDVKNRLDEYARGLVTGGTVF  
EELGFFYVGPIDGHDLENMVPILEKLRDSDSSKPVLLHIKTTKGYGYPPAMAASDRMH  
GVAKFDIGTVQVKGKAAAPSFTSILANSMIDAATEHREIVGITAAMPGGTGMDIFGRR  
FPKRTFDVGIAEQHAVTFAAGMACEGLKPFCCYSTFMQRGYDQVVHDVAIQNLPVR  
MILDRAGVVGNDGPTHHGICYDLAYLGCIPNLTIMSPSDEIELRNMVMTCAEFDDGPTV  
RLRYPRGVGYGAEKQLDLFGYDLENGEIPAKGVPMIEIGKGRILRLPGGVRAMSKQPPAN  
GSSDSKLSGAGSRGQKKEDRVAILSFGTRLSEALKADEVEESDPDLAVTVADARFMK  
PLDVDLIRDALAKDNGIITVEEGSIGGFGDHVLFHFLSLDGALDDGELKFRPMVLPD  
TYFEAATQYEQYEMAGLNAHHRGTVMLRAKKEVPEVLESE

>*Thalassiosira\_pseudonana/DXS*

MKFLSSTIAAAILATSSAFTTTLPRTSSIQPTSTRLHDTTNGPPYSGPASKPILDTVNYP  
HDMKRLDMKELKQLAHELWELIEAVSKTGGHFSSSLGVIELTVLHYVFDMPEDDII  
WDVAHQCYPHKMLTGRDRFPTRLQNLGGISGFCKRKESEYDSFGAGHSSTSISAAQGM



SVAKSLLNKRKNNCIAVIGDGAITGGMAYEAMNNAGYLSSRMIVVLNDNGQVSLPTG  
TQTAGGVRPASQLSAYTSNLIVSKPFQDFRDFAKSLNRLFPENVQDVNKRLEDEYARGL  
VTGGTVFEELGFFYVGPIDGHDLNMPVILEKLRDSDSTKPVLLHIKTNKGQGYPPALA  
ASDKMHGVGKFDIATGIQVKGKAAAPSLTSIFANAMIDAATEDRAIVGITAAMPGGTG  
MDIFGRRFPKRTYDVGIAEQHAVTFAAGMACEGLKPFCCIYSTFMQRGYDQVVHDVAI  
QNLPMILDRAGVVGNDGPTHHGCDLALYLCIPNLTIMSPSDEIELKNMVMTCAAF  
DDGPTVLRYPYPRGVGYGAEKLNLFYELNGEVPKGEVIKIGKGRIRRPGGIHALKE  
ELTANGSSDSRMGAGSRGLLKENRIAILTLGTRLSEALVAADEIEGANPDLA VTVADAR  
FMKPLDEDLIRELSDNNGIIVTVEEGSIGGFGDHVLFHFLSLDGDALDDGELKFRPMLVLPD  
TYFEAATQYEQYEMAGLNSHIRDTVLRRLANVKVEAAPVLAEE

>*Thalassiosira\_puntigera/DXS*

MKCIEILSIAAAATTASAFTLPKTPVSKTSTTRLNEAINRVQPQGPYPYSGPSSKPILDSVQ  
FPRDMNRLTMQELEQLAHELWEVIEAVSKTGGHFSSSLGVIELTVALHYVFDMPEDD  
IIWDVAHQCYPHKMLTGRDRFPTLRQLGGISGFCKRKESEYDSFGAGHSSTISAAQG  
MSIGKSLNKRKNNCIAVIGDGAITGGMAYEAMNNAGYLSSRMIVVLNDNGQVSLPT  
GTQSAGGVRPASQLSAYTSNLLVSKPFQDFRDFAKGMNRLFPEGVQDVNKRIDEYAR  
GLVTGGTIFEELGFFYVGPIDGHDLNMPVILEKLRDSDSNKPVLLHIKTNKGYGYPPA  
LAASDRMHGVAKFNVGTGVQVKGKAAAPSLTSIMANAMIDAATEHREIVGITAAMPG  
GTGMDIFGRRFPKRTFDVGIAEQHAVTFAAGMACEGLKPFCCIYSTFMQRGYDQVVH  
DVAIQNLPMILDRAGVVGNDGPTHHGCDLALYLCIPNLTIMSPSDEMELRNMVM  
TCAEFDEGPSVLRYPYPRGVGYGPEKLQELFGYELENGEIPAKGTPIEIGKGRIRLPGGVR  
ALSSSEPAANGSTDSKLGASKSRGKKKTDRVAILSFGTRLSEALIAADEVEESDPDLAV  
TVADARFMKPLDIDLVRDLAKDNGIVITVEEGSIGGFGDHVLFHFLSLDGDALDDGELKFR  
PMLVLPD TYFEAATQYEQYEMAGLNAHHRGTVLRRLAKKVEVPVLS

>*Thalassiosira\_rotula/DXS*

MKCILAAVLIAALATSATAFTIPTTPITKSSTSSTTQLNDAINRREPAGPPYAGPASKPILD  
SVKFPRDMNRLDMKELEQLAHELWEVIEAVSKTGGHFSSSLGVIELTVALHYVFDMP  
EDDIIWDVAHQCYPHKMLTGRDRFPTLRQLNGMSGFCKRKESEYDSFGAGHSSTIS  
AAQMSIGKSLNKRKNNCIAVIGDGAITGGMAYEAMNNAGYLSSRMIVVLNDNGQV  
SLPTGTQTAGGVRPASQLSAYTSNLLVSKPFQDFRDFAKGLNRLFPENVQDVNKRLE  
YARGLVTGGTIFEELGFFYVGPIDGHDLNMPVILEKLRDSDSSKPVLLHIKTNKGYGY  
PPALAASDRMHGVAKFDIGTGVQVKGKVAAPSFTSIMANAMIDAATEHREIVGITAAM  
PGGTGMDIFGRRFPKRTFDVGIAEQHAVTFAAGMACEGLKPFCCIYSTFMQRGYDQVV  
HDVAIQNLPMILDRAGVVGNDGPTHHGCDLALYLCIPNLTIMSPSDEMELRNMIM  
TCADFDDGPTVLRYPYPRGVGYGPEKLQELFGYELENGEIPSKGTPIEIGKGRIRLPGGVR  
ALSEPAANGSTDSKLGAGSRGKKKEDRVAILTFGTRLSEALIAADEVEEMDPDLAVT  
VADARFMKPLDIDLIRELAEDNNGIITVEEGSIGGFGDHVLFHFLSLDGDALDKGDLKFRP  
MVLPSDYFEAATQYEQYEMAGLNSHIRGTVLRRLAKKIEVPVLS

>*Thalassiosira\_weissflogii/DXS*

MKFIQSSIAALLLAAAAPTSAFTTFAPRTTPASTQLQSTRDAAASSTSGPPYSGPASKPIL  
DSVNYPRDMKRLDMRELKQLAHELWEVIEAVSKTGGHFSSSLGVIELTVALHYVFD  
MPEDDIIWDVAHQCYPHKMLTGRDRFPTLRQLGGISGFCKRKESEYDSFGAGHSSTSI  
SAAQMSVAKSMLNKRNNCIAVIGDGAITGGMAYEAMNNAGYLSSRMIVVLNDNG  
QVSLPTGTQTAGGVRPASQLSAYTSNLIVSKPFQDFRDFAKNLNKLFPENVQDVNKRID  
EYARGVVTGGTLFEELGFFYVGPIDGHDLNLPVILEKLRDSDSNKPVLLHIKTEKGGH  
YPPALAASDRMHGVGKFDIATGIQVKGKAAAPSLTSIFANSLIDAATEDRAIVGITAAM  
PGGTGMDIFGRRFPKRTFDVGIAEQHAVTFAAGMACEGLKPFCCIYSTFMQRGYDQVV  
HDVAIQNLPMILDRAGVVGNDGPTHHGCDLALYMGIPNLTIMSPSDEIELKNMV  
MTCAGIDDGPSVLRYPYPRGVGYGAEKLNLFYELNGEIPKGTPIEIGKGRIRRPGG  
VRALSEDTEPANGSKDGKMGTSRGLLKEHRVAILTLGTRLSEALVAADEVEELDPDL

AVTVADARFMKPLDVDMIRELAENNGILITVEEGSIGGFGDHVLHFLSLDGALDEGD  
L KFRPMVLPDITYFEAATQYEQYEMAGLNAHHIRGTMRLTKKVDVPVLVD

### CMK protein sequences

#### >*Amphipora\_sp*/CMK

MRGLLTLLFVTLGAGASAFTVGSRDGTARYSATRILRQSTGSSATNTDVSTASSETDST  
LSLFSPCKINLFLRIIRKREDGFHDLASLFQAIGFGDSLELTPIDGTADEFTCNMPGVPVD  
SSNLVLRALELMRTKTGVEQYFKANLIKQVPAQAGLGGGSANAATAMWGANQLMG  
DPASLEEMVEWSGDLGSDITFFLSRGTA YCTGRGEIMDSIDPPLPSGTRLCIVKPNVGLS  
TPAVFKALEYDKLSTLDADEELLPTFLKGIDDPDGYFINDLEFPAPFKCLPELDEL  
LQEEGFDHVMMSGSGTSIFCLGEPSPDKKAFEEKFAKR

#### >*Bolidomonas\_pacifica*/CMK

LYYFVICHVSNTEGFHQRLPPQTPHTPRRSTSLNCAGSKDPSLASLTLFSPSKVNLFLRII  
RKRDDGFHDLASLFCAIGLGDTLNLESADTDVFTCNMPGVPVDSSNLVVRALNLVREK  
TGNEGKKFKINLHKECPAQAGLGGGSGNAATAMFAVNELLGRPATMDQLVEWSGEL  
GSDITFFLSKGTA YCTGRGEVX

#### >*Chaetoceros\_neogracile*/CMK

TLAFAPNKGYNLQFQSHVSSRNAQMASRLFSTVDESEYDLSLFSPCKINLFLRIIRKP  
DGFHDLASLFQTIGFGDMLHLKYEESSTADIFECNMDGVPTDKSNLVIRALDLVRSKTG  
NEDKFFKANLVKQVPAQAGLGGGSGNAAAAMWGANELLGRPATLEQLVEWSGALG  
SDITFFLSQGTAYCTGRGEIMTPVDLLPDGTVKVCIVKPDIGLSTPEVFRALDYDQLSTIDP  
EELLKTFMEKGAVDAGDDAYVNDLEQPAFDCLPELGKLKEELKSVGFDHVMMSG  
GTSIFCIGEPX

#### >*Chaetoceros\_sp*/CMK

MRGDSVLLGLQTGVKWDWDYRAKVS DHDA YVNDLEQPAFDALPELKRLKDELK  
SIEGFDHVMMSGSGTSIFCIGEPKDKDAFMKEFGQRKGINVFPAEFTSRKQGTWFQKPF  
N

#### >*Coscinodiscus\_wailesii*/CMK

MEDVPTDKSNLVKALDLMRDKTGCQKFFKAKLVKQVPAQAGLGGGSANAATAMF  
GANELMGRPATLDQLIEWSGDLGSDITFFLSQGTAYCTGRGEILDPIPTSPFSSETKICIV  
KPKNIGLSTPSVFRALDYDKLSSRDPQELDTFLEKGVLGVEKNDYVNDLEDPAFRVV  
PELLALKNEVEECGFDFVMMSGSGTSLFGLGGGDDAIEKTRERFTGRDDVSIFETNFLS  
RGEWFEGEIME

#### >*Cylindrotheca\_closterium*/CMK

XGSSFGFAGPSATLQQYDTSTLLRASRAAEKKASRTKWVKS RGFDDDIAGGSDAMTL  
FSPXKINLFLRIIRKREDGFHDLASLFQAVGYGDTLELKKLDDATEDAFSCNMEGVPTD  
MTNLVLRALELMREKTGVKQYFDANLIKQVPAQAGLGGGSANAATAMWGANELMG  
NPASLED MIEWSGDLGSDITFFLSRGTA FCTGRGEIMTPDPLLEAGTKLCIVKPNVGLS  
TPSVFKALDYDELSEL DADK VLLPEFLNAESVEAVS GEYYINDLEPPAFRCVPELDELK  
DKLTAVDGFKHVMMSGSGTSIFCIGEPSPDPEQFQKDFGEDEELKVMFTEFINREEGTWF  
ERP

#### >*Fragilariopsis\_kerguelensis*/CMK

XSSHKEFSAEKRNKEFYTIPSDMLLRFEKIAKKRRMKLSTGVYFSFGFWCYMYXLCNW  
ISVSSVRGFSVVKSNHIFSGSSTESSRKQYTTIKSTRVGQAKISTGHQAWRGS LHF  
DASL KSSSTSIPELTQQENEQKTLVLFSPCKINLFLRIIGKREDGFHDLASLFQAIGFGDTLELAP  
LKEEKSSIEKRNDEFTCNMPGVPVDSTNLVLRALS LMRQKTGVKMYFKTNLIKQVPAQ  
AGLGGGSANAATAMWGANELMGRPASLEQMIEWSAELGSDITFFLSRGTA FCTGRGEI  
LTPIDPPLPTGTVKVCIVKPNIGLSTPSVFKALDYDELSDLPDDVLLPAFLDSVGVENVA  
REYYVNDLEPPAFRCVQKLGELKDSLLKVDGFGQHVMMMSGSGTSIFCIGEPEDTESFLKE  
FGESEDIQVFFSEFIDRESNEWFQNPKNAA

**>Leptocylindrus\_apora/CMK**

IASLLLLTSTTTAFSSWVPKRKSVSIHQNYRITTRQVTADAHSSVTTASSASDDASVEW  
DMSLFSPAKINLFLRIIRKREDGFHDLASLFQITIGFDYLHIRKLGSDAEDDVLTCNMDG  
VPTDRTNLVMRAVDLVREKTGIKDQFFQMDLVKYVPHQAGLGGGSGNAAAALYGVN  
ELGPNATLEELVEWSGELGSDITFFLSEGTAYCTGRGEIMTPVPLPNSAGTKLCIVKP  
DIGLSTPSVFKALDYDKLSKADPENVLQIFQTEGMSPSEDEKFINDLERPAFKVLPDLEE  
LKNDLISCGFETVMMMSGSGTSIFCIGSTS DPEKFSNLIESRDDVNVYSTEFINRDGNGWY  
TSPDDTSTKIQ

**>Leptocylindrus\_danicus/CMK**

XSILLTTTQAFQIPSTFGTARRTALTVKSADAIDETLTKDENSETWDLSLFSPAKINIFLRI  
IRKREDGFHDLASLFQITIGFDYLHLKLLDDDEEAKEDILTCNMEGVPTDKSNLVMRAI  
DLVRQNTGITDKFFKCDLVKHVPHQAGLGGGSGNAAAALYGANELLGRPATLEQLVE  
WSGDLGSDITFFLSEGTAYCTGRGEIMTPVDPLPAGQTVYIVKPDIGLSTPSVFKALDYD  
LLSTLDPEEILNVFMTEGVTTSNQSVFVNDLEPPAFKVVPELCELKDDLISCGFDTVMM  
SGSGTSIFCLGKPGDEEKFEELIGSRDDVQVYPTFINRVGCAADGGWYQDPNA

**>Nannochloropsis\_gaditana/CMK**

MTHSFSVGGGGAGGGGGAGAMNRAHSSAVAGRSLVKSKEKVGAVEARELTAGKEG  
GGVATARALAAPEVQDMTRRVSTHCGRPLFFLTRTPSINCLTCNVDLGMNLFRIIRRR  
EDGYHELASLFQTVAFGDTLLKVLPAIAVQDEFSSSYTSLVTDQSNLVLRALDLFREK  
TGKTQRFEVYLEKRTPVQAGLGGGSSNAATTLFAANQLLGT PATDEELIEWSKELGSDI  
TFFLSKGTAYCTGRGELLEPPVPLPSSIKLYVVKPPVGLSTPAVFKALDLDALNGEDPR  
DLLAAFTQHLSRGELGPASMYINDLEPPAFQCVPRKAIKEALLARGFDCVMMMSGSGT  
AIFAMAPAPSSLPSSAGLSFDAAAFAREMEVDIWA TTTFTNRPSDDPKAWYPNPHA

**>Nitzschia\_punctata/CMK**

XTTTESITEDEFTCNMPGVVPVDSTNLVLRALELMRSKTGVQKYFKANLIXQVPAQALG  
GGSANAATAMWAANTLMGNPATLEQMIEWSGELGSDITFFLSRGTAFACTGRGEIMTPI  
EPPLPTGTKLCIVKPDIGLSTPSVFKALDYDELSDLNADEVLLPAFLNANGVENVPSEYY  
INDLEPPAFRCVPEL GELKKS LQKVTGFKHVMMMSGSGTSIFCIGEPEDKEAFMKEFGET  
EDLQVFFSEFIGREEGEWFRPE

**>Nitzschia\_sp/CMK**

XTNLLFILKDDDDDEREDERSYGLLEVSVKDFVRILLQLCRTFGDGDGVGLFQQFPNN  
DDGFVPTITVSSSSSTIASHRHHRHRTGVFLDAVSSSTESSTTPSTAQSQDDGSLTLFSP  
CKINLFLRIVRKREDGFHDLASLFQAIGFGDTLELKLQDDSEEDGDEEEKTTTTTADS  
FTCNMPGVVPVDDTNLVLRALQLMRTKTGKTDQHFAANLIKECPAQAGLGGGSANAAT  
AMWGANELMGRPASLEQMIEWSGELGSDITFFLSQGTAYCTGRGEIMTPTNPPLKSGT  
KLCIVKPNVGLSTPSVFKALDYDELSTLDADKTLLPAFLETGVENVSDEYYINDLEPP  
AFRCVPEL GELKSSLQKIKGFDHVMMMSGSGTSIFCIGEPDDKDAFLKEFGNRQDELQVF  
FSEFISREEGQW

**>Phaeodactylum\_tricornutum/CMK**

MYFASAHVFLAALSREA VAFTDNRFPFHNHPSAKVTTTSSLSAYTETNESSRGKISERT  
LTLFSPCKINLFLRIIRKREDGFHDLASLFQAIGFGDTLELTSIDSDADEF TCNMPGVVPD  
NSNLVLRAIQLMREKTGEKQSFANLIKQVPAQAGLGGGSANAATAMWGVNELMGR  
PASLDQMVEWSGALGSDITFFLSRGTA YCTGRGEIMSSIDPPLPSGTKICIVKIDIGLSTPS  
VFKALDYDKLSNLNADDVLLPTFLRGIDQVPDSYFVNDLETPAFKCIPELRLSKEELLG  
VAGFDHVMMMSGSGTSIFCLGEPKDNKDFHDFVSRKGLQVFFSEFISRPEGVWFEKP

**>Pseudonitzschia\_australis/CMK**

MSNPFISFSFWCCYFGLFIATSSVEGFVVEPCRVL RQSSLSASTQENQETLELFSPCKIN  
LFLRIIRKREDGFHDLASLFQAIGFGDTLELTPLVETTEIGEGSDEF TCNMPGVVPVDSTNL  
VLRALDLMREKTGVDMFFKTNLIKQVPAQAGLGGGSANAATAMWGANELMGRPASI  
EQIEWSGALGSDITFFLSQGTAYCTGRGEILTPIDPPLPTGTKVCIVKPDIGLSTPSVFKA

LDYDELSMDPDDVLLPAFLDTVGVENVSKHYVNDLESPAFCVPRLGELKESLLKV  
AGFKHVMMSGSGTSIFCVGEPDDRESFVKEFGEMDDVQVFFSEFIDREKEIWFQKP

>*Pseudonitzschia\_arenysensis*/CMK

MSSLHKISSLCIFYCFLCFFVATALVQGFASTXQNSALARSXKQTSLFGSSSEEPKATLK  
LFSPCKINLFLRIIRKREDGFHDLASLFQAIGFGDTLELSPLDDDDADADEFNMPGVP  
VDSTNLVLRALMREKAGVDMYFKSNLIKQVPAQAGLGGGSANAATAMWGANEL  
MGRPASVEQMIEWSGALGSDITFFLSQGTAYCTGRGEILDPIDPPLPTGTKVCIVKPNIG  
LSTPSVFKALNYDELSDKDADKDLLPAFLDTEGVENVPKDCYVNDLEPPAFRCVPRLG  
ELKESLLKVKGFEHVMMSGSGTSIFCIGEPEDKDAFLEEFAMDDVKVFFSEFIDREKGE  
EW

>*Pseudonitzschia\_multistriata*/CMK

MSNLISSFILWCCYFGLTTTSSVQGFVVEPNHVFRQSSLSASSQENQETLELFSPCKIN  
LFLRIIRKREDGFHDLASLFQAIGFGDTLELTPGETADEGEGSDEFNMPGVPVDSTN  
LVLRALDLMREKTGVNMYFKTNLIKCPAQAGLGGGSANAATAMWGANELMGRPA  
SLEQMIEWSGALGSDITFFLSQGTAYCTGRGEILTPIDPPLPTGTKVCIVKPDVGLSTPSV  
FKALDYDELSNMDPDDVLLPAFLDAEGVENVSAEYVNDLEPPAFRCVPRLGELKESL  
LKVSGFKHVMMSGSGTSIFCIGEPDDHESFMKKFGEMDDVQVFFSEFIDREKGEWFQK  
P

>*Pseudonitzschia\_pungens*/CMK

XSESPRVVSMVTSNSIRELIFYVETSKHSENTDHLVQELKLIYRVRPXDLYPFELILREK  
QIASQRIGLQDKKMANASLSFLLWCCYFGIFAATSVVQGFVVDSTRFARQSSLSASTQ  
QDVATLELFSPCKINLFLRIIRKREDGFHDLASLFQAIGFGDSLELTPLEETSEDEEGSDE  
FTCNMPGVPVDSTNLVLRALDLMREKTGVKMYFKTNLVKQVPAQAGLGGGSANAAT  
AMWGANELMGRPASLEQMIEWSGALGSDITFFLSQGTAYCTGRGEILTPIDPPLPTGTK  
ICIVKPDIGLSTPSVFKALDYDELSKMDPDDVLLPAFLDTEGVENVSEDCYVNDLEPPA  
FRCVPRLGELKESLQKVAGFQHVMMSGSGTSIFCIGEPEDREAFLDEFGKMEDVQVFFS  
EFIDREKQWFQKP

>*Pseudonitzschia\_pungens\_fr\_cingulata*/CMK

MANASLSFLLWCCYFGIFAATSVVQGFVVDSTRFARQSSLSASTQQDVATLELFSPCK  
INLFLRIIRKREDGFHDLASLFQAIGFGDSLELTPLEETSEDEEGSDEFNMPGVPVDST  
NLVLRALDLMREKTGVKMYFKTNLVKQVPAQAGLGGGSANAATAMWGANELMGRP  
ASLEQLIEWSGALGSDITFFLSQGTAYCTGRGEILTPIDPPLPTGTKICIVKPDIGLSTPSV  
KALDYDELSKMDPDDVLLPAFLDTKGVENVSEDCYVNDLEPPAFRCVPRLGELKESLQ  
KVAGFQHVMMSGSGTSIFCIGEPEDREAFLDEFGKMEDVQVFFSEFIDREKQWFQKP

>*Skeletonema\_costatum*/CMK

MISTKVFALLXAAACGVXSVDGFVSSSTSRLGNXQKSNLHATTAESAPDSLALFSPCKI  
NLFLRIIRKRPDGFHDLASLFQAVGFGDTLTLNLNYGEXDEFNMEGVPTXSSNLVI  
RALELMREKTGKTSVYFNADLFKRCPAQAGLGGGSAXAATAMWGANELMGRPAXLE  
QLIEWSGALGSDITFFLSRGTAYCTGRGEXMTPIDPPLPSGAKLSIVKPNLGLSTPKVFK  
ALDYDQLSTEDPQMLCDLFLKEGVVNTDDKYIINDLEQPAFDCLPELKALKDELLEVX  
GFDHVMMSGSGTSIFCIGEPEDKDAFMKEFDERDGVSVFNAAFISREEXSWFQDPSQPL  
AVVDESEEEVVAPIIEEGMTAEEAERVRESRLKAAEAMKESIPAEPDNSEIVVEDEGSSS  
NFSKGLRSILNGXRKFEESRFGEKASYPSPKWL

>*Skeletonema\_marinoi*/CMK

XTMISTKVFALLXAAACGVXSVDGFVSSSTSRLGNIQKSNLHATTAESAPDSLALFSPC  
KINLFLRXIRKRPXGFHDLASLFQAVGFGDTLTLNLNYGEXDEFNMEGVPTXSSNL  
VIRALELMREKTGKTSVYFNADLFKRCPAQAGLGGGSANAATAMWGANELMGRPAT  
LEQLIEWSGALGSDITFFLSRGTAYCTGRGEIMTPIDPPLPSGAKLSIVKPNLGLSTPKV  
KALDYDQLSTXDPQMLCDLFLKEGVVNTDDKYIINDLEQPAFDCLPELKALKDELLE  
VEGFDHVMMSGSGTSIFCIGEPEDKDAFMKEFDERDGVSVFNAAFISREEGSWFQDPSQ

PLAVVDESEEEVVAPIIEEGMTAEEAERVRESRLKAAEAMKESIPAEPDNSEIVVEDEGS  
SSNFSKGLRSILNGQRKFEESERFGEKGYPT

>*Skeletonema\_menzelii*/CMK

MISTKMFALLFAAACGVSPVNGFVSSSTSRLGTIQKSHLHATTAESESLSLFSPCKINLFL  
RIIRKRPDGFHDLASLFQAVGFGDTLTLSLNNHGDADDEFTCNMEGVPTDSSNLVLRAL  
QLMREKTGKTNVHFNADLFRCPAQAGLGGGSANAATAMWGANELMGRPATLEQLI  
EWSGALGSDITFFLSRGTA YCTGRGEIMTPIDPPLPSGAKLSIVKPNLGLSTPKVFKALD  
YDQLSTEDPQMLCDVFLREGVVNTDDKYIINDLEQPAFDCLPELKALDELQVEGFD  
HVMMSGSGTSIFCIGEPTDKDAFMKEFDERDGVSVFNAAFVREEGSWFQDPSQPLAIV  
DEPEEEEEEEVAQIIEEGMTAEEAEKVRESRLRAAEAMKESPSETDNNIATKDSGTNFS  
KGLRSILNGERKFEESERFGEKGTAYPTPSKW

>*Thalassionema\_frauenfeldii*/CMK

MEWRILMICVQFLSVVSSFGTQRPGKSSLFHPSTQLHAKPNVKLALFSPCKINLFLRIIRK  
REDGFHDLASLFQAIGFGDTLELTILEGADADEFTCNMEGVVPVDQTNLVLRALTLMRE  
KTGNHNTYFYANLIKECPAQAGLGGGSANAATAMFGANELMGNPATLEEMVEWSAD  
LGSDITFFLSRGTA YCTGRGEIMTPIDPPIEASTKL VIVKPNVGLSTPTVFKSLEYDKLSD  
LDPENLLQQFMEESENVA MVDES AFVNDLEPPAFKNLPELMKMKEDLQKVKGQFQHV  
MSGSGTSIFCIGTPADAEAFEKQFSYRDDLVFFAEFIN

>*Thalassiosira\_minuscola*/CMK

MTVSKALATLLLATIGALPTHGFAPSTHAHRMPTKNDVFSATSQKSAVYAATDSDVTS  
KSSTDALTLFSPCKINLFLRIIRKRPDGFHDLASLFQAVGFGDTLHLSLTDKDADEFTCN  
MEGVPTDSSNLVLRALRLMRQKTDRKNVFFKADLVKRVPAQAGLGGGSANAATAM  
WGANELMGNPATLDELIEWSGELGSDITFFLSRGTA YCTGRGEIMTAIDPPLPSGAKLSI  
VKPDLGLSTPKVFGALDYDGLSEENPEGLRDTFLKEGVVNTDDKYIINDLEQPAFDC

>*Thalassiosira\_pseudonana*/CMK

MNLLSHALTTILFATCSILPSDAFVSSPIGQQTTLTKNFKNPFLRTQTRVAASDDSSSDT  
LKLFSCKINLFLRIIRKRPDGFHDLASLFQAVGFGDTLTLSLSEQPNDEFTCNMEGVPT  
DSSNLVLRALRLMRQKTGKTDVYFKADLFRVPAQAGLGGGSANAATAMWGANEL  
MGKPATLGELIEWSGELGSDITFFLSRGTA YCTGRGEVMTPIDPPLPSGSKLSIVKPDGL  
LSTPKVFSALDYEQLSAEDPEELLNTFLEFGVVDTDDEYYIINDLEQPAFDCLPELKAL  
EDLLKVKGFDHVMMSGSGTSIFCIGEADKDAFMKEFDSREGVSVFNAEFINREDGCW  
FESP

>*Thalassiosira\_puntigera*/CMK

MAISKALATVLLATFCASPVRGFVPPTHTHRSPSQGVYSATSEHSPSYATVDDADEATK  
STSSESLTLFSPCKINLFLRIIRKRPDGFHDLASLFQAVGFGDTLRLSPTEGDADEFSCNM  
EGVPTDSTNLVLRALRLMRRRTGRDDAFFEADLVKRVPAQAGLGGGSANAATAMWG  
ANELMGRPATLDQLIEWSGELGSDITFFLSRGTA YCTGRGEVMTPIDPPIPSGARLSIVK  
PDLGLSTPRVFAALDYDRLSEEDPEGLRDAFLREGVVDAPDGCYVNDLEQPAFDCLPQ  
LKALKEELLEVEGFDHVMMSGSGTSIFCIGEPADAEAFMKEFDGRDGVSVFNSEFINRE  
EGCWFVDPEEN

>*Thalassiosira\_rotula*/CMK

ALATVLLATSCALPTHGFIPSSSHRQPPSQGLLSSATSQKSLHATDSAADDEATKSAP  
STESLTLFSPCKINLFLRIIRKRPDGFHDLASLFQAVGFGDTLHLSLATKSDDEFCNM  
AGVPTDSSNLVLRALRLMRQKTDRKDVFNADLVKRVPAQAGLGGGSANAATAMW  
GANELMGRPATLDQLIEWSGELGSDITFFLSRGTA YCTGRGEVMTPIDPPIPSGAKLSIV  
KPDGLGLSTPKVQALDYDQLSTEDPQMLCDTFXKGVDTKDEYYIINDLEQPAFDCLPQ  
KALKEELLGVEGFDHVMMSGSGTSIFCIGEAADKEGFMKEFDSRKGISVFNSEFINRED  
GCWFEDPEN

>*Thalassiosira\_weissflogii*/CMK

MTSSGFCFLSKAILGLIATSVILSAQCFALESIAGRHLRASTTEDSSETLTLFSPCKI  
NLFLRIIRKRPDGYHDLASLFQAVGFGDTLHLSLSDNDGKDEFTCNMEGVPTDSTNLVL

RALELMRQKTGKTDVFFHADLFKRVPQAQAGLGGGSANAATAMWGANVLMGNPATL  
EQIIDWSADLGS DITFFLSRGTAYCTGRGEIMTPIDPPLPAGAKLTIVKPD LGLSTPKVFG  
ALDYDQLSVEDPETLLNIFLKDGVINTDDKYI NDLEQPAFDCLPELEALKDEL LCVKG  
FDHVMMSSGSGTSIFCIGEPNDKDSFMKEFDEREGVSVFNTEFISREEGSWFVNPEK

## IPP1 protein sequences

### >*Amphiprora* *sp*/IPP1

XSRLLVQAFQGTTPFRRRPTMMTTKATTATLAARIRSSNRMNHHPSRWFSSTTRITLSS  
SSSSSYGDNMNQHDMMESDLLIAVNELDQAILPNNNNHHNKSDSRLTKRLGHTFGPAT  
PRATLHRAFSLFLFDSSGRRMLLTQRAASKLTFPNVWTNACCSHPLQDMAVDEVD DGI  
RDYPTFPGIKQAAWRKCRHELGLDLRALMATTEDGDDSDDDPTQQRQLQLQDKMQFI  
TRFHYWAADSVTYGPSTEWGEHEVDYVLFQLPPGIETSGSLLQPNPEEVGDYKWVTI  
DELQAMLKSGNNKNLLWSPWFV GILERGGFEWWKDL SGALSGKYTNDQITFFDPPPE  
HVADYNLPSHNRWTGVWKWNPGKNEPGSASSKTKDETMSTSS

### >*Bolidomonas* *pacifica*/IPP1

XLSTFGIFFLLLPILNAFTFLRFHVPPMQFARGSKRFLPSTPSFLRYRASVPLLSSCGY  
RVESRLFSAYGAGLSQEDMMKNDMLILTGNDDSIIGPESKKVAHTFNESSPRGGVHRA  
FSVFLFNSKNEMLITKRAACKITFPNVWTNACCSHQLYNQSPEEVDHNSPEDVASRAG  
TVESNGSKNAALRKL YHEL GIRSIPHSSFRFLTRFWY WASDTGTYGEEAPWGEHEIDYI  
LFARQDGV EVMNEEEVDDIAWVSKDKLREMMNDPELKWSPWFV GIMNMKGFTYW  
DNLDSIIDGDGND FQDDTIX

### >*Chaetoceros* *sp*/IPP1

MSRNNWNGNSMSQQDFMEKDTVLVLDYDDK VIGSESKRASHEFNLEQPRGVLHRAFS  
VFIFDESTNELLLQQRAKEKITFPSVWTNTCCSHPLHGMSPK EVDTPEDVAAGNVPGA  
KYAAIRKLEHELGI PPSQLPLEKFKFLTRLHYWAADTVTHGPKSPWGEHEIDYV LFFV  
KSKNKITIVPNREEVEDVKWVTQQQLDMMNDSSLLFSPWFRIIAKKWLLGANGWWI  
DLEETIKTDKYCDYVTIHRXDPPKEHLGGLGNAACQFTSTGSTM TTDGDVCKKQGAYG  
KIKTHKESKLSQLSHLDEVWSAITLLYKPLKSNLQSDHV VNTFDASDLEFCDDILCKVS  
RSFAAVIRQLPATLLVDVLIFYLVLRALDTVEDDMTAFESNEIKIQYLLKFHETALVDPT  
WSMDGVGEADEKRL LQEFPKCHSVYAALKSESKDVIANITKRMAEGMAEFV GKDLG  
QGTVDVKEYNRYCHFVAGLVGEGLSRLFAASGLEDA SLAKELYLS DQMGLFLQKTNII  
RDYLEDYVDGRAFWPQSVWKKYSKTGDLGYFANQSD EDAKYHALCCLNELVTDALH  
LVPDCLKYLSKLQCCEVFRFCAIPQVMAIATLDKCYH NKDVFTGVVKIRKGM SCKLIN  
NTHDVXGVHAVFYEF AKSIANKARHAKASGFVDPTFESTIQICKKICDLTQGEVQKMK  
KMRTVELSLALVAGW

### >*Coscinodiscus* *wailesii*/IPP1

MTTWTW DGNAMTQTDMMIKDTVLVLDENDNVIGSASKKKSHDFSIENPRGILHRAFSVF  
MFDEDSGELL LQKRASSKITFPNVWTNTCCSHPLHGMSPP EIDDPSTILTGQTTGVINGA  
IRKLEHELGI PPSQLRADKFKFLTRFHYWAADTVTHGPDAPWGEHEIDYV LFLYLCDDK  
ELTVVPHPDEVD DVKWSREALVKEME VGLWSPWFRILANKWLLNEGGW WEDLAK  
TMAMPDESILRFDPPAEHMGKGKAGAWLGTV

### >*Cylindrotheca* *closterium*/IPP1

KDTVVL DDEDNIIGSASKKESHVFNTEQPRGILHRAFSVFLFDESTGELL LQQRASKI  
TFPNVWTNTCCSHPLHGMEPAEVDTPEDVANGTVMGAKYAAIRKLDHEL GIPSNQLP  
VEKFKFLTRLHYWAADTVTHGRDSPWGEHEIDYV LFFCVKNKSELTIKPHPDEVD DIK  
WVSTQKLDEM MKDSSLLFSPWFR LICKKWLVTSSWWKNL KQTMTTNDHCDYATIHEF  
DPPKEHLGNGKAGPLFEEKQGD ESKKQGAYGKVKTHKEPMLKQLSHLDEVFAAFYF  
LKVNPLKSNLETPFIEENFGADDLAFCDDILVKVRSFAAVIRQLPSSMLVDILIFYLVL R  
ALDTIEDDTTAFESNDIKIKHLANFRKTALGDPTWTMDGVGEGDEKR XKISTSVTXVY

AKLSEKSQSVIADITQRMADGMAEFVGGKDLGQGTVDVEQYNRYCHFVAGLVGEGLSR  
LFAASGLEKQSFAGELFLSDQMGLFLQKTNIIRDYLEDYVDHRAFWPQTIWKKYSKTG  
ELGYFTRQEDEEVRERSLQCLNELVSDALELVPDCLAYLSKLQCKQVFRFCAIPQVMAI  
ATLDKCYSNSNVFTGVVKIRKGTSCKLILRTNTLDEVHDTFYQFSQAIVGKAMAKRRM  
GVVDPSYERTMKACETIQELTASGHRRQTSKRRMPFVLGGLLLAASLGSKHIPVDFPG  
MQQSLLLLGIGMLSFPGPYAVSQESDLKDANALQ

>*Fragilariopsis kerguelensis*/IPP1

MKTSILLGILLRTNHCVTSFSSSLITPKSRSRHAIVTAKASTLLPALAISTPVDSTSEKGF  
ASSLSHAQENIDKYDNDASAVYGADMNQEDMMESDLLVVVDHNDNVIELSDKQPISK  
KAAHTFSEEQPRGVLHRAFSLFCFNEDGKLLTRRAADKITFPAVWTNTACSHPLLKM  
PLSEVDYWPDAFPQMPGIKRAAARKSKHELGLNIQPFLMDVQFVSRFHYWASDVQTH  
GLEAPWGEHEVDYILFLQCPVQEFEEKVNPEEVADIKYVSIVELKDMLYNQPELKWSP  
WVFGIMERGGFTWWEDLENTLQGGKYTNRDIQFFDPLPNHVAKYNLPQHTRDKTGWW

>*Leptocylindrus apora*/IPP1

MTQEQFMQKDTVLLVDEQDNVIGSASKKDSHVHDEHPRGILHRAFSVFLFNEKDELL  
LQQRASSKITFPYVWTNTCCSHPLHGLEPGEVDGPKDVADGSVMGVKHA AVRKLWH  
ELGIPAEQLPVEKFKFLTRLHYWAADTVTHGKGSWGEHEIDYVLFIKVPQDSL TINPR  
PDEVEAVQWVSKKQLLHLLSVASDETTITEGNKNSLWSPWFRLIAYRWLLPDGTGWW  
DNVDDVLHTDKFNDYENVHKFDPPKEHMGGGGNAGEWLGQLQ

>*Leptocylindrus danicus*/IPP1

MTNNTTNTATTAATWTGSGLTQSDLMTKDTVLLVDEAXNVIGSASKKESHVFDATT  
NTQRGILHRAFSVFLFNERGELLQQRASKITFPNVWTNTCCSHPLHGLEPGEVDGPQ  
DVQDGSVMGVKHA AVRKLWHELGIPAEQVPIEKFKFLTRLHYWAADTVTHGKESPW  
GEHEIDYVLFIQVEQSKLTINPHPEVDAIKWVSKEQLIEEMKVGLWSPWFRLIVNKW  
VIPVWVNDLKM TLETDVHCDYENIHRFDPPEEHMGGGGNAGPWLKLEIGSTGTISS  
SSTS NFLKLFATLGXAIAAVSSCVLHLYPELIESLPFEFSPVFEQTVALCITSLLVSTFYV  
SVQSQLKRQIMNMPPEVPEIPVLGSIKSFLTNPWDLMENWHKQYGGIYRYTLLGRTC  
VALADPAHLRIVLQSKIKHVKKDVFAYKPFPLILGKGIVTSEGKSWMRQLSVSSVLR  
IDVLDEIPRITLGAVQRLCAKLDAAAESGEVIEMAEELRHLTLQVISETFFSLPAEESNTT  
FAQMYLPIVEEGNKR VWHPERNFAFFLPFFWKHKAAVARLNSYVSKLVVERWEQRRS  
EKTSNDPKRKEDVLDRIISA YDKEDAPKWGKAVVNQLRDELKTFMLAGHETSAAMM  
TWALYELVSNEELMETVTKEGASVFGADYDWCKDGVNADYLP SREELGKLELSEGCL  
RESLRKYSVVPTVTRKA VENIQVGSHTIPKGSTIMINIQA VHHD AKLWPDPMKFDPTRF  
CDPSPQPKPYTFIPFIEGPRNCLGQYLALLESKMVLSMLLQRY SFTLQSGGEP RHPYMV  
PVIPKTGVFVTVKRKF

>*Nannochloropsis gaditana*/IPP1

HRYDHLNMPRFKLLSIFSPSRKMTQEDFMRKDECLVVD DRDLLLRTENKYNCHRFTP  
EQPRGLLHRAFSVFLFDKTGRLLLQQRASKITFPQVWTNTCCSHPLSGFTPT EVDMP  
DLANGKVKGVKRAAIRKLQHELGIPSFQVPLRKFKFLTRLHYWAADAATHGPAAEWG  
EHEIDYVLFIQVPTVTLRPNPEEVNAV K FVSLPELQEMMHPSSGLRWSPWFRIIAERFLP  
LWWADLKLTLRTNTYFDPTTIHRFEPPTDPTGNISPSEFAQDTAASAPSGGLVPMGRAS  
ASPTLPPAFGKGDEGKKQGAYGKIKSHSESLVSQLLRVDEVGAALRLKFARPPETDV  
GHAGDADVRFCADMLGRVSRFAAVIRQLPKGLCLETVVFYLVLRALD TVEDDMEAF  
QKDPGAKLRHLRGFYHVALEDPEFKMHGVGKGDEARLLENFDRVVRVFQGLTPGMR  
TVIKDITRRMGAGMADFVDKDLGQGTISLQEYDLYCHYVAGLVGEGLTRMFAESGLE  
APDLATSRQELANS MGLFLQKTNIIRDYLEDVFDGRAFWPQEVWKRHAPSAVPAQSLG  
AFAEAHRERALACLDELVTNALEHVPDCLAYMSRLHTPQIFRCAIPQVMAIATLEKV  
YHNPDVFTGVVKIRKGLACKLIMEANSMDLYTCFSSFAGKIARRAAALEKRGDSVSA  
QTLRVCEEIRVLCKEGEASSTNTKKYARALHA AFAALLLAIALVPVLALWGTESSRGG  
FFARPIESPAGLSGADYVQMAFNVLVSVVATLALS YVVSKVWRPLGGQGNRSRLAA

>*Nitzschia punctata*/IPP1

AKELPIEKFKFLTRLHYWAADTVTHGNESPWGEHEIDYVLFVCVKSDELTIKPHPDEV  
DAYKWVTRQGLIDMMEDKSLLFSPWFRICKKWLXSWWKDLNNTMTTDKYCDYV  
NIHCFDPPKEHLGGGGKAGPLFEKEAEGDESKKQGAYGKVKTHKESTLKQLSHIDEVF  
X

>*Nitzschia\_sp/IPP1*

MSTKTWDGNAMSQEDMMKKDVTVLVDDKDNVVGSAKRTSHEFEDGQPRGILHRAF  
SVFLFDESTGELLQQRASKITFPNVWTNTCCSHPLHGMEPNEVDKPEAVADGSVLG  
AKYAAIRKLDHELGIASDLPVEKFKFLTRLHYWAADTVTHGDESPWGEHEIDYVLF  
CVKSKSELTVPKPHPDEVDACKWVTRQGLIDMMQDKSLLFSPWFRICKKWLDSWW  
ANLETTMTTDKYCDFVNIHRFDPPKEHLGGLGKAGPLFDAPTEGDASKKQGAYGK  
LTHKESKQISHIDEVFAAVYFLYVKPLKSNLDTEYIRESFGADDLAFCEILVKVSR  
AAVIRQLPSTMLVDILIFYLVRALDTIEDDTTSFPSQEVKIKHLLNFHKTALGDPT  
WMDGVEADEKRLQNFDKCHRVFACLREPSRVIADIAQRMATGMAEFVSKDLGQGT  
DQYNDYRCHYFVAGLVGGLSRLFSASGLEKASFASELVLSQDQMGFLQKTNIRDY  
LVEDYVDHRAFWPQTIWKKYSKSGDLGYFTNQDDPEVRKNSLACLNELVTDAL  
ELVPDCLSYMSKLQCAEIRFCAIPQVMAIATLEYCYANSNVFTGVVKIRKGSSCK  
LILNTNLDVHDTFYHFANCLLKTANTNRAKGVDPXYARTVKICETIMETTSPAK  
TRLDKARRMPLIASSILVAGVGATMLKSDRPSSTPLWQQGLLLALGVATFGPSMIR  
KPSPSNLIKADLEEKLMSSPQSRSKTNSSRNDETDSDEKKEAK

>*Phaeodactylum\_tricornutum/IPP1*

MRRMPGKIGAMILGAAWIGRVAGLAPSAKSRQFMVAQSWGLRAIKDRCATCLFSS  
NAKSNDAKAYGVNMDQEAMMESDRLIAVNENDRLVRNVNLSKRNGHTFNKETPRA  
ALHRAFSFFLFDQDRMLLTQRAGTKITFPNVWTNTCCSHPLYDMTPNEVDVADDA  
YPMFPGIKHAAVRKCKHELGIAPENLPKEDIQFLTRFHYWAADTVTYGDDTAWGE  
HEIDYILFLQVNGQVPVDANPDEVSDYKYVSMEEMKDMMEDPDLLWSPWFRGIM  
NRGGFEWWADLRGSLQGGKYTNLDVTFFDPPKEHYASYNLPSHERKTGVLSSTVE  
QT

>*Pseudonitzschia\_australis/IPP1/IPP1*

MQLPLQISVAFLLQKHSLAFGLLSPHSKSSPLSPLSVAAPLTEIDEKNIEIFNID  
SLDYGADMDQEDMMESDLLVIVDENDNVIELSEDQSVSKKAAHSFSNEQPRGILH  
RAFSLFCFNE DGKLLLTQRALDKITFGVWTNTACSHPLMEMPLSEVDVWPGAYPE  
MFGIKNAAVRKAHELGLDIRSLIPDVQFVSRFHYWASDVQTHGIEAPWGEHEVDY  
ILFLKSSVKEFKLEVNPEEVGDVMYVVISSELKDMIYNQPDLTWSPWFIGIMERGG  
FDWWEDLENTLQGGKNTNEDIQFFDALPDHVATYNIPEHSRQTGVWKN

>*Pseudonitzschia\_arenysensis/IPP1*

MRLHSRIPLSLLLGGCSKGLLSRHNNPRTFSLGSAPAAESSYGADMNQEDMMESD  
FLVIVDHNDNVIEVSEEVSKKAAHSFSSDQPRGILHRAFSLFCFNQDGKLLLTRR  
ASDKITFGVWTNTACSHPLLQMAKSEVDVWPEAFPELSGIKNAATRKAQHELGLS  
LEQADMQFVSRFHYWASDVETHGSETPWGEHEVDYILFLQAPLAEFELDVNAEEV  
AEVRYVSIDELKDMLYNQADLTWSPWFGIMERGGFDWWQDLENTLAGKNTNEDI  
VFFDPLPDHVASYNKPDHTRQSTGIW

>*Pseudonitzschia\_delicatissima/IPP1*

MRLPLRISLSLLLGGCSLTFGLIPHHTPRSLTTTSSLSSAAESSLEYGADMNQED  
MME SDLLVIVDHNDNVIELSEDXSVSKKAAHSFSSDQPRGILHRAFSLFCFNQD  
GKLLLTRRATDKITFGVWTNTACSHPLLQMPNSEVDHWPAAFPEMFGIKNAARKA  
KHELGLVVEPKDMQFVSRFHYWASDVETHGPEAPWGEHEVDYILFLQAPVQEFPL  
NINSEEVGDAEYVSIAELKDMLYNQPELTWSPWFGIMERGGFDWWEDLENTLAG  
KNTNQDITFFDPLPDHVASYNTPEHTRQTTGVWKNE

>*Pseudonitzschia\_multistriata/IPP1*

MATQTDWDGNAMSQEDMMMMDKDTVLVDDDDNVIGSASKRKSHEFTPEQPRAIL  
HRAF SVFIFDESTGELLQQRASSKITFPNVWTNTCCSHPLHGMEPPEVDGPKDV  
ADGTVMGVKNAAIRKLDHELGIPELSDINKFKFLTRLHYWAADTITHGPDSPWGE  
HEIDYVLFYT



VSSKSDLTIKGHPDEVDDVKWVTRDVLIEMMEDKSLLFSPWFRICKKWLLDHWKDLKRTMTTDDHVDLKSIIHCFDPPKEHLGGGGKAGPLFGAEVEGDER

>*Pseudonitzschia pungens*/IPP1

MRLTIKILVAFMLQTQSRGFGIVPRRHESRSLPLSASAMTKDEKNIGSVNIDSADYGA  
DMNQEDMMESDLLVIVDHNDNVVELSEDQTVSKKAAHTFSEEQPRGILHRAFSLFCFN  
DDGKLLLQTRASDKITFPGVWTNTACSHPLMQMPASEVDIWPDAFPEMPGIKRAAAR  
KAKHELGLDIQPLPGVQFVSRFHYWASDVQTHGPDAPWGEHEVDYILFLKAPVKEFV  
LDVNPEEVTAVKYVSISELKDMLYNRSDLTWSPWFVGMERGGFDWWEDLENTLQG  
KNTNTEIQFFDPLPSHVATYNSPEHTRQTGVWKNK

>*Pseudonitzschia pungens fr. cingulata*/IPP1

FSLTLKNADEYSIMRLTIKILVAFMLQTQSRGFGIVPRRHESRSLPLSASAMTKDEKNI  
GSVNIDSADYGADMNQEDMMESDLLVIVDHNDNVVELSEDQTVSKKAAHTFSEEQPR  
GILHRAFSLFCFNDDGKLLLQTRASDKITFPGVWTNTACSHPLMQMPASEVDVWPDAF  
PEMPGIKRAAARKAKHELGLDIQPLPGVQFVSRFHYWASDVQTHGPDAPWGEHEVD  
YILFLKAPVKEFVLDVNPEEVTAVKYVSISELKDMLYNRSDLTWSPWFVGMERGGFD  
WWEDLENTLQKNTNTEIQFFDPLPYHVATYNSPEHTRQTGV

>*Skeletonema costatum*/IPP1

MGNKQSSSSSSSTTNKKWDGNAMSQNDFMERDTVLVLNNDIVIGSESKKASHVFST  
DQSRGVLHRAFSVCLFDSSTGEMLLQKRASTKITFPNVWTNTCCSHPLHGMEVNEVDG  
PEDVKDGSVKGVKAAAVRKLEHELGIPIVVGELPIENFKFLTRLHYWAADTVTHGEKSP  
WGEHEIDYVLFITVPTKDSITLKPHPDEVDDVCWVTQSKLLEMFANDDLLFSPWFRLIA  
NRWMIGGGENGKGGWWDDLDRTMNTDDFCDYENIHRFDPPTEHLGGGGDAGPWL  
TKEAPSANGTVEDTVGDASKKQGA YGKIKTHKESKLSQLMHIDEVCSAISLLYINPLKS  
NLESDNIKKSFDTDDXAFCDILCKVSRSF AA VIRQLPSELLVDVLIFYLVLRALDTIED  
DMTFFASNDEKVKVLLAFHKALVDPEWKMMGC GEGDERRLLEEFPKCHSVFAALRE  
DSRRVISDITLRMATGMAEFVKGDLGQGTVDVQQYNYRCHYFVAGLVGEGLSRLFSAS  
GLEKPALATELYLSDQMGLFLQKTNIRDYLEDYVDGRAFWPQT VVKKYSASGDLGY  
FANPATVEHETAALNCIDELVTDALVDPCLS YLSKLRCAEVFRFCAIPQVMAIATLD  
KCYHNKDVFSGVVKIRKGMSCALINGTTDINGVHSIFHRFASSILRQAELSRKKG FEDPS  
YERTVKA CTILELTETEAKRSGSSVVGSTLMLSTAAVGCLTYKYPNGSNTTALTTAIA  
SXGMFSLFKTSFMKYVGKKG VSSLMPASKLTEKKEG

>*Skeletonema marinoi*/IPP1

MSQSDFMERDTVLVLSDNDEVIGSASKKESHQFTPKNPRGILHRAFSVFLYDLSTQXAI  
EDVMDGTVKGVKAAAVRKLEHELGIPIVVGQLPMDSFKFLTRLHYWAADTVTHGEKSP  
WGENEVDYVLFVTVLNKDKITIEPHVDEVDDVCWVTKAKLFEMFADKDLLFSPWFRII  
ANRWMISSGEETGKGGWWDDLDRTMNT

>*Skeletonema menzeli*/IPP1

MSQSDFMERDTVLVLSDNDEVLG SASKKESH LFTPTNPRGILHRAFSVFLYDASTKELL  
LQQRAS TKITFPECMDEYLLFXXXXXXXXXXXXXHGMDVNEVDS AEDVKG GTVNGVKA  
AAIRKLEHELGIPIVVGQLPMDSFKFLTRLHYWAADTVTHGEKSPWGENEIDYVLFATVP  
CKDSITINPRPDEVDDVCWVTQSKLLEMFADKDLLFSPWFRIIANRWMISKGGWWDDL  
DRTMTTNDFCDYESIHRFDPPQEHLGGGGNAGPLFVAEK

>*Thalassionema frauenfeldii*/IPP1

MMESDLLIPVGSNDELISSMDVTKKKAHIFSKEGQPRGVLHRAFSFFLFDQNRMLVT  
KRASSKITFPGVWTNTCCSHPLQKMKPEEVDDPLEEKNSIEFRYDGVKHAARRKLRHE  
LGIDPIHIPHEHIQFITKFHYWAADTLNYGTDNPPWGEHEVDYILFFKHKGSEPIVLPCE  
DEVEDYRYVTQDEFKDM LKDIPAKNWSPWFRGILDRGLWQWWDDLDHGTLTGKHS  
NNDVIFFDPPKEYYASFNNDPSHNRQTGV LNAASV

>*Thalassiosira minuscola*/IPP1

MPLSFSRSPLRRAPQRRNTAPLLLGLSILLVIYYARRKIIDIDTPQLRKYEVPSSSNHAQ  
GGELLN VYELPPSHATEPAHLTPVDPTTLVPKGTHPIDVAHRNGYLHTGHILYIMDDSG

SLLFLQRTPGVVTC PSTWSLLGEHANANESPRETVLRALEEELGFVGLNFDDFKNAGA  
FTAELHPSHNLKNSLTVTINTVTDLPLYIRHYGPRNDDRDRQLTYLWVVNFPKKHAE  
IDWKFDDDEVADHKWVTLEEAASWLSGDGRNGKEDGEAKGGTGSDSAKEYGSDWKE  
VKDDGPDEGEFCHHTIRSLYEVL SHMI

>*Thalassiosira\_pseudonana/IPP1*

MAPTTKKWDGTAMTQTD FMERDTVLVLSNDDTVIGSESKRSSH EFTPTQPRGVLHRA  
FSVFLFDSSTNELLLQKRASTKITFPNVWTNTCCSHPLHGMDVNEVDGPEDVKDGSVM  
GVKAAALRKLEHEL GIPV GELPVEGFKFLTRLHYWAADTVTHGKKSPWGEHEIDYVL  
FITIPSKSTLTLKPHPDEVD DTRWVTQS QLLEMFAD ELLFSPWFRLITNRWMIGEGSEK  
NVNGGKGGWDD LKRTMETNDFCDYVNIHRFDPPSEHMGGGGDAGPMFDGAAVVG  
DASKKQGA YGKV KTHKESTLKQLVHLDEVCSAITLLYLQPLKSNLESDVIKQTFDPSD  
LAFCD EILCKVSR SFAAVIRQLPPTLLMDVLIFYLVLRALDTIEDDMTFFDSNEDKV KIL  
QSFHKTALVDPNWKMMGCGEGDERRLLEEFKCHSVYAALPEGSRRVISDITQRMAT  
GMAEFVDKDLGQGTVDVQ QY

>*Thalassiosira\_puntigera/IPP1*

MRSRISLPAAMMMQRRHYRFFYGGISSTATGSRSNYHEAFASLALRKRQIASRTRRHV  
TCSAIVSNWDGTSMSQAQFMERDTVLVTDENDILIDSASKKSSHEFTPSQPRGILHRAFS  
VFLFDASTSELLLQKRASTKITFPNVWTNTCCSHPLHGMEKDEV DGPDDV RDGSVGGV  
RHAAVRKLAHEL GIPMEELWAMEREFKFLTRVHYWAADTVTHGEKSPWGEHEIDYIL  
FATIPSKNKLTLKHPPEEVD DTRWVTQS QLLGMFDDG TLLFSPWFRLIAQRWMIGKTT  
NVGQEGWWD

>*Thalassiosira\_rotula/IPP1*

XASANEDEDEDDEDDASMATGGSRSKFCGAFGSPAAWN RQIIASTSRRNHFTSTAIAS  
VSNTNINMNNNNQASSPSPSTSWDGTSM SQAQFMERDTVLVTDHNDIVTSSATKKDS  
HIFNPTNPRGILHRAFSVFLFDASTSELLLQKRASTKITFPNVWTNTCCSHPLHGMEPNE  
VDGVEDVKDGSVGGVRNA AVRKLGHELGM DKG ELET MSEEFKFLTRVHYWAADTV  
THGEKSPWGEHEIDYILFATIPTKNSLALNPHPEEVD DIRWVTQS QLLQMFDDETLLFSP  
WFRIIANKWMVGGGGWDDLDRTMNTNAFCDYETIHRFDPP EEHMGGEGDAGPWL  
CDPE